

**UNIVERSIDAD NACIONAL DE SAN CRISTÓBAL
DE HUAMANGA**

**FACULTAD DE INGENIERÍA DE MINAS, GEOLOGÍA Y
CIVIL**

ESCUELA PROFESIONAL DE INGENIERÍA CIVIL



TESIS:

**Modelo inteligente adaptativo fundamentado en redes
neuronales convolucionales para mejorar la seguridad
vial y prevenir accidentes de tránsito, en las avenidas,
Arenales y Cusco, Ayacucho, 2025**

Para optar el título profesional de:

INGENIERO CIVIL

PRESENTADO POR:

Bach. Abel DE LA CRUZ MORAN

ASESOR:

Dr. Ing. Hemerson LIZARBE ALARCÓN

AYACUCHO - PERÚ

2026

RESUMEN

La presente investigación diseñó, entrenó y validó un modelo inteligente adaptativo de segmentación de instancias basado en redes neuronales convolucionales para la detección en tiempo real de doce categorías de elementos críticos de la vía pública en las avenidas Arenales y Cusco de Ayacucho, entorno urbano andino a 2,761 m sobre el nivel del mar caracterizado por reductores de velocidad artesanales que incumplen la Directiva N° 01-2011-MTC/14, señalización deteriorada, semáforos desprogramados, imprudencia peatonal y transporte informal no regulado que incluye al mototaxi como vehículo predominante. Se construyó un conjunto de datos híbrido de 25,602 imágenes con 127,525 instancias integrando 4,598 capturas originales del entorno ayacuchano —con el mototaxi como clase local exclusiva— y tres conjuntos de datos internacionales (BDD100K, BSTLD y RLMD), etiquetados mediante LabelMe asistido por SAM2 en doce categorías y divididos en proporciones 70/20/10. Se implementó YOLOv8L-seg con entrenamiento progresivo en tres fases de resolución creciente (640, 800 y 1,024 píxeles) mediante aprendizaje por transferencia desde COCO, alcanzando mAP₅₀ Box de 0.810 y mAP₅₀ Mask de 0.778, con una mejora acumulada de 30.7 puntos porcentuales respecto a la primera fase. La comparación multi-arquitectura con la familia YOLO26 (nano, small y large) identificó a YOLO26L-seg como el modelo de mayor desempeño, con mAP₅₀ Box de 0.829 y mAP₅₀ Mask de 0.788, superando a YOLOv8L-seg en nueve de las doce clases. Se desarrolló el software QHAWAY ADAS, que integra estimación de distancias por modelo geométrico de cámara monocular, cálculo del tiempo estimado de colisión, seguimiento de múltiples objetos y alertas de voz en español peruano, operando entre 19.2 y 25.4 fotogramas por segundo. Se concluye que las arquitecturas de aprendizaje profundo con entrenamiento progresivo constituyen una herramienta viable para la detección automatizada de elementos viales en el entorno urbano andino informal de Ayacucho.

Palabras clave: Seguridad vial; segmentación de instancias; redes neuronales convolucionales; YOLOv8; YOLO26; sistema de asistencia avanzada al conductor; mototaxi; entorno urbano andino; estimación de distancias monoculares.

SUMMARY

This research designed, trained, and validated an adaptive intelligent model for instance segmentation based on convolutional neural networks for the real-time detection of twelve categories of critical road elements along Avenidas Arenales and Cusco in Ayacucho, an Andean urban environment at 2,761 m above sea level characterised by artisanal speed bumps non-compliant with MTC Directive No. 01-2011-MTC/14, deteriorated road markings, malfunctioning traffic lights, pedestrian recklessness, and informal transportation dominated by the mototaxi—a locally prevalent three-wheeled vehicle absent from all major international road datasets. A hybrid dataset of 25,602 images with 127,525 annotated instances was constructed combining 4,598 original Ayacucho captures—featuring the mototaxi as a locally exclusive class—with three international datasets (BDD100K, BSTLD, and RLMD), annotated using LabelMe assisted by SAM2 across twelve categories with 70/20/10 partitioning. The YOLOv8L-seg architecture was trained through a three-phase progressive strategy with increasing input resolution (640, 800, and 1,024 pixels) using COCO transfer learning, achieving mAP₅₀ Box of 0.810 and mAP₅₀ Mask of 0.778, with a cumulative improvement of 30.7 percentage points over the Phase 1 baseline. A multi-architecture comparison across the YOLO26 family (nano, small, and large) identified YOLO26L-seg as the best-performing model, with mAP₅₀ Box of 0.829 and mAP₅₀ Mask of 0.788, surpassing YOLOv8L-seg in nine of the twelve classes. The QHAWAY ADAS software was developed integrating monocular pinhole distance estimation, time-to-collision computation, multi-object tracking, and Peruvian Spanish voice alerts, sustaining 19.2–25.4 frames per second. It is concluded that deep learning architectures with progressive training constitute a viable tool for automated road element detection in the informal Andean urban environment of Ayacucho.

Keywords: Road safety; instance segmentation; convolutional neural networks; YOLOv8; YOLO26; advanced driver assistance system; mototaxi; Andean urban environment; monocular distance estimation.

INTRODUCCIÓN

La seguridad vial constituye una problemática multidimensional cuya complejidad trasciende las cifras de mortalidad y lesiones que habitualmente la caracterizan. Detrás de cada siniestro de tránsito existe una cadena causal donde confluyen factores humanos, infraestructurales y tecnológicos: Peatones que cruzan las vías de manera temeraria ignorando los cruces señalizados, conductores que exceden los límites de velocidad o conducen en estado de ebriedad, señalización horizontal borrada por el desgaste, semáforos desprogramados, reductores de velocidad improvisados construidos por las propias pobladores sin especificaciones técnicas, y calles invadidas por el comercio ambulante que obliga a peatones y vehículos a compartir espacios no diseñados para esa convivencia. Estas condiciones se potencian mutuamente y configuran un escenario de riesgo permanente que cobra la vida de más de un millón de personas cada año a escala global, con una concentración desproporcionada en los países de ingresos bajos y medios.

En el Perú, los registros oficiales documentan cifras que superan las tres mil víctimas fatales anuales, con una tasa de mortalidad que duplica el promedio europeo. No obstante, reducir la problemática a las estadísticas de siniestralidad sería desconocer sus verdaderos orígenes. Las calles peruanas exhiben un nivel de informalidad que transforma los espacios públicos en escenarios de riesgo cotidiano: Vendedores ambulantes que invaden aceras y cruces peatonales, mototaxis que operan sin regulación efectiva, vehículos estacionados en doble fila que obstruyen la visibilidad, y una proporción alarmante de peatones que nunca cruzan por el cruce peatonal ni respetan las indicaciones semafóricas. Esta realidad anula la función de la señalización horizontal existente, cuyos cruces y líneas de carril quedan ocultos bajo el comercio ambulante, los vehículos mal estacionados y el deterioro acumulado.

En la ciudad de Ayacucho, situada a 2,761 metros sobre el nivel del mar, todos estos factores se concentran y amplifican. El trazado urbano de herencia colonial resulta incompatible con las demandas del tránsito motorizado contemporáneo. Las avenidas principales presentan señalización horizontal severamente deteriorada, semáforos con desprogramación crónica, y una proliferación de reductores de velocidad improvisados que las comunidades vecinales construyen con montículos de concreto sin cumplir la normativa vigente, generando obstáculos peligrosos que provocan maniobras evasivas, daños vehiculares y caídas de motociclistas. A ello se agrega la presencia masiva de mototaxis sin regulación, el comercio ambulante que desborda hacia la calzada, y las aglomeraciones peatonales en horarios punta. Las avenidas Arenales y Cusco concentran estas problemáticas de manera emblemática, constituyendo nodos de alta confluencia donde los datos revelan que la gran mayoría de siniestros fatales ocurrieron en vías sin señalización horizontal, y que la provincia de Huamanga concentra aproximadamente la mitad de los siniestros fatales del departamento, validando la pertinencia de la zona de estudio seleccionada.

Ante esta convergencia de factores, la presente investigación propone el desarrollo de un sistema de detección de seguridad vial fundamentado en segmentación de instancias mediante redes neuronales convolucionales profundas. A diferencia de los enfoques convencionales que generan cajas delimitadoras rectangulares, la segmentación de instancias produce máscaras a nivel de píxel para cada elemento detectado, permitiendo delimitar con precisión la geometría irregular de señalizaciones horizontales, reductores de velocidad, siluetas de peatones y tipología de cada vehículo. El modelo fue entrenado con un conjunto de datos construido a partir de múltiples fuentes de adquisición de la zona de estudio, etiquetado con herramientas asistidas por inteligencia artificial y organizado en doce categorías que abarcan los elementos viales más relevantes del contexto ayacuchano, incluyendo el mototaxi como elemento distintivo del transporte informal peruano.

La motivación que impulsa esta investigación se ancla en la necesidad de preservar la integridad física de quienes transitan por las vías ayacuchanas. La inexistencia de sistemas inteligentes de asistencia al conductor en la ciudad de Ayacucho representa una brecha tecnológica que la presente investigación busca cubrir, demostrando que es posible adaptar arquitecturas avanzadas de aprendizaje profundo a las condiciones específicas de una ciudad andina peruana. El sistema integra el modelo de segmentación en un software funcional denominado QHAWAY ADAS, vocablo quechua ayacuchano que significa “mirar” u “observar con atención”. Este software incorpora estimación monocular de distancias, cálculo del tiempo de colisión, seguimiento multi-objeto y generación de alertas preventivas por voz, abordando de forma integral los múltiples factores que originan la problemática vial: detecta la señalización existente, identifica los reductores de velocidad, reconoce el estado operativo de los semáforos, monitorea la presencia de peatones y clasifica los distintos tipos de vehículos del parque automotor local.

El trabajo de investigación se encuentra organizado de la siguiente manera:

1. PLANTEAMIENTO DEL PROBLEMA

- Se formula el problema de investigación a nivel mundial, nacional y regional, analizando las causas multifactoriales de la problemática vial: Imprudencia peatonal, conductas inapropiadas de conductores, reductores de velocidad improvisados, comercio ambulatorio, señalización deteriorada, semáforos desprogramados e informalidad del transporte. Se establece la delimitación espacial, temporal y la unidad de análisis.

2. MARCO TEÓRICO

- Se presentan los antecedentes internacionales y nacionales, las bases teóricas de la seguridad vial e ingeniería de tránsito, y los marcos conceptuales de las redes neuronales convolucionales, la segmentación de instancias, el aprendizaje por transferencia y las arquitecturas de detección en tiempo real.

3. MARCO METODOLÓGICO

- Se detalla la metodología, la población y muestra, las hipótesis y la operacionalización de variables. Se describe la construcción del conjunto de datos, el etiquetado asistido por inteligencia artificial, la división estratificada, el entrenamiento progresivo del modelo, los criterios de selección de la arquitectura y el desarrollo del software de asistencia al conductor.

4. RESULTADOS

- Se presentan las métricas de rendimiento del modelo, matrices de confusión, curvas de precisión-recuperación, análisis de explicabilidad mediante mapas de calor y la validación del software. Los resultados se contrastan con las hipótesis específicas de la investigación.

5. CONCLUSIONES

- Se presentan las conclusiones en relación con los objetivos e hipótesis, las recomendaciones derivadas del estudio y las sugerencias para futuras líneas de investigación.

A mis padres, por su sacrificio silencioso y su amor incondicional que hicieron posible cada paso de este camino. A mis hermanos, compañeros de vida, por su apoyo constante, su paciencia y sus palabras de aliento en los momentos más difíciles. Todo lo que soy se lo debo a ustedes; este logro les pertenece tanto como a mí.

AGRADECIMIENTOS

A los ingenieros de la Escuela de Formación Profesional de Ingeniería Civil de la Universidad Nacional de San Cristóbal de Huamanga, cuya dedicación en la enseñanza y compromiso con la formación integral de sus estudiantes sentaron las bases que me permitieron encaminar mi desarrollo profesional.

Mi especial reconocimiento al Dr. Ing. Hemerson Lizarbe Alarcón, asesor de esta tesis, cuya orientación constante, exigencia académica y confianza depositada en este trabajo fueron fundamentales para la culminación exitosa de la investigación.

A los distinguidos miembros del jurado evaluador, Mg. Ing. Alex Sander Ircañaupa Huamaní e Mg. Ing. Víctor Bellido Aedo, por su rigurosa revisión, observaciones pertinentes y valiosas sugerencias que contribuyeron significativamente a mejorar la calidad de este trabajo.

A mis compañeros y amigos que me acompañaron a lo largo de este camino, por las jornadas compartidas, el apoyo mutuo y las palabras de aliento en los momentos más exigentes. Su compañerismo hizo más llevadero este proceso.

Finalmente, mi gratitud a todos quienes de una u otra forma contribuyeron a que esta meta se hiciera realidad. Su respaldo y confianza fueron el impulso necesario para perseverar hasta el final.

ÍNDICE GENERAL

ÍNDICE GENERAL	IX
LISTA DE FIGURAS	XII
LISTA DE CUADROS	XVI
1 PLANTEAMIENTO DEL PROBLEMA	1
1.1 Descripción del problema	1
1.1.1 A nivel mundial	1
1.1.2 A nivel país	2
1.1.3 A nivel regional	4
1.2 Delimitación del problema	5
1.2.1 Espacial (geográfica)	5
1.2.2 Temporal	7
1.2.3 Temática y unidad de análisis	8
1.3 Formulación del problema	9
1.3.1 Problema general	9
1.3.2 Problemas específicos	9
1.4 Justificación e importancia	9
1.5 Limitaciones de la investigación	12
1.6 Objetivos de la investigación	13
1.6.1 Objetivo general	13
1.6.2 Objetivos específicos	13
2 MARCO TEÓRICO	14
2.1 Antecedentes	14
2.1.1 Investigaciones internacionales.	14
2.1.2 Investigaciones nacionales	24
2.2 Bases teóricas	29
2.2.1 Seguridad vial y prevención de accidentes de tránsito.	29
2.2.2 Accidentes de tránsito	34
2.3 Marco conceptual	46
2.3.1 Inteligencia artificial	46
2.3.2 Aprendizaje automático (<i>machine learning</i>)	47
2.3.3 Aprendizaje profundo (<i>deep learning</i>)	49
2.3.4 Redes neuronales convolucionales (CNN).	52
2.3.5 Detección de objetos	54
2.3.6 Familia YOLO: Evolución de YOLOv1 a YOLOv8	56
2.3.7 Aprendizaje por transferencia (<i>transfer learning</i>).	59
2.3.8 Aumento de datos (<i>data augmentation</i>)	60

2.3.9	Métricas de evaluación para detección de objetos y segmentación de instancias	61
2.3.10	Herramientas y <i>frameworks</i> de implementación	64
3	MARCO METODOLÓGICO	65
3.1	Método	65
3.2	Orientación	65
3.3	Enfoque.	66
3.4	Instrumento de recolección de datos	66
3.5	Tipo o alcance de la investigación	67
3.6	Nivel de investigación	67
3.7	Diseño de investigación	68
3.7.1	Propósito de la investigación.	68
3.7.2	Número de mediciones	69
3.7.3	Cronología de las observaciones	69
3.8	Estudio del diseño de investigación	69
3.9	Población y muestra.	70
3.9.1	Población	70
3.9.2	Muestra	71
3.10	Hipótesis	72
3.10.1	Hipótesis general	72
3.10.2	Hipótesis específicas	72
3.11	Operacionalización de variables	73
3.11.1	Variables.	73
3.11.2	Dimensiones	73
3.11.3	Indicadores.	74
3.12	Técnicas e instrumentos	76
3.12.1	Técnicas de investigación	76
3.12.2	Instrumentos de investigación	78
3.12.3	Herramientas de software utilizadas.	79
3.13	Desarrollo de trabajo de tesis	80
3.13.1	Recolección de datos	80
3.13.2	Características viales de la Avenida Arenales	80
3.13.3	Características viales de la Avenida 09 de diciembre	85
3.13.4	Definición de clases y etiquetado en Labelme	99
3.13.5	Entrenamiento del modelo.	102
3.13.6	Especificaciones del hardware de entrenamiento y despliegue del sistema	104
3.13.7	Resultados de la Fase 3: Optimización a resolución 1024.	106
3.13.8	Comparación multi-arquitectura: YOLOv8L-seg frente a la familia YOLO26 (nano, small, large)	114

3.13.9	Análisis de explicabilidad del modelo mediante Grad-CAM	118
3.13.10	Desarrollo del software QHAWAY ADAS	122
4	RESULTADOS	141
4.1	Análisis y discusión de resultados	141
4.1.1	Desempeño del modelo en detección y segmentación de instancias . .	141
4.1.2	Comparación con investigaciones previas en detección de elementos viales	141
4.1.3	Análisis de la estimación de distancias.	144
4.1.4	Análisis de la explicabilidad visual del modelo	144
4.1.5	Comparación con el modelo base de referencia	145
4.1.6	Posicionamiento de QHAWAY frente a sistemas ADAS comerciales . . .	145
4.2	Contrastación de hipótesis	146
4.2.1	Contrastación de hipótesis específica 01	146
4.2.2	Contrastación de hipótesis específica 02	155
4.2.3	Contrastación de hipótesis específica 03	161
5	CONCLUSIONES	169
5.1	Conclusiones de la investigación.	169
5.1.1	Conclusion 01	169
5.1.2	Conclusion 02	169
5.1.3	Conclusion 03	170
5.2	Limitaciones del estudio	170
5.3	Recomendaciones	173
5.4	Trabajos futuros	174
	REFERENCIAS BIBLIOGRÁFICAS	175
6	ANEXOS	188
6.1	Panel fotográfico	188
	Anexo 2: Matriz de operacionalización de variables	196

LISTA DE FIGURAS

1	Vista satelital y delimitación espacial de la zona de estudio	6
2	Cronograma de actividades de la investigación	8
3	Evolución cronológica de investigaciones sobre Deep Learning en detección de elementos viales	24
4	Evolución cronológica de investigaciones nacionales sobre seguridad vial y detección automatizada	28
5	Distribución de defunciones causadas por siniestros de tránsito a nivel mundial	29
6	Sistema-Visión Cero	30
7	Siniestralidad Vial en función del nivel de ingreso en 87-88 países	32
8	Resultados de la Encuesta Nacional de Programas Presupuestales sobre conductas viales	33
9	Road Safety Fundamentals: Concepts, Strategies, and Practices that Reduce Fatalities and Injuries on the Road	36
10	Portada del Manual de Dispositivos de Control del Tránsito Automotor	40
11	Patrón dimensional y señalización horizontal en el pavimento	42
12	Especificaciones dimensionales de la línea de cruce peatonal	42
13	Reductor de velocidad en el pavimento	44
14	Portada del Manual de Seguridad Vial del MTC	45
15	Sistema inteligente de transporte	46
16	Tipos de aprendizaje de la inteligencia artificial	48
17	Modelo matemático de una neurona artificial: entradas ponderadas, función de suma y función de activación no lineal	49
18	Esquema arquitectónico típico de una red neuronal convolucional	53
19	Detección de objetos	55
20	Evolución cronológica de la familia YOLO: de YOLOv1 a YOLO26	59
21	Representación del transfer learning	59
22	Entorno PyTorch	64
23	Punto de inicio de la recolección de datos: Intersección Av. Arenales y Av. Cusco	80
24	Señalización vertical en estado de deterioro – Av. Arenales	81
25	Estado del pavimento en la Av. Arenales – Tramo 1	81
26	Estado del pavimento en la Av. Arenales – Tramo 2	82
27	Talleres automotrices y estacionamiento informal – Av. Arenales	82
28	Medición del ancho de carril – Av. Arenales	83
29	Reductor de velocidad sin señalización vertical – Av. Arenales	83
30	Reductor de velocidad sin pintura ni señalización – Av. Arenales	84
31	Reductor de velocidad en la Av. Arenales	84
32	Placa de inauguración de la Av. 9 de Diciembre	85
33	Pasos peatonales en la Av. 9 de Diciembre	85
34	Reductor de velocidad en la Av. 9 de Diciembre	86
35	Reductores de velocidad con dimensiones normativas – Av. Cusco	87

36	Reductores de velocidad con dimensiones insuficientes – Av. Cusco	87
37	Reductores de velocidad destruidos por acción de terceros – Av. Cusco	88
38	Reductor de velocidad subdimensionado – frente a SENCICO	89
39	Baden sin señalización en la Av. Cusco	89
40	Entorno vial en las inmediaciones de Plaza Vea – Av. Cusco	90
41	Reductor de velocidad trapezoidal conforme a normativa – Av. Cusco	91
42	Señalización horizontal – línea central continua en la Av. Cusco	91
43	Pasos peatonales en la Av. Las Américas	92
44	Reductores de velocidad en la Av. Las Américas	92
45	Dinámica de circulación en el centro histórico de Ayacucho	93
46	Señalización horizontal en el centro histórico de Ayacucho	94
47	Capturas de Google Earth Pro – Av. Independencia	95
48	Capturas de Google Earth Pro – Inmediaciones de la UNSCH	95
49	Instalación del dispositivo de captura en el parabrisas del vehículo	96
50	Fotogramas extraídos de las videograbaciones en diferentes condiciones	97
51	Etiquetado de reductor de velocidad y vehículos en LabelMe	101
52	Etiquetado de escenas con alta densidad de objetos en LabelMe	101
53	Etiquetado de escenas con dispositivos de control y datos georreferenciados	101
54	Evolución de las funciones de pérdida y métricas de rendimiento durante la Fase 3	110
55	Curva Precisión-Exhaustividad para detección de cajas delimitadoras (<i>Box</i>) – mAP ₅₀ = 0.810	111
56	Curva Precisión-Exhaustividad para segmentación de instancias (<i>Mask</i>) – mAP ₅₀ = 0.778	111
57	Curva Precisión-Confianza para segmentación de instancias (<i>Mask</i>)	112
58	Curva F1-Confianza para segmentación de instancias (<i>Mask</i>)	112
59	Mapa de calor de activación – Av. Mariscal Cáceres (imagen 118)	118
60	Mapa de calor de activación – Jirón Libertad (imagen 146)	119
61	Mapa de calor de activación – Av. Las Américas (imagen 429)	119
62	Panel de explicabilidad Grad-CAM – Intersección con contraluz solar (imagen 82, 12 detecciones)	120
63	Panel de explicabilidad Grad-CAM – Plaza con arquitectura colonial (imagen 192, 17 detecciones)	121
64	Pantalla de bienvenida del software QHAWAY ADAS	125
65	Selector de escenarios de validación del software QHAWAY ADAS	125
66	Geometría del modelo de cámara Pinhole para estimación de distancia	127
67	Escenario 1: Detección de peatones en condiciones de iluminación nocturna	134
68	Escenario 1: Estimación simultánea de distancias a múltiples peatones	134
69	Escenario 1: Alerta de PELIGRO por niños cruzando intempestivamente	135
70	Escenario 2: Detección simultánea de múltiples categorías en Av. Cusco	136
71	Escenario 2: Peatón detectado detrás de un camión estacionado	136
72	Escenario 2: Peatón cruzando detrás de un camión	137

73	Escenario 2: Peatón cruzando distraído con celular	137
74	Escenario 3: Detección de semáforo en rojo en condiciones nocturnas	138
75	Escenario 3: Emisión de la alerta “Semáforo en rojo, deténgase”	139
76	Escenario 4: Detección de reductor de velocidad con banner de alerta	139
77	Escenario 4: Detección de reductor de velocidad a 3.6 metros	140
78	Escenario 4: Detección múltiple con reductor de velocidad y tráfico denso	140
79	Detección en Av. Naciones Unidas 110: Reductor de velocidad, persona y automóviles	147
80	Detección multi-clase en intersección: Paso peatonal, mototaxi, motocicleta y persona	147
81	Detección en Av. Cusco (Jr. Venezuela 140): Paso peatonal y peatones cruzando .	148
82	Detección en intersección con semáforo en rojo: Múltiples clases simultáneas .	148
83	Detección en Av. Las Américas: Bus de transporte, persona, automóvil y paso peatonal	149
84	Detección en Av. Mariscal Cáceres: Bus, semáforos en verde y múltiples pasos peatonales	150
85	Detección en Jr. 9 de Diciembre: Semáforo, mototaxis, motocicletas y peatones .	150
86	Detección en Av. San Paulo: Reductor de velocidad, camiones, automóviles y motocicleta	151
87	Detección en Plaza de Armas: Automóvil, motocicleta, personas y pasos peatonales	151
88	Detección en centro histórico: Escena congestionada con múltiples vehículos y peatones	152
89	Detección en intersección del centro histórico: Automóvil, motocicletas, camión, peatones y paso peatonal	152
90	Detección en cruce peatonal concurrido: Múltiples peatones y pasos peatonales	153
91	Pantalla de bienvenida del software integral QHAWAY ADAS	161
92	Selector de escenarios: gestión por tipo de riesgo vial	162
93	Gestión automatizada: panel de monitoreo en tiempo real con estadísticas en vivo — Escenario 1	163
94	Alerta de máxima prioridad: peatón cruzando intempestivamente — Escenario 1	164
95	Alerta de reductor de velocidad con distancia anticipada — Escenario 4	164
96	Alerta de precaución: peatón distraído con celular en zona de precaución — Escenario 2	165
97	Coordinación de respuestas: semáforo en rojo, peatones y alertas simultáneas — Escenario 3	166
98	Punto de inicio de la recolección de datos – Intersección Av. Arenales y Av. Cusco	188
99	Estado del pavimento con presencia de baches – Av. Arenales	188
100	Comercio de automotrices y estacionamientos informales – Av. Arenales	189
101	Reductor de velocidad improvisado sin señalización vertical – Av. Arenales . . .	189
102	Reductor de velocidad sin pintura ni señalización – Av. Arenales	190
103	Placa de inauguración y paso peatonal – Av. 9 de Diciembre	190

104	Reductor de velocidad destruido por acción de terceros – Av. Cusco	191
105	Baden sin señalización en la intersección Av. Cusco y Av. Arenales	191
106	Entorno vial en inmediaciones de Plaza Vea – Av. Cusco	192
107	Reductor de velocidad trapezoidal conforme a normativa – Av. Cusco	192
108	Paso peatonal frente al Hospital Regional – Av. Las Américas	193
109	Reductor de velocidad no normativo – Av. Las Américas	193

LISTA DE CUADROS

1	Antecedentes internacionales sobre detección de elementos viales con Deep Learning	23
2	Antecedentes nacionales sobre seguridad vial e inteligencia artificial	27
3	Población que cumple las reglas de seguridad vial	33
4	Metodología de investigación	71
5	Herramientas de software utilizadas en el desarrollo del sistema QHAWAY ADAS	79
6	Parámetros de calibración de la cámara de videograbación	96
7	Clases definidas para el modelo de segmentación de instancias	99
8	Parámetros de configuración del modelo SAM en LabelMe	100
9	Hiperparámetros de entrenamiento – Etapa 01	102
10	Hiperparámetros de entrenamiento – Etapa 02	103
11	Especificaciones del hardware de entrenamiento del modelo YOLOv8L-seg . . .	104
12	Especificaciones del hardware de recolección de datos en campo	105
13	Métricas de rendimiento del hardware durante la operación del sistema QHAWAY ADAS en tiempo real	106
14	Configuración de hiperparámetros de la Fase 3 (1024 px)	107
15	Métricas globales de la Fase 3	108
16	Rendimiento por clase — Fase 3	109
17	Progresión del entrenamiento en tres fases	113
18	Comparación con modelo base	113
19	Comparación multi-arquitectura sobre el conjunto de validación QHAWAY. Los modelos YOLO26 reportan el mejor punto de control en 300 épocas a 1024 px; YOLOv8L-seg reporta el mejor punto de control en la época 173.	115
20	Comparación del desempeño por clase entre YOLOv8L-seg (época 173) y YOLO26L-seg (época 179). Δ indica la ganancia (positiva) o pérdida (negativa) de YOLO26L-seg respecto a YOLOv8L-seg. Los valores en negrita indican el mayor mAP_{50} por clase.	116
21	Estudio de escalado de la familia YOLO26-seg: desempeño y contexto de despliegue recomendado	117
22	Stack tecnológico del software	123
23	Parámetros de calibración	128
24	Alturas de referencia por clase	129
25	Lógica del sistema de alertas	131
26	Parámetros prosódicos	132
27	Escenarios de validación	133
28	Resumen de detecciones por clase en las imágenes de validación — Hipótesis específica 1	153
29	Alturas de referencia por clase para la estimación de distancias monoculares .	156
30	Protocolos de respuesta según el tipo de elemento crítico detectado	157
31	Definición de zonas de riesgo vial del sistema QHAWAY ADAS	158

32	Correspondencia entre elementos detectados y avisos preventivos generados por el sistema QHAWAY ADAS	159
33	Correspondencia entre los componentes de la hipótesis específica 3 y la evidencia documentada	167
34	Matriz de consistencia	195
35	Matriz de operacionalización de variables	197

GLOSARIO

Backbone	Componente principal de la arquitectura de una red neuronal encargado de la extracción de características visuales a partir de la imagen de entrada.
Bounding Box	Rectángulo delimitador que localiza y enmarca un objeto detectado en una imagen, definido por las coordenadas (x_1, y_1, x_2, y_2) .
ByteTrack	Algoritmo de seguimiento multi-objeto que asigna identificadores persistentes a cada detección a lo largo de cuadros consecutivos de video.
Cooldown	Intervalo de tiempo mínimo entre la emisión de alertas consecutivas, diseñado para evitar la saturación auditiva del conductor.
Early Stopping	Técnica que detiene automáticamente el entrenamiento del modelo cuando la métrica de validación deja de mejorar, previniendo el sobreajuste.
Edge-TTS	Servicio de síntesis de voz neuronal de Microsoft Azure utilizado para generar las alertas auditivas con voz peruana en el software QHAWAY ADAS.
Filtro Temporal	Mecanismo que descarta detecciones que persisten menos de 2 cuadros consecutivos, reduciendo falsos positivos transitorios.
Inferencia	Proceso de aplicar un modelo entrenado a datos nuevos para obtener predicciones, ejecutado en cada cuadro de video por el software QHAWAY ADAS.
Kernel	Matriz numérica que se desliza sobre una imagen para extraer características durante la operación de convolución en redes neuronales.
Máscara de segmentación	Contorno a nivel de píxel que delimita la forma exacta de un objeto detectado, generado por la cabeza de segmentación del modelo YOLOv8L-seg.
Modelo Pinhole	Modelo geométrico de cámara que establece la relación proyectiva entre las dimensiones reales de un objeto y su representación en el plano de imagen.
Multithreading	Técnica de programación concurrente que permite ejecutar múltiples hilos simultáneamente, empleada en PyQt5 para separar la inferencia de la interfaz gráfica.
Overfitting	Sobreajuste del modelo a los datos de entrenamiento, donde memoriza patrones específicos perdiendo capacidad de generalización ante datos nuevos.
Recall	Exhaustividad del modelo: proporción de objetos reales presentes en la escena que fueron correctamente detectados por el sistema.
Señalización horizontal	Marcas viales pintadas sobre la calzada, como pasos peatonales, líneas de carril y flechas direccionales, detectadas por el modelo.
Track ID	Identificador numérico único asignado por ByteTrack a cada objeto detectado, que se mantiene persistente entre cuadros consecutivos de video.
Warm-up	Fase inicial de 3 iteraciones de inferencia ejecutadas al cargar el modelo para estabilizar la latencia de la GPU antes del procesamiento real.
YOLO	You Only Look Once, familia de algoritmos de detección de objetos en tiempo real que procesan la imagen completa en una única pasada de la red.

ACRÓNIMOS

ADAS	Advanced Driver Assistance System (Sistema Avanzado de Asistencia al Conductor), conjunto de tecnologías electrónicas que asisten al conductor en la conducción y el estacionamiento.
CNN	Convolutional Neural Network (Red Neuronal Convolucional), arquitectura de aprendizaje profundo especializada en el procesamiento de imágenes y video.
CSV	Comma-Separated Values (Valores Separados por Comas), formato de archivo utilizado para el registro histórico de detecciones y alertas del software QHAWAY ADAS.
CUDA	Compute Unified Device Architecture, plataforma de computación paralela de NVIDIA para la aceleración de inferencia en GPU.
FCW	Forward Collision Warning (Alerta de Colisión Frontal), sistema que advierte al conductor cuando existe riesgo de colisión con un objeto al frente del vehículo.
FOV	Field of View (Campo de Visión), ángulo de cobertura de la cámara utilizada para la captura de video, expresado en grados.
FPS	Frames Per Second (Cuadros Por Segundo), medida de velocidad de procesamiento del software QHAWAY ADAS en tiempo real.
GPU	Graphics Processing Unit (Unidad de Procesamiento Gráfico), acelerador de hardware para inferencia y entrenamiento de redes neuronales profundas.
IoU	Intersection over Union (Intersección sobre Unión), métrica para evaluar la superposición entre la predicción del modelo y la anotación de referencia.
mAP	mean Average Precision (Precisión Media Promedio), métrica estándar para evaluar el rendimiento global de modelos de detección de objetos.
MTC	Ministerio de Transportes y Comunicaciones del Perú, entidad que establece las normativas de señalización vial y dispositivos de control de tránsito.
NCAP	New Car Assessment Program (Programa de Evaluación de Vehículos Nuevos), protocolo de la NHTSA que define umbrales para sistemas FCW.
NHTSA	National Highway Traffic Safety Administration, agencia federal de los Estados Unidos que establece estándares de seguridad vehicular.
NMS	Non-Maximum Suppression (Supresión de No Máximos), algoritmo para eliminar detecciones redundantes en modelos de visión artificial.
SGD	Stochastic Gradient Descent (Descenso de Gradiente Estocástico), algoritmo de optimización empleado en el entrenamiento del modelo YOLOv8L-seg.
TTC	Time-To-Collision (Tiempo de Colisión), tiempo estimado para que un objeto en aproximación alcance la posición de la cámara.
TTS	Text-To-Speech (Texto a Voz), tecnología de síntesis de voz neuronal empleada para la emisión de alertas auditivas en el software QHAWAY ADAS.
YOLO	You Only Look Once, familia de algoritmos de detección de objetos en tiempo real que procesan la imagen completa en una única pasada de la red neuronal.

LISTA DE SÍMBOLOS

α	Ángulo desde el centro óptico de la cámara hasta el punto inferior de la caja delimitadora, utilizado en la estimación por punto de apoyo (<i>footpoint</i>).
c	Clase predicha por el modelo YOLOv8L-seg para cada objeto detectado, donde $c \in \{0, 1, \dots, 11\}$.
c_y	Coordenada vertical del centro óptico de la imagen, calculada como $c_y = h_{img}/2$.
D	Distancia estimada entre la cámara y el objeto detectado, calculada mediante el modelo Pinhole: $D = \frac{H_{real} \times f}{h_{bbox}}$.
Δt	Intervalo temporal entre cuadros consecutivos de video, utilizado para el cálculo de la velocidad relativa de aproximación.
η	Tasa de aprendizaje (<i>learning rate</i>) del optimizador SGD durante el entrenamiento del modelo YOLOv8L-seg.
f	Distancia focal de la cámara expresada en píxeles ($f = 950$ px), parámetro del modelo Pinhole.
$f(x)$	Función de activación en redes neuronales, como SiLU: $f(x) = x \cdot \sigma(x)$, empleada en YOLOv8.
$F1$	Medida F1-score, media armónica entre precisión y recall: $F1 = 2 \times \frac{P \times R}{P + R}$.
h_{bbox}	Altura de la caja delimitadora (<i>bounding box</i>) del objeto detectado en la imagen, en píxeles.
h_{cam}	Altura de montaje de la cámara respecto al suelo ($h_{cam} = 1,50$ m).
H_{real}	Altura real de referencia del objeto según su clase, utilizada en la ecuación del modelo Pinhole.
IoU	Intersección sobre Unión (<i>Intersection over Union</i>): $IoU = \frac{\text{área de intersección}}{\text{área de unión}}$.
K	Kernel o filtro convolucional utilizado en las capas de la arquitectura YOLOv8L-seg.
\mathcal{L}	Función de pérdida (<i>loss function</i>) compuesta por las pérdidas de caja, clasificación y segmentación.
mAP	Precisión media promedio (<i>mean Average Precision</i>), métrica principal de evaluación del modelo.
p	Puntaje de confianza de la detección, donde $p \in [0, 1]$. Umbral mínimo: $p \geq 0,50$.
P	Precisión (<i>Precision</i>) del modelo: proporción de detecciones correctas sobre el total de detecciones.
R	Recall o exhaustividad del modelo: proporción de objetos reales correctamente detectados.
$\sigma(x)$	Función sigmoide: $\sigma(x) = \frac{1}{1 + e^{-x}}$, utilizada en las cabezas de detección y clasificación.
θ_{tilt}	Ángulo de inclinación de la cámara respecto al eje horizontal ($\theta_{tilt} = 5^\circ$).
TTC	Tiempo de colisión (<i>Time-To-Collision</i>): $TTC = \frac{D}{v_{rel}}$, umbral FCW $\leq 2,4$ s.
v_{rel}	Velocidad relativa de aproximación entre el vehículo y el objeto detectado, en m/s.

TESIS

1. PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción del problema

1.1.1. A nivel mundial

La siniestralidad vial constituye una de las crisis de salud pública más devastadoras del siglo XXI, originada por una convergencia de factores humanos, infraestructurales y tecnológicos que interactúan de manera compleja en el entorno vial urbano. De acuerdo con el *Global Status Report on Road Safety 2023* de la Organización Mundial de la Salud (Organización Mundial de la Salud., 2024), aproximadamente 1.19 millones de personas fallecen anualmente como consecuencia de siniestros de tránsito, cifra que equivale a más de 3,200 muertes diarias. Si bien esta cifra representa una reducción del 5 % respecto a los 1.25 millones registrados en 2010, entre 20 y 50 millones de personas sufren traumatismos no mortales cada año, muchos de los cuales derivan en discapacidades permanentes. El 92 % de estas muertes ocurren en países de ingresos bajos y medios, donde las deficiencias de infraestructura vial, la ausencia de sistemas inteligentes de gestión del tráfico y la limitada fiscalización de las normas de tránsito configuran un entorno de riesgo estructural (Organización Mundial de la Salud., 2024).

Sin embargo, las estadísticas de mortalidad vial no son fenómenos aislados, sino la manifestación visible de una cadena causal donde confluyen múltiples factores de origen. A nivel global, el 74 % de los siniestros de tránsito son atribuibles a factores humanos que incluyen tanto las conductas imprudentes de conductores como las de peatones . La distribución geográfica de la mortalidad refleja profundas desigualdades: La Región Africana de la OMS presenta la tasa más elevada con 19.4 defunciones por cada 100,000 habitantes, siendo además la única región donde las muertes se incrementaron en un 17 % desde 2010, mientras que Europa registra la tasa más baja con aproximadamente 6.5 por cada 100,000 habitantes y la Región de las Américas se sitúa en 14.09 por cada 100,000 habitantes con 145,090 muertes en 2021 (Organización Panamericana de la Salud, 2025). Esta disparidad se explica no solo por la cantidad de vehículos, sino fundamentalmente por la calidad de la infraestructura vial, la señalización, la presencia de sistemas de monitoreo inteligente y el grado de cumplimiento normativo.

Los peatones representan uno de los segmentos más vulnerables de esta crisis, constituyendo el 23 % de las víctimas fatales a nivel global, equivalente a aproximadamente 274,000 muertes anuales (Organización Mundial de la Salud., 2024). No obstante, la vulnerabilidad peatonal no se limita a la condición de víctima pasiva; la evidencia científica documenta ampliamente que las conductas peatonales de riesgo constituyen un factor determinante en la ocurrencia de siniestros. El metaanálisis de Simmons et al. (2020), publicado en *Injury Prevention*, demostró que el uso de teléfonos celulares al cruzar la calle incrementa el tiempo de reacción entre un 50 % y 67 %, mientras que estudios

observacionales revelan que entre el 12 % y 45 % de los peatones se encuentran distraídos al momento de cruzar. El cruce indebido o *jaywalking* contribuye aproximadamente al 59 % de los choques vehículo-peatón según Ansariyar et al. (2025), y datos del *Fatality Analysis Reporting System* de Estados Unidos indican que el 60 % de las muertes peatonales urbanas ocurrieron fuera de los cruces señalizados. La ebriedad peatonal agrava el panorama: la National Highway Traffic Safety Administration (2019) reporta que el 32 % de los peatones fallecidos presentaban niveles de alcohol iguales o superiores a 0.08 g/dL, mientras que Dultz y Frangos (2013) encontraron que el 36 % de todas las muertes peatonales están asociadas al consumo de alcohol.

Sin embargo, estas conductas de riesgo no operan en el vacío: Se potencian exponencialmente cuando la infraestructura vial es deficiente, la señalización está deteriorada o ausente, y no existen sistemas tecnológicos de monitoreo y alerta. La revisión sistemática publicada en el *Journal of Road Safety* (Journal of Road Safety, 2025) establece que la ausencia de aceras, cruces señalizados e isletas de refugio eleva significativamente el riesgo de lesión y muerte peatonal. El 80 % de las carreteras del mundo no cumplen con los estándares de seguridad peatonal según la OMS e iRAP . La relación entre velocidad e impacto es determinante: a 30 km/h la probabilidad de muerte del peatón es aproximadamente del 10 %, mientras que a 50 km/h supera el 80 % (Organización Mundial de la Salud., 2024). El estudio de Quistberg et al. (2014), específicamente sobre Lima, Perú, demostró que las colisiones peatonales eran un 81 % menos probables cuando existían aceras a ambos lados de la vía (OR=0.19), que las barreras peatonales reducían el riesgo en un 89 % (OR=0.11), pero que la presencia de vendedores ambulantes obstruyendo las vías incrementaba el riesgo en un 182 % (OR=2.82) y los vehículos estacionados informalmente en un 267 % (OR=3.67). Este último hallazgo es particularmente relevante para ciudades latinoamericanas donde la informalidad comercial invade sistemáticamente el espacio peatonal y vehicular.

En paralelo, las conductas inapropiadas de los conductores constituyen el otro pilar causal de la siniestralidad. El exceso de velocidad, la conducción en estado de ebriedad, el desacato de señales de tránsito y semáforos, y las maniobras imprudentes son responsables de la mayor proporción de accidentes fatales a nivel mundial (Organización Mundial de la Salud., 2024). La ausencia de reductores de velocidad adecuados en zonas de alto riesgo, o peor aún, la proliferación de reductores improvisados construidos por las propias comunidades sin estándares técnicos, genera un doble riesgo: no cumplen su función preventiva efectiva y pueden constituir un peligro adicional para los usuarios de la vía. Estas problemáticas convergen en un ciclo que solo puede romperse mediante la implementación de sistemas inteligentes de monitoreo y gestión del tráfico basados en tecnologías de inteligencia artificial.

1.1.2. A nivel país

En el Perú, la siniestralidad vial refleja la convergencia de deficiencias estructurales profundas: infraestructura vial precaria, señalización deteriorada o inexistente,

informalidad en el transporte público, conductas imprudentes tanto de conductores como de peatones, y la ausencia casi total de sistemas tecnológicos inteligentes de gestión del tráfico. Los datos del Observatorio Nacional de Seguridad Vial (ONSV), basados en registros de la Policía Nacional del Perú, revelan que en 2023 se registraron 87,083 siniestros de tránsito, con 58,000 lesionados y 3,316 fallecidos, manteniéndose las víctimas fatales consistentemente por encima de 3,000 anuales desde 2018 (ONSV, 2025). En el acumulado del periodo 2016-2020, se registraron más de 420,000 accidentes, 14,000 muertes y 272,000 heridos (Defensoría del Pueblo, 2023). La tasa de mortalidad vial del Perú se estima en 13 por cada 100,000 habitantes, más del doble del promedio europeo (Organización Mundial de la Salud., 2024), con un costo económico estimado en 4.6 % del PIB según el Fondo Mundial para la Seguridad Vial (Global Road Safety Facility, 2018).

Las causas de esta crisis tienen raíces profundas en el comportamiento humano y la deficiencia institucional. La Defensoría del Pueblo estableció que el 74 % de los accidentes son causados por factores humanos (Defensoría del Pueblo, 2023). El ONSV desagregó las causas identificando: Imprudencia del conductor (28.2%), exceso de velocidad (26.6%), ebriedad del conductor (7.2%), imprudencia del peatón (3.2%) e invasión del carril contrario (2.4%). No obstante, estas cifras registradas por la policía subestiman significativamente la contribución peatonal, ya que según la Asociación de Víctimas de Accidentes de Tránsito, el 40 % de los accidentes fatales involucran responsabilidad directa del peatón (Asociación de Víctimas de Accidentes de Tránsito, 2025). Para el periodo enero a octubre de 2025, los datos preliminares del ONSV (Observatorio Nacional de Seguridad Vial, 2025) registran 72,873 siniestros a nivel nacional, con el exceso de velocidad (16,136 casos) como causa predominante, seguido de ebriedad del conductor, confirmando la persistencia de los factores humanos como detonantes principales.

La imprudencia peatonal en el Perú alcanza dimensiones alarmantes. Según el Instituto Nacional de Estadística e Informática (2021), el 60.9 % de la población urbana nunca o casi nunca cruza las vías por el cruceo peatonal, mientras que el 70.1 % no respeta los semáforos cuando están en rojo. Estas conductas se manifiestan de múltiples formas: Cruce temerario a media calle, caminata por la calzada vehicular, uso de teléfonos celulares y auriculares al cruzar, consumo de alcohol y baja percepción del riesgo. Los peatones constituyen el 25 % de las víctimas fatales a nivel nacional y el 50 % en zonas urbanas, con 2,327 peatones fallecidos entre 2021 y 2023 (Observatorio Nacional de Seguridad Vial, 2024). Los datos acumulados 2021-2025 del ONSV revelan que los peatones representan el 27.2 % de los fallecidos totales (Observatorio Nacional de Seguridad Vial, 2025). El estudio de Quistberg et al. (2015) estableció que los peatones comprendían el 78 % de las muertes por tránsito en el Perú, una de las proporciones más elevadas a nivel mundial.

Sin embargo, la imprudencia peatonal no puede desligarse del entorno urbano que la propicia. Las calles peruanas presentan un nivel de informalidad que convierte los espacios públicos en escenarios de riesgo permanente. Los vendedores ambulantes invaden sistemáticamente aceras, cruceos peatonales y bermas, obligando a los peatones a

desplazarse por la calzada vehicular. Quistberg et al. (2014) demostraron cuantitativamente que la presencia de comercio ambulante incrementa el riesgo de colisión peatonal en un 182 %. El transporte informal, particularmente mototaxis y motos lineales que operan sin rutas establecidas ni regulación efectiva, genera un caos vehicular que incrementa exponencialmente los conflictos entre vehículos y peatones. Los vehículos estacionados en doble fila, sobre las veredas y en zonas prohibidas obstruyen la visibilidad y reducen el espacio de circulación. Esta informalidad generalizada anula la función de la señalización horizontal existente, cuyos cruceros peatonales y líneas de carril quedan ocultos bajo el comercio ambulante, los vehículos mal estacionados y el deterioro por falta de mantenimiento.

1.1.3. A nivel regional

En el departamento de Ayacucho, la problemática vial concentra y amplifica todos los factores causales identificados a nivel mundial y nacional, configurando un escenario de riesgo particularmente crítico. Los datos históricos del ONSV (ONSV, 2025) revelan que Ayacucho registró 549 siniestros de tránsito en 2024, un incremento del 32.3 % respecto a los 415 de 2023. Paralelamente, los datos del INEI para el tercer trimestre de 2024 registraron 207 accidentes, representando un incremento del 97.1 % respecto al mismo periodo del año anterior, con 202 personas lesionadas (Instituto Nacional de Estadística e Informática, 2024). Para el periodo enero a octubre de 2025, los datos preliminares del ONSV registran 586 siniestros, proyectando que la cifra anual superará significativamente la de 2024 (Observatorio Nacional de Seguridad Vial, 2025). Los tipos predominantes son choques (184), despistes (94), atropellos (82) y choques y fuga (64), con el exceso de velocidad (75 casos), imprudencia del conductor (27) y ebriedad (23) como causas principales.

El análisis detallado de la base de datos de siniestros fatales del ONSV para Ayacucho (2021-2023) (Observatorio Nacional de Seguridad Vial, 2025) revela 116 siniestros con consecuencias fatales que dejaron 182 fallecidos y 269 lesionados. La evolución interanual muestra una tendencia creciente sostenida: 33 en 2021, 39 en 2022 y 44 en 2023. La provincia de Huamanga concentra el 49.1 % de los siniestros fatales (57 de 116), validando la pertinencia de las avenidas Arenales y Cusco como zona de estudio. Las causas específicas revelan que el exceso de velocidad (24.1 %) y la conducción en estado de ebriedad (19.8 %) representan conjuntamente el 44 % de las causas fatales. La fatiga y cansancio (9 casos), la invasión de carril (6) y las fallas mecánicas de frenos (6) complementan el perfil causal. La imprudencia del peatón se registra formalmente en 4 casos (3.4 %), aunque esta cifra subestima la contribución peatonal real considerando las limitaciones del registro policial y los hallazgos de estudios nacionales que atribuyen hasta el 40 % de responsabilidad al peatón (Asociación de Víctimas de Accidentes de Tránsito, 2025).

Un hallazgo de particular relevancia para la presente investigación emerge del análisis de las condiciones de señalización en los siniestros fatales de Ayacucho. Los datos del ONSV (Observatorio Nacional de Seguridad Vial, 2025) revelan que, de los siniestros fatales con registro de señalización, el 80 % carecían de señalización horizontal (40 de 50 casos

registrados), mientras que el 50.9% ocurrieron en vías con características de curva y el 62% en vías con perfil longitudinal inclinado. Este dato es contundente: la ausencia de líneas de carril, cruces peatonales pintados, flechas direccionales y demarcaciones en el pavimento está directamente asociada a la ocurrencia de siniestros fatales. En el ámbito urbano de Huamanga, la situación se replica con agravantes: la señalización horizontal en las avenidas principales se encuentra deteriorada por el desgaste vehicular, las condiciones climáticas y la nula reposición, mientras que los cruces peatonales son invisibles en numerosas intersecciones. Los semáforos presentan desprogramación crónica y deterioro técnico, como lo documentó el diagnóstico del MTC-Promovilidad de marzo de 2023 que identificó zonas con alta congestión afectadas por la falta de señalización y deterioro de equipos de semaforización en cinco ejes viales clave (MTC-Promovilidad, 2023). El Plan de Desarrollo Urbano de Ayacucho confirma que la ciudad carece de señalización vial adecuada, con solo algunas señales dispersas que requieren urgentemente un Plan de Señalización integral (Equipo técnico PDU Ayacucho, 2020).

La problemática de los reductores de velocidad improvisados merece especial atención en el contexto ayacuchano. Ante la ausencia de rompemuelleres normados por las autoridades competentes, las comunidades vecinales de zonas residenciales, escolares y periurbanas construyen reductores artesanales utilizando montículos de concreto sin especificaciones técnicas. Estos dispositivos improvisados no cumplen con las dimensiones, geometría, señalización ni materiales establecidos en el Manual de Dispositivos de Control del Tránsito Automotor del MTC, generando un doble riesgo: por un lado, no reducen efectivamente la velocidad de forma gradual y segura; por otro, constituyen obstáculos peligrosos que provocan maniobras evasivas súbitas, daños vehiculares y caídas de motociclistas. La detección y clasificación automatizada de estos elementos mediante redes neuronales convolucionales permitiría tanto su inventariado como la generación de alertas preventivas para los conductores.

La convergencia de peatones imprudentes, calles invadidas por la informalidad, señalización deteriorada, semáforos deficientes, reductores improvisados y conductores irresponsables en un contexto geográfico extremo fundamenta la necesidad imperiosa de desarrollar un modelo inteligente adaptativo basado en redes neuronales convolucionales que permita detectar, clasificar y monitorear en tiempo real la señalización horizontal, los semáforos, los reductores de velocidad, los peatones y los vehículos en las avenidas Arenales y Cusco, generando alertas preventivas y contribuyendo a la gestión automatizada de la seguridad vial en la ciudad de Ayacucho.

1.2. Delimitación del problema

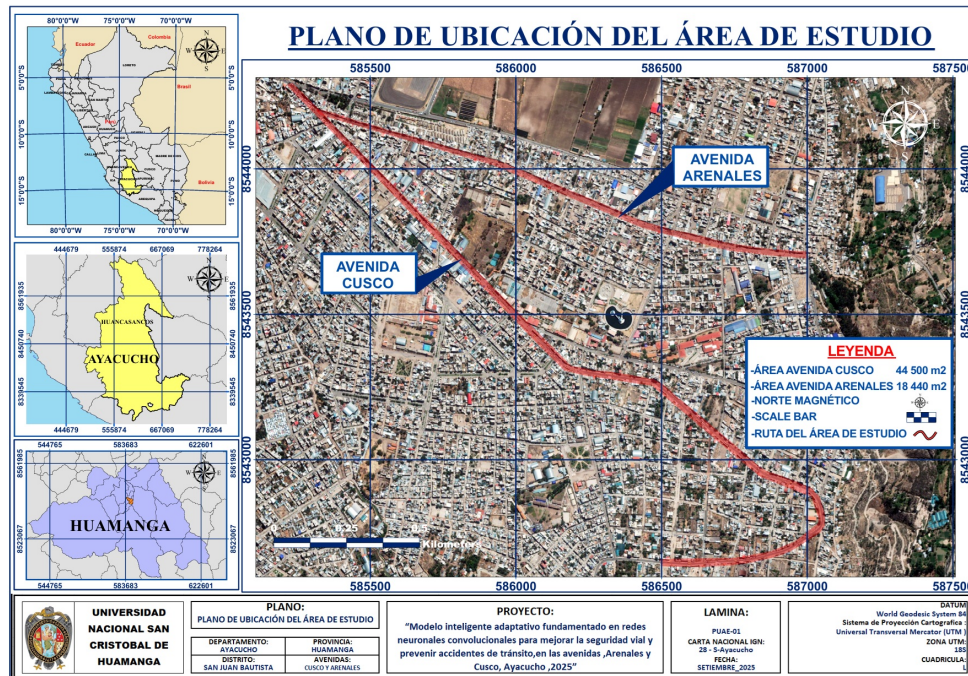
1.2.1. Espacial (geográfica)

El área de investigación se encuentra localizada en el distrito de San Juan Bautista, provincia de Huamanga, departamento de Ayacucho, ubicado estratégicamente en la sierra sur del Perú. La zona de estudio comprende dos avenidas principales que constituyen corredores

viales de alta confluencia vehicular y peatonal, con una extensión total de **62,940 m²** de superficie vial analizada.

Figura 1

Vista satelital y delimitación espacial de la zona de estudio



Nota. Elaborado con el software ArcGIS Pro.

El ámbito espacial de la investigación está definido por las siguientes vías urbanas:

- **Avenida Arenales:** Vía arterial con una superficie de 18,440 m², caracterizada por tráfico mixto (vehículos livianos, transporte público, motocicletas y peatones) y presencia de señalización horizontal, semáforos y reductores de velocidad en condiciones variables.
- **Avenida Cusco:** Vía principal con una superficie de 44,500 m², presenta alto flujo vehicular y peatonal, intersecciones críticas, señalización horizontal y dispositivos de control de tránsito.

La selección de estas avenidas se fundamenta en los siguientes criterios técnicos:

- Representatividad de elementos viales:** Las avenidas Arenales y Cusco presentan diversidad de componentes de seguridad vial (señalizaciones horizontales, semáforos, reductores de velocidad, cruces peatonales, peatones y vehículos), lo que permite obtener un dataset representativo para el entrenamiento del modelo.
- Variabilidad de condiciones viales:** Las características de ambas avenidas incluyen zonas comerciales, residenciales e intersecciones de alto tráfico, con diferentes configuraciones de calzada y elementos de control, lo que garantiza la robustez del modelo ante diversos escenarios de operación.
- Problemática de seguridad vial identificada:** Estas zonas presentan deficiencias en señalización vial, reportándose que los usuarios no respetan los semáforos en rojo y

muy pocas veces cruzan por el cruce peatonal.

- d) **Accesibilidad para recolección de datos:** Las avenidas permiten la captura sistemática de imágenes mediante cámara Android Pro en recorridos peatonales y vehiculares, garantizando la obtención de datos en diferentes horarios y condiciones de iluminación.

La georreferenciación de la zona de estudio se realizó mediante el sistema de coordenadas UTM Zona 18S, datum WGS84, permitiendo la ubicación precisa de los elementos viales detectados por el modelo.

1.2.2. Temporal

El horizonte temporal del estudio abarca un periodo de cuatro meses, desde octubre de 2025 hasta enero de 2026, intervalo en el cual se realizó la captura sistemática de imágenes de las avenidas Arenales y Cusco, el desarrollo y entrenamiento del modelo de detección basado en redes neuronales convolucionales, así como la validación del sistema de alertas preventivas para la seguridad vial.

Este periodo temporal se ha establecido considerando las siguientes etapas:

- **Octubre 2025:** Recolección de datos mediante recorridos sistemáticos en las avenidas Arenales (18,440 m²) y Cusco (44,500 m²). Captura de aproximadamente 3,500 imágenes de elementos viales (señalizaciones horizontales, semáforos, reductores de velocidad, peatones y vehículos) mediante cámara Android Pro. Registro de condiciones variables de iluminación y tráfico en diferentes horarios.
- **Noviembre 2025:** Etiquetado del dataset mediante herramientas especializadas (Labelme/Roboflow). Preprocesamiento y aumento de datos para el entrenamiento del modelo. Desarrollo de la arquitectura del modelo de detección basado en YOLOv11. Configuración del entorno de desarrollo en Python con librerías especializadas (Ultralytics, OpenCV, PyTorch).
- **Diciembre 2025:** Entrenamiento del modelo con el dataset etiquetado. Optimización de hiperparámetros mediante validación cruzada. Evaluación de métricas de rendimiento (mAP@0.5, precisión, recall, F1-Score). Pruebas de adaptabilidad del modelo bajo diferentes condiciones de iluminación y oclusión.
- **Enero 2026:** Validación del sistema de alertas preventivas. Análisis de tiempos de detección y generación de alertas. Evaluación de efectividad de clasificación por clase de elemento vial. Documentación de resultados y elaboración del informe final de tesis.

Figura 2
Cronograma de actividades de la investigación



Nota. Elaboración propia.

1.2.3. Temática y unidad de análisis

- **Temática:** La temática de esta investigación se circunscribe al campo de la ingeniería de transportes con enfoque en seguridad vial y aplicación de inteligencia artificial. El estudio se centra en la implementación de técnicas de aprendizaje profundo (Deep Learning), específicamente redes neuronales convolucionales mediante las arquitecturas YOLOv8 y YOLO26, para el desarrollo de un modelo inteligente adaptativo de detección, clasificación y segmentación de instancias de 12 clases de elementos viales en tiempo real. La investigación integra conocimientos de visión artificial, procesamiento de imágenes y normativa de señalización vial peruana, orientados a mejorar la seguridad vial y prevenir accidentes de tránsito mediante la generación de alertas preventivas en las avenidas Arenales y Cusco del distrito de San Juan Bautista, Ayacucho
- **Unidad de análisis:** La unidad de análisis de esta investigación está constituida por los elementos de seguridad vial detectados en imágenes capturadas en las avenidas Arenales (18,440 m²) y Cusco (44,500 m²). El modelo detecta, clasifica y segmenta 12 clases de elementos viales, organizados en las siguientes categorías:

 - 1) **Señalizaciones horizontales (3 clases):** Paso peatonal, línea recta y línea recta y derecha; clasificadas según el Manual de Dispositivos de Control del Tránsito

Automotor.

- ii) **Reductores de velocidad (1 clase):** Elementos físicos de moderación del tráfico instalados en la calzada para reducir la velocidad vehicular en zonas críticas.
- iii) **Semáforos (2 clases):** Semáforo en verde y semáforo en rojo; detectando el estado operativo del dispositivo de control de tránsito.
- iv) **Peatones (1 clase):** Persona; usuarios vulnerables del sistema vial detectados en cruces peatonales, aceras y zonas de circulación peatonal.
- v) **Vehículos (5 clases):** Motocicleta, automóvil particular, camión, bus de transporte y mototaxi; detectados para la evaluación del flujo vehicular y la generación de alertas preventivas.

1.3. Formulación del problema

1.3.1. Problema general

¿De qué manera el modelo inteligente adaptativo, fundamentado en redes neuronales convolucionales, permite detectar y clasificar en tiempo real los elementos críticos de la vía pública usuarios vulnerables, vehículos, infraestructura vial informal y dispositivos de control del tránsito y generar alertas tempranas orientadas a la prevención de accidentes de tránsito en las avenidas Arenales y Cusco, Ayacucho, 2025?

1.3.2. Problemas específicos

1. ¿Cómo la arquitectura y el entrenamiento del modelo inteligente adaptativo, fundamentado en redes neuronales convolucionales, permite detectar y segmentar las 12 clases de elementos críticos de la vía pública personas, vehículos, reductores de velocidad, señalizaciones horizontales y semáforos bajo condiciones variables de iluminación, perspectiva y densidad de tráfico del entorno andino informal de Ayacucho?
2. ¿En qué medida el algoritmo del modelo inteligente adaptativo permite estimar en tiempo real la distancia a los elementos críticos detectados, clasificar zonas de riesgo vial y generar alertas de voz preventivas en español para el conductor en las avenidas Arenales y Cusco de Ayacucho?
3. ¿De qué manera la implementación del sistema QHAWAY ADAS permite gestionar automáticamente la detección de elementos críticos de la vía pública, generar alertas inteligentes con síntesis de voz, registrar datos históricos de detección y coordinar respuestas de emergencia bajo condiciones reales de conducción en las avenidas Arenales y Cusco, Ayacucho, 2025?

1.4. Justificación e importancia

- **Conveniencia:** La presente investigación responde a una necesidad concreta del entorno vial ayacuchano, dado que las avenidas Arenales y Cusco representan nodos de alta confluencia vehicular y peatonal donde los incidentes de tránsito ocurren con

frecuencia considerable. El desarrollo de un sistema inteligente fundamentado en segmentación de instancias mediante arquitecturas de aprendizaje profundo (YOLOv8 y YOLO26) resulta viable gracias al acceso a dispositivos de adquisición de imágenes de costo accesible y a frameworks de visión artificial de distribución libre, permitiendo su implementación sin requerir infraestructura especializada de alto costo.

La selección estratégica de estas arterias viales posibilita la obtención de muestras representativas de los elementos que integran el escenario de estudio, abarcando las 12 categorías de detección y segmentación.

- **Relevancia social:** El estudio confronta una problemática de elevado impacto social en Ayacucho, ciudad donde las colisiones y atropellos vehiculares constituyen una amenaza persistente para la integridad de transeúntes y conductores, particularmente en cruces de elevada concentración de tráfico. De acuerdo con el Plan de Desarrollo Urbano de Ayacucho, los usuarios viales no respetan la señalización semafórica en fase roja y omiten el uso del paso peatonal, lo que evidencia la urgencia de implementar mecanismos inteligentes de prevención y alerta.

El sistema propuesto, basado en segmentación de instancias con modelos YOLOv8 y YOLO26, contribuirá de manera significativa a la mitigación de riesgos viales a través de la identificación automática y la delimitación precisa (a nivel de píxel) de 12 categorías de elementos viales críticos, junto con la emisión de alertas preventivas contextualizadas. Los beneficiarios directos comprenden a los miles de peatones, conductores de vehículos particulares, usuarios del transporte público, motociclistas y operadores de mototaxi que circulan diariamente por estas vías, con especial atención a grupos vulnerables como estudiantes, adultos mayores y personas con movilidad reducida.

- **Aplicaciones prácticas:** El sistema desarrollado ofrece aplicabilidad directa en la gestión inteligente de la movilidad urbana ayacuchana. La arquitectura de segmentación de instancias, implementada mediante YOLOv8L-seg y YOLO26, posibilita la identificación y delineación precisa en tiempo real de las 12 categorías de elementos viales, proporcionando no solo la localización espacial de cada objeto sino también su contorno exacto mediante máscaras de segmentación, lo cual supera las capacidades de los sistemas convencionales basados únicamente en cajas delimitadoras.

El software resultante facilitará la generación automatizada de estadísticas de detección por categoría, el análisis comparativo de rendimiento entre diferentes arquitecturas de segmentación (YOLOv8 versus YOLO26) bajo condiciones operativas variables, y la producción de informes sobre la efectividad del módulo de alertas preventivas. La metodología diseñada posee potencial de replicabilidad en otras intersecciones críticas del sistema vial ayacuchano, proyectando su alcance hacia una red integral de monitoreo de seguridad vial urbana. Complementariamente, los organismos municipales podrán emplear los resultados para diagnosticar deficiencias

en la señalización existente y establecer prioridades de intervención para la mejora de la infraestructura vial.

- **Unidad metodológica:** La investigación aporta un marco metodológico innovador que articula técnicas avanzadas de visión artificial —específicamente segmentación de instancias mediante las arquitecturas YOLOv8 y YOLO26— con un enfoque de observación sistemática no participante, configurando un protocolo estandarizado para la detección, clasificación y segmentación automática de 12 categorías de elementos de seguridad vial. El proceso de entrenamiento progresivo en tres fases (transfer learning desde COCO, fine-tuning a 640px y optimización a 1024px), junto con la estrategia de aumento de datos mediante 22 técnicas especializadas, constituye una contribución metodológica replicable y adaptable a contextos urbanos con características análogas (Goodfellow, 2016).
- **Valor teórico:** Desde el plano teórico, esta investigación contribuye a la expansión del conocimiento en la convergencia entre inteligencia artificial, ingeniería de transportes y seguridad vial urbana. El desarrollo de un sistema de segmentación de instancias, entrenado específicamente para la identificación de 12 categorías de elementos viales en el contexto peruano, aporta evidencia original sobre la aplicabilidad de arquitecturas de aprendizaje profundo en ciudades intermedias de economías en desarrollo (LeCun et al., 2015a).

Adicionalmente, la investigación contribuye al desarrollo de marcos conceptuales sobre la implementación de tecnologías emergentes en infraestructuras urbanas con recursos limitados, demostrando cómo la integración de dispositivos de captura accesibles, algoritmos de segmentación de instancias de código abierto y estrategias de entrenamiento progresivo puede generar soluciones efectivas y sostenibles para problemáticas complejas de seguridad vial en contextos urbanos de países en vías de desarrollo.

1.5. Limitaciones de la investigación

Si bien la presente investigación emplea metodologías avanzadas de aprendizaje profundo para la segmentación de elementos viales, es pertinente reconocer las siguientes limitaciones inherentes al desarrollo del estudio:

- **Limitaciones tecnológicas:** Las restricciones presupuestarias impiden el despliegue de un sistema integral de monitoreo con hardware embebido especializado (como dispositivos de inferencia en el borde, GPU dedicadas o servidores de procesamiento de alto rendimiento) ni la integración directa con infraestructura semafórica existente. En consecuencia, el sistema se desarrolla como un prototipo funcional cuyo procesamiento se ejecuta en un entorno controlado de escritorio, lo que limita la evaluación de su desempeño bajo condiciones de operación continua y prolongada en escenarios reales de tráfico.
- **Limitaciones en el instrumento de recolección de datos:** Si bien se emplearon múltiples fuentes de adquisición de imágenes (capturas terrestres mediante cámara Android Pro e imágenes satelitales y de perspectiva aérea a través de Google Earth Pro), cada una presenta restricciones particulares. Las capturas terrestres están sujetas a limitaciones en estabilización óptica, velocidad de obturación frente a objetos en desplazamiento rápido y capacidad de aproximación para registrar elementos viales distantes. Por su parte, las imágenes obtenidas de Google Earth Pro presentan variaciones en resolución y actualización temporal.
- **Complejidad del entorno vial:** La dinámica vehicular y peatonal en las avenidas Arenales y Cusco comprende múltiples situaciones que se manifiestan de forma simultánea: Adelantamientos irregulares, estacionamiento en doble fila, motocicletas y mototaxis circulando entre carriles, presencia de comercio ambulatorio y aglomeraciones peatonales en horarios punta. Cada uno de estos microescenarios posee particularidades que podrían no estar completamente representadas en el conjunto de datos de entrenamiento, lo cual podría incidir en la capacidad de generalización del modelo ante situaciones no contempladas durante su fase de aprendizaje.
- **Limitaciones en el proceso de anotación:** El etiquetado del conjunto de datos para la segmentación de las categorías de elementos viales, aunque asistido por herramientas de anotación semiautomática con modelos de segmentación auxiliares, permanece sujeto a criterios subjetivos del anotador, particularmente en la delineación de contornos de objetos parcialmente ocluidos o con geometrías irregulares. Las posibles inconsistencias en los criterios de clasificación y la delimitación de fronteras entre instancias superpuestas pueden introducir ruido en el proceso de entrenamiento, repercutiendo en las métricas de evaluación del sistema.
- **Limitaciones de representatividad temporal:** La adquisición de imágenes se efectuó durante el periodo comprendido entre octubre de 2025 y enero de 2026, intervalo que podría no capturar la totalidad de las variaciones estacionales del flujo vehicular y

peatonal, ni situaciones atípicas como festividades locales, manifestaciones cívicas o fenómenos climáticos extremos que modifiquen los patrones habituales de tránsito en la zona de estudio. Esta ventana temporal, si bien abarca la temporada de mayor actividad pluvial, no contempla los meses de menor precipitación ni periodos vacacionales que alteran sustancialmente la densidad vehicular.

1.6. Objetivos de la investigación

1.6.1. Objetivo general

Desarrollar e implementar el modelo inteligente adaptativo QHAWAY, fundamentado en redes neuronales convolucionales, para la detección y clasificación en tiempo real de 12 clases de elementos críticos de la vía pública (usuarios vulnerables, vehículos, infraestructura vial informal y dispositivos de control del tránsito) y la generación de alertas tempranas orientadas a la prevención de accidentes de tránsito en las avenidas Arenales y Cusco, Ayacucho, 2025.

1.6.2. Objetivos específicos

1. Desarrollar la arquitectura y entrenamiento del modelo inteligente adaptativo mediante una estrategia progresiva en tres fases de resolución creciente (640, 800 y 1024 px) sobre un conjunto de datos híbrido de 25,602 imágenes con 127,525 instancias anotadas en 12 clases de elementos críticos de la vía pública, evaluando el desempeño con las métricas de precisión media (mAP_{50}) de detección y segmentación por clase y de forma agregada.
2. Implementar el algoritmo del modelo inteligente adaptativo integrando la estimación de distancias mediante el modelo geométrico de cámara monocular, el seguimiento de múltiples objetos en movimiento, la clasificación de zonas de riesgo vial (zona de peligro: Distancia menor a 3 m; zona de precaución: entre 3 m y 7 m) y la generación de alertas de voz preventivas en español peruano, validando la exactitud del estimador de distancias mediante referencias métricas independientes del entorno de estudio.
3. Validar la implementación del sistema QHAWAY ADAS mediante pruebas operacionales en cuatro escenarios de riesgo vial (niños en la vía, peatones ocultos o distraídos, semáforos y reductores de velocidad) bajo condiciones reales de conducción en las avenidas de estudio, evaluando la velocidad de procesamiento en tiempo real, la coherencia semántica del modelo mediante análisis de explicabilidad visual y el registro histórico de detecciones y alertas para análisis posterior.

2. MARCO TEÓRICO

2.1. Antecedentes

2.1.1. Investigaciones internacionales

A continuación, se exponen las diferentes investigaciones internacionales que detallan cómo las redes neuronales convolucionales y los modelos de detección de objetos han sido aplicados en estudios orientados a la detección de peatones, vehículos, señalizaciones horizontales, semáforos, reductores de velocidad y cruces peatonales, así como su implementación en sistemas embebidos y el desarrollo de aplicaciones de seguridad vial en tiempo real. El Cuadro 1 presenta de manera sintética los hallazgos más representativos de esta selección.

Bao et al. (2026) estructuraron EMD-YOLOv8, un algoritmo de detección de peatones en la vía pública fundamentado en mejoras arquitectónicas sobre YOLOv8, integrando restauración de color Multi-Scale Retinex, un bloque residual mejorado, una red de recalibración espacial multi-escala y una cabeza de detección especializada en objetos pequeños. El modelo redujo los parámetros en un 47.3 % respecto a YOLOv8s, incrementando simultáneamente la precisión, el recall, el mAP50 y el mAP50-95 en 2.2 %, 5.7 %, 7.5 % y 4.9 % respectivamente, abordando eficazmente la detección de peatones en escenarios de baja luminosidad y alta densidad peatonal, elementos críticos para la seguridad vial urbana.

Unar et al. (2026) configuraron un framework de atención visual impulsado por inteligencia artificial para la detección y priorización dinámica de la interacción vehículo-peatón en sistemas inteligentes de transporte (ITS). Mediante la incorporación de mecanismos de atención contextual en un modelo de aprendizaje profundo, el enfoque optimiza el reconocimiento de las relaciones dinámicas entre vehículos y peatones en escenarios urbanos complejos con oclusiones y congestión. Los resultados experimentales sobre dos datasets de referencia evidenciaron un desempeño superior frente a métodos convencionales, garantizando eficiencia computacional y escalabilidad para el despliegue a gran escala en infraestructuras de transporte inteligente.

Zhu et al. (2026) propusieron DFEM-Net, una red de extracción dinámica adaptativa de características para la detección simultánea de peatones y vehículos en imágenes viales urbanas. La arquitectura incorpora un módulo de fusión multi-escala (Scalseq) que ajusta dinámicamente la forma y el tamaño de los kernels convolucionales, junto con un módulo de codificación triple de características (Zoomcat) que refuerza la detección de objetos pequeños y densamente superpuestos. Los resultados experimentales registraron una mejora de 7.8 % en precisión y 2.9 % en mAP@0.5, contribuyendo a la gestión del tráfico urbano y a los sistemas inteligentes de transporte.

D. Du et al. (2026) articularon MLE-YOLO, un framework multimodal ligero construido sobre YOLOv11 y optimizado para la detección robusta de vehículos y peatones bajo condiciones climáticas adversas en escenarios de conducción autónoma. La arquitectura integra cuatro

innovaciones clave: un módulo transformador parcial multi-etapa (M-SPTM), una red de agregación mixta (MANet), un módulo de submuestreo ligero (LDM) y una cabeza de detección eficiente (ELDH). Los experimentos demostraron mejoras del 3.6 % en F1 y 3.0 % en mAP con una reducción del 15.8 % en parámetros, validando su viabilidad como solución ligera para la percepción en tiempo real dentro de los sistemas de transporte autónomo.

Dai et al. (2026) concibieron DCExYOLO, un método orientado a sistemas inteligentes de transporte que articula el realce de imágenes por multi-exposición con la arquitectura YOLO para la detección de peatones en tiempo real bajo escenarios de conducción nocturna. Los autores perfeccionaron el algoritmo Zero-DCE++ para generar imágenes realzadas en múltiples niveles de exposición, fusionándolas con la imagen original como entrada al detector. Los experimentos sobre múltiples conjuntos de datos evidenciaron que DCExYOLO reduce significativamente la tasa logarítmica promedio de fallos (MR^2) respecto a la línea base de YOLO, contribuyendo al fortalecimiento de la seguridad vial mediante la detección robusta de usuarios vulnerables en entornos de iluminación adversa.

Ma et al. (2026) elaboraron un sistema asistente de reconocimiento de semáforos con retroalimentación de audio multilingüe (indonesio, mandarín e inglés) basado en YOLOv12, diseñado específicamente para conductores con deficiencia de visión cromática (daltonismo) dentro del marco de los sistemas inteligentes de transporte (ITS). El enfoque emplea una estrategia de inferencia por posición espacial que reconoce el estado del semáforo mediante la posición relativa de la luz activa en lugar de depender exclusivamente de la información cromática, alcanzando una confianza promedio de detección de 0.73 y una máxima de 0.95 con una latencia de 0.214 s en CPU, contribuyendo a la accesibilidad e inclusión en la seguridad vial.

M. Chaman et al. (2026) consolidaron un benchmarking exhaustivo de las arquitecturas YOLOv8 a YOLOv12 para la detección de objetos en tiempo real en sistemas híbridos ADAS e ITS, evaluando su idoneidad en subsistemas de reconocimiento de señales de tránsito (TSR), detección de peatones (PDS), detección de vehículos (VDS) y asistencia de velocidad inteligente (ISA). Un dataset unificado de 42,000 imágenes anotadas con 16 clases relacionadas al tránsito fue construido integrando BDD100K, CCTSDB y muestras propias. Los resultados posicionaron a YOLOv12 como el modelo con mejor balance entre precisión (97.3 %), recall (96.2 %), mAP50-95 (82.2 %) y velocidad de inferencia (70–73 FPS).

Zhumadillayeva et al. (2026) integraron un framework híbrido YOLO–CNN–LSTM con sensores IoT multimodales y modelado de gemelo digital para el monitoreo en tiempo real de la infraestructura vial, enfocándose en la detección consciente del contexto de grietas, baches, marcas deterioradas y superficies cubiertas de nieve. La validación experimental registró una precisión de 96.26 %, sensibilidad de 95.52 % y especificidad de 95.73 % con una eficiencia temporal de 11.02 ms, demostrando una solución fiable y escalable para la gestión inteligente del mantenimiento de la infraestructura de transporte.

Kamjula et al. (2025) realizaron un análisis comparativo del desempeño de los modelos YOLOv5 y YOLOv8 para la detección de baches y señales de tráfico en carreteras

de la India, donde las condiciones viales heterogéneas y los elevados volúmenes de tráfico representan desafíos significativos para la seguridad. Los autores entrenaron ambas arquitecturas con un dataset extenso de carreteras indias y señales de tráfico recopilado de Kaggle y Roboflow, orientando el sistema hacia su integración en plataformas avanzadas de asistencia al conductor (ADAS). Los resultados evidenciaron que YOLOv5 obtuvo un rendimiento superior frente a YOLOv8 bajo las métricas y datasets utilizados, contribuyendo al desarrollo de sistemas de alerta preventiva para conductores en el contexto vial indio.

Gupta et al. (2025) propusieron un modelo optimizado basado en YOLOv7 para la detección de marcas viales en entornos complejos, incorporando el optimizador AdamW y una arquitectura denominada Multi-Stage Cross-Convolutional Bottleneck Network (MSCCBN). El sistema fue evaluado con un dataset propio de 2,887 imágenes de alta resolución distribuidas en 11 categorías de marcas viales. La implementación sobre el dataset Cityscape alcanzó un mAP de 0.88 con una precisión de 0.886 y recall de 0.51, demostrando efectividad en la minimización de falsos positivos y negativos para aplicaciones de seguridad vial.

Acharya et al. (2025) presentaron un sistema integrado de identificación de carriles y detección de objetos orientado a mejorar la seguridad en vehículos autónomos mediante arquitecturas de aprendizaje profundo. El sistema combina detección avanzada de bordes con análisis de regiones de interés, empleando redes neuronales convolucionales para la identificación de límites viales y YOLO para la detección en tiempo real de vehículos, peatones y señales de tráfico. Las evaluaciones realizadas sobre diversos datasets demostraron una precisión del 84.86 % con latencia reducida, posibilitando la navegación segura en entornos con infraestructura vial mínima.

Rahman et al. (2025) desarrollaron BanglaRoadSignNet, un framework basado en YOLOv11 para la detección y reconocimiento de señales viales en Bangladesh, abordando los desafíos particulares de señalización no estandarizada y multilingüe en países en desarrollo. La arquitectura fue adaptada a un dataset local que contempla oclusiones y condiciones ambientales adversas propias del contexto bangladesí. El modelo alcanzó una precisión de 94.3 %, recall de 98.8 %, mAP50 de 99.6 % y mAP50-95 de 94.6 %, superando versiones previas de YOLO y estableciendo una solución viable para sistemas de transporte inteligente en tiempo real.

Sunidhi et al. (2025) desarrollaron un sistema de detección de carriles, objetos y estimación de velocidad para ego-vehículos mediante algoritmos de percepción visual. La propuesta integra múltiples módulos especializados —YOLO-NAS, Gaussian Blur, transformada de Hough y detección de bordes Canny— para lograr reconocimiento simultáneo de peatones, vehículos, señales de tráfico y límites de carril con baja latencia. El análisis experimental demostró la efectividad del sistema integrado como solución ADAS con capacidades avanzadas de percepción del entorno.

Bhuvaneshwari et al. (2025) implementaron un sistema de clasificación de señales de

tráfico con alerta por voz, integrando YOLOv4 para detección y redes convolucionales para clasificación precisa. El sistema incorpora un módulo de generación de alertas auditivas que notifica al conductor sobre regulaciones viales relevantes (límites de velocidad, señales de alto, cruces peatonales) al detectar señalización en el campo visual. La solución se destaca por su practicidad de implementación en vehículos convencionales sin requerir modificaciones sustanciales.

Anees y Agrawal (2025) propusieron un sistema integrado de detección de carriles, alineación vehicular y frenado automático mediante retroalimentación visual y técnicas de aprendizaje profundo. El sistema ejecuta análisis continuo de marcas de carril para garantizar la adherencia del vehículo a su trayectoria, incorporando un mecanismo de frenado adaptativo activado por detección visual de vehículos precedentes. Los resultados de simulación evidenciaron mejoras significativas en prevención de colisiones y precisión de mantenimiento de carril bajo diversas condiciones de conducción.

Gupta et al. (2025) diseñaron un modelo optimizado basado en YOLOv7 para la detección de señalizaciones horizontales en entornos viales complejos, incorporando el optimizador AdamW y una arquitectura denominada Multi-Stage Cross-Convolutional Bottleneck Network (MSCCBN) orientada a describir patrones y formas de marcas viales con mayor eficiencia. El sistema fue evaluado sobre un dataset de referencia compuesto por 2,887 imágenes de alta resolución con 11 categorías de marcas viales, incluyendo cruces peatonales y señales de carril, alcanzando un mAP50 de 0.889, una precisión de 0.886 y un recall de 0.83, lo cual contribuye al fortalecimiento de los sistemas inteligentes de transporte mediante la identificación automatizada de la infraestructura vial horizontal.

Rahman et al. (2025) formularon BanglaRoadSignNet, un framework robusto fundamentado en la arquitectura YOLOv11 para la detección y el reconocimiento simultáneo de señales de tránsito en Bangladesh, abordando los desafíos propios de un país en desarrollo como la señalización no estandarizada, las oclusiones y las condiciones ambientales adversas. El modelo fue entrenado con un dataset local de 28 tipos de señales viales, obteniendo una precisión de 94.3 %, recall de 98.8 %, mAP50 de 99.0 % y mAP50-95 de 92.6 %, superando versiones previas de YOLO y posicionándose como una solución práctica para aplicaciones de sistemas inteligentes de transporte (ITS) en tiempo real.

Bhuvaneshwari et al. (2025) implementaron un sistema de asistencia al conductor que integra la detección de señales de tránsito mediante YOLOv4, la clasificación a través de redes neuronales convolucionales y un módulo de alertas por voz, entrenado con el dataset GTSRB (German Traffic Sign Recognition Benchmark). El sistema genera notificaciones auditivas en tiempo real al identificar regulaciones viales como límites de velocidad, señales de pare y cruces peatonales, contribuyendo directamente a la prevención de accidentes de tránsito y al fortalecimiento de la seguridad vial para conductores y usuarios vulnerables de la vía.

X. Zhang et al. (2025) diseñaron el algoritmo YOLO-TSR, basado en el framework YOLOv8, para mejorar el reconocimiento de señales de tráfico de pequeño tamaño. La propuesta

integra tres contribuciones: un mecanismo de atención eficiente en el backbone, un proceso de submuestreo mediante operaciones de slicing para preservar información de objetos pequeños, y el mecanismo content-aware CARAFE para optimizar el upsampling. Los experimentos sobre los datasets TT100K y CCTSDB2021 reportaron un mAP50 de 72.75 % y 89.86 % respectivamente, superando al YOLOv8 original en ambos conjuntos de datos.

Choi et al. (2025) desarrollaron un sistema de repintado automático de líneas de carril basado en visión artificial, integrando una red neuronal YOLO para detección de carriles con el algoritmo Pure Pursuit para seguimiento de trayectorias. El sistema demuestra robustez en la identificación de marcaciones desvanecidas, discontinuas o parcialmente deterioradas, empleando una configuración multi-cámara que optimiza el reconocimiento de carriles y el posicionamiento de boquillas de pintado para operaciones de mantenimiento vial inteligente.

Russon et al. (2025) evaluaron la seguridad de los cruces peatonales mediante la implementación de modelos de redes neuronales convolucionales (ConvNeXtV2, ResNet50 y ResNeXt50) entrenados con imágenes aéreas de Google Maps e imágenes a nivel de calle de Google Street View en Francia. La metodología incorpora Mask R-CNN para la segmentación precisa de pasos de cebra y rotondas, analizando nueve factores de riesgo identificados en entornos urbanos y rurales. Los hallazgos demuestran el potencial de las redes convolucionales para automatizar las auditorías de infraestructura peatonal, contribuyendo a intervenciones más eficaces en materia de seguridad vial.

Said et al. (2025) optimizaron un framework de redes neuronales convolucionales con integración de mapas piramidales de características multi-escala para la detección eficiente de semáforos en sistemas inteligentes de transporte (ITS) y sistemas avanzados de asistencia al conductor (ADAS). El modelo aborda la detección de semáforos pequeños, ocluidos y de baja resolución en fondos urbanos complejos, mediante convoluciones dilatadas, alineación de regiones de interés (ROI) y Soft Non-Maximum Suppression. Los resultados experimentales sobre datasets reales demostraron un desempeño superior frente a métodos existentes, ofreciendo una solución escalable para el reconocimiento automatizado de dispositivos de control del tránsito.

Olszewski et al. (2025) investigaron el rendimiento de arquitecturas CNN ligeras para la detección y el reconocimiento de señales de tránsito en sistemas vehiculares, evaluando dos arquitecturas para detección y dos para reconocimiento con el objetivo de minimizar la complejidad computacional sin sacrificar la precisión. Los experimentos confirmaron que los modelos ligeros ejecutan ambas tareas de manera satisfactoria, alcanzando una efectividad comparable a la de modelos complejos reportados en la literatura, lo cual facilita su despliegue en dispositivos embebidos orientados a sistemas ADAS con recursos computacionales limitados.

Ayachi, Said, Afif, Alshammari, Hleili y Ben Abdelali (2025) examinaron de manera comprensiva diversas iteraciones de la familia YOLO mediante transfer learning para evaluar su eficacia en tareas de detección de objetos urbanos pertinentes a sistemas

avanzados de asistencia al conductor (ADAS), utilizando el dataset BDD100K como referencia. Los hallazgos indicaron que YOLOv9 ofrece un equilibrio óptimo entre velocidad y precisión para aplicaciones ADAS en tiempo real, mientras que YOLOv1, pese a su rapidez de procesamiento, carece de la exactitud alcanzada por sus sucesores, subrayando la importancia de seleccionar arquitecturas que armonicen velocidad, precisión y tamaño del modelo.

Dhatrika et al. (2025) desplegaron un sistema de reconocimiento de objetos en tiempo real para ADAS sobre dispositivos de borde (edge), capaz de detectar automóviles, bicicletas, camiones, buses, conos y peatones mediante aprendizaje profundo implementado en una placa de inteligencia artificial. El modelo alcanzó un mAP de 91.9 % con una precisión de 98.6 % y un recall de 96 %, demostrando la viabilidad del despliegue de modelos YOLO en hardware embebido para aplicaciones de seguridad vial vehicular en escenarios operativos reales.

H. Shah et al. (2025) sintetizaron una revisión comprehensiva de los algoritmos críticos empleados en tecnologías ADAS, abarcando detección facial, estimación de pose, estimación de mirada y detección de enlaces, analizando la evolución histórica de estos sistemas y el impacto del aprendizaje automático en la mejora de su precisión. El estudio evalúa sistemáticamente cada algoritmo con y sin machine learning, evidenciando que la integración de técnicas de aprendizaje profundo incrementa sustancialmente la exactitud de las decisiones tomadas por los sistemas de asistencia al conductor a costa de mayor carga computacional.

H. Liu et al. (2025) formularon MDTs-ADNet, una red de detección de accidentes de tránsito de doble flujo temporal-espacial que integra información espacial de frames RGB con dinámicas temporales extraídas del flujo óptico, posibilitando la caracterización comprehensiva de patrones de accidentes bajo condiciones variables. El framework fue validado sobre el dataset propio 4M-TAD y el dataset público AI City Challenge, alcanzando un AUC de 84.33 % y demostrando fuerte capacidad de generalización, lo cual aporta avances significativos a las tecnologías automatizadas de detección de accidentes orientadas a la reducción de fatalidades viales.

Kulawik y Kuczyński (2025b) entrenaron tres modelos YOLO (YOLOv7, YOLOv8n y YOLOv9t) para la detección y clasificación de señalizaciones horizontales en imágenes digitales orientadas a sistemas de asistencia al conductor (ADAS), empleando un dataset propio de 6,250 imágenes con 13,360 anotaciones distribuidas en nueve clases de marcas viales —incluyendo cruces peatonales, flechas direccionales y límites de carril— capturadas en calles urbanas, rurales y autopistas de Polonia. YOLOv7 alcanzó una precisión de 0.95, recall de 0.91 y mAP@0.5 de 0.98, confirmando que los detectores basados en YOLO gestionan eficazmente la variabilidad de las marcas horizontales bajo diversas condiciones de visibilidad y deterioro.

Gowri et al. (2025) elaboraron una revisión sistemática sobre sensores y técnicas de aprendizaje profundo para detección de semáforos orientada a seguridad vehicular y

sistemas de frenado automático de emergencia. El estudio analiza arquitecturas como detectores Single-Shot, Faster R-CNN y YOLO aplicadas a detección de estados semafóricos, examinando el pipeline completo desde la recolección de datos hasta la inferencia en tiempo real. La investigación establece un marco referencial para el desarrollo de sistemas inteligentes de transporte, identificando desafíos en fusión multi-modal de sensores y adaptabilidad a configuraciones semafóricas diversas.

Shovo et al. (2024) construyeron un framework basado en YOLO para detección de señales viales en Bangladesh, desarrollando el dataset BRSD con 9,133 imágenes anotadas en 29 categorías de señalización local. La evaluación con YOLOv8 reportó una precisión de 0.986 y recall de 0.999, superando a YOLOv5 y YOLOv7 en precisión promedio media. La investigación establece las bases para integrar visión artificial avanzada en sistemas de transporte inteligente adaptados a condiciones viales locales de países en desarrollo.

Chen et al. (2024) introdujeron CFG_SI_YOLO, un modelo especializado en la detección de defectos en señalizaciones horizontales que incorpora un módulo CoordConv multicanal, la función de pérdida Focal-EIoU, la activación GELU y un módulo de atención por similitud (SimAM) para mitigar las detecciones falsas por similitud inter-clase. Los experimentos sobre un dataset propio con múltiples tipos de marcas viales —incluyendo cruces peatonales, flechas direccionales y líneas de carril— alcanzaron una precisión de 85.7%, recall de 85.8%, F1 de 85.7% y mAP de 82.8%, resultados relevantes para el mantenimiento preventivo de la infraestructura vial horizontal.

Lu et al. (2024) concibieron X-CDNet, un detector de cruces peatonales en tiempo real fundamentado en YOLOX, integrando un módulo básico basado en ConvNeXt con convolución RepSLK (Reparameterizable Sparse Large-Kernel) para expandir el campo receptivo sin incrementar el tiempo de inferencia. Adicionalmente, construyeron el dataset CD9K con escenas de conducción realistas aumentadas mediante lluvia y niebla sintéticas. El modelo alcanzó un AP50 de 93.3 a una velocidad de 123 FPS, superando a YOLOX en precisión y velocidad, lo cual resulta pertinente para los sistemas de asistencia al conductor orientados a la protección de peatones.

Palwe et al. (2024) desarrollaron un prototipo inteligente para la vigilancia de baches y reductores de velocidad empleando YOLOv5 con almacenamiento en la nube, orientado a la identificación de deterioros en la superficie de rodadura que comprometen la seguridad de conductores y peatones. El sistema fue entrenado con el 70% del dataset y genera alertas automáticas dirigidas a usuarios y autoridades de mantenimiento vial, demostrando la viabilidad del aprendizaje profundo para la gestión proactiva de la infraestructura vial urbana.

G. Wang et al. (2024) diseñaron CGADNet, una red ligera derivada de YOLOv8 para la detección en tiempo real de cruces peatonales y flechas direccionales en escenarios complejos de conducción autónoma, incorporando un módulo C2f_Van basado en VanillaBlock, convolución parcial (PConv) y una función de pérdida con mecanismo de enfoque dinámico WIoU_{v3}. El modelo redujo los parámetros en 63.81%, los FLOPs en

70.07 % y los pesos en 63.11 %, alcanzando 50.35 FPS en Jetson Orin Nano, lo cual valida su aplicabilidad para el despliegue de sistemas de detección de señalizaciones horizontales en dispositivos de computación de borde.

Wu et al. (2024) idearon un sistema de inspección de marcas viales basado en imágenes de Street View que combina segmentación semántica, mapeo de perspectiva inversa y umbralización de imagen para cuantificar el ratio de deterioro de las señalizaciones horizontales, complementándolo con un detector de daños construido sobre YOLOv11x. El modelo alcanzó un mAP de 73.5 %, automatizando el proceso de inspección de la infraestructura vial horizontal y proporcionando datos objetivos para la planificación del mantenimiento de marcas viales, un componente esencial de la seguridad en el entorno de los vehículos autónomos.

R. Wang et al. (2024a) perfeccionaron YOLOv5-FPNet, un framework ligero para la percepción visual de reductores de velocidad en vehículos inteligentes conectados, integrando FasterNet con Dynamic Snake Convolution para la extracción adaptativa de características estructurales, el módulo de atención SimAM y la función de pérdida adaptativa Inner-WiseloU. Los resultados experimentales sobre un dataset propio de reductores registraron mejoras del 38.76 % en mAP, 143.15 % en mAP50-95 y 51.23 % en FPS respecto a redes ligeras convencionales, proporcionando una solución rápida y precisa para el reconocimiento de obstáculos en la calzada dentro de los sistemas vehiculares inteligentes.

Dewi et al. (2024) investigaron técnicas de mejoramiento de imágenes combinadas con modelos YOLO para el reconocimiento de marcas viales en condiciones nocturnas, compilando un dataset propio mediante conducción nocturna en Taiwán. El estudio evaluó las técnicas de Contrast Stretching, Histogram Equalization y CLAHE sobre imágenes nocturnas, determinando que la combinación de CLAHE con YOLOv8 produce los mejores resultados con una precisión del 90 % y recall del 90 %, contribuyendo significativamente a la detección confiable de señalización en escenarios de baja luminosidad.

Li y Zhao (2023) propusieron el algoritmo PMM-YOLO para detección de señales de tráfico, abordando los desafíos de tamaño reducido, oclusiones y detección errónea en entornos complejos. La arquitectura incorpora un módulo de convolución atrous paralelo adaptativo (APA), muestreo adaptativo multi-rama (MAS) y un módulo de fusión de características multi-escala (MSF), junto con un módulo de reorganización de salida para optimizar la detección de objetivos pequeños. Los resultados experimentales alcanzaron un mAP@0.5 de 86.4 % en el dataset TT100K, mejorando en 2.55 puntos porcentuales al YOLOv5 original.

Bhavsar et al. (2023) examinaron el tráfico vehicular y las infracciones de tránsito en una rotonda urbana de Ahmedabad, India, mediante video capturado por vehículo aéreo no tripulado (UAV), empleando YOLOv7 para la detección de vehículos y SORT para el seguimiento en tiempo real. El análisis reveló que aproximadamente el 23.26 % de los vehículos cometieron infracciones relacionadas con indisciplina de carril, conducción en contravía y exceso de velocidad, evidenciando la utilidad de la visión artificial para el

diagnóstico de la seguridad vial y la formulación de políticas de gestión del tránsito en países en desarrollo.

Kaya et al. (2023) compararon el desempeño de Faster R-CNN (R101-FPN y X101-FPN) y YOLOv7 para la detección automática de cruces peatonales en redes viales urbanas, con el propósito de prevenir colisiones entre vehículos, peatones y micro-vehículos en zonas de confluencia perpendicular. Los resultados sobre un dataset propio evidenciaron que YOLOv7 alcanzó una precisión de 98.6 % frente al 98.29 % de Faster R-CNN, posicionándose como una solución eficaz para la detección de infraestructura peatonal dentro de los sistemas inteligentes de transporte orientados a la reducción de siniestros viales.

Peralta-López, Morales-Viscaya, Lázaro-Mata et al. (2023) abordaron la detección de reductores de velocidad y baches mediante una red neuronal profunda con imágenes capturadas por cámara ZED estéreo montada en vehículo en la ciudad de Celaya, Guanajuato, México, constituyendo una de las escasas investigaciones latinoamericanas en esta temática. La arquitectura ligera con 37 filtros de convolución 3×3 fue entrenada con 714 imágenes etiquetadas manualmente, alcanzando una precisión de 98.13 % en la discriminación entre baches y reductores, demostrando la viabilidad del aprendizaje profundo para la detección de anomalías en la superficie de rodadura orientada a sistemas ADAS en contextos viales de países en desarrollo.

Dewi et al. (2023) construyeron el Taiwan Road Marking Sign Dataset (TRMSD), un conjunto de datos público para el reconocimiento de señales de tránsito en la superficie de rodadura —incluyendo flechas direccionales, cruces peatonales y límites de velocidad— mediante modelos YOLO (v2, v3, v4 y v4-tiny), evaluando el efecto del volteo de imagen para la distinción entre objetos izquierdos y derechos. YOLO v4 sin volteo obtuvo la mejor precisión de test (95.43 %) y un IoU de 66.12 %, evidenciando la importancia de la configuración de aumentación de datos en la detección de elementos viales horizontales para sistemas de asistencia al conductor.

Ben Elallid et al. (2022) realizaron una encuesta comprehensiva sobre la aplicación de aprendizaje profundo y por reforzamiento en conducción autónoma. La revisión construye una taxonomía de algoritmos orientados a las funciones principales de vehículos autónomos: comprensión de escenas, planificación de movimiento, toma de decisiones y control vehicular. El estudio identifica desafíos vigentes y direcciones de investigación futura, proporcionando un marco referencial sobre técnicas de inteligencia artificial de vanguardia aplicadas a sistemas de transporte inteligente.

Z.-D. Zhang et al. (2022) materializaron CDNet, una red de detección de cruces peatonales en tiempo real basada en YOLOv5 e implementada en Jetson Nano, incorporando módulos de atención squeeze-and-excitation (SE), entrenamiento con muestras negativas (NST), algoritmo de región de interés (ROI) y aumentación sintética de niebla para escenarios climáticos adversos. El sistema alcanzó un F1 promedio de 94.83 % a 33.1 FPS en el dispositivo de borde bajo escenarios complejos (nublado, lluvioso, con niebla y nocturno), constituyendo una referencia para la optimización de algoritmos de detección

de elementos viales en dispositivos de computación de borde dentro de los sistemas inteligentes de gestión del tránsito (ITMS).

A continuación, se exponen las diferentes investigaciones internacionales que detallan cómo las redes neuronales convolucionales y los modelos de detección de objetos han sido aplicados en estudios orientados a la detección de peatones, vehículos, señalizaciones horizontales, semáforos y cruces peatonales, así como su implementación en sistemas embebidos y el desarrollo de aplicaciones de seguridad vial en tiempo real. El Cuadro 1 presenta de manera sintética los hallazgos más representativos de esta selección.

Cuadro 1

Comparación analítica de antecedentes internacionales sobre detección de elementos viales con Deep Learning

Autor(es)	Año	Aporte principal	Brecha identificada
Kim et al.	2018	Faster R-CNN con fusión de features de frames sucesivos para peatones nocturnos.	Solo peatones, sin señalización vial ni vehículos.
Dai et al.	2021	Faster R-CNN multitarea para peatones y distancia nocturna con cámara NIR; precisión ~80 %.	Sin señales horizontales ni semáforos.
Zhang et al.	2022	CDNet (YOLOv5) en Jetson Nano para cruces peatonales; F1=94.83 % a 33.1 FPS.	Solo cruces peatonales, sin vehículos.
Bhavsar et al.	2023	YOLOv7 + SORT con UAV para tráfico en rotonda urbana de India.	Sin alertas preventivas en tiempo real.
Kaya et al.	2023	Faster R-CNN vs YOLOv7 para cruces peatonales; YOLOv7 alcanza 98.6 %.	Sin detección multiclase ni alertas.
Lu et al.	2024	X-CDNet (YOLOX) para cruces peatonales; AP50=93.3 a 123 FPS.	Sin señalización horizontal ni semáforos.
Kulawik & Kuczyński	2025	YOLOv7/v8n/v9t con 9 clases de marcas horizontales; mAP@0.5=0.98.	Sin peatones ni semáforos; contexto europeo.
Chaman et al.	2026	Benchmark YOLOv8-v12 con 16 clases para ADAS/ITS; precisión 97.3 %.	Sin validación en Latinoamérica.
Bao et al.	2026	EMD-YOLOv8 para peatones con -47.3 % de parámetros; +7.5 % mAP50.	Solo peatones, sin infraestructura vial.
Du et al.	2026	MLE-YOLO (YOLOv11) para vehículos y peatones en clima adverso.	Sin señalización horizontal ni semáforos.

Nota. Elaboración propia basada en revisión sistemática de investigaciones internacionales (2018–2026).

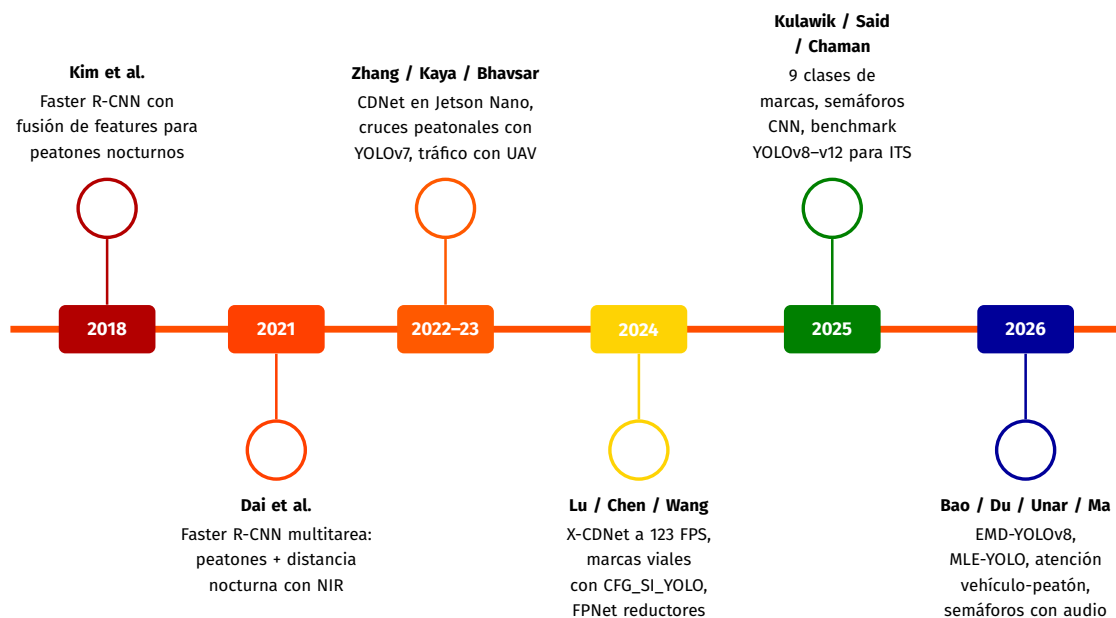
El análisis de los antecedentes internacionales evidencia una tendencia creciente hacia la optimización de arquitecturas YOLO y Faster R-CNN para la detección de elementos viales en tiempo real, con énfasis en la reducción de parámetros para dispositivos edge y la mejora

de la robustez bajo condiciones adversas. No obstante, se identifican brechas que justifican la presente investigación:

- Los estudios se centran en un tipo específico de elemento vial sin abordar la detección y segmentación simultánea de las 12 categorías contempladas en este trabajo.
- Los datasets empleados corresponden a señalización europea o asiática, sin validación en contextos latinoamericanos.
- No se integra segmentación de instancias con alertas preventivas por voz.
- Las investigaciones en ciudades intermedias andinas a más de 2,700 m.s.n.m., con mototaxis y señalización deteriorada, son prácticamente inexistentes.

Figura 3

Evolución cronológica de investigaciones sobre Deep Learning en detección de elementos viales



Nota. Elaboración propia basada en revisión de literatura de investigaciones internacionales (2018–2026).

2.1.2. Investigaciones nacionales

En el contexto de la investigación tecnológica peruana orientada al transporte y la seguridad vial, múltiples estudios han explorado el despliegue de sistemas inteligentes para optimizar la gestión del tráfico urbano y mitigar la siniestralidad vehicular. Estos antecedentes conforman el sustento científico nacional que respalda la aplicación de modelos inteligentes para afrontar las problemáticas vigentes de seguridad vial. El Cuadro 2 sintetiza los hallazgos más relevantes de esta revisión.

Urbina Domínguez (2025) concibió un sistema de mejora de la seguridad vial en las arterias principales de Ayacucho mediante redes neuronales convolucionales, orientado a la identificación automatizada de 14 categorías de señalizaciones verticales alineadas al Manual de Dispositivos de Control del Tránsito Automotor. Empleó 2,250 fotografías de señales verticales caprocesadas mediante la arquitectura YOLOv8m durante 104 épocas

con EarlyStopping en GPU NVIDIA A100. Las métricas obtenidas registraron una precisión de 0.997, recall de 0.997 y mAP@0.5 de 0.993, integrando además un módulo de alertas auditivas validado en cinco circuitos vehiculares urbanos de la ciudad, lo cual confirma la viabilidad de aplicar modelos de detección de objetos para la asistencia al conductor en el contexto ayacuchano.

Taype Delgado y Aranibar Huayhua (2023) plantearon un diagnóstico y una propuesta de fortalecimiento de la seguridad vial mediante la metodología del Manual de Seguridad Vial (MSV) en la vía vecinal Vista Alegre, tramo que conecta con la carretera nacional PE-3S de acceso a Abancay. La investigación identificó sectores críticos con alta frecuencia de siniestros y presencia de vehículos pesados mediante el análisis de factores de mitigación conforme al manual DG-2018, determinando la necesidad de incrementar la señalización, los agentes viales y los dispositivos de demarcación en la zona residencial estudiada. El estudio configuró un marco metodológico replicable para el análisis de seguridad vial en vías vecinales del territorio peruano.

Ramos Zuñiga (2021a) articuló un algoritmo de detección y reconocimiento automatizado de señales preventivas peruanas orientado a la sistematización de inventarios viales, tomando como área de estudio la Av. Eduardo Habich en Lima Metropolitana. La metodología se estructuró en dos fases: recolección de datos en campo y procesamiento en gabinete, empleando 2,026 imágenes de señales preventivas capturadas con cámaras de distintas resoluciones y aplicando filtros de color y forma basados en componentes RGB normalizados. El algoritmo alcanzó un porcentaje de reconocimiento de 82.5 %, sentando las bases para la automatización del inventario de señalización vial mediante técnicas de procesamiento de imágenes en el ámbito nacional.

Díaz Calle (2021) analizó la accesibilidad peatonal y el flujo vehicular en la intersección de la Av. La Marina con la Av. Universitaria en Lima, evaluando las condiciones existentes y las modificaciones que generaría la construcción de un viaducto. La metodología incorporó una inspección de seguridad vial conforme a la Guía de Auditorías de Seguridad Vial Urbana (ASVU) y una micro-simulación en el software Vissim con dos escenarios comparativos. Los resultados evidenciaron que el proyecto del viaducto prioriza la circulación vehicular sin garantizar condiciones adecuadas de desplazamiento peatonal seguro, subrayando la necesidad de incorporar la accesibilidad de usuarios vulnerables en la planificación de infraestructura vial urbana.

Leiva Ottone (2021) diseñó un sistema automático de grabación, detección, seguimiento y conteo de peatones en intersecciones semaforizadas de Lima, abordando la carencia de información sobre flujos peatonales que dificulta la planificación de infraestructura segura. El sistema integró el algoritmo YOLOv3 para la detección de peatones, el filtro de Kalman para el seguimiento temporal y un módulo de conteo automatizado, complementado con el diseño de la estructura física para grabación continua. Las simulaciones estáticas y dinámicas validaron la factibilidad del sistema, estableciendo bases tecnológicas para el monitoreo automatizado de peatones en intersecciones urbanas del contexto peruano

mediante visión por computadora.

Fabian Ramos (2020) examinó la relación entre la señalización vial y la incidencia de accidentes de tránsito en la intersección de las avenidas Nicolás Ayllón y Rivaguero en el distrito del Agustino, Lima. El estudio cuantitativo correlacional aplicó un cuestionario a 100 conductores para evaluar la percepción sobre el uso y respeto de la señalización existente. Los hallazgos confirmaron la existencia de relaciones positivas y estadísticamente significativas entre la implementación adecuada de señalización vial y la reducción de siniestros en intersecciones críticas de la capital, reforzando la importancia de la sensibilización al conductor como complemento a la infraestructura de control del tránsito.

Rodriguez Delgado (2019) efectuó un diagnóstico de seguridad vial en la carretera nacional PE-28B en el tramo del poblado de Pisac al poblado de Calca, Cusco, aplicando la metodología de Inspección de Seguridad Vial (ISV) del Manual de Seguridad Vial MSV-2017. Mediante fichas de inspección se identificaron deficiencias en señales verticales, señales horizontales, delineación, semáforos, iluminación y estado del pavimento a lo largo del tramo estudiado. Las conclusiones determinaron que la aplicación sistemática de la ISV permite proponer medidas correctivas específicas para cada deficiencia detectada, reduciendo los factores de riesgo en vías rurales y urbanas de la región andina.

Jhoffel Solano (2018) formuló una propuesta de implementación de señalización de tránsito para la prevención de accidentes en las avenidas Mesones Muro y Pakamuros de la ciudad de Jaén, utilizando como referencia normativa el Manual de Dispositivos de Control del Tránsito (MTC-2016). La investigación identificó que el tramo VI registra el mayor índice de accidentalidad (17.56 mv/k) y que el 50 % de la población evaluada desconoce la normativa de tránsito, mientras que el 88 % de los encuestados atribuye la ocurrencia de siniestros a la deficiente señalización. El estudio concluyó que la dotación de señalización adecuada, complementada con programas de educación vial, contribuiría significativamente a la disminución de accidentes en ciudades peruanas de tamaño intermedio.

Chambillo Ango (2016) implementó tres metodologías de evaluación de seguridad vial sin registro previo de accidentes en intersecciones urbanas de Ayacucho: evaluación del diseño geométrico, índice de severidad mediante la técnica sueca de conflictos de tráfico (TSCT) y frecuencia de accidentes previstos según el Highway Safety Manual (HSM). El análisis de la intersección Av. Independencia con Av. Universitaria, sustentado en estudios de inventario, velocidad, volumen y conflictos de tráfico, determinó un índice de severidad de 8.23 conflictos por cada 100,000 vehículos y una probabilidad de 5.27 accidentes anuales previstos, constituyendo un precedente metodológico relevante para la evaluación sistemática de la seguridad vial en intersecciones de la ciudad de Ayacucho.

Marino Ayuque (2016) configuró un sistema de clasificación de señales de tránsito vehicular basado en redes neuronales convolucionales, motivado por los 1,041,704 accidentes de tránsito registrados en el período 2004–2015 donde el exceso de velocidad constituye uno de los principales factores causales. El modelo fue entrenado con la base de datos de señales de tránsito alemanas (GTSRB) y alcanzó un porcentaje de aciertos de 95.29 %,

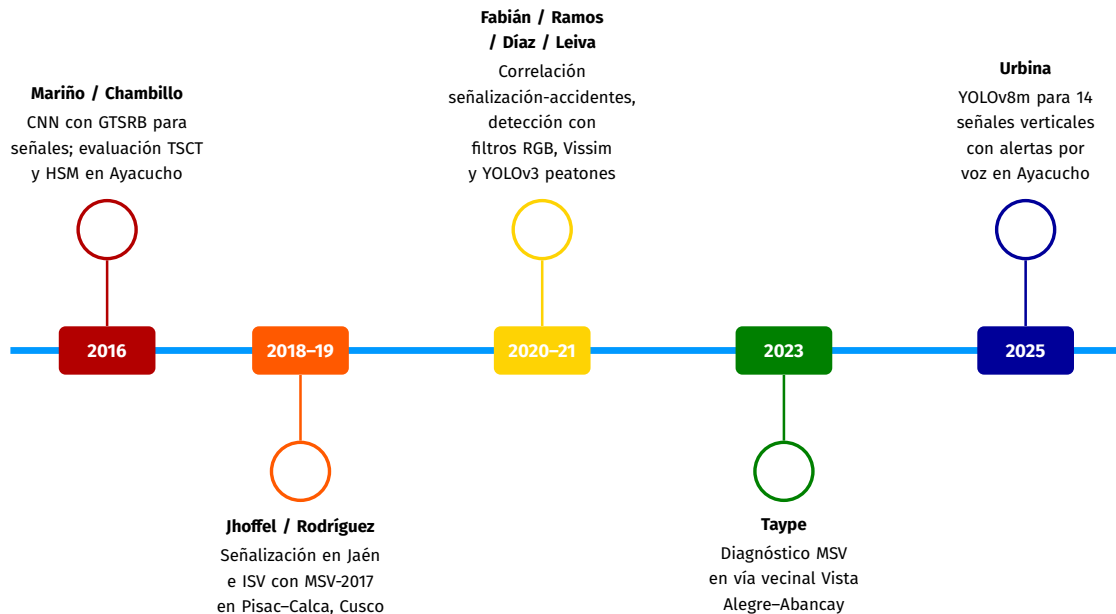
próximo al estado del arte de la época (99.46 %), siendo posteriormente implementado en una aplicación para iPhone que clasifica señales mediante la cámara del dispositivo. Esta investigación sentó uno de los primeros precedentes en el Perú sobre la aplicación de técnicas de aprendizaje profundo en sistemas inteligentes de transporte.

Cuadro 2

Comparación analítica de antecedentes nacionales sobre seguridad vial y tecnologías de detección

Autor(es)	Año	Aporte principal	Brecha identificada
Urbina	2025	YOLOv8m para 14 señales verticales en Ayacucho con alertas por voz; precisión 0.997 y mAP@0.5=0.993.	Solo señales verticales, sin peatones, vehículos ni señalización horizontal.
Taype	2023	Diagnóstico de seguridad vial con MSV en vía vecinal Vista Alegre–Abancay; identificación de sectores críticos.	Enfoque normativo-manual, sin tecnología de detección automatizada.
Ramos	2021	Algoritmo de detección de señales preventivas con filtros RGB en Lima; reconocimiento de 82.5 %.	Técnicas clásicas de procesamiento, sin deep learning ni tiempo real.
Díaz	2021	Inspección ASVU y micro-simulación Vissim en intersección Av. La Marina–Universitaria, Lima.	Simulación sin componente de visión artificial ni detección automática.
Leiva	2021	Sistema YOLOv3 + filtro de Kalman para conteo de peatones en intersecciones de Lima.	Solo peatones, sin detección de vehículos ni señalización vial.
Fabián	2020	Estudio correlacional señalización–accidentes en Av. Nicolás Ayllón, Lima (100 conductores).	Enfoque estadístico sin implementación tecnológica de detección.
Rodríguez	2019	ISV con MSV-2017 en carretera PE-28B tramo Pisac–Calca, Cusco; fichas de inspección.	Inspección manual sin automatización ni visión por computadora.
Jhoffel	2018	Propuesta de señalización en Av. Mesones Muro y Pakamuros de Jaén con MTC-2016.	Propuesta normativa sin implementación tecnológica de monitoreo.
Chambillo	2016	Tres métodos de evaluación (TSCT, HSM) en intersección Av. Independencia–Universitaria, Ayacucho.	Métodos estadísticos sin detección automatizada en tiempo real.
Mariño	2016	CNN para clasificación de señales con GTSRB; 95.29 % de aciertos en app iPhone.	Dataset europeo, sin señalización peruana ni segmentación de instancias.

Nota. Elaboración propia basada en revisión de investigaciones nacionales (2016–2025).

Figura 4**Evolución cronológica de investigaciones nacionales sobre seguridad vial y detección automatizada**

Nota. Elaboración propia basada en revisión de investigaciones nacionales (2016–2025).

La revisión de los antecedentes nacionales revela una transición gradual desde enfoques normativos-manuales de inspección vial hacia la incorporación incipiente de técnicas de inteligencia artificial para la detección automatizada de elementos del tránsito. Sin embargo, se identifican brechas relevantes:

- La mayoría de investigaciones peruanas emplean métodos convencionales de inspección sin automatización mediante visión por computadora.
- los escasos estudios que incorporan deep learning se limitan a la detección de un único tipo de elemento, señales verticales o peatones.
- No se reportan investigaciones nacionales que implementen segmentación de instancias a nivel de píxel para elementos viales
- El único antecedente local en Ayacucho (Urbina Domínguez, 2025) aborda exclusivamente señales verticales, dejando sin cobertura la detección de peatones, vehículos, señalización horizontal, semáforos, reductores de velocidad y mototaxis que constituyen el alcance integral de la presente investigación.

2.2. Bases teóricas

2.2.1. Seguridad vial y prevención de accidentes de tránsito

2.2.1.1 Conceptualización de la seguridad vial

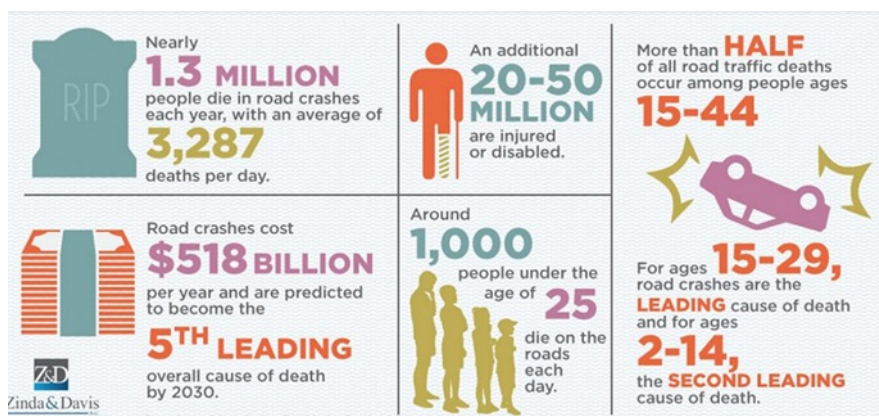
La seguridad vial constituye una disciplina transversal que articula principios de ingeniería, salud pública y gestión institucional con el propósito de mitigar la ocurrencia y severidad de los siniestros en las redes de transporte terrestre. De acuerdo con la World Health Organization y World Bank (2004), esta disciplina aborda el fenómeno de las lesiones ocasionadas por el tránsito desde una perspectiva epidemiológica, reconociendo que dichos eventos responden a causas identificables y, en consecuencia, susceptibles de intervención sistemática.

En su informe más reciente, la World Health Organization (2023) cuantifica la magnitud del problema: Aproximadamente 1,19 millones de personas pierden la vida cada año como consecuencia de colisiones y atropellos en las vías, lo que equivale a más de 3 200 decesos diarios. El noventa por ciento de estas muertes se concentra en naciones de ingresos bajos y medios, donde la motorización crece a un ritmo superior a la capacidad de los gobiernos para mejorar la infraestructura y la regulación. Los peatones representan el grupo más afectado, con un incremento del tres por ciento en sus cifras de mortalidad hasta alcanzar 274 000 víctimas, mientras que los ciclistas registraron un aumento cercano al veinte por ciento, sumando 71 000 fallecidos.

El mismo informe identifica cinco factores de riesgo fundamentales cuya regulación legislativa resulta determinante: Velocidad excesiva, conducción bajo los efectos del alcohol, uso del casco en motociclistas, empleo del cinturón de seguridad y sistemas de retención infantil. No obstante, únicamente seis países a nivel global disponen de legislación que cumple las mejores prácticas recomendadas para los cinco factores simultáneamente (World Health Organization, 2023).

Figura 5

Distribución de defunciones causadas por siniestros de tránsito a nivel mundial



Nota. Adaptado de Kavanagh (2017)

2.2.1.2 Enfoque de Sistema Seguro y Visión Cero

El paradigma de *Sistema Seguro* (Safe System Approach) representa la evolución conceptual más significativa en la gestión de la seguridad vial durante las últimas tres décadas. Sus raíces se encuentran en la política sueca denominada Visión Cero, adoptada oficialmente por el Parlamento de Suecia en octubre de 1997, cuyo principio rector establece que ninguna muerte ni lesión grave resulta éticamente aceptable dentro del sistema de transporte por carretera (Johansson, 2009). A diferencia del enfoque tradicional que atribuye la responsabilidad principalmente al usuario, Visión Cero redistribuye dicha responsabilidad hacia quienes diseñan, construyen, mantienen y gestionan el sistema vial.

Björnberg et al. (2022) profundizan en la dimensión ética y tecnológica de Visión Cero, señalando que la premisa central descansa sobre dos principios biomecánicos: Primero, los seres humanos cometen errores de manera inevitable como parte de su naturaleza cognitiva; segundo, el cuerpo humano posee umbrales finitos de tolerancia a las fuerzas de impacto. En consecuencia, el sistema de transporte debe diseñarse con redundancia suficiente para que, cuando un componente falle, otro impida que las consecuencias resulten fatales o catastróficas. Los umbrales biomecánicos de referencia establecen que la probabilidad de muerte de un peatón atropellado es inferior al diez por ciento cuando la velocidad de impacto no supera los 30 km/h, pero se eleva drásticamente al ochenta por ciento cuando dicha velocidad alcanza los 50 km/h.

Figura 6
Sistema-Visión Cero



Nota. Adaptado de Baron Road Drafting Consultants (2017)

2.2.1.3 Factores determinantes de la siniestralidad vial

La literatura especializada identifica tres categorías interdependientes de factores que configuran la siniestralidad en las redes viales, esquema que la doctrina latinoamericana de tránsito denomina “triángulo de la accidentalidad” y que constituye una simplificación pedagógica de la Matriz de Haddon.

- **Factor humano.** Cal y Mayor Reyes Spíndola y Cárdenas Grisales (2018) señalan que el comportamiento del usuario constituye el elemento más frecuentemente asociado a los siniestros de tránsito, aunque este hecho refleja una deficiencia del diseño

sistémico más que una responsabilidad exclusiva del individuo. La World Health Organization (2023) precisa que la velocidad excesiva contribuye a cerca de un tercio de los fallecimientos viales en países de altos ingresos y a la mitad en naciones de ingresos bajos y medios. La conducción bajo los efectos del alcohol multiplica el riesgo de colisión fatal, mientras que el uso de teléfonos móviles durante la conducción cuadruplica la probabilidad de verse involucrado en un evento. Otros factores humanos relevantes incluyen la fatiga, la inexperiencia del conductor, el incumplimiento de normas de prioridad de paso y la omisión del uso de dispositivos de seguridad como el cinturón o el casco.

- **Factor vehicular.** Las condiciones mecánicas del vehículo y sus dispositivos de seguridad activa (frenos ABS, control electrónico de estabilidad, asistencia de frenado de emergencia) y pasiva (cinturones de tres puntos, bolsas de aire frontales y laterales, zonas de deformación programada, reposacabezas activos) inciden de manera directa en la probabilidad de ocurrencia y la severidad de las lesiones. Garber y Hoel (2015) enfatizan que los estándares vehiculares han evolucionado significativamente en las últimas décadas, incorporando tecnologías de asistencia avanzada al conductor (ADAS) que intervienen proactivamente para evitar colisiones mediante sistemas de detección de peatones, alerta de cambio involuntario de carril, frenado autónomo de emergencia y control adaptativo de cruce. Adicionalmente, el estado de conservación de los neumáticos, el sistema de frenos, la iluminación y los limpiaparabrisas constituyen factores técnicos cuya deficiencia incrementa sustancialmente el riesgo de siniestro.
- **Factor infraestructural y ambiental.** El diseño geométrico de la vía, la calidad y regularidad superficial del pavimento, la iluminación, la señalización horizontal y vertical, las condiciones meteorológicas y el mobiliario urbano conforman el entorno físico donde se desarrolla la movilidad. O'Flaherty (1997) establece que la ingeniería vial debe concebir cada elemento de la infraestructura considerando la interacción con el usuario y el vehículo, de modo que el diseño físico induzca comportamientos seguros y perdone los errores inevitables. La American Association of State Highway and Transportation Officials (2018) incorpora este principio en su séptima edición del *Green Book*, introduciendo clasificaciones por contexto (rural, suburbano, urbano y centro urbano) que complementan las categorías funcionales tradicionales. Los factores ambientales como la lluvia, la niebla, el deslumbramiento solar y la reducción de visibilidad nocturna agravan las condiciones de riesgo, especialmente en vías con deficiente señalización o pavimentos deteriorados.

2.2.1.4 Magnitud del problema: Panorama mundial

De acuerdo con la World Health Organization (2023), a pesar de una reducción del cinco por ciento en la mortalidad vial global entre 2010 y 2021, las lesiones causadas por el tránsito continúan representando la principal causa de muerte entre niños y jóvenes de 5 a 29 años. La carga económica de los siniestros viales se estima en aproximadamente el tres

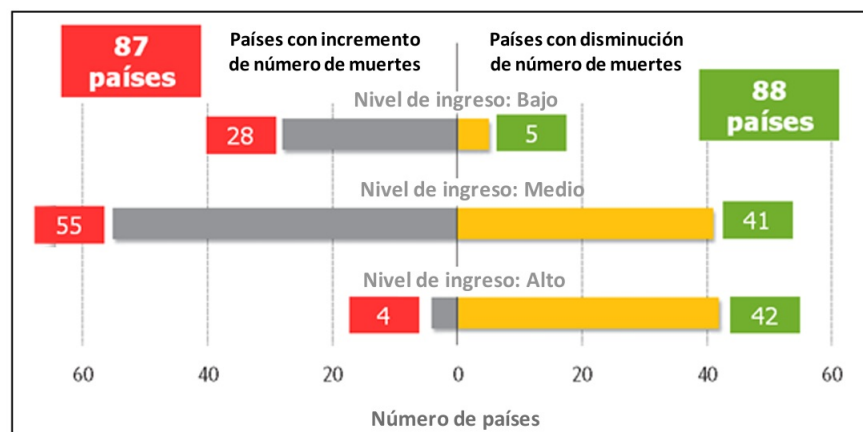
por ciento del producto interno bruto de los países, cifra que asciende al cinco por ciento en las naciones de ingresos bajos y medios (World Health Organization & World Bank, 2004). Los usuarios vulnerables de la vía peatones, ciclistas y motociclistas representan más de la mitad de todas las víctimas mortales a nivel mundial, proporción que se incrementa en contextos urbanos de países en desarrollo donde la infraestructura peatonal resulta deficiente o inexistente.

2.2.1.5 Magnitud del problema: Contexto peruano

En el Perú, la siniestralidad vial constituye un problema de salud pública con dimensiones crecientes que impacta directamente en el desarrollo socioeconómico del país. Según datos del Observatorio Nacional de Seguridad Vial (2023), sustentados en registros de la Policía Nacional del Perú, durante el año 2023 se contabilizaron 87 083 siniestros de tránsito a nivel nacional, que dejaron un saldo de 58 000 heridos y 3 316 fallecidos. Estas cifras representan un incremento sostenido respecto a las décadas anteriores, concentrándose aproximadamente la mitad de los eventos en el departamento de Lima. Las principales causas registradas incluyen el exceso de velocidad, la imprudencia del conductor, el estado de ebriedad, la imprudencia del peatón y las fallas mecánicas.

Figura 7

Siniestralidad Vial en función del nivel de ingreso en 87-88 países



Nota. Adaptado de Observatorio Nacional de Seguridad Vial (2023)

El Ministerio de Transportes y Comunicaciones reconoce que, si bien en países con niveles bajos y medios de ingreso como el Perú la siniestralidad vial ha sido percibida históricamente como un costo inevitable asociado al crecimiento vehicular, la experiencia internacional demuestra que los accidentes son eventos prevenibles mediante la aplicación sistemática de medidas de ingeniería, fiscalización, educación y atención sanitaria. El mismo manual establece que las administraciones públicas tienen la responsabilidad de trabajar para reducir la siniestralidad empleando todos los medios posibles y disponibles, debiendo implementar auditorías e inspecciones de seguridad vial como herramientas proactivas de gestión.

2.2.1.6 Cumplimiento de normas de seguridad vial en el Perú

De acuerdo con la Instituto Nacional de Estadística e Informática (2021), del total de la población urbana peruana, el 60,9% declara cruzar siempre por el cruceo peatonal, mientras que un 13,7% reconoce no hacerlo nunca o casi nunca. Respecto al uso de puentes peatonales, el 41,3% manifiesta utilizarlos siempre incluso cuando existe un cruceo peatonal disponible. En cuanto al respeto semafórico, el 70,1% afirma cruzar la calzada únicamente cuando la señalización vehicular se encuentra en rojo. El 77,3% de la población declara esperar siempre al vehículo de transporte público sobre la acera sin invadir la calzada; sin embargo, un 3,9% reconoce no mirar previamente a ambos lados antes de cruzar la vía.

Cuadro 3

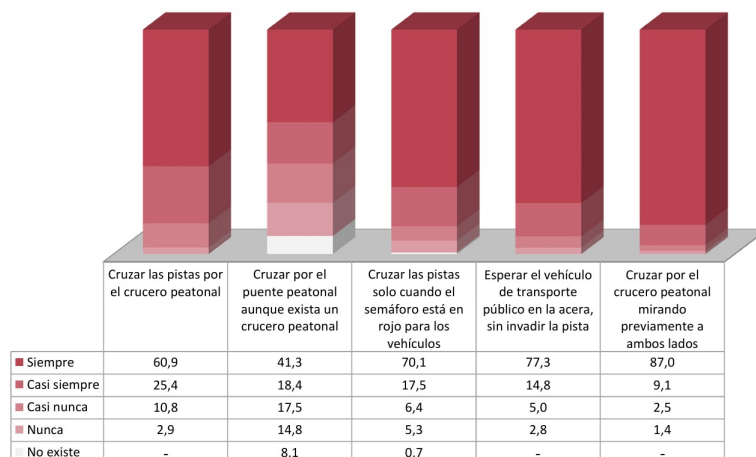
Perú urbano: Población que cumple las reglas de seguridad vial, según habitualidad, 2021 (Porcentaje)

Reglas de seguridad vial	Nunca	Casi nunca	Casi siempre	Siempre	No existe
Cruzar las pistas por el cruceo peatonal	2,9	10,8	25,4	60,9	-
Cruzar por el puente peatonal aunque exista un cruceo peatonal	14,8	17,5	18,4	41,3	8,1
Cruzar las pistas solo cuando el semáforo está en rojo para los vehículos	5,3	6,4	17,5	70,1	0,7
Esperar el vehículo de transporte público en la acera, sin invadir la pista	2,8	5,0	14,8	77,3	-
Cruzar por el cruceo peatonal mirando previamente a ambos lados	1,4	2,5	9,1	87,0	-

Nota. Adaptado de Instituto Nacional de Estadística e Informática (2021).

Figura 8

Resultados de la Encuesta Nacional de Programas Presupuestales sobre conductas viales



Nota. Adaptado de Instituto Nacional de Estadística e Informática (2021).

2.2.2. Accidentes de tránsito

2.2.2.1 Definición y clasificación desde la ingeniería de transportes

Desde la perspectiva de la ingeniería de transportes, un accidente o siniestro de tránsito se define como un evento no intencional que se produce en una vía de circulación pública, en el que participa al menos un vehículo en movimiento y que ocasiona daños materiales, lesiones corporales o la muerte de alguno de los involucrados. Baker y Fricke (1986) sistematizaron la investigación técnica de siniestros viales estableciendo protocolos rigurosos para la reconstrucción cinemática del evento, la documentación de evidencia física, marcas de frenado, posición final de vehículos, dispersión de escombros y el análisis multifactorial de las causas contribuyentes, sentando las bases de la accidentología moderna como disciplina técnico-científica.

Roess et al. (2019) clasifican los siniestros de tránsito según múltiples criterios técnicos.

- Por su **tipología dinámica**, se distinguen: Colisiones frontales (dos vehículos en sentidos opuestos), colisiones laterales o de ángulo (impacto perpendicular, frecuente en intersecciones), colisiones por alcance (impacto trasero por seguimiento a distancia insuficiente), volcaduras (pérdida de estabilidad vertical del vehículo), atropellos a peatones (colisión entre vehículo y persona a pie), choques contra objetos fijos (postes, árboles, barreras, muros) y caídas de ocupantes (especialmente en vehículos de transporte público).
- Por su **gravedad**, se categorizan empleando la Escala de Severidad de Lesiones (AIS): eventos con daños materiales exclusivamente, eventos con heridas leves (AIS 1–2), eventos con heridas graves (AIS 3–5) y eventos fatales (AIS 6).
- Por su **ubicación**, se diferencian los siniestros urbanos —caracterizados por la multiplicidad de usuarios vulnerables, la complejidad geométrica de las intersecciones, la presencia de semáforos y la diversidad de movimientos vehiculares y peatonales— de los rurales o interurbanos, donde predominan las velocidades de operación elevadas, las distancias de visibilidad variables según la geometría del trazado y los tiempos de respuesta de los servicios de emergencia significativamente mayores. Mannering y Washburn (2013) precisan que los siniestros en intersecciones urbanas representan entre el cuarenta y el sesenta por ciento del total de eventos en áreas urbanas, lo que subraya la criticidad de estos puntos de conflicto para la seguridad vial.

2.2.2.2 Análisis causal del modelo secuencial al sistémico

La comprensión de las causas de los siniestros viales ha transitado desde modelos lineales hacia enfoques sistémicos más comprensivos. El modelo secuencial de Heinrich, basado en la metáfora del dominó, establece que la eliminación de cualquier factor intermedio de la cadena causal interrumpe la secuencia y previene el accidente. Su investigación en el ámbito de la seguridad industrial lo llevó a formular la proporción 300:29:1, configurando la denominada “pirámide de la seguridad”. Este principio resulta relevante para la ingeniería

vial porque implica que por cada siniestro grave registrado en un punto de la red, ocurren decenas de incidentes menores y centenares de situaciones de riesgo no documentadas que revelan las deficiencias latentes del sistema.

El tránsito hacia modelos sistémicos se consolida con Reason (1997), quien argumenta que los siniestros rara vez obedecen a una causa única, sino que emergen de la convergencia de fallas latentes en múltiples capas del sistema. En el contexto vial, estas capas abarcan la planificación institucional (políticas de transporte, asignación presupuestaria), el diseño de la infraestructura (geometría, drenaje, señalización), la regulación y fiscalización (límites de velocidad, control de alcoholemia), el mantenimiento (estado del pavimento, reposición de señales deterioradas) y las acciones del usuario (velocidad de operación, atención a la vía, respeto a las señales).

2.2.2.3 Puntos críticos y zonas de concentración de siniestros

Las denominadas zonas de concentración de siniestros, puntos negros o *hotspots* constituyen localizaciones específicas de la red vial donde la frecuencia o gravedad de los eventos resulta estadísticamente superior a la esperada para condiciones similares de tránsito e infraestructura. Mannering y Washburn (2013) indican que la identificación de estos puntos constituye el primer paso del ciclo de gestión de la seguridad vial y requiere la integración de análisis estadísticos (tasa de accidentes, índice de severidad, métodos bayesianos empíricos) con inspecciones de campo que evalúen las condiciones geométricas, de señalización y de operación del punto.

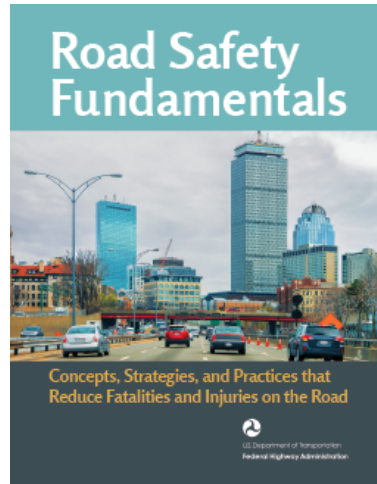
Gold (1998), en su metodología desarrollada para el Banco Interamericano de Desarrollo y aplicable al contexto latinoamericano, propone un ciclo iterativo de cuatro fases para abordar la seguridad vial desde la ingeniería: identificación de ubicaciones con problemas de seguridad, diagnóstico de los factores contribuyentes mediante inspección detallada, selección e implementación de contramedidas de ingeniería y evaluación de la efectividad de las intervenciones. Este ciclo constituye la base conceptual de la Inspección de Seguridad Vial (ISV) recogida en el Ministerio de Transportes y Comunicaciones (2017b), herramienta que permite evaluar la infraestructura existente para detectar deficiencias antes de que se produzcan siniestros.

2.2.2.4 Estrategias de prevención integradas

La Federal Highway Administration (2017), en su material de formación *Road Safety Fundamentals*, organiza las estrategias de prevención en cinco módulos interrelacionados: Fundamentos de la seguridad vial, comportamiento humano, medición de la seguridad, resolución de problemas y gestión integral. Este enfoque reconoce que la prevención efectiva requiere combinar intervenciones de ingeniería como el rediseño de intersecciones conflictivas, la instalación de reductores de velocidad, la mejora de la iluminación y la implementación de señalización clara con medidas de educación, fiscalización y atención post-siniestro.

Figura 9

Road Safety Fundamentals: Concepts, Strategies, and Practices that Reduce Fatalities and Injuries on the Road



Nota. Adaptado de Federal Highway Administration (2017).

Las contramedidas de ingeniería de bajo costo constituyen intervenciones particularmente relevantes para contextos con recursos limitados, como ciudades intermedias de países en desarrollo. Estas incluyen: Mejora de la señalización horizontal y vertical, instalación de reductores de velocidad en zonas escolares y residenciales, ensanchamiento de bermas, mejoramiento de la iluminación en intersecciones, instalación de barreras de contención, canalización del tráfico mediante isletas y demarcación de cruces peatonales con pintura retrorreflectiva. La efectividad de estas medidas ha sido documentada extensamente en la literatura técnica; por ejemplo, la correcta demarcación de cruces peatonales tipo “cebra” reduce los atropellos entre un veinticinco y un cuarenta por ciento según estudios compilados por la International Transport Forum y OECD (2022).

2.2.2.5 Ingeniería de transportes: Fundamentos y alcance

La ingeniería de transportes constituye la rama de la ingeniería civil dedicada al estudio, planificación, diseño, construcción, operación y mantenimiento de los sistemas de transporte para el movimiento eficiente y seguro de personas y mercancías. Cal y Mayor Reyes Spíndola y Cárdenas Grisales (2018), en la novena edición de su obra referencial ampliamente utilizada en la formación de ingenieros civiles en Iberoamérica, definen la ingeniería de tránsito como la fase operativa de la ingeniería de transportes que trata de la planificación, diseño y operación del tráfico por las calles y carreteras, sus redes, terminales, tierras adyacentes y su relación con otros modos de transporte. Esta definición subraya que la ingeniería de tránsito no se limita al movimiento vehicular, sino que abarca la totalidad de las interacciones entre los componentes del sistema de transporte, incluyendo peatones, ciclistas, transporte público y vehículos de carga.

Garber y Hoel (2015) amplían esta perspectiva al precisar que la ingeniería de tránsito y transportes comprende cuatro áreas funcionales interrelacionadas: la planificación del transporte, que evalúa la demanda futura y define las inversiones necesarias en

infraestructura; el diseño geométrico, que establece las dimensiones y características físicas de las vías; la ingeniería de tránsito propiamente dicha, que optimiza la operación de las redes existentes mediante señalización, semaforización y gestión; y la seguridad vial, que busca minimizar la frecuencia y severidad de los siniestros mediante intervenciones sobre todos los componentes del sistema.

2.2.2.6 Flujo vehicular y relación fundamental del tránsito

El flujo vehicular constituye el fenómeno central de estudio de la ingeniería de tránsito y se describe mediante tres variables macroscópicas fundamentales interrelacionadas: el flujo o volumen (q), la velocidad (v) y la densidad o concentración (k). La relación fundamental del tránsito expresa que el flujo es igual al producto de la velocidad por la densidad: $q = v \times k$, ecuación que Cal y Mayor Reyes Spíndola y Cárdenas Grisales (2018) denominan la “ecuación fundamental del flujo vehicular” y que rige el comportamiento de la corriente de tránsito bajo cualquier condición operativa.

- El **volumen de tránsito** se define como el número de vehículos que pasan por un punto o sección transversal de una vía durante un período determinado, expresándose comúnmente en vehículos por hora (veh/h). Su variante más utilizada en el análisis operacional es la tasa de flujo, calculada para intervalos de quince minutos y expandida a la hora equivalente mediante el factor de hora pico (FHP). El **Tránsito Promedio Diario Anual (TPDA)** constituye el indicador fundamental para la planificación y el diseño de nuevas vías, obteniéndose como el cociente entre el volumen total anual y los 365 días del año (Cal y Mayor Reyes Spíndola & Cárdenas Grisales, 2018).
- La **velocidad** admite múltiples definiciones operacionales según el propósito del análisis. La velocidad de punto mide la rapidez instantánea de un vehículo al pasar por una ubicación específica. La velocidad media temporal es el promedio aritmético de las velocidades de todos los vehículos observados en un punto durante un intervalo. La velocidad media espacial, también denominada velocidad de recorrido, corresponde al promedio de las velocidades de todos los vehículos que ocupan un tramo de vía en un instante dado y es la variable que se emplea en la relación fundamental del tránsito. La diferencia entre ambos promedios resulta relevante para el ingeniero civil, pues la velocidad media temporal siempre sobreestima la velocidad media espacial (Roess et al., 2019).
- La **densidad** cuantifica el número de vehículos presentes en una unidad de longitud de la vía en un instante determinado, expresándose en vehículos por kilómetro (veh/km). Valores bajos de densidad corresponden a condiciones de flujo libre donde los conductores pueden seleccionar su velocidad deseada; conforme la densidad aumenta, las interacciones entre vehículos se intensifican, reduciendo progresivamente la velocidad hasta alcanzar la densidad de congestión, punto en el cual el flujo se detiene completamente (Garber & Hoel, 2015).

2.2.2.7 Capacidad vial y nivel de servicio

La **capacidad vial** se define como la tasa máxima de flujo razonablemente esperable que puede soportar una sección de vía o carril bajo condiciones prevalecientes de infraestructura, tránsito y control. El Transportation Research Board (2022), en la séptima edición del *Highway Capacity Manual* (HCM-7), establece las metodologías estandarizadas para su determinación en diferentes configuraciones: segmentos de autopista, segmentos de carreteras de dos carriles, vías urbanas de múltiples carriles, intersecciones semaforizadas, intersecciones no semaforizadas, rotondas y facilidades peatonales y ciclistas. Esta edición incorpora por primera vez procedimientos para evaluar el impacto de vehículos conectados y automatizados (CAV) sobre la capacidad y el nivel de servicio.

El **nivel de servicio** (LOS, por sus siglas en inglés *Level of Service*) constituye una medida cualitativa que describe las condiciones operativas de una corriente de tránsito desde la perspectiva del usuario, empleando una escala de seis niveles designados con las letras A a F. El nivel A representa condiciones de flujo libre con total libertad de maniobra; el nivel B describe un flujo estable con ligera restricción; el nivel C corresponde a flujo estable pero con limitaciones perceptibles en la selección de velocidad y maniobra; el nivel D indica flujo cercano a la inestabilidad con reducciones significativas de velocidad; el nivel E representa operación a capacidad con flujo inestable; y el nivel F describe condiciones de congestión con flujo forzado y velocidades extremadamente bajas (Transportation Research Board, 2022).

Para intersecciones semaforizadas, el nivel de servicio se evalúa mediante la demora promedio por vehículo, medida en segundos por vehículo (s/veh). Una demora inferior a 10 s/veh corresponde al nivel A, mientras que demoras superiores a 80 s/veh definen el nivel F. La relación volumen/capacidad (v/c) complementa este análisis, considerándose que valores superiores a 0,85 indican condiciones de saturación inminente que requieren intervención. Estas métricas resultan fundamentales para el ingeniero civil al dimensionar intersecciones, programar ciclos semafóricos y justificar proyectos de mejoramiento vial (Roess et al., 2019).

2.2.2.8 Diseño geométrico de vías

El diseño geométrico de vías constituye una de las actividades centrales de la ingeniería civil de transporte, pues define las dimensiones físicas y la configuración espacial de la infraestructura que determina las condiciones de circulación, visibilidad y seguridad. La American Association of State Highway and Transportation Officials (2018) (AASHTO), en la séptima edición de su publicación conocida como el *Green Book*, provee las directrices de diseño adoptadas internacionalmente y adaptadas a las normativas nacionales de cada país.

Los elementos fundamentales del diseño geométrico comprenden tres componentes interrelacionados.

- El **alineamiento horizontal** define la trayectoria de la vía en planta mediante la combinación de tangentes (tramos rectos), curvas circulares (definidas por su radio de curvatura) y curvas de transición (espirales que facilitan el paso gradual de una tangente a una curva circular). El radio mínimo de curvatura se determina en función de la velocidad de diseño, el peralte máximo admisible y el coeficiente de fricción lateral, garantizando que las fuerzas centrífugas no superen la capacidad de adherencia del vehículo.
- El **alineamiento vertical** describe el perfil longitudinal de la vía mediante rasantes rectas conectadas por curvas verticales convexas (cambio de pendiente de ascendente a descendente) y cóncavas (cambio de descendente a ascendente), cuya longitud mínima se calcula para asegurar la distancia de visibilidad de parada requerida.
- La **sección transversal** define la geometría perpendicular al eje de la vía, incluyendo la calzada (superficie destinada a la circulación vehicular), las bermas (franjas laterales que proporcionan soporte estructural, espacio de emergencia y protección del borde del pavimento), las cunetas (elementos de drenaje superficial), las veredas o aceras (superficies peatonales), los separadores centrales y las pendientes de los taludes (American Association of State Highway and Transportation Officials, 2018).

2.2.2.9 Marco normativo del tránsito en el Perú

El ordenamiento jurídico peruano en materia de transporte terrestre se estructura a partir de la Congreso de la República del Perú (1999), que establece los lineamientos generales para la gestión del transporte y el tránsito, orientando la acción estatal hacia la seguridad y la protección de los usuarios. Esta norma se reglamenta mediante el Ministerio de Transportes y Comunicaciones (2009), que define las reglas de circulación, las obligaciones de conductores y peatones, las sanciones por infracciones y los procedimientos de fiscalización. Ambos instrumentos jurídicos constituyen el marco legal dentro del cual se inserta cualquier intervención tecnológica orientada a la seguridad vial en el territorio nacional, incluido el modelo inteligente adaptativo propuesto en esta investigación.

2.2.2.10 Manual de Dispositivos de Control del Tránsito Automotor

El Ministerio de Transportes y Comunicaciones (2016) constituye el instrumento técnico oficial que regula el diseño, fabricación, instalación y mantenimiento de todos los dispositivos destinados a controlar y guiar el tránsito vehicular y peatonal en las vías del territorio nacional. Aprobado originalmente por R.D. N° 016-2016-MTC/14 y actualizado por R.D. N° 26-2024-MTC/18, este documento normativo establece los criterios de uniformidad en el diseño, ubicación, tamaño, forma, color y uso de los dispositivos de control, garantizando que los usuarios reciban mensajes consistentes e inequívocos en toda la red vial.

El Manual se sustenta en el principio de que ninguna vía debe ponerse en servicio sin contar con los dispositivos de señalización correspondientes, dado que la señalización transmite

instrucciones y mensajes que otros medios no pueden comunicar de manera igualmente efectiva. Los dispositivos regulados se organizan en cuatro categorías principales: señales verticales, señalización horizontal o marcas en el pavimento, semáforos y dispositivos auxiliares.

Figura 10

Portada del Manual de Dispositivos de Control del Tránsito Automotor



Nota. Adaptado de Manual de Dispositivos de Control del Tránsito Automotor (2024).

Como referencia internacional, el Federal Highway Administration (2023), en su undécima edición publicada por la Federal Highway Administration de los Estados Unidos, establece estándares uniformes para los dispositivos de control del tránsito, constituyendo el referente técnico más citado a nivel mundial. Esta edición, la primera actualización en catorce años, incorpora especificaciones para carriles exclusivos de autobuses pintados en rojo, carriles verdes para bicicletas, señalización de estaciones de carga para vehículos eléctricos y balizas intermitentes rectangulares de advertencia rápida (RRFB).

2.2.2.11 Señalización vertical

Las señales verticales son dispositivos de control del tránsito instalados a un costado de la vía o sobre ella, que transmiten un mensaje al usuario mediante palabras, símbolos o una combinación de ambos, regulando, previniendo o informando sobre las condiciones de circulación. El Ministerio de Transportes y Comunicaciones (2016) clasifica las señales verticales en tres grupos funcionales:

- **Señales reglamentarias (R).** Informan al usuario sobre las limitaciones, prohibiciones y restricciones que regulan el uso de la vía. Su cumplimiento es obligatorio. Incluyen señales de prioridad (PARE R-1, Ceda el Paso R-2), señales de prohibición (prohibido adelantar R-16, prohibido voltear R-10), señales de restricción (velocidad máxima R-30, peso máximo R-34) y señales de obligación (dirección obligada R-14). Se

caracterizan por su forma circular con fondo blanco, orla y símbolo en rojo para prohibiciones y en azul para obligaciones.

- **Señales preventivas (P).** Advierten al usuario sobre la existencia de peligros potenciales y las condiciones que requieren precaución. Tienen forma de rombo con fondo amarillo y símbolo en negro. Incluyen señales de curvatura (P-1 a P-7), señales de intersección (P-8 a P-15), señales de pendiente (P-35, P-36), señales de zona escolar (P-49), señales de peatones (P-48) y señales de superficie de rodadura (P-31 a P-34).
- **Señales informativas (I).** Guían al usuario proporcionando información sobre destinos, distancias, servicios, puntos de interés y otros datos que facilitan la navegación. Presentan fondo verde con texto y flechas en blanco para señales de dirección, y fondo azul para señales de servicios generales.

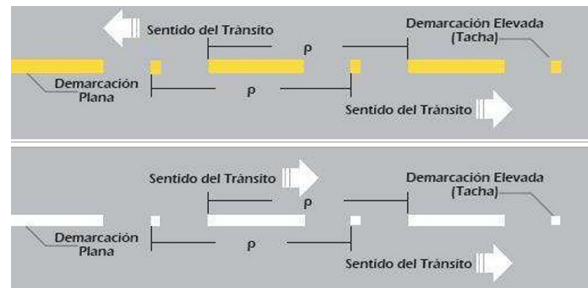
Los criterios técnicos de diseño incluyen: **Retroreflectividad** (capacidad de la señal para reflejar la luz de los faros vehiculares, garantizada mediante láminas retrorreflectivas tipo I, III o XI según el tipo de vía), **altura de montaje** (mínimo 1,50 m en zona urbana y 1,20 m en zona rural, medidos desde el borde inferior de la señal hasta el nivel de la calzada), **distancia de ubicación** (función de la velocidad de diseño para asegurar el tiempo de percepción-reacción del conductor) y **orientación** (ligera rotación respecto a la perpendicular de la vía para optimizar la retroreflexión) (Ministerio de Transportes y Comunicaciones, 2016).

2.2.2.12 Señalización horizontal: Marcas en el pavimento

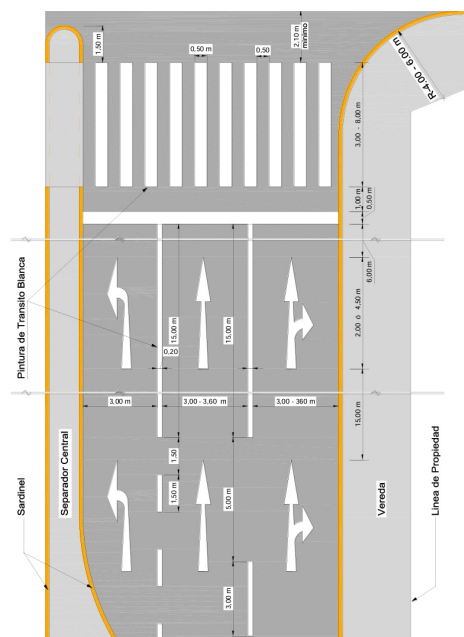
Las marcas en el pavimento, también denominadas demarcaciones, constituyen la señalización horizontal y están conformadas por dispositivos planos aplicados o adheridos sobre la superficie de rodadura, sardineles y otras estructuras viales. De acuerdo con el Ministerio de Transportes y Comunicaciones (2016), estas marcas cumplen funciones esenciales de canalización, regulación y advertencia que complementan a los demás dispositivos de control del tránsito. El Manual clasifica las marcas en el pavimento en dos categorías principales:

Marcas planas en el pavimento. Comprenden líneas longitudinales (de borde de calzada, de carril, canalizadoras, de prohibición de adelantamiento), líneas transversales (de pare, de cruce peatonal, de ceda el paso), flechas direccionales, símbolos (triángulos de prioridad, marcas de velocidad) y textos (PARE, SOLO, BUS).

Las principales marcas planas reguladas por el Manual incluyen: Líneas de borde de calzada (continuas, color blanco, ancho 10 o 15 cm), líneas de carril (segmentadas, color blanco, patrón 4,50 m pintado y 7,50 m sin pintar), líneas de prohibición de adelantamiento (continuas, color amarillo), líneas canalizadoras de tránsito, líneas demarcadoras de entradas y salidas, líneas de transición por reducción de carriles, líneas de pare (continuas, color blanco, ancho 50 cm, ubicadas a 1,20 m antes del cruce peatonal), líneas de cruce peatonal tipo “cebra” (franjas blancas paralelas de 50 cm de ancho con separación de 50 cm).

Figura 11*Patrón dimensional y señalización horizontal en el pavimento*

Nota. Adaptado de Manual de Dispositivos de Control del Tránsito Automotor (2024).

Figura 12*Especificaciones dimensionales de la línea de cruce peatonal*

Nota. Adaptado de Manual de Dispositivos de Control del Tránsito Automotor (2024).

2.2.2.13 Semáforos

Los semáforos son dispositivos de señalización luminosa que asignan de manera alternada el derecho de paso a las diferentes corrientes vehiculares y peatonales que confluyen en una intersección, operando mediante la sucesión cíclica de indicaciones de color: Verde (paso permitido), ámbar (transición, preparación para detenerse) y rojo (paso prohibido). Roess et al. (2019) precisan que el diseño del ciclo semafórico involucra la determinación de cuatro parámetros fundamentales: La **longitud del ciclo** (tiempo total para completar una secuencia de fases, típicamente entre 60 y 120 segundos en intersecciones urbanas), el **número de fases** (secuencias de movimientos no conflictivos atendidos simultáneamente), la **distribución de tiempos de verde** (proporcional a los volúmenes críticos de cada fase) y los **intervalos de cambio y despeje** (ámbar más todo-rojo, calculados en función de la velocidad de aproximación y las dimensiones de la intersección).

El Ministerio de Transportes y Comunicaciones (2016) clasifica los semáforos en:

- **Semáforos vehiculares** (de tres luces circulares o con flechas direccionales)
- **Semáforos peatonales** (que emplean las indicaciones de “hombre caminando” en verde y “mano levantada” en rojo, con cuenta regresiva del tiempo de cruce),
- **Señales informativas** Guían al usuario proporcionando información sobre destinos, distancias, servicios, puntos de interés y otros datos que facilitan la navegación. Presentan fondo verde con texto y flechas en blanco para señales de dirección, y fondo azul para señales de servicios generales.
- **Semáforos intermitentes** (ámbar intermitente para precaución, rojo intermitente para detención obligatoria)
- **Semáforos especiales** (para carriles reversibles, pasos a nivel ferroviario y zonas de peaje). La ubicación de los semáforos debe garantizar su visibilidad desde una distancia mínima de 60 m para velocidades de hasta 60 km/h, considerando la altura de montaje, el alineamiento horizontal y la ausencia de obstrucciones visuales.

La semaforización moderna incorpora tecnologías de control adaptativo que ajustan los tiempos de verde en tiempo real según las condiciones de demanda, empleando detectores de lazo inductivo, detectores de video, detectores de radar o detectores de infrarrojos. Estos sistemas representan una de las aplicaciones más consolidadas de los Sistemas Inteligentes de Transporte en el ámbito urbano.

2.2.2.14 Reductores de velocidad

Los reductores de velocidad constituyen dispositivos físicos de moderación del tránsito (*traffic calming*) instalados sobre la calzada con el propósito de obligar a los conductores a reducir su velocidad en zonas donde la seguridad de los usuarios vulnerables lo requiere, tales como entornos escolares, áreas residenciales, zonas hospitalarias y aproximaciones a intersecciones conflictivas. Cal y Mayor Reyes Spíndola y Cárdenas Grisales (2018) identifican los principales tipos: **Resaltos o gibas** (elevaciones trapezoidales o sinusoidales de la calzada, con alturas de 7,5 a 10 cm y longitudes de rampa de 1,0 a 3,7 m según la velocidad deseada de paso), **mesas de velocidad** (plataformas elevadas de mayor longitud que los resaltos, con superficie plana en la parte superior que puede funcionar simultáneamente como cruce peatonal elevado), **cojines o almohadas de velocidad** (reductores parciales que cubren solo una porción del ancho del carril, permitiendo el paso de vehículos de emergencia y autobuses sin reducción drástica) y **bandas transversales alertadoras** (franjas rugosas que producen vibración y ruido al ser atravesadas).

El Ministerio de Transportes y Comunicaciones (2016) especifica que los reductores de velocidad en el Perú deben cumplir dimensiones normalizadas, contar con señalización preventiva anticipada (señal P-33a “Reductor de velocidad adelante”), demarcación horizontal mediante franjas blancas sobre el dispositivo para garantizar su visibilidad diurna y nocturna, y señalización reglamentaria de velocidad máxima. La correcta identificación de estos dispositivos resulta crítica para el modelo inteligente adaptativo propuesto, pues su detección oportuna permite generar alertas preventivas al conductor,

evitando tanto el paso a velocidad excesiva (riesgo de pérdida de control) como el frenado brusco (riesgo de colisión por alcance).

Figura 13

Reductor de velocidad en el pavimento



Nota. Adaptado de Motor Maintenance (2024).

2.2.2.15 Peatones y usuarios vulnerables de la vía

Los peatones, ciclistas y motociclistas conforman la categoría de usuarios vulnerables de la vía, cuya protección constituye una prioridad del enfoque de sistema seguro dado que carecen de la protección estructural que brinda la carrocería de un vehículo. Garber y Hoel (2015) indican que la seguridad peatonal depende de la interacción entre tres factores: Las características del peatón (velocidad de caminata, capacidad visual y auditiva, edad, condición física), las características de la infraestructura (existencia y estado de veredas, cruces peatonales, refugios peatonales, iluminación) y el comportamiento vehicular (velocidad de operación, respeto a la señalización, distancia de frenado).

La velocidad de caminata de los peatones constituye un parámetro de diseño fundamental en ingeniería de tránsito. El Transportation Research Board (2022) recomienda emplear una velocidad de caminata de 1,0 m/s para el cálculo del tiempo mínimo de cruce peatonal en semáforos, considerando la presencia de adultos mayores y personas con movilidad reducida. Para poblaciones con predominancia de peatones jóvenes y activos, puede emplearse un valor de 1,2 m/s. El ancho efectivo de la vereda necesario para un flujo peatonal adecuado se determina según el nivel de servicio peatonal, que al igual que su equivalente vehicular emplea una escala de A a F basada en la densidad peatonal y la velocidad de desplazamiento.

El Ministerio de Transportes y Comunicaciones (2016) establece que los cruces peatonales deben demarcarse en todas las intersecciones de vías arteriales y colectoras, así como en zonas escolares, centros comerciales y otros generadores importantes de tránsito peatonal. Las dimensiones normativas especifican franjas blancas de 50 cm de ancho con separaciones de 50 cm, extendidas a todo el ancho de la calzada. En intersecciones semaforizadas, la fase peatonal debe calcularse para permitir que un peatón cruce la

totalidad de la calzada a la velocidad de diseño antes de que la indicación peatonal cambie a rojo intermitente.

2.2.2.16 Manual de Seguridad Vial (MSV-2017)

El Manual de Seguridad Vial (2017) forma parte del sistema de Manuales de Carreteras establecido por el Reglamento Nacional de Gestión de Infraestructura Vial (D.S. N° 034-2008-MTC). Su alcance se dirige a los órganos responsables de la gestión vial en los tres niveles de gobierno nacional, regional y local y tiene por finalidad contribuir a la mejora de las características de la infraestructura vial y su entorno para incrementar la seguridad intrínseca y la calidad de protección de las redes de carreteras. El Manual constituye el referente normativo para la realización de inspecciones de seguridad vial (ISV), auditorías de seguridad vial (ASV) y evaluaciones de la consistencia del diseño geométrico.

Figura 14

Portada del Manual de Seguridad Vial del MTC



Nota. Adaptado de Manual de Seguridad Vial (2017).

2.2.2.17 Sistemas Inteligentes de Transporte (ITS)

Los Sistemas Inteligentes de Transporte (ITS, por sus siglas en inglés) representan la aplicación integrada de tecnologías de la información, las comunicaciones, la electrónica y la automatización al ámbito del transporte, con el objetivo de mejorar la eficiencia, la seguridad y la sostenibilidad de las redes viales. Sussman (2005) conceptualiza los ITS como un paradigma emergente que transforma los sistemas de transporte tradicionales mediante la incorporación de capacidades de percepción, procesamiento y comunicación que permiten la toma de decisiones en tiempo real tanto por parte de los gestores del sistema como de los usuarios individuales.

Figura 15
Sistema inteligente de transporte



Nota. Adaptado de Modeshift (2023)

2.3. Marco conceptual

2.3.1. Inteligencia artificial

2.3.1.1 Definición y evolución histórica

La inteligencia artificial (IA) constituye la rama de la informática dedicada al desarrollo de sistemas capaces de realizar tareas que, cuando son ejecutadas por seres humanos, requieren inteligencia.

Los orígenes formales de la IA se remontan a la propuesta de Russell y Norvig (2021) conocida como el Taller de Dartmouth (1956), donde John McCarthy, Marvin Minsky, Nathaniel Rochester y Claude Shannon acuñaron el término “inteligencia artificial” como disciplina científica independiente. Desde entonces, el campo ha experimentado ciclos alternados de entusiasmo y retracción: los primeros sistemas expertos de la década de 1970, los “inviernos” de la IA en los años 1980 y 1990 por las limitaciones computacionales, y el renacimiento actual impulsado por la convergencia de tres factores: la disponibilidad de enormes volúmenes de datos digitalizados, el incremento exponencial de la capacidad de cómputo mediante unidades de procesamiento gráfico (GPU) y el desarrollo de algoritmos de aprendizaje profundo que explotan eficientemente ambos recursos.

Russell y Norvig (2021) clasifican los métodos de IA en tres grandes categorías según la naturaleza de la tarea: métodos de búsqueda y planificación, métodos de representación del conocimiento y razonamiento, y métodos de aprendizaje automático. Esta última categoría, que engloba las técnicas mediante las cuales un sistema mejora su desempeño a partir de la experiencia acumulada en datos, constituye el fundamento computacional del modelo propuesto en esta investigación y se desarrolla en las secciones siguientes.

2.3.2. Aprendizaje automático (*machine learning*)

2.3.2.1 Definición formal

El aprendizaje automático (*machine learning*) se define como el estudio de algoritmos que mejoran automáticamente su rendimiento en una tarea determinada a medida que acumulan experiencia. Mitchell (1997) formaliza esta definición estableciendo que un programa aprende de la experiencia E con respecto a una clase de tareas T y una medida de rendimiento P , si su desempeño en las tareas de T , medido por P , mejora con la experiencia E .

A diferencia de la programación convencional, donde el desarrollador especifica explícitamente las reglas que transforman las entradas en salidas, el aprendizaje automático invierte este proceso: Dado un conjunto de entradas y sus salidas deseadas, el algoritmo descubre automáticamente las reglas que gobiernan la relación entre ambas. Esta capacidad resulta especialmente valiosa para tareas como la detección de objetos en imágenes, donde la variabilidad visual de las escenas (iluminación, ángulo, oclusión, escala) hace inviable la formulación explícita de reglas deterministas (Mitchell, 1997).

2.3.2.2 Paradigmas del aprendizaje automático

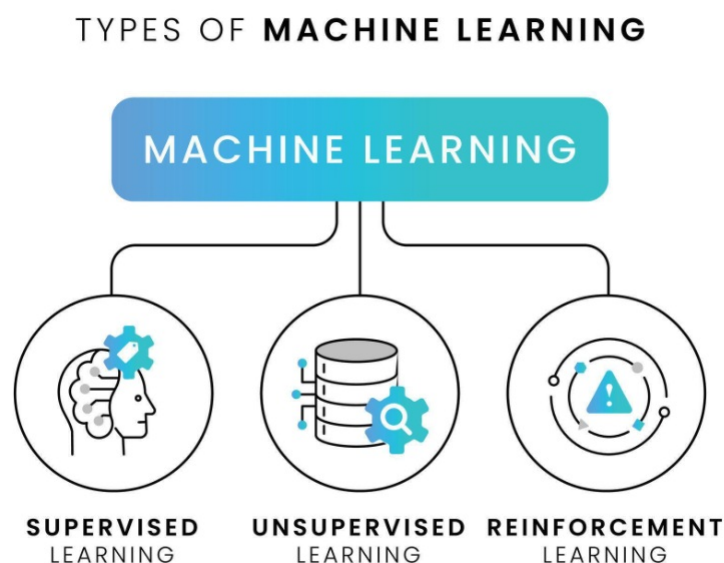
Bishop (2006) organiza los paradigmas del aprendizaje automático en tres modalidades fundamentales según la naturaleza de la supervisión disponible durante el entrenamiento:

- **Aprendizaje supervisado.** El modelo se entrena a partir de un conjunto de datos donde cada muestra de entrada está asociada a una etiqueta o valor objetivo conocido, conformando pares $\{(x_i, y_i)\}_{i=1}^N$. El algoritmo ajusta iterativamente sus parámetros internos para minimizar la discrepancia entre sus predicciones \hat{y}_i y las etiquetas reales y_i , empleando una función de pérdida $\mathcal{L}(\hat{y}, y)$ como criterio de optimización. Las tareas supervisadas se dividen en dos categorías: *clasificación*, donde la variable de salida es una categoría discreta (por ejemplo, clasificar una imagen como “peatón” o “vehículo”), y *regresión*, donde la salida es un valor continuo (por ejemplo, predecir las coordenadas de un *bounding box*). Los algoritmos supervisados clásicos incluyen la regresión logística, las máquinas de vectores de soporte (SVM), los árboles de decisión, los bosques aleatorios (*Random Forest*) y las redes neuronales artificiales. En el contexto de la detección de objetos, el aprendizaje supervisado es el paradigma dominante.

- **Aprendizaje no supervisado.** El modelo recibe datos sin etiquetas $\{x_i\}_{i=1}^N$ y busca descubrir estructuras, patrones o agrupaciones inherentes a la distribución subyacente de los datos. Las tareas principales incluyen el *clustering* (agrupamiento de observaciones con características similares en conjuntos homogéneos, como *k*-means, DBSCAN y agrupamiento jerárquico), la reducción de dimensionalidad (compresión de la información preservando las relaciones relevantes, mediante técnicas como el análisis de componentes principales PCA y los autoencoders), la estimación de densidad y la detección de anomalías. A diferencia del aprendizaje supervisado, no existe una “respuesta correcta” predefinida contra la cual evaluar las predicciones; el modelo debe identificar la estructura por sí mismo. En el ámbito del transporte, las técnicas no supervisadas se emplean, por ejemplo, para identificar patrones de congestión vehicular o segmentar zonas de concentración de siniestros sin conocimiento previo de las categorías (Bishop, 2006).
- **Aprendizaje por refuerzo.** Un agente interactúa con un entorno dinámico, ejecutando acciones que generan recompensas o penalizaciones. El objetivo es aprender una política de decisión que maximice la recompensa acumulada esperada a largo plazo, donde r_t es la recompensa en el paso t y γ es un factor de descuento temporal. A diferencia del aprendizaje supervisado, el agente no recibe la respuesta correcta, sino una señal de retroalimentación diferida que puede estar separada temporalmente de la acción que la originó, lo que genera el denominado “problema de asignación de crédito”. Sus aplicaciones comprenden la navegación autónoma de vehículos, el control robótico, la optimización de semáforos adaptativos y los juegos estratégicos (Russell & Norvig, 2021).

Figura 16

Tipos de aprendizaje de la inteligencia artificial



Nota. Adaptado de Noohani (2023)

2.3.3. Aprendizaje profundo (*deep learning*)

2.3.3.1 Estructura de una red neuronal artificial

Una red neuronal artificial (RNA) constituye un modelo computacional inspirado en la organización del sistema nervioso biológico, compuesto por unidades de procesamiento elementales denominadas neuronas artificiales, organizadas en capas e interconectadas mediante conexiones ponderadas. De manera análoga a las neuronas biológicas donde las dendritas reciben señales electroquímicas, el soma las procesa y el axón transmite la respuesta a otras neuronas, una neurona artificial recibe entradas numéricas, las pondera, las suma y aplica una transformación no lineal para producir una salida.

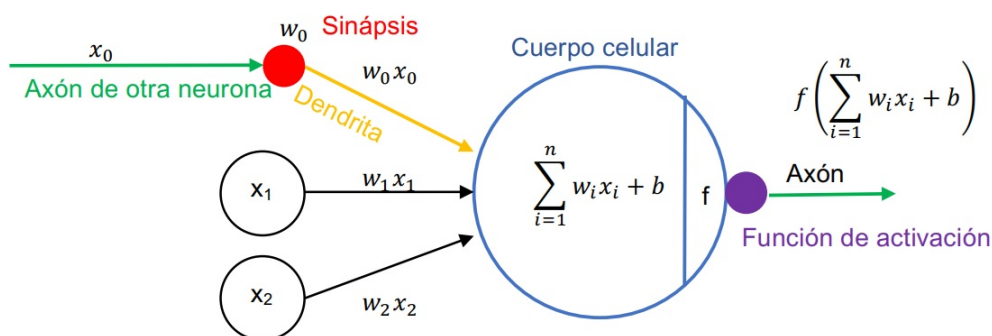
Goodfellow et al. (2016) formalizan la operación de una neurona artificial mediante la expresión:

$$y = f \left(\sum_{i=1}^n w_i x_i + b \right) = f(\mathbf{w}^T \mathbf{x} + b) \quad (1)$$

donde $\mathbf{x} = (x_1, x_2, \dots, x_n)$ representa el vector de entradas, $\mathbf{w} = (w_1, w_2, \dots, w_n)$ el vector de pesos sinápticos que pondera la importancia relativa de cada entrada, b el sesgo o umbral (*bias*) que desplaza la función de activación, y $f(\cdot)$ la función de activación no lineal que introduce capacidad de modelar relaciones complejas. Los pesos sinápticos constituyen los parámetros entrenables del modelo: valores positivos amplifican la señal de entrada (sinapsis excitatoria) mientras que valores negativos la atenúan (sinapsis inhibitoria).

Figura 17

Modelo matemático de una neurona artificial: entradas ponderadas, función de suma y función de activación no lineal



Nota. Adaptado de Ramos Zuñiga (2021b).

La arquitectura de una RNA se organiza en tres tipos de capas. La **capa de entrada** recibe los datos brutos; para imágenes en color, cada neurona de entrada corresponde a un valor de intensidad de un canal cromático (rojo, verde o azul) de un píxel específico, de modo que una imagen de 640×640 píxeles con 3 canales genera 1228800 entradas. Las **capas ocultas** ejecutan las transformaciones sucesivas mediante combinaciones lineales seguidas de funciones de activación no lineales; el número y tamaño de estas capas define la profundidad y capacidad expresiva del modelo. Redes "poco profundas" con una o

dos capas ocultas pueden aproximar funciones simples, pero las redes “profundas” con decenas o centenares de capas son necesarias para capturar las abstracciones jerárquicas requeridas en tareas complejas como la detección de objetos. La **capa de salida** produce el resultado final, cuyo formato depende de la tarea: un valor continuo para regresión, una distribución de probabilidades sobre clases para clasificación, o en el caso de la detección de objetos, coordenadas de *bounding boxes*, puntuaciones de confianza, probabilidades de clase y coeficientes de máscara (Goodfellow et al., 2016).

2.3.3.2 Funciones de activación

Las funciones de activación introducen no linealidad en la red, permitiendo que el modelo aprenda relaciones complejas que una simple transformación lineal no podría capturar. Sin funciones de activación no lineales, una red de cualquier profundidad sería equivalente a una única capa lineal, pues la composición de funciones lineales es lineal. Goodfellow et al. (2016) describen las funciones de activación más utilizadas en la práctica:

La **función sigmoide** $\sigma(z) = 1/(1 + e^{-z})$ transforma cualquier valor real en el intervalo $(0, 1)$, interpretable como probabilidad. Resulta útil en la capa de salida para clasificación binaria, pero presenta dos limitaciones para capas profundas: el *desvanecimiento de gradiente* (*vanishing gradient*), pues su derivada se aproxima a cero para valores de entrada grandes en magnitud absoluta, dificultando el aprendizaje en las capas tempranas de la red; y la *saturación*, donde las neuronas con activaciones cercanas a 0 o 1 actualizan sus pesos con gradientes prácticamente nulos.

La función **ReLU** (*Rectified Linear Unit*), definida como $f(z) = \max(0, z)$, fue popularizada por Nair y Hinton (2010) y constituyó un avance determinante para el entrenamiento de redes profundas. Su derivada es 1 para entradas positivas y 0 para entradas negativas, lo que elimina el desvanecimiento de gradiente en la mitad del espacio de activación y acelera significativamente la convergencia. Krizhevsky et al. (2017) demostraron que AlexNet con ReLU convergía seis veces más rápido que con funciones sigmoide equivalentes. La limitación de ReLU es el problema de las “neuronas muertas” (*dying ReLU*): neuronas cuyas entradas son persistentemente negativas nunca se activan y sus pesos dejan de actualizarse.

Las variantes mejoradas incluyen **Leaky ReLU** $f(z) = \max(\alpha z, z)$ con $\alpha \approx 0,01$, que asigna una pequeña pendiente a las entradas negativas para mantener el flujo de gradientes; y **SiLU/Swish** $f(z) = z \cdot \sigma(z)$, una función suave y no monótona que ha demostrado rendimiento superior en redes profundas. YOLOv8 emplea la función **SiLU** como activación por defecto en su arquitectura, combinando las ventajas de suavidad de la sigmoide con la escalabilidad de ReLU.

La función **Softmax** $\sigma(z_i) = e^{z_i} / \sum_{j=1}^K e^{z_j}$ generaliza la sigmoide para clasificación multiclase, produciendo una distribución de probabilidades sobre K clases cuya suma es unitaria. En la cabeza de detección de YOLOv8, Softmax se emplea en la rama de clasificación para determinar la probabilidad de pertenencia a cada una de las doce clases

de elementos viales.

2.3.3.3 Proceso de entrenamiento: Propagación hacia adelante y retropropagación

El entrenamiento de una red neuronal profunda consiste en un proceso iterativo de optimización que ajusta los millones de pesos sinápticos para minimizar una función de pérdida (*loss function*) que cuantifica la discrepancia entre las predicciones del modelo y las etiquetas reales. Rumelhart et al. (1986), en su artículo fundacional publicado en *Nature*, formalizaron el algoritmo de retropropagación (*backpropagation*) que hizo viable el entrenamiento de redes multicapa al proporcionar un método eficiente para calcular los gradientes de la función de pérdida respecto a cada peso de la red.

El proceso opera en dos fases alternadas. La **propagación hacia adelante** (*forward propagation*) consiste en pasar los datos de entrada a través de todas las capas de la red, aplicando en cada capa l la transformación $\mathbf{h}^{(l)} = f(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)})$, donde $\mathbf{h}^{(0)} = \mathbf{x}$ es la entrada original. La información fluye capa por capa hasta obtener la predicción \hat{y} en la capa de salida. Seguidamente, se evalúa la función de pérdida $\mathcal{L}(\hat{y}, y)$ comparando la predicción con la etiqueta real.

La **retropropagación** (*backpropagation*) calcula el gradiente de la función de pérdida con respecto a cada peso de la red aplicando recursivamente la regla de la cadena del cálculo diferencial, desde la capa de salida hacia la capa de entrada. Para cada peso $w_{ij}^{(l)}$ de la capa l , se calcula $\partial\mathcal{L}/\partial w_{ij}^{(l)}$, que indica la dirección y magnitud del ajuste necesario para reducir el error. Finalmente, un algoritmo de optimización actualiza los pesos empleando la regla:

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}_t) \quad (2)$$

donde η es la tasa de aprendizaje (*learning rate*) y $\nabla_{\mathbf{w}} \mathcal{L}$ el gradiente de la función de pérdida respecto a los pesos. La tasa de aprendizaje controla el tamaño de cada paso de actualización: valores demasiado grandes provocan oscilaciones que impiden la convergencia, mientras que valores demasiado pequeños ralentizan el entrenamiento excesivamente (Goodfellow et al., 2016).

2.3.3.4 Hiperparámetros del entrenamiento

El proceso de entrenamiento de una red neuronal profunda se gobierna por un conjunto de hiperparámetros cuya configuración adecuada resulta determinante para el rendimiento del modelo:

El **tamaño de lote** (*batch size*) define el número de muestras procesadas simultáneamente antes de cada actualización de pesos. Lotes grandes proporcionan estimaciones más estables del gradiente pero requieren más memoria GPU; lotes pequeños introducen ruido que puede ayudar a escapar de mínimos locales pero incrementan la varianza de las actualizaciones.

Una **época** (*epoch*) corresponde a una pasada completa del algoritmo de entrenamiento por todo el conjunto de datos. Cada época procesa todos los lotes secuencialmente, actualizando los pesos tras cada lote. El entrenamiento se ejecuta durante múltiples

épocas, permitiendo que el modelo refine progresivamente sus representaciones. La monitorización de la función de pérdida en el conjunto de validación a lo largo de las épocas permite detectar el *sobreajuste*: Cuando la pérdida de entrenamiento continúa descendiendo pero la pérdida de validación comienza a ascender, el modelo está memorizando los datos en lugar de aprender patrones generalizables.

2.3.3.5 Técnicas de regularización

Las redes neuronales profundas, al disponer de millones de parámetros entrenables, son susceptibles al **sobreajuste** (*overfitting*): El modelo memoriza los datos de entrenamiento incluyendo el ruido y las particularidades irrelevantes en lugar de aprender patrones generalizables a datos nuevos. El fenómeno opuesto, el **subajuste** (*underfitting*), ocurre cuando el modelo carece de capacidad suficiente para capturar los patrones del dato. Las técnicas de regularización mitigan el sobreajuste penalizando la complejidad del modelo o introduciendo perturbaciones controladas durante el entrenamiento.

- Srivastava et al. (2014) introdujeron el **dropout**, técnica que durante el entrenamiento desactiva aleatoriamente una fracción p de las neuronas en cada iteración, estableciendo sus salidas en cero. Esta operación obliga a la red a desarrollar representaciones redundantes y robustas, pues ninguna neurona puede depender excesivamente de la presencia de otra. Conceptualmente, el dropout equivale a entrenar un ensamble implícito de 2^n subredes (donde n es el número de neuronas) y promediar sus predicciones durante la inferencia.
- Ioffe y Szegedy (2015) propusieron la **normalización por lotes** (*batch normalization*, BN), que normaliza las activaciones de cada capa restando la media μ_B y dividiendo por la desviación estándar σ_B del mini-batch, seguido de una transformación lineal aprendible con parámetros γ (escala) y β (desplazamiento):
- La **regularización por weight decay** añade un término de penalización $\lambda \|\mathbf{w}\|_2^2$ a la función de pérdida, desincentivando pesos de gran magnitud y favoreciendo soluciones más suaves. Como se mencionó anteriormente, AdamW implementa esta regularización de forma desacoplada para mayor efectividad.
- La **detención temprana** (*early stopping*) monitoriza la función de pérdida en el conjunto de validación durante el entrenamiento y detiene el proceso cuando dicha pérdida deja de mejorar durante un número predefinido de épocas consecutivas (*patience*), evitando que el modelo se sobreajuste al continuar entrenando innecesariamente. YOLOv8 implementa esta técnica con una paciencia configurable, típicamente de 50 épocas.

2.3.4. Redes neuronales convolucionales (CNN)

2.3.4.1 Fundamentos y motivación

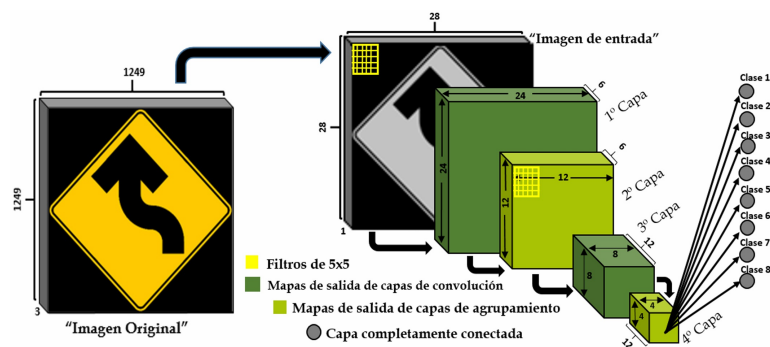
Las redes neuronales convolucionales (CNN, *Convolutional Neural Networks*) constituyen una clase especializada de redes neuronales profundas diseñadas para procesar datos con topología de rejilla, como las imágenes, que pueden conceptualizarse como rejillas

bidimensionales de píxeles. LeCun et al. (2015b) señalan que las CNN se distinguen de las redes completamente conectadas por tres principios arquitectónicos fundamentales.

Para ilustrar la eficiencia de las CNN, considérese una imagen de entrada de 640×640 píxeles con 3 canales de color (resolución estándar de YOLOv8): Esto equivale a 1228 800 valores de entrada. Si se empleara una red completamente conectada con una capa oculta de 1000 neuronas, se requerirían más de mil millones de pesos solo para la primera capa, un número inmanejable. Una CNN con un filtro de 3×3 aplicado sobre los 3 canales requiere únicamente 27 pesos (más un sesgo) por filtro, que se reutilizan en todas las posiciones de la imagen, reduciendo los parámetros en varios órdenes de magnitud (Goodfellow et al., 2016).

Figura 18

Esquema arquitectónico típico de una red neuronal convolucional



Nota. Adaptado de Vizcaya et al. (2017).

2.3.4.2 Operación de convolución

La operación de convolución constituye el mecanismo central de extracción de características en las CNN. Matemáticamente, la convolución discreta bidimensional entre una imagen de entrada \mathbf{I} y un kernel o filtro \mathbf{K} de dimensiones $k_h \times k_w$ produce un mapa de características (*feature map*) \mathbf{S} definido por:

$$S(i, j) = (\mathbf{I} * \mathbf{K})(i, j) = \sum_{m=0}^{k_h-1} \sum_{n=0}^{k_w-1} I(i+m, j+n) \cdot K(m, n) \quad (3)$$

En la práctica, el kernel es una matriz de tamaño reducido (típicamente 3×3 , 5×5 o 1×1) que se desplaza sobre toda la imagen de entrada realizando en cada posición el producto punto entre los valores del kernel y la región correspondiente de la imagen. Considérese un ejemplo concreto: dada una imagen de 6×6 píxeles y un kernel de 3×3 con stride 1, el kernel se posiciona en la esquina superior izquierda, multiplica cada uno de sus 9 valores por los 9 píxeles correspondientes de la imagen, suma los productos y genera el primer valor del mapa de características de salida. Luego se desplaza una posición a la derecha y repite la operación. Al completar una fila, desciende una posición y reinicia desde la izquierda, barriendo toda la imagen hasta producir un mapa de características de $(6-3+1) \times (6-3+1) = 4 \times 4$ (Goodfellow et al., 2016).

Los valores del kernel no se diseñan manualmente, sino que se **aprenden automáticamente** durante el entrenamiento mediante retropropagación. Cada capa convolucional aplica

múltiples kernels en paralelo (por ejemplo, 64, 128 o 256 filtros), generando múltiples mapas de características que codifican distintos patrones. Las primeras capas de una CNN aprenden a detectar características de bajo nivel como bordes horizontales, verticales y diagonales, gradientes de color y texturas elementales. Conforme se avanza en la profundidad de la red, los mapas de características codifican patrones progresivamente más abstractos.

2.3.5. Detección de objetos

2.3.5.1 Definición y formulación del problema

La detección de objetos constituye una de las tareas fundamentales de la visión por computadora cuyo objetivo es localizar e identificar simultáneamente todos los objetos de interés presentes en una imagen. A diferencia de la clasificación de imágenes (que asigna una única etiqueta a la imagen completa), la detección debe responder tres preguntas para cada objeto: *¿qué es?* (clasificación), *¿dónde está?* (localización mediante *bounding box*) y *¿con qué confianza?* (puntuación de certeza). Formalmente, dado una imagen I , el detector produce un conjunto de detecciones $\{(c_i, s_i, b_i)\}_{i=1}^N$, donde c_i es la clase predicha, s_i la puntuación de confianza y $b_i = (x, y, w, h)$ las coordenadas del *bounding box* (centro x, y ; ancho w ; alto h) que encuadra al objeto (Zou et al., 2023).

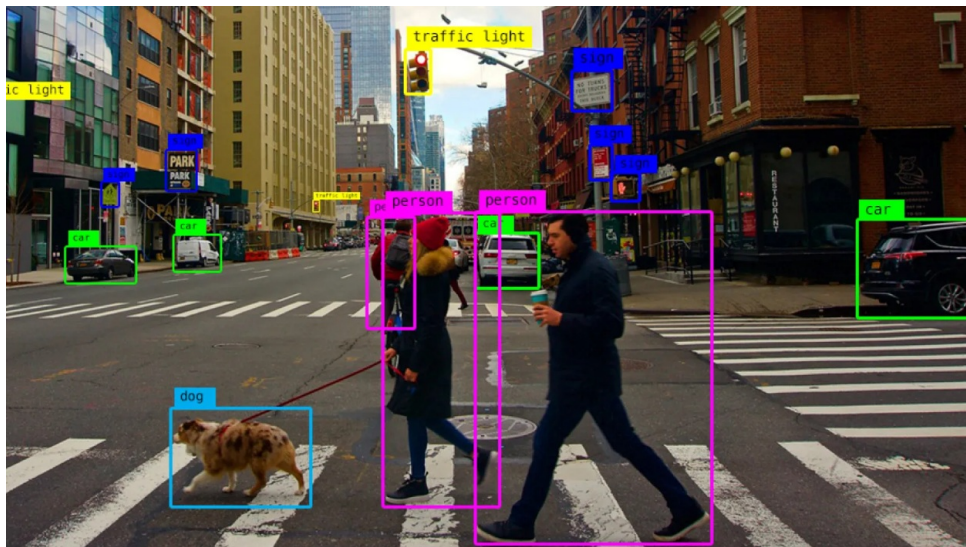
Zou et al. (2023), en su encuesta comprehensiva publicada en *Proceedings of the IEEE* que abarca dos décadas de investigación, clasifican los detectores de objetos basados en aprendizaje profundo en dos paradigmas fundamentales:

- **Detectores de dos etapas (*two-stage detectors*).** Estos métodos descomponen la detección en dos fases secuenciales: generación de propuestas de regiones y clasificación/refinamiento de cada propuesta. La familia R-CNN representa este paradigma. Girshick et al. (2014) propusieron **R-CNN**, que extraía aproximadamente 2000 regiones candidatas mediante el algoritmo *Selective Search* y clasificaba cada una con una CNN independiente, requiriendo 47 segundos por imagen. Girshick (2015) aceleraron significativamente el proceso con **Fast R-CNN** al procesar la imagen completa una sola vez con la CNN y compartir los mapas de características entre todas las regiones mediante una capa de *RoI Pooling*, reduciendo el tiempo a 0,3 segundos. Ren et al. (2017) alcanzaron velocidades cercanas al tiempo real con **Faster R-CNN** al reemplazar el algoritmo externo de propuestas por una *Region Proposal Network* (RPN) integrada en la propia red convolucional, que genera propuestas directamente a partir de los mapas de características del backbone, logrando 5 cuadros por segundo con un mAP del 73,2% en PASCAL VOC.
- **Detectores de una etapa (*one-stage detectors*).** Estos métodos eliminan la fase de propuestas de regiones y predicen directamente las clases y localizaciones de todos los objetos en una única pasada por la red. W. Liu et al. (2016) propusieron **SSD** (*Single Shot MultiBox Detector*), que realiza predicciones a múltiples escalas utilizando mapas de características de diferentes capas de la red, detectando objetos grandes en capas profundas (baja resolución) y objetos pequeños en capas tempranas (alta resolución).

Lin, Goyal et al. (2017) abordaron el problema del desbalance entre *foreground* (objetos) y *background* (fondo) en detectores de una etapa mediante la **Focal Loss**, función que reduce la contribución de los ejemplos fáciles correctamente clasificados y concentra el entrenamiento en los ejemplos difíciles, proponiendo **RetinaNet** como primer detector de una etapa que superó en precisión a los detectores de dos etapas. La familia YOLO, iniciada por Redmon et al. (2016a), reformuló la detección como un problema de regresión directa unificado, logrando velocidades de procesamiento en tiempo real que la hacen idónea para aplicaciones como la detección de elementos viales.

Figura 19

Detección de objetos



Nota. Adaptado de Augmented AI (2023).

2.3.5.2 Arquitectura backbone neck head

Los detectores de objetos modernos comparten una arquitectura modular de tres componentes que Zou et al. (2023) sistematizan de la siguiente manera:

- El **backbone** (columna vertebral) es la red convolucional encargada de extraer características jerárquicas multi-escala a partir de la imagen de entrada. Típicamente se emplea una red de clasificación pre-entrenada en ImageNet como ResNet, VGG o DarkNet a la que se le eliminan las capas de clasificación finales. El backbone genera mapas de características en múltiples resoluciones: las capas tempranas producen mapas de alta resolución espacial (80×80 para una entrada de 640×640) con características de bajo nivel semántico (bordes, texturas), mientras que las capas profundas producen mapas de baja resolución (20×20) con características semánticas de alto nivel (objetos completos, contexto de escena). Esta jerarquía resulta fundamental: los objetos pequeños como señales de tránsito distantes se detectan mejor en los mapas de alta resolución, mientras que los objetos grandes como camiones se detectan en los mapas de baja resolución.
- El **neck** (cuello) agrega y fusiona los mapas de características multi-escala

del backbone para crear representaciones enriquecidas. Lin, Dollár et al. (2017) propusieron la **Feature Pyramid Network** (FPN), que construye una pirámide de características mediante un camino descendente (*top-down*) con conexiones laterales: los mapas profundos, ricos en semántica pero pobres en resolución, se sobremuestran y se combinan con los mapas tempranos de alta resolución, logrando que cada nivel de la pirámide contenga información semántica fuerte independientemente de su resolución. S. Liu et al. (2018) extendieron FPN con **PANet** (*Path Aggregation Network*), añadiendo un camino ascendente (*bottom-up*) complementario que permite que las capas profundas también se beneficien de la información de alta resolución de las capas tempranas.

- La **head** (cabeza) genera las predicciones finales a partir de los mapas fusionados. En los detectores *anchor-based*, la cabeza predice ajustes (*offsets*) respecto a cajas predefinidas (*anchors*) de diferentes proporciones y tamaños, distribuidas sobre cada posición del mapa de características. En los detectores *anchor-free* como YOLOv8, la cabeza predice directamente los centros de los objetos y sus dimensiones sin depender de cajas predefinidas, simplificando el proceso y eliminando los hiperparámetros de diseño de anchors. YOLOv8 emplea una **cabeza desacoplada** (*decoupled head*) que procesa la clasificación y la regresión en ramas convolucionales independientes, eliminando la competencia por parámetros compartidos.

2.3.6. Familia YOLO: Evolución de YOLOv1 a YOLOv8

2.3.6.1 YOLOv1: La reformulación como regresión unificada

Redmon et al. (2016a) propusieron **YOLO** (*You Only Look Once*), reformulando radicalmente la detección de objetos como un problema de regresión unificado. La imagen de entrada se divide en una cuadrícula de $S \times S$ celdas ($S = 7$); cada celda es responsable de detectar los objetos cuyo centro geométrico cae dentro de ella, prediciendo $B = 2$ *bounding boxes* con sus respectivas puntuaciones de confianza (probabilidad de contener un objeto multiplicada por el IoU predicho) y $C = 20$ probabilidades de clase (para las clases de PASCAL VOC). El vector de salida de toda la imagen es un tensor de dimensiones $S \times S \times (B \times 5 + C) = 7 \times 7 \times 30$, producido en una única evaluación (*forward pass*) de la red. Esta formulación alcanzó 45 cuadros por segundo con un mAP del 63,4 % en PASCAL VOC, velocidad sin precedentes que habilitó la detección en tiempo real. Sus limitaciones incluían dificultades con objetos pequeños, objetos agrupados y la detección de múltiples objetos en una misma celda.

2.3.6.2 YOLOv2 (YOLO9000): *Anchor boxes* y entrenamiento multi-escala

Redmon y Farhadi (2017) mejoraron la arquitectura con **YOLOv2**, incorporando normalización por lotes en todas las capas convolucionales (mejorando el mAP en 2 puntos porcentuales), un backbone más profundo (Darknet-19 con 19 capas convolucionales), resoluciones de entrada múltiples durante el entrenamiento (*multi-scale training*), y *anchor boxes* predefinidas cuyas dimensiones se determinaron mediante clustering

k -means sobre las anotaciones del conjunto de entrenamiento.

2.3.6.3 YOLOv3: Predicción a tres escalas y Darknet-53

Redmon y Farhadi (2018) desarrollaron **YOLOv3**, cuya contribución principal fue la **predicción a tres escalas** distintas empleando un backbone Darknet-53 (53 capas convolucionales con conexiones residuales) y una estructura de fusión tipo FPN simplificada. Cada escala predice tres *bounding boxes* por posición, para un total de $3 \times (80 + 5) \times (N_1 + N_2 + N_3)$ predicciones, donde N_i es el número de celdas en cada escala. YOLOv3 mejoró significativamente la detección de objetos pequeños y medianos, alcanzando un mAP@0.5 del 57,9% en COCO con 51 ms por imagen. La clasificación multiclase adoptó clasificadores sigmoide independientes por clase en lugar de Softmax, permitiendo la asignación de múltiples etiquetas a un mismo objeto.

2.3.6.4 YOLOv4: Sistematización de mejoras

Bochkovskiy et al. (2020) sistematizaron con **YOLOv4** las técnicas de mejora clasificándolas en dos categorías: *bag of freebies* (BoF, técnicas que mejoran la precisión sin costo adicional en inferencia, como aumento de datos Mosaic, regularización DropBlock, pérdida CloU y label smoothing) y *bag of specials* (BoS, módulos que incrementan ligeramente el costo de inferencia pero mejoran sustancialmente la precisión, como el módulo SPP, bloques de atención SAM y conexiones PAN). La contribución más influyente fue el **Mosaic augmentation**, que combina cuatro imágenes de entrenamiento en una sola imagen compuesta, reduciendo la necesidad de lotes grandes al exponer a la red a mayor diversidad contextual en cada iteración. YOLOv4 alcanzó 43,5 % mAP en COCO a 65 FPS en una GPU Tesla V100.

2.3.6.5 YOLOv5: Democratización del acceso

Jocher et al. (2020) lanzaron **YOLOv5** como implementación nativa en PyTorch, facilitando la adopción industrial y académica mediante: una interfaz de entrenamiento de línea de comandos simplificada, exportación multi-formato (ONNX, TensorRT, CoreML, TFLite), *auto-anchor* (cálculo automático de dimensiones de anchors), autoajuste de hiperparámetros, y escalabilidad en cinco variantes (Nano, Small, Medium, Large, Extra-large) que cubren el espectro desde dispositivos embebidos hasta servidores de alto rendimiento.

2.3.6.6 YOLOv7: Reparametrización y entrenamiento eficiente

C.-Y. Wang et al. (2023) propusieron **YOLOv7**, introduciendo la noción de *trainable bag-of-freebies* y la técnica de reparametrización de modelos: durante el entrenamiento se emplean arquitecturas con múltiples ramas paralelas que enriquecen la capacidad de aprendizaje, pero durante la inferencia estas ramas se fusionan matemáticamente en una estructura secuencial equivalente, eliminando el costo computacional adicional sin pérdida de precisión.

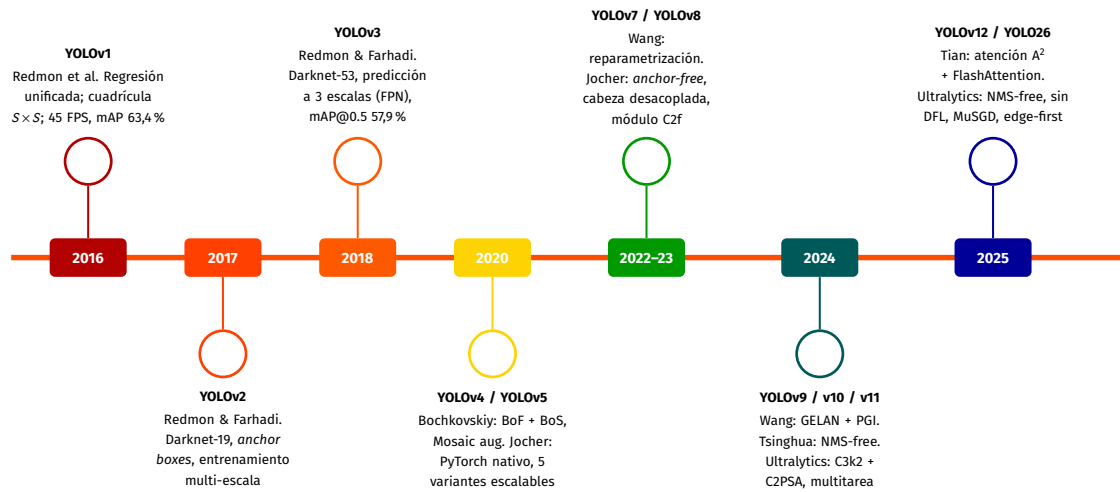
2.3.6.7 YOLOv8: Arquitectura del modelo empleado en esta investigación

Jocher et al. (2023a) presentaron **YOLOv8** como la evolución de la plataforma Ultralytics, incorporando las siguientes mejoras arquitectónicas:

- **Backbone: CSPDarknet con módulos C2f.** El backbone emplea una variante del CSPDarknet (*Cross Stage Partial Darknet*) donde el módulo C3 de YOLOv5 fue reemplazado por el módulo **C2f** (*Cross Stage Partial bottleneck con 2 convoluciones*). El módulo C2f divide los mapas de características en dos ramas paralelas: Una rama pasa directamente sin procesamiento adicional (identidad) y la otra atraviesa una secuencia de n *bottleneck blocks* con conexiones residuales. Las salidas de ambas ramas y las salidas intermedias de cada bottleneck se concatenan, combinando gradientes ricos de la rama procesada con la información preservada de la rama directa. Esta estrategia mejora el flujo de gradientes durante el entrenamiento y la eficiencia del aprendizaje. El backbone incluye además un módulo **SPPF** (*Spatial Pyramid Pooling Fast*) que aplica max pooling con kernels de 5×5 en cascada ($5 \rightarrow 5 \rightarrow 5$, equivalente a 5, 9, 13) para capturar contexto a múltiples escalas receptivas.
- **Head: detección anchor-free con cabeza desacoplada.** YOLOv8 abandona el paradigma de *anchor boxes* adoptando un enfoque **anchor-free** donde cada posición del mapa de características predice directamente el centro del objeto y sus cuatro límites (distancias del centro a cada borde del *bounding box*), sin necesidad de ajustes respecto a cajas predefinidas. La cabeza de detección está **desacoplada** (*decoupled*), separando las predicciones de clasificación (rama convolucional independiente con activación sigmoide para la probabilidad de cada clase) y de regresión (rama convolucional independiente con *Distribution Focal Loss* para las coordenadas), mejorando el rendimiento de ambas tareas.
- **Funciones de pérdida.** YOLOv8 emplea una combinación ponderada de tres funciones de pérdida: **Binary Cross-Entropy** (BCE) para la rama de clasificación, que calcula la entropía cruzada binaria independiente para cada clase; **Complete IoU Loss** (CIoU) para la rama de regresión, que considera simultáneamente la distancia entre centros, la superposición de áreas y la consistencia de proporciones (*aspect ratio*); y **Distribution Focal Loss** (DFL) que modela la distribución de las coordenadas del *bounding box* como una distribución de probabilidad discreta sobre valores posibles, mejorando la precisión de la localización para objetos ambiguos.
- **Neck: FPN + PAN bidireccional.** El neck implementa la combinación de Feature Pyramid Network (camino descendente *top-down* para enriquecer semánticamente las capas de alta resolución) y Path Aggregation Network (camino ascendente *bottom-up* para enriquecer espacialmente las capas de baja resolución), fusionando mapas de características a tres escalas: P3 (stride 8, resolución 80×80 para entrada de 640×640 , óptima para objetos pequeños como señales de tránsito distantes y tachas retroreflectivas), P4 (stride 16, resolución 40×40 , para objetos medianos como peatones y motocicletas) y P5 (stride 32, resolución 20×20 , para objetos grandes como

camiones y autobuses).

Figura 20
Evolución cronológica de la familia YOLO: de YOLOv1 a YOLOv6



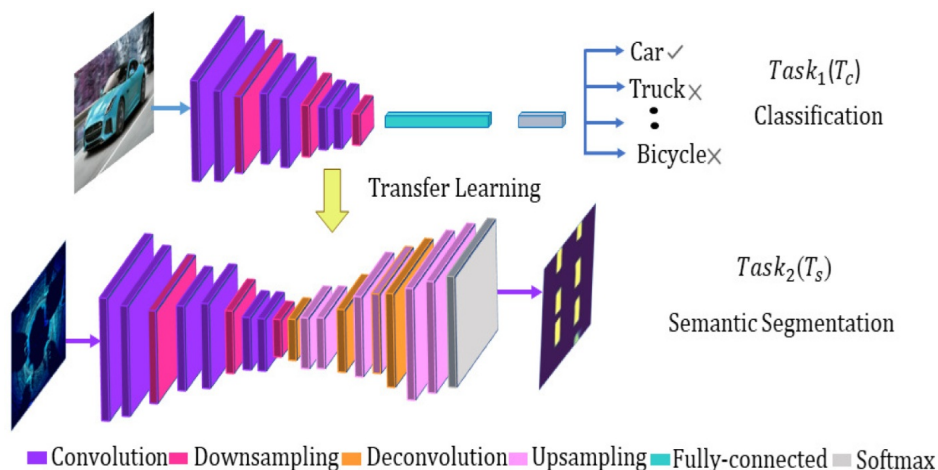
Nota. Elaboración propia basada en la revisión de publicaciones de la familia YOLO (2016–2025).

2.3.7. Aprendizaje por transferencia (*transfer learning*)

2.3.7.1 Fundamentos y formalización

El aprendizaje por transferencia (*transfer learning*) constituye una estrategia de entrenamiento que aprovecha el conocimiento adquirido por un modelo en una tarea fuente (generalmente con abundantes datos) para mejorar el aprendizaje en una tarea objetivo (frecuentemente con datos limitados). S. J. Pan y Yang (2010), en su encuesta fundacional con más de 22 000 citas, formalizan el transfer learning como un problema donde, dados un dominio fuente \mathcal{D}_S con una tarea de aprendizaje \mathcal{T}_S y un dominio objetivo \mathcal{D}_T con una tarea \mathcal{T}_T , se busca mejorar el aprendizaje de la función predictiva $f_T(\cdot)$ en \mathcal{D}_T usando el conocimiento de \mathcal{D}_S y \mathcal{T}_S , donde $\mathcal{D}_S \neq \mathcal{D}_T$ o $\mathcal{T}_S \neq \mathcal{T}_T$.

Figura 21
Representación del transfer learning



Nota. Adaptado de Imad et al. (2021).

La necesidad del transfer learning se fundamenta en una limitación práctica recurrente: La compilación y anotación de conjuntos de datos masivos y específicos para cada nueva tarea resulta costosa en tiempo, recursos humanos y financieros. El entrenamiento de una red neuronal profunda desde cero (*training from scratch*) con un dataset limitado conduce típicamente a sobreajuste severo. El transfer learning mitiga este problema proporcionando al modelo un punto de partida informado pesos pre-entrenados que ya codifican características visuales genéricas , reduciendo drásticamente la cantidad de datos y el tiempo de entrenamiento necesarios para alcanzar un rendimiento competitivo (S. J. Pan & Yang, 2010).

2.3.7.2 Transferencia de características en redes convolucionales

Yosinski et al. (2014) demostraron empíricamente que las capas de las CNN aprenden características de generalidad decreciente conforme aumenta la profundidad. Las **primeras capas** (capas 1 a 3 en una red típica) aprenden características genéricas y altamente transferibles: detectores de bordes orientados (filtros de Gabor), detectores de manchas cromáticas (*color blobs*), patrones de textura y gradientes direccionales. Estas características son prácticamente idénticas independientemente de si la red fue entrenada para clasificar animales, vehículos o escenas naturales. Las **capas intermedias** (capas 4 a 6) desarrollan características semi-transferibles que comienzan a especializarse en la tarea pero mantienen relevancia para tareas relacionadas: combinaciones de bordes que forman esquinas, patrones repetitivos de texturas específicas, formas geométricas elementales. Las **capas profundas** (capas 7 en adelante) aprenden características altamente específicas para la tarea de entrenamiento: partes de objetos, configuraciones espaciales particulares, patrones contextuales propios del dominio.

2.3.8. Aumento de datos (*data augmentation*)

2.3.8.1 Fundamentos y justificación

El aumento de datos (*data augmentation*) comprende un conjunto de técnicas que generan variaciones artificiales de las imágenes de entrenamiento aplicando transformaciones que preservan la semántica de las anotaciones (las etiquetas de clase y la localización de los objetos se ajustan consistentemente con la transformación aplicada), con el propósito de incrementar la diversidad efectiva del conjunto de datos y mejorar la capacidad de generalización del modelo. Shorten y Khoshgoftaar (2019), en su encuesta exhaustiva con más de 7000 citas, demuestran que el aumento de datos constituye la técnica de regularización más efectiva cuando el tamaño del dataset es limitado, superando incluso al dropout y al weight decay en muchos escenarios.

La justificación teórica del aumento de datos se fundamenta en la teoría de la generalización: Un modelo generaliza mejor cuando ha sido expuesto a una mayor diversidad de variaciones de los datos durante el entrenamiento. Si el conjunto de datos original contiene únicamente imágenes capturadas bajo condiciones de iluminación diurna, el modelo tendrá dificultades para detectar objetos en condiciones de atardecer,

amanecer o iluminación artificial. Al generar artificialmente variaciones de brillo, contraste y tonalidad, se simula esta diversidad sin necesidad de capturar físicamente imágenes adicionales.

2.3.9. Métricas de evaluación para detección de objetos y segmentación de instancias

2.3.9.1 Intersección sobre Unión (IoU)

La métrica fundamental para evaluar la calidad de la localización en detección de objetos es la **Intersección sobre Unión** (*Intersection over Union*, IoU), definida como la razón entre el área de intersección y el área de unión entre el *bounding box* predicho B_p y el *bounding box* de referencia (*ground truth*) B_{gt} :

$$\text{IoU} = \frac{|B_p \cap B_{gt}|}{|B_p \cup B_{gt}|} = \frac{\text{Área de intersección}}{\text{Área de unión}} \quad (4)$$

El IoU varía entre 0 (sin solapamiento alguno entre las cajas) y 1 (coincidencia perfecta donde ambas cajas son idénticas). Un umbral de IoU determina si una detección se considera correcta o incorrecta: típicamente, $\text{IoU} \geq 0,5$ para el estándar PASCAL VOC (una detección se acepta si al menos la mitad de su área se solapa con el objeto real) y umbrales variables de 0,5 a 0,95 en incrementos de 0,05 para el estándar más exigente de COCO. Para la segmentación de instancias, el IoU se calcula sobre las **máscaras binarias** en lugar de los *bounding boxes*, comparando la coincidencia pixel a pixel entre la máscara predicha y la máscara de referencia, lo que constituye una evaluación más rigurosa de la calidad de la segmentación (Padilla et al., 2020).

Variantes mejoradas del IoU empleadas como funciones de pérdida durante el entrenamiento incluyen: **GIoU** (*Generalized IoU*) que penaliza la distancia entre cajas que no se solapan, **DIoU** (*Distance IoU*) que minimiza la distancia entre los centros de las cajas, y **CIoU** (*Complete IoU*) que adicionalmente considera la consistencia de proporciones (*aspect ratio*), esta última empleada por YOLOv8 en su función de pérdida de regresión.

2.3.9.2 Clasificación de detecciones: TP, FP, FN

En el contexto de la detección de objetos, las métricas clásicas de clasificación se redefinen incorporando el umbral de IoU. Dada una predicción del modelo y un conjunto de objetos reales anotados (*ground truth*), cada detección se clasifica como:

- **Verdadero Positivo (TP):** Detección cuyo IoU con algún objeto real de la misma clase es \geq al umbral. Es decir, el modelo detectó correctamente un objeto que realmente existe. Si un mismo objeto real es detectado múltiples veces, solo la detección con mayor IoU se cuenta como TP; las restantes se cuentan como FP (detecciones duplicadas).
- **Falso Positivo (FP):** Detección que no corresponde a ningún objeto real con $\text{IoU} \geq$ al umbral, o detección duplicada de un objeto ya asociado a otra detección de mayor confianza. Los FP representan “falsas alarmas” del modelo: detectar un vehículo donde no hay ninguno, confundir una sombra con un peatón, o generar detecciones redundantes.

- **Falso Negativo (FN):** Objeto real presente en la imagen que el modelo no detectó. Los FN representan “omisiones” que en el contexto de la seguridad vial pueden tener consecuencias graves: no detectar un peatón en un cruce, ignorar un semáforo en rojo o pasar por alto un reductor de velocidad.

2.3.9.3 Precisión, Recall y F1-Score

$$\text{Precisión} = \frac{TP}{TP + FP} \quad (5)$$

La **precisión** (*precision*) cuantifica la proporción de detecciones emitidas por el modelo que son correctas. Una precisión alta indica que cuando el modelo anuncia la presencia de un objeto, generalmente tiene razón. En el contexto de la seguridad vial, una precisión baja generaría excesivas falsas alarmas que saturarían al sistema o al conductor con información errónea, reduciendo la confianza en el modelo.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

El **recall** (*sensibilidad*) cuantifica la proporción de los objetos reales presentes en la imagen que fueron efectivamente detectados por el modelo. Un recall alto indica que el modelo no omite objetos. En el contexto de la seguridad vial, un recall bajo es particularmente peligroso: no detectar un peatón o un semáforo puede generar una situación de riesgo.

$$\text{F1-Score} = 2 \times \frac{\text{Precisión} \times \text{Recall}}{\text{Precisión} + \text{Recall}} \quad (7)$$

El **F1-Score** es la media armónica de precisión y recall, proporcionando una métrica única que equilibra ambos criterios. Toma valores entre 0 y 1, donde 1 indica un detector perfecto sin falsos positivos ni falsos negativos. La media armónica penaliza severamente los valores extremos: si la precisión es 0,95 pero el recall es 0,10, el F1-Score es apenas 0,18, reflejando que un detector que omite el 90 % de los objetos es inaceptable independientemente de que sus pocas detecciones sean precisas (Padilla et al., 2020).

2.3.9.4 Curva Precisión-Recall y Average Precision (AP)

La relación entre precisión y recall no es fija, sino que varía según el umbral de confianza aplicado a las detecciones. Un umbral de confianza alto (por ejemplo, 0,9) produce pocas detecciones, todas de alta certeza: alta precisión pero bajo recall, pues muchos objetos reales quedan por debajo del umbral y no se detectan. Un umbral bajo (por ejemplo, 0,1) produce muchas detecciones, incluyendo numerosas falsas alarmas: bajo precisión pero alto recall. La **curva Precisión-Recall** (P-R) grafica esta relación paramétrica a medida que el umbral de confianza varía de 1 a 0, y el **Average Precision** (AP) se define como el área bajo esta curva:

$$\text{AP} = \int_0^1 p(r) dr \quad (8)$$

donde $p(r)$ es la precisión en función del recall r . En la práctica, la integral se aproxima mediante la interpolación de 101 puntos (*101-point interpolation*) empleada por el

benchmark COCO. Un AP de 1,0 indica un detector perfecto (precisión de 100 % en todos los niveles de recall), mientras que un AP de 0,0 indica un detector completamente inútil. El AP resume en un único valor escalar el rendimiento completo del detector para una clase específica a un umbral de IoU dado, capturando tanto la capacidad de evitar falsas alarmas (precisión) como la capacidad de detectar todos los objetos (recall) (Everingham et al., 2010).

2.3.9.5 Mean Average Precision (mAP)

El **mean Average Precision** (mAP) es el promedio del AP calculado sobre todas las clases del problema:

$$\text{mAP} = \frac{1}{N_c} \sum_{c=1}^{N_c} \text{AP}_c \quad (9)$$

donde N_c es el número de clases (doce en la presente investigación). Se emplean dos protocolos estándar según el benchmark de referencia:

El **mAP@0.5** (protocolo PASCAL VOC, Everingham et al. (2010)) calcula el AP utilizando un umbral de IoU fijo de 0,5. Este protocolo es relativamente indulgente, aceptando detecciones cuya localización es aproximada. Resulta útil para evaluar la capacidad del modelo de identificar la presencia y clase de los objetos.

El **mAP@0.5:0.95** (protocolo COCO, Lin et al. (2014)) promedia los valores de mAP calculados a diez umbrales de IoU equidistantes $\{0,50, 0,55, 0,60, \dots, 0,95\}$, constituyendo una métrica más exigente que penaliza tanto las detecciones imprecisas (bajo IoU) como las omisiones. Este protocolo, adoptado como estándar de la comunidad de visión por computadora, es el empleado como métrica principal en la presente investigación.

Para la segmentación de instancias, se calcula adicionalmente el **mAP^{mask}** (o **mAP_{seg}**), donde el IoU se computa sobre las máscaras de segmentación binarias en lugar de los *bounding boxes*, evaluando así la precisión de la delineación pixel a pixel de cada objeto. Típicamente, **mAP^{mask}** es inferior al mAP de detección (*bounding box*) porque la segmentación requiere una localización más precisa a nivel de píxel.

2.3.9.6 Matriz de confusión en detección multiclase

La **matriz de confusión** para un detector de objetos multiclase es una tabla de dimensiones $(N_c+1) \times (N_c+1)$ —las N_c clases de objetos más una fila/columna de “fondo” (*background*)—que registra la correspondencia entre las clases predichas (columnas) y las clases reales (filas) de cada detección. El elemento (i, j) de la matriz indica cuántas veces un objeto de clase i fue predicho como clase j . Los elementos diagonales representan las predicciones correctas; los elementos fuera de la diagonal revelan los patrones de confusión entre clases.

La interpretación de la matriz de confusión resulta esencial para dirigir las mejoras iterativas del modelo en la presente investigación. Por ejemplo, si la matriz revela confusiones frecuentes entre mototaxis y motocicletas (clases visualmente similares

en perspectiva frontal), puede indicar la necesidad de aumentar las muestras de entrenamiento que muestren las diferencias estructurales (tres ruedas vs. dos ruedas, cabina vs. ausencia de cabina). Si cierto tipo de señal vertical es sistemáticamente ignorado (alta concentración en la fila “fondo”), puede sugerir que la señal está insuficientemente representada en el dataset o que su tamaño en las imágenes es demasiado pequeño para la resolución del modelo (Padilla et al., 2020).

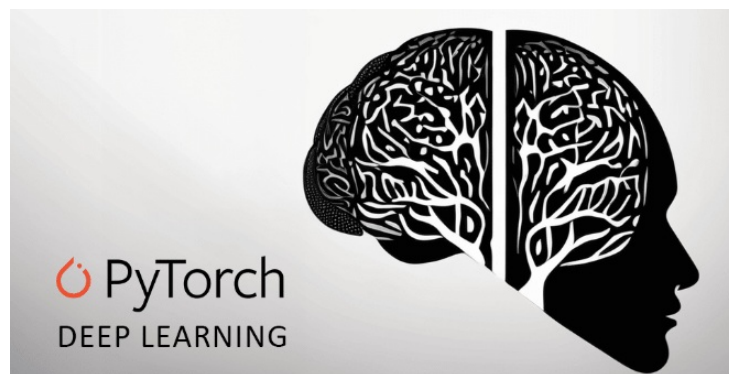
2.3.10. Herramientas y *frameworks* de implementación

2.3.10.1 PyTorch y Ultralytics

PyTorch es el framework de aprendizaje profundo de código abierto desarrollado por Meta AI Research. Su paradigma de grafo computacional dinámico (*define-by-run*) permite definir y modificar la arquitectura de la red durante la ejecución, facilitando la depuración y la experimentación. PyTorch gestiona automáticamente la diferenciación automática (cálculo eficiente de gradientes), la transferencia de datos entre CPU y GPU, y la distribución del entrenamiento en múltiples dispositivos.

Figura 22

Entorno PyTorch



Nota. Adaptado de Padilla et al. (2020).

2.3.10.2 OpenCV

OpenCV (*Open Source Computer Vision Library*) es la biblioteca de visión por computadora más utilizada a nivel mundial, con más de 2500 algoritmos optimizados para el procesamiento de imágenes y video en tiempo real. En el contexto de la presente investigación, OpenCV se emplea para la captura y decodificación de frames de video, la transformación de espacios de color, el redimensionamiento de imágenes, la visualización de detecciones (dibujo de *bounding boxes* y máscaras sobre las imágenes) y la escritura de los resultados en formato de video.

3. MARCO METODOLÓGICO

3.1. Método

La presente investigación adopta un método **deductivo** debido a que su estructura metodológica parte de teorías generales consolidadas sobre inteligencia artificial, redes neuronales convolucionales y sistemas de transporte inteligente para derivar una solución tecnológica específica aplicada al entorno vial de las avenidas Arenales y Cusco en Ayacucho. Este enfoque se fundamenta en los siguientes principios científicos validados:

- Principios del modelo de cámara pinhole para estimación de distancias monoculares en sistemas ADAS.
- Fundamentos de redes neuronales convolucionales (CNN) aplicadas a tareas de detección y segmentación de objetos en tiempo real.
- Estándares internacionales de sistemas de alerta de colisión frontal establecidos por la NHTSA (*National Highway Traffic Safety Administration*).
- Marco teórico del seguimiento multi-objeto mediante el algoritmo ByteTrack para estimación de velocidad y cómputo de Tiempo de Colisión (TTC).

La investigación procede **deductivamente** al tomar estos fundamentos teóricos universales y aplicarlos sistemáticamente al problema localizado de detección de usuarios vulnerables de la vía pública en el entorno urbano andino de Ayacucho.

3.2. Orientación

La investigación se clasifica como de orientación **aplicada** porque desarrolla una solución tecnológica concreta — el sistema QHAWAY ADAS — para abordar el problema específico de detección de usuarios vulnerables en las avenidas Arenales y Cusco de Ayacucho. El estudio trasciende el ámbito teórico al generar un sistema funcional que integra los siguientes componentes:

- Segmentación de instancias en tiempo real mediante YOLOv8L-seg y YOLO26L-seg para 12 clases de elementos viales críticos del entorno ayacuchano.
- Estimación de distancias monoculares mediante el modelo de cámara pinhole calibrado para la configuración de montaje en vehículo Toyota Hilux.
- Seguimiento multi-objeto mediante ByteTrack con asignación persistente de identidades de trayectoria.
- Clasificación de zonas de riesgo operacional: zona de peligro ($D < 3,0$ m o $TTC < 2.5$ s) y zona de precaución ($3,0 \leq D < 7,0$ m).
- Generación de alertas de voz en español mediante Microsoft Azure Neural TTS (es-PE-AlexNeural) con gestor de prioridades y cooldown.
- Interfaz gráfica de usuario (PyQt5) con superposición de máscaras de segmentación, etiquetas de distancia y banners de alerta en tiempo real.

La orientación **aplicada** se confirma por el enfoque en la implementación práctica de un sistema tecnológico validado bajo condiciones reales de conducción naturalista en el

entorno vial ayacuchano, sin escenarios artificiales ni situaciones provocadas.

3.3. Enfoque

La investigación adopta un enfoque **cuantitativo** porque se fundamenta en la recolección y análisis de datos numéricos medibles para evaluar el desempeño del sistema desarrollado. Las variables cuantificables del estudio se organizan en tres categorías:

Métricas de detección y segmentación:

- **mAP50 Box:** precisión media de detección por bounding box al umbral IoU = 0.50.
- **mAP50 Mask:** precisión media de segmentación de instancias al umbral IoU = 0.50.
- **mAP50-95 Box y Mask:** métricas estrictas multi-umbral (IoU de 0.50 a 0.95).
- **Precisión:** proporción de detecciones correctas sobre el total de detecciones realizadas.
- **Recall:** proporción de objetos presentes correctamente identificados por el modelo.
- **F1-score:** media armónica de precisión y recall por clase y agregado.

Métricas de rendimiento en tiempo real:

- **FPS (Frames Per Second):** velocidad de procesamiento del sistema (rango obtenido: 19.2–25.4 FPS).
- **Latencia (ms/frame):** tiempo de respuesta extremo a extremo del pipeline (rango obtenido: 17–20 ms/frame).

Métricas de validación geométrica de distancias:

- D_{geo} : distancia de referencia geométrica independiente (m).
- D_{est} : distancia estimada por el sistema QHAWAY (m).
- ε : residual absoluto entre referencia y estimación (m).

Estas métricas son procesadas estadísticamente para establecer conclusiones objetivas y replicables sobre el desempeño del sistema desarrollado, siguiendo los estándares de evaluación del conjunto de datos COCO ampliamente reconocidos en la literatura científica internacional de visión computacional.

3.4. Instrumento de recolección de datos

La presente investigación adopta un enfoque **prolectivo** debido a que los datos fueron recopilados directamente mediante sistemas de monitoreo implementados específicamente para el estudio. Este enfoque se manifiesta en cuatro métodos complementarios de adquisición de datos:

- **Método 1 — Grabación de video desde vehículo en movimiento:** registros en Full HD (1920 × 1080 px, 30 fps) mediante smartphone Redmi Note 12 Pro (sensor nativo 108 MP) montado en soporte de ventosa sobre el parabrisas de un vehículo Toyota Hilux a 1.50 m de altura, en múltiples rutas bajo condiciones diurnas y nocturnas.
- **Método 2 — Fotografías georreferenciadas:** capturadas mediante la aplicación GPS Map Camera, que estampa automáticamente coordenadas geográficas, fecha y hora en cada fotografía, constituyendo el registro documental del levantamiento métrico de reductores de velocidad informales.

- **Método 3 — Fotografías convencionales:** capturadas en el centro histórico de Ayacucho, el campus de la UNSCH y diversas intersecciones urbanas de los cinco distritos de la provincia de Huamanga.
- **Método 4 — Capturas de Google Earth Pro:** en modo Street View y capturas a nivel de calle, permitiendo cobertura de zonas urbanas extensas incluyendo áreas de acceso vehicular restringido.

La naturaleza **prolectiva** del instrumento garantiza que los datos recolectados sean específicos para los objetivos de la investigación, permitiendo un control directo sobre la calidad, precisión y relevancia de la información obtenida.

3.5. Tipo o alcance de la investigación

La presente investigación adopta un alcance de **desarrollo tecnológico** debido a que su propósito fundamental es la creación, implementación y validación empírica del sistema QHAWAY ADAS, una solución tecnológica innovadora basada en segmentación de instancias y estimación de distancias monoculares para la detección de usuarios vulnerables de la vía pública en el entorno urbano andino de Ayacucho.

A diferencia de un estudio explicativo-causal, el presente trabajo no pretende demostrar estadísticamente la reducción de accidentes de tránsito — lo cual requeriría un diseño experimental controlado con grupos de tratamiento y control, medición longitudinal de siniestros y análisis de impacto causal. En cambio, el alcance de la investigación se circunscribe a los siguientes objetivos tecnológicos:

- Construir un dataset híbrido representativo del entorno vial andino (25,602 imágenes, 127,525 instancias anotadas en 12 clases).
- Documentar cuantitativamente la no conformidad de los reductores de velocidad informales respecto a la Directiva MTC N°01-2011-MTC/14.
- Entrenar y comparar arquitecturas de detección y segmentación: YOLOv8L-seg y familia YOLO26 (nano, small, large).
- Evaluar el desempeño del sistema mediante métricas estándar internacionales (mAP50 Box, mAP50 Mask, Precisión, Recall, F1).
- Demostrar la operación del sistema en tiempo real bajo condiciones naturalistas de conducción en Ayacucho.
- Validar geoméricamente el estimador de distancias monoculares mediante referencias métricas independientes del entorno de estudio.

3.6. Nivel de investigación

La presente investigación corresponde a un nivel **tecnológico-aplicativo** debido a que está dirigida fundamentalmente hacia el desarrollo y validación de una solución tecnológica funcional — el sistema QHAWAY ADAS — que aplica directamente conocimientos consolidados en redes neuronales convolucionales, segmentación de instancias y sistemas ADAS al entorno vial específico de Ayacucho.

Los productos tecnológicos generados por la investigación son:

- **Dataset QHAWAY:** 25,602 imágenes con 127,525 instancias anotadas en 12 clases, el primero en representar el entorno vial andino informal con el mototaxi como clase independiente.
- **Modelo YOLO26L-seg entrenado:** mejor rendimiento con mAP50 Box = 0.829 y mAP50 Mask = 0.788 a época 179, operando a 19.2–25.4 FPS.
- **Sistema QHAWAY ADAS:** pipeline completo de detección, segmentación, seguimiento, estimación de distancias y alerta de voz en español, validado bajo condiciones naturalistas.
- **Registro métrico de reductores informales:** primera documentación cuantitativa de no conformidad (96 % de incumplimiento) respecto a la Directiva MTC N°01-2011-MTC/14 en la región andina.

Este nivel tecnológico-aplicativo fue reconocido por la revista **MDPI Sensors** con los siguientes indicadores de impacto:

- Factor de Impacto JCR 2024: **3.5**
- CiteScore Scopus 2024: **8.2**
- Cuartil JCR (Web of Science): **Q2**
- Cuartil SCImago (Scopus): **Q1**
- Índice h: **273**
- Bases de indexación: Web of Science (SCIE), Scopus, PubMed, DOAJ

3.7. Diseño de investigación

3.7.1. Propósito de la investigación

La investigación se fundamenta en un propósito **no experimental** porque el diseño metodológico no involucra la manipulación controlada de variables independientes en condiciones de laboratorio. El sistema QHAWAY fue implementado y validado en las avenidas Arenales y Cusco bajo condiciones **naturalistas de conducción**, registrando eventos auténticos del tráfico ayacuchano sin escenarios provocados ni situaciones artificiales. Los eventos capturados durante la validación incluyeron:

- Detección de dos niños que cruzaron inesperadamente la calzada para recuperar una pelota.
- Peatón distraído consultando un dispositivo móvil al emerger detrás de un camión de carga estacionado.
- Reductor de velocidad no conforme en vía sin señalización previa.
- Gestión simultánea de semáforo en rojo con cinco peatones en escena nocturna (FPS: 19.2).

Este enfoque naturalista garantiza la validez ecológica de los resultados obtenidos, representando condiciones reales de operación del sistema en el entorno urbano andino.

3.7.2. Número de mediciones

La investigación adopta un diseño **transversal** porque la recolección de datos se realizó durante un período temporal específico y delimitado, capturando información del comportamiento del tráfico vehicular y peatonal en las avenidas estudiadas. Los datos recolectados fueron utilizados para:

- Construcción del dataset híbrido local (4,598 imágenes, 10,701 instancias).
- Entrenamiento del modelo en tres fases progresivas (640, 800 y 1024 px).
- Validación del desempeño en el conjunto de validación (5,120 imágenes).
- Validación operacional bajo condiciones naturalistas de conducción.

3.7.3. Cronología de las observaciones

La investigación adopta un enfoque **prospectivo** porque los datos recolectados están destinados a entrenar un modelo que opera en tiempo real sobre nuevas escenas no vistas durante el entrenamiento, generando las siguientes capacidades prospectivas:

- Detección y segmentación de objetos viales en fotogramas nuevos no pertenecientes al conjunto de entrenamiento.
- Estimación de distancias en tiempo real para objetos detectados en escenas futuras.
- Generación de alertas de voz anticipatorias basadas en la clasificación de zonas de riesgo.
- Generalización del modelo a condiciones de iluminación, clima y densidad de tráfico no vistas durante el entrenamiento.

3.8. Estudio del diseño de investigación

La investigación corresponde a un **estudio de desarrollo tecnológico** debido a que su objetivo central es crear, implementar y validar el sistema QHAWAY ADAS, una solución tecnológica innovadora que integra segmentación de instancias (YOLOv8L-seg / YOLO26L-seg), estimación de distancias monoculares (modelo pinhole), seguimiento multi-objeto (ByteTrack) y síntesis de voz (Microsoft Azure Neural TTS) para la detección de usuarios vulnerables de la vía pública en el entorno urbano andino de Ayacucho.

La metodología de desarrollo tecnológico comprende las siguientes etapas:

- **Etap 1 — Levantamiento de campo:** encuesta métrica sistemática de reductores de velocidad informales en las avenidas de estudio, con registro GPS, altura, longitud de cuerda, condición de pintura y señalización.
- **Etap 2 — Construcción del dataset:** recolección de 4,598 imágenes locales mediante cuatro métodos complementarios e integración con tres datasets internacionales (BDD100K, BSTLD, RLMD) para conformar el dataset híbrido QHAWAY de 25,602 imágenes.
- **Etap 3 — Anotación:** etiquetado de 127,525 instancias en 12 clases mediante la plataforma LabelMe con flujo de trabajo asistido por SAM2, con revisión cruzada independiente del 20 % de las anotaciones.

- **Etap 4 — Entrenamiento progresivo:** estrategia en tres fases con resolución creciente (640, 800 y 1024 px) evaluada como estudio de ablación, comparando YOLOv8L-seg y la familia YOLO26 (nano, small, large).
- **Etap 5 — Integración del sistema:** desarrollo del pipeline completo QHAWAY ADAS con los módulos de captura, detección, seguimiento, estimación de distancias, gestión de alertas e interfaz gráfica.
- **Etap 6 — Validación operacional:** prueba del sistema bajo condiciones naturalistas de conducción en las avenidas Arenales y Cusco, evaluando FPS, latencia y comportamiento ante eventos reales del tráfico ayacuchano.
- **Etap 7 — Validación geométrica:** verificación cuantitativa del estimador de distancias mediante referencias métricas independientes del entorno de estudio (juntas de losas, ancho de calzada, geometría vehicular).
- **Etap 8 — Análisis de explicabilidad:** aplicación de tres algoritmos Grad-CAM (Grad-CAM, Grad-CAM++, EigenCAM) para validar la coherencia semántica de las representaciones aprendidas por el modelo.

3.9. Población y muestra

3.9.1. Población

La población de estudio está constituida por el conjunto total de imágenes del entorno vial de las avenidas Arenales y Cusco de Ayacucho y sus zonas de influencia, conformado por **25,602 imágenes** que representan los elementos críticos de la vía pública: Usuarios vulnerables (personas), vehículos (motocicletas, automóviles, camiones, buses y mototaxis), infraestructura vial informal (reductores de velocidad), señalizaciones horizontales (pasos peatonales, líneas de carril) y dispositivos de control del tránsito (semáforos).

Esta población incluye dos componentes complementarios:

- **Componente local (4,598 imágenes):** imágenes originales del entorno vial ayacuchano, recolectadas mediante cuatro métodos de captura complementarios (video desde vehículo en movimiento, fotografías georreferenciadas con GPS Map Camera, fotografías convencionales y capturas de Google Earth Pro en modo Street View) a lo largo de las cinco avenidas de estudio: Arenales, Cusco, 9 de Diciembre, Las Américas e Independencia, cubriendo los cinco distritos urbanos de la provincia de Huamanga.
- **Componente internacional (21,004 imágenes):** imágenes provenientes de tres conjuntos de datos públicos de reconocimiento internacional — BDD100K, BSTLD y RLMD — incorporadas para ampliar la variabilidad de condiciones de iluminación, perspectiva y densidad de tráfico, y fortalecer la capacidad de generalización del modelo en las clases compartidas con el entorno local.

La población abarca una totalidad de **127,525 instancias anotadas** en 12 clases de elementos viales, recolectadas bajo diversas condiciones de iluminación (diurna, crepuscular y

nocturna), densidad de tráfico y estado de conservación de la infraestructura vial, representando el universo de escenarios documentados en el área de estudio.

3.9.2. Muestra

La muestra corresponde al **subconjunto local** del dataset híbrido QHAWAY, conformado por **4,598 imágenes originales** capturadas directamente en el entorno vial ayacuchano, con un total de **10,701 instancias anotadas** en las 12 clases definidas para el modelo. Este subconjunto constituye la contribución original de la presente investigación al conjunto de datos híbrido y representa las características específicas del entorno vial andino informal que no se encuentran en repositorios internacionales.

- **Criterio de selección:** muestreo intencional por diversidad de condiciones, garantizando la representación de las 12 clases de elementos viales en variedad de escenarios, ángulos de captura, condiciones de iluminación y estados de conservación de la infraestructura.
- **Partición del dataset híbrido completo:** 70 % para entrenamiento (17,921 imágenes), 20 % para validación (5,120 imágenes) y 10 % para prueba (2,561 imágenes).
- **Clase exclusiva del entorno local:** el mototaxi (357 instancias locales) constituye la clase de mayor relevancia contextual, ausente en todos los repositorios internacionales incorporados al dataset híbrido.

Cuadro 4: Metodología de investigación

Aspecto metodológico	Descripción
Método	Deductivo
Orientación	Aplicada
Enfoque	Cuantitativo
Instrumento de recolección	Prolectivo
Tipo/Alcance de investigación	Desarrollo tecnológico
Nivel de investigación	Tecnológico-aplicativo
Propósito de la investigación	No experimental
Número de mediciones	Transversal
Cronología de las observaciones	Prospectivo
Estudio del diseño de investigación	Estudio de desarrollo tecnológico
Población	25,602 imágenes del entorno vial ayacuchano y datasets internacionales (BDD100K, BSTLD, RLMD)

Continúa en la siguiente hoja

Cuadro 4: Metodología de investigación (Continued)

Aspecto metodológico	Descripción
Muestra	4,598 imágenes locales recolectadas mediante cuatro métodos de captura complementarios (10,701 instancias anotadas en 12 clases)
Instrumento de recolección de datos	Smartphone Redmi Note 12 Pro, GPS Map Camera, Google Earth Pro

Fuente: Elaboración propia

3.10. Hipótesis

3.10.1. Hipótesis general

El modelo inteligente adaptativo QHAWAY, fundamentado en redes neuronales convolucionales, detecta y clasifica en tiempo real las 12 clases de elementos críticos de la vía pública con una precisión media $mAP_{50} \geq 0.75$ y una velocidad de procesamiento mayor o igual a 15 fotogramas por segundo, constituyendo una herramienta tecnológica viable para la prevención de accidentes de tránsito en las avenidas Arenales y Cusco, Ayacucho, 2025.

3.10.2. Hipótesis específicas

1. La estrategia de entrenamiento progresivo en tres fases con resolución creciente produce una mejora acumulada superior a 30 puntos porcentuales en la precisión media de detección y segmentación, logrando el modelo una precisión media mAP_{50} Box igual a 0.810 y mAP_{50} Mask igual a 0.778, con la clase reductor de velocidad como la de mayor desempeño ($mAP_{50} = 0.923$) y el mototaxi alcanzando $mAP_{50} = 0.769$ como clase representativa del entorno vial andino local.
2. El algoritmo del modelo estima distancias con margen de error acotado y comportamiento conservador en todos los casos evaluados (errores entre 0.20 m y 3.10 m), clasifica correctamente las zonas de riesgo para las 12 clases detectadas y genera alertas de voz preventivas en español con un tiempo de respuesta inferior a 20 milisegundos por fotograma, activando protocolos diferenciados según el tipo de elemento vial detectado (peatón, semáforo en rojo y reductor de velocidad sin señalización).
3. El sistema QHAWAY ADAS opera de forma autónoma entre 19.2 y 25.4 fotogramas por segundo en los cuatro escenarios de riesgo vial, detectando correctamente los 12 tipos de elementos críticos de la vía pública, generando alertas de voz diferenciadas según el nivel de riesgo y exportando el registro histórico de detecciones y alertas para análisis posterior; el análisis de explicabilidad visual confirma que el modelo concentra su atención en los elementos críticos de la

vía pública, validando su viabilidad tecnológica como herramienta de apoyo a la prevención de accidentes en el entorno urbano andino informal de Ayacucho.

3.11. Operacionalización de variables

3.11.1. Variables

- **Variable independiente (X): Modelo inteligente adaptativo.** Constituye un sistema tecnológico fundamentado en redes neuronales convolucionales que integra algoritmos de aprendizaje profundo y técnicas de visión artificial para el procesamiento automático de información visual en tiempo real. Este modelo se caracteriza por su capacidad de detectar, clasificar y segmentar a nivel de instancia los 12 elementos críticos de la vía pública presentes en el entorno andino informal de Ayacucho: personas, motocicletas, automóviles, camiones, buses, mototaxis, reductores de velocidad, pasos peatonales, líneas de carril y semáforos. La naturaleza adaptativa del sistema se expresa en la estrategia de entrenamiento progresivo en tres fases con resolución creciente, que permite al modelo optimizar progresivamente su precisión y capacidad de segmentación de instancias conforme se incrementa la resolución de entrada de 640 a 1024 píxeles (Goodfellow, 2016).
- **Variable dependiente (Y): Desempeño del sistema ADAS para la detección de elementos críticos de la vía pública.** Representa el conjunto de métricas cuantitativas que evalúan la capacidad del sistema QHAWAY ADAS para detectar, segmentar y estimar distancias a los elementos críticos de la vía pública en tiempo real, orientadas a la generación de alertas tempranas para la prevención de accidentes de tránsito. Esta variable abarca la precisión de detección y segmentación de instancias (mAP_{50}), la velocidad de procesamiento en fotogramas por segundo (FPS) y la exactitud del estimador de distancias monoculares expresada en términos del error absoluto residual (ε) obtenido bajo condiciones reales de conducción en las avenidas Arenales y Cusco de Ayacucho (Organización Mundial de la Salud, 2023).

3.11.2. Dimensiones

Las dimensiones representan los componentes específicos que conforman cada variable, permitiendo su análisis de manera más precisa y detallada (Hernández Sampieri et al., 2014).

3.11.2.1 Variable independiente (X): Modelo inteligente adaptativo

- **Dimensión X1: Arquitectura y entrenamiento del modelo.** Estructura y proceso de aprendizaje de la red neuronal convolucional YOLOv8L-seg utilizada para la detección y segmentación de instancias de los elementos viales críticos. Incluye la configuración del backbone CSPDarknet con módulos C2f, la cabeza de segmentación de instancias, los parámetros de entrenamiento (tasa de aprendizaje, épocas, tamaño de lote) y la estrategia progresiva en tres fases de resolución creciente (640, 800 y 1024 px) evaluada como estudio de ablación (Redmon et al., 2016b).

- **Dimensión X2: Algoritmo del modelo.** Conjunto de procedimientos computacionales que integran la inferencia del modelo de segmentación, el seguimiento de múltiples objetos en movimiento mediante el algoritmo de asociación de detecciones, la estimación de distancias por el modelo geométrico de cámara monocular ($D = H_{\text{real}} \times f / h_{\text{bbox}}$), el cálculo del tiempo estimado de colisión y la clasificación de zonas de riesgo vial (zona de peligro: $D < 3$ m; zona de precaución: $3 \leq D < 7$ m) (LeCun et al., 2015a).
- **Dimensión X3: Implementación del modelo.** Integración de todos los componentes del sistema QHAWAY ADAS en una aplicación funcional con interfaz gráfica de usuario, que procesa video en tiempo real, genera alertas de voz preventivas en español peruano diferenciadas por nivel de riesgo, y exporta registros históricos de detecciones y alertas. La implementación fue validada bajo condiciones reales de conducción en cuatro escenarios de riesgo vial en las avenidas de estudio (Powers, 2011).

3.11.2.2 Variable dependiente (Y): Desempeño del sistema ADAS para la detección

- **Dimensión Y1: Precisión de detección y segmentación (mAP₅₀ Box/Mask).** Capacidad del modelo para identificar correctamente los elementos críticos de la vía pública y delimitar con precisión sus contornos a nivel de instancia. Se cuantifica mediante la métrica de precisión media al umbral de intersección sobre unión de 0.50 (mAP₅₀), tanto para la tarea de detección por caja delimitadora (Box) como para la tarea de segmentación de instancias (Mask), evaluadas de forma agregada y por cada una de las 12 clases del modelo (World Health Organization, 2018).
- **Dimensión Y2: Velocidad de procesamiento (FPS y latencia).** Capacidad del sistema para procesar los fotogramas de video en tiempo real a una velocidad suficiente para la generación de alertas preventivas oportunas. Se mide en fotogramas por segundo (FPS) y en milisegundos de latencia por fotograma, considerando el tiempo total del ciclo de procesamiento que incluye la inferencia del modelo, el seguimiento de objetos, la estimación de distancias y la gestión de alertas (Ministerio de Transportes y Comunicaciones, 2017a).
- **Dimensión Y3: Exactitud de estimación de distancias (residual ε).** Precisión del estimador monocular de distancias basado en el modelo geométrico de cámara para determinar la separación entre el vehículo y los elementos críticos detectados. Se cuantifica mediante el error absoluto residual (ε) entre la distancia estimada por el sistema y la distancia de referencia obtenida mediante referencias métricas independientes del entorno de estudio (juntas de losas de pavimento, ancho de calzada y geometría vehicular) (Dodge & Karam, 2016).

3.11.3. Indicadores

Los indicadores son elementos específicos y medibles que permiten cuantificar las dimensiones de las variables de estudio (Caballero Romero, 2014).

3.11.3.1 Indicadores de la variable independiente (X):

1. Arquitectura y entrenamiento del modelo (X1)

- Arquitectura base del modelo [Tipo (categórica): YOLOv8L-seg]
- Número de parámetros del modelo [Millones (M): 43.7 M]
- Número de fases de entrenamiento progresivo [Cantidad (und): 3 fases]
- Resolución de entrada en fase final [Píxeles (px × px): 1024 × 1024]
- Número total de épocas ejecutadas [Cantidad (und): 473 épocas]
- Tiempo total de entrenamiento [Horas (h): 24.1 h]

2. Algoritmo del modelo (X2)

- Umbral de confianza mínima para detección [Decimal: 0.55]
- Umbral de zona de peligro [Metros (m): $D < 3,0$ m]
- Umbral de zona de precaución [Metros (m): $3,0 \leq D < 7,0$ m]
- Distancia focal calibrada de la cámara [Píxeles (px): 950 px]
- Altura de montaje de la cámara [Metros (m): 1.50 m]

3. Implementación del modelo (X3)

- Total de imágenes del dataset híbrido [Cantidad (und): 25,602 imágenes]
- Total de instancias anotadas [Cantidad (und): 127,525 instancias]
- Número de clases entrenadas [Cantidad (und): 12 clases]
- Número de escenarios de validación [Cantidad (und): 4 escenarios]

3.11.3.2 Indicadores de la variable dependiente (Y):

1. Precisión de detección y segmentación — mAP_{50} Box/Mask (Y1)

- Precisión media agregada de detección mAP_{50} Box [Decimal: 0.810]
- Precisión media agregada de segmentación mAP_{50} Mask [Decimal: 0.778]
- mAP_{50} Box de la clase reductor de velocidad [Decimal: 0.923]
- mAP_{50} Box de la clase persona [Decimal: 0.866]
- mAP_{50} Box de la clase mototaxi [Decimal: 0.769]
- Precisión global del modelo (Precision) [Decimal: 0.885]
- Sensibilidad global del modelo (Recall) [Decimal: 0.724]

2. Velocidad de procesamiento — FPS y latencia (Y2)

- Velocidad de procesamiento en tiempo real [Fotogramas por segundo (FPS): 19.2–25.4]
- Latencia de inferencia por fotograma [Milisegundos (ms): 17–20 ms]
- Tiempo de respuesta del sistema de alertas [Milisegundos (ms): < 20 ms/fotograma]

3. Exactitud de estimación de distancias — residual ε (Y3)

- Error absoluto residual mínimo [Metros (m): 0.20 m]

- Error absoluto residual máximo [Metros (m): 3.10 m]
- Comportamiento del estimador ante variaciones de postura y posición [Tipo (categórica): conservador – sobreestima la distancia]

En el Cuadro 35 se presenta la matriz de operacionalización de las variables identificadas, donde se detallan las variables, dimensiones, indicadores, unidades de medida y valores obtenidos en el sistema QHAWAY ADAS. Esta matriz establece la relación sistemática entre los constructos teóricos y los resultados empíricamente medibles, facilitando la evaluación rigurosa del desempeño del modelo inteligente adaptativo desarrollado para la detección de elementos críticos de la vía pública en el entorno urbano andino informal de las avenidas Arenales y Cusco de Ayacucho.

3.12. Técnicas e instrumentos

3.12.1. Técnicas de investigación

La presente investigación emplea diversas técnicas especializadas para la recopilación y análisis de datos, así como para el desarrollo y validación del modelo inteligente adaptativo QHAWAY, fundamentado en redes neuronales convolucionales para la detección de elementos críticos de la vía pública en las avenidas Arenales y Cusco de Ayacucho. A continuación se detallan las técnicas utilizadas:

1. **Revisión bibliográfica y documental:** Recopilación y análisis crítico de la literatura científica existente sobre modelos de visión computacional aplicados a sistemas ADAS, técnicas de segmentación de instancias, estimación de distancias monoculares y detección de usuarios vulnerables en entornos viales urbanos. La revisión abarcó fuentes primarias indexadas en bases de datos especializadas como IEEE Xplore, Google Scholar, Scopus y Web of Science, así como la normativa vial peruana vigente (Directiva MTC N° 01-2011-MTC/14) (Goodfellow, 2016).
2. **Observación sistemática no participante:** Recolección de datos mediante la captura planificada de fotografías y videos del comportamiento del tráfico vehicular y peatonal en las avenidas de estudio, sin intervenir en el desarrollo natural de los eventos observados. La observación se realizó bajo diversas condiciones de iluminación (diurna, crepuscular y nocturna) y densidad de tráfico, empleando cuatro métodos de captura complementarios:
 - Grabación de video desde vehículo en movimiento mediante smartphone Redmi Note 12 Pro montado en soporte sobre el parabrisas a 1.50 m de altura, a resolución Full HD (1920 × 1080 px, 30 fps).
 - Fotografías georreferenciadas mediante la aplicación GPS Map Camera, que registra coordenadas, fecha y hora en cada imagen.
 - Fotografías convencionales en intersecciones urbanas, el campus de la UNSCH y el centro histórico de Ayacucho.
 - Capturas de Google Earth Pro en modo Street View para cobertura de zonas de

acceso restringido.

3. **Construcción del conjunto de datos híbrido:** Conformación del dataset QHAWAY mediante la integración de 4,598 imágenes locales originales del entorno vial ayacuchano con tres conjuntos de datos internacionales de acceso público (BDD100K, BSTLD y RLMD), alcanzando un total de 25,602 imágenes con 127,525 instancias anotadas en 12 clases (Yu, Chen, Wang et al., 2020a).
4. **Anotación de imágenes mediante segmentación de instancias:** Etiquetado de las imágenes del conjunto local mediante la herramienta LabelMe con flujo de trabajo asistido por el modelo de segmentación automática SAM2 (*Segment Anything Model 2*), generando polígonos de segmentación a nivel de píxel para cada instancia de las 12 clases definidas. Se aplicó un protocolo de control de calidad que incluyó la revisión cruzada independiente del 20 % de las anotaciones (Kirillov et al., 2023).
5. **Entrenamiento progresivo del modelo de segmentación de instancias:** Entrenamiento del modelo YOLOv8L-seg mediante una estrategia en tres fases de resolución creciente (640, 800 y 1024 px) con transferencia de aprendizaje a partir de pesos preentrenados en el conjunto de datos COCO. Cada fase parte de los mejores pesos de la fase anterior, con tasas de aprendizaje decrecientes y aumentación de datos mediante mosaico, mezcla de imágenes y copia y pegado de instancias. El entrenamiento se ejecutó en GPU NVIDIA GeForce RTX 5080 (16 GB VRAM) con precisión mixta automática (Jocher et al., 2023b).
6. **Evaluación del modelo mediante métricas estándar de visión computacional:** Cuantificación del desempeño del modelo entrenado mediante las métricas de precisión media al umbral de intersección sobre unión de 0.50 (mAP_{50}) para detección por caja delimitadora y segmentación de instancias, calculadas tanto de forma agregada como por cada una de las 12 clases. Se emplearon adicionalmente las métricas de precisión global, sensibilidad global, puntuación F1 y curvas Precisión-Sensibilidad (Powers, 2011).
7. **Validación operacional bajo condiciones naturalistas:** Prueba del sistema QHAWAY ADAS bajo condiciones reales de conducción en las avenidas Arenales y Cusco de Ayacucho, sin escenarios provocados ni situaciones artificiales. Se registraron cuatro escenarios de riesgo vial representativos: niños cruzando intempestivamente, peatones ocultos o distraídos, semáforos en rojo y reductores de velocidad sin señalización.
8. **Validación geométrica del estimador de distancias:** Verificación cuantitativa de la exactitud del modelo geométrico de cámara monocular mediante referencias métricas independientes del entorno de estudio: juntas de losas de pavimento (cuadrícula de 3.0 m \times 3.0 m), ancho de calzada verificado in situ (6.0 m) y geometría del vehículo Toyota Hilux empleado en la recolección de datos (S. Park et al., 2014).
9. **Análisis de explicabilidad del modelo mediante mapas de calor de activación:** Generación de mapas de calor de activación sobre imágenes representativas del

entorno vial ayacuchano empleando tres algoritmos complementarios de la familia Grad-CAM: el algoritmo original, su versión mejorada Grad-CAM++ y EigenCAM. Este análisis permite verificar que las decisiones del modelo se fundamentan en características visuales relevantes de los elementos viales, y no en correlaciones espurias del fondo de la imagen (Selvaraju, Cogswell, Das et al., 2017a).

3.12.2. Instrumentos de investigación

Para la implementación de las técnicas mencionadas se utilizaron los siguientes instrumentos:

■ Sistema de captura de imágenes y video:

- **Smartphone Redmi Note 12 Pro** (sensor nativo de 108 MP; video grabado en modo Full HD: 1920×1080 px, 30 fps), montado mediante soporte con ventosa sobre el parabrisas del vehículo a 1.50 m de altura, simulando la perspectiva frontal del conductor.
- **Aplicación GPS Map Camera**, que registra automáticamente las coordenadas geográficas, fecha y hora en cada fotografía, proporcionando un registro georreferenciado del levantamiento de campo.
- **Google Earth Pro** (modo Street View), empleado para la captura de imágenes virtuales del entorno vial de Ayacucho en zonas de acceso restringido.
- **Vehículo Toyota Hilux**, utilizado como plataforma de grabación, cuya altura de parabrisas (1.50 m) proporciona la perspectiva frontal del conductor para los parámetros de calibración del estimador de distancias.

■ Herramientas de anotación:

- **LabelMe** (versión con integración SAM2), plataforma de código abierto para la anotación de imágenes mediante polígonos de segmentación a nivel de píxel, con flujo de trabajo asistido por el modelo de segmentación automática SAM2.

■ Hardware de procesamiento:

- **GPU NVIDIA GeForce RTX 5080** (16 GB VRAM), empleada para el entrenamiento del modelo con precisión mixta automática (AMP) y para la inferencia en tiempo real durante la validación operacional.

■ Software de desarrollo del sistema QHAWAY ADAS:

- **Python 3.11**, lenguaje principal de desarrollo del software.
- **Ultralytics YOLOv8L-seg**, modelo de segmentación de instancias utilizado para la detección en tiempo real (Jocher et al., 2023b).
- **PyTorch 2.x con CUDA**, para la inferencia acelerada en GPU.
- **OpenCV 4.x**, para la lectura, procesamiento y escritura de fotogramas de video.
- **ByteTrack**, algoritmo de seguimiento de múltiples objetos para la asignación de identidades persistentes entre fotogramas consecutivos (Y. Zhang, Sun, Jiang et al., 2022a).

- **PyQt5 5.15**, para la interfaz gráfica de usuario con tema oscuro profesional.
- **Microsoft Azure Neural TTS** (voz es-PE-AlexNeural), para la síntesis de alertas de voz en español peruano (Microsoft Corporation, 2024).

3.12.3. Herramientas de software utilizadas

En la presente investigación se emplearon las herramientas de software detalladas en el Cuadro 5.

Cuadro 5: Herramientas de software utilizadas en el desarrollo del sistema QHAWAY ADAS

Herramienta	Descripción
LaTeX	Preparación y edición del informe final de investigación con tipografía científica de alta calidad.
Python 3.11	Lenguaje principal de desarrollo del software QHAWAY ADAS, integración de módulos y lógica de procesamiento.
Ultralytics YOLOv8L-seg	Modelo de segmentación de instancias para la detección en tiempo real de las 12 clases de elementos viales críticos.
PyTorch 2.x + CUDA	Inferencia acelerada en GPU y entrenamiento del modelo con precisión mixta automática.
OpenCV 4.x	Procesamiento digital de fotogramas: lectura, conversión de espacios de color y escritura de video procesado.
LabelMe + SAM2	Plataforma de anotación de imágenes con flujo de trabajo asistido por segmentación automática para la generación del dataset local.
PyQt5 5.15	Interfaz gráfica de usuario de la aplicación QHAWAY ADAS con tema oscuro y controles interactivos en tiempo real.
ByteTrack	Algoritmo de seguimiento de múltiples objetos para la asignación de identidades persistentes entre fotogramas consecutivos.
edge-tts (Azure Neural TTS)	Síntesis de alertas de voz en español peruano (voz es-PE-AlexNeural) con parámetros prosódicos diferenciados por nivel de riesgo.
GPS Map Camera	Aplicación móvil para captura de fotografías georreferenciadas con registro automático de coordenadas, fecha y hora.

Continúa en la siguiente hoja

Cuadro 5: Herramientas de software utilizadas en el desarrollo del sistema QHAWAY ADAS (Continued)

Herramienta	Descripción
Google Earth Pro	Herramienta de captura de imágenes virtuales del entorno vial en modo Street View para ampliación del dataset local.
Anaconda / Jupyter	Entorno de desarrollo y análisis exploratorio del dataset durante las etapas de preprocesamiento y evaluación de resultados.

Fuente: Elaboración propia

3.13. Desarrollo de trabajo de tesis

3.13.1. Recolección de datos

Para la obtención del conjunto de datos, se realizó un recorrido a lo largo de las avenidas Arenales y Cusco en la ciudad de Ayacucho, empleando un dispositivo móvil con la aplicación GPS Map Camera, la cual permitió registrar fotográficamente las condiciones reales de la vía junto con información georreferenciada. Durante el recorrido, se capturaron las características fundamentales de la infraestructura vial, tales como: Señalización horizontal y vertical, reductores de velocidad, estado del pavimento, presencia de cruces peatonales, dispositivos de control del tránsito, así como las condiciones del entorno urbano que inciden en la seguridad vial.

Figura 23

Punto de inicio de la recolección de datos: Intersección Av. Arenales y Av. Cusco



Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

3.13.2. Características viales de la Avenida Arenales

Durante el recorrido por la Av. Arenales, se identificaron diversas deficiencias en la infraestructura vial que comprometen la seguridad tanto de conductores como de

peatones. A continuación, se describen los hallazgos más relevantes registrados en campo. En primer lugar, se evidenció la presencia de señalización vertical en estado de deterioro. En la Figura 24a se observa una señal reglamentaria de “No estacionar” (R-27) con desgaste visible y soporte oxidado, ubicada a la altura de la cuadra 2 de la Av. Arenales. De igual forma, en la Figura 24b.

Figura 24

Señalización vertical en estado de deterioro – Av. Arenales



(a) Señal reglamentaria R-27 deteriorada



(b) Señal informativa

Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Estado del pavimento en la Av. Arenales

Asimismo, se verificó el estado de la superficie de rodadura a lo largo de la Av. Arenales, evidenciándose un notable deterioro del pavimento por falta de mantenimiento. En las Figuras 25 y 26 se aprecian baches, fisuras y desprendimiento de la carpeta asfáltica en diversos tramos de la vía, condiciones que representan un riesgo para la circulación vehicular y peatonal.

Figura 25

Estado del pavimento en la Av. Arenales – Tramo 1



(a) Deterioro de pavimento



(b) Superficie de rodadura irregular

Nota. Imagenes propia

Figura 26

Estado del pavimento en la Av. Arenales – Tramo 2



(a) Presencia de baches en la calzada



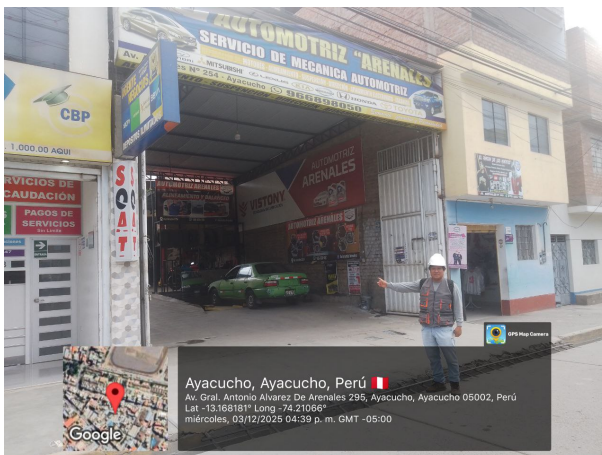
(b) Falta de mantenimiento en la vía

Nota. Imagenes propia.

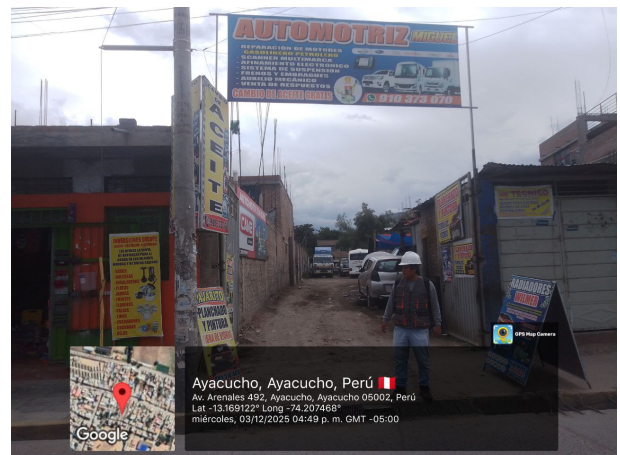
Se identificó la presencia constante de talleres de servicio automotriz a lo largo de la Av. Arenales, los cuales generan que diversos vehículos se estacionen de manera informal sobre la calzada y las veredas en zonas que no corresponden a paraderos autorizados. Esta situación provoca oclusión en los cruces peatonales y reduce la visibilidad tanto para conductores como para peatones, tal como se evidencia en la Figura 27. Asimismo, se observó la proliferación de avisos publicitarios que invaden el espacio de la vía y la presencia de comercio ambulatorio que obstruye el libre tránsito peatonal y vehicular.

Figura 27

Talleres automotrices y estacionamiento informal – Av. Arenales



(a) Taller automotriz



(b) Avisos publicitarios

Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Se ejecutó la medición del ancho de la calzada, obteniéndose un carril de 4.50 m, tal como se muestra en la Figura 28.

Figura 28*Medición del ancho de carril – Av. Arenales*

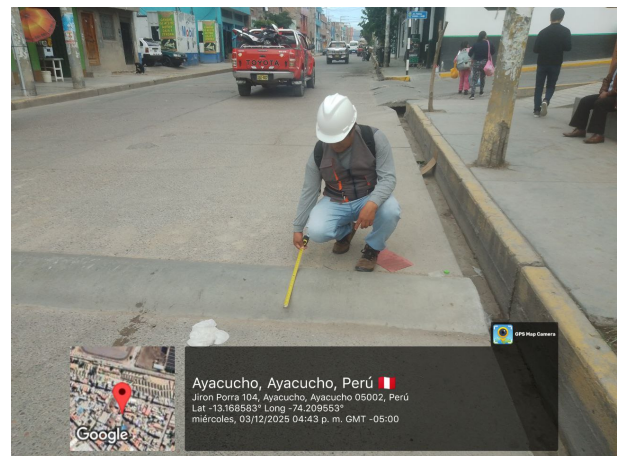
Nota. Imagen propia

Características de los reductores de velocidad

Se identificaron reductores de velocidad que no cumplen con las especificaciones técnicas establecidas en el Manual de Dispositivos de Control del Tránsito Automotor. En la Figura 29 se observa un reductor de velocidad de tan solo 25 cm de longitud y 10 cm de altura, el cual carece de señalización vertical que alerte a los conductores sobre su presencia. Dicho reductor fue construido sin ningún criterio de diseño ingenieril, lo que ocasiona saltos bruscos al ser transitado por los vehículos, generando riesgo tanto para los ocupantes como para la integridad mecánica de las unidades. Asimismo, en la Figura 30 se registró un segundo reductor de velocidad con una longitud de 2.40 m y una altura de 10 cm, el cual tampoco se encuentra pintado ni cuenta con señalización vertical de advertencia, incumpliendo lo estipulado en la normativa vigente.

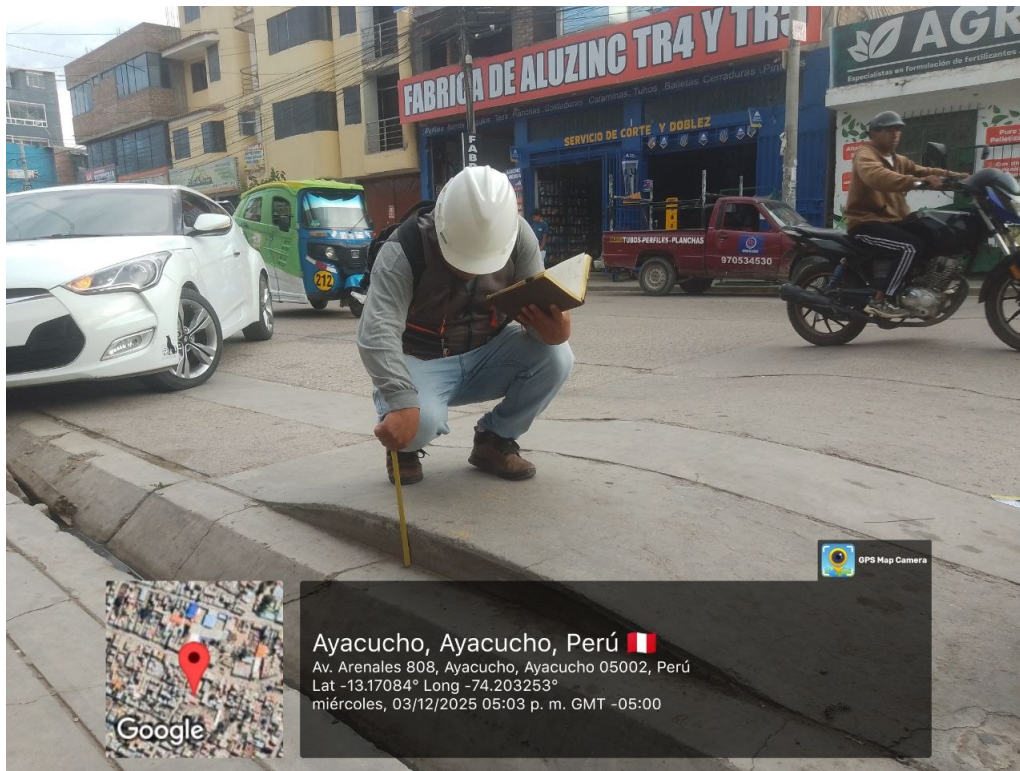
Figura 29*Reductor de velocidad sin señalización vertical – Av. Arenales*

(a) Vista general del reductor de velocidad



(b) Dimensiones: 25 cm de L y 10 cm de H

Nota. Imágenes propias. Registro fotográfico con GPS Map Camera.

Figura 30*Reductor de velocidad sin pintura ni señalización – Av. Arenales*

Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

En un tramo intermedio de la Av. Arenales se registró otro reductor de velocidad con dimensiones de 1.50 m de longitud y 10 cm de altura, el cual tampoco cumple con la longitud mínima de cuerda de 3.50 m establecida en la Directiva N° 01-2011-MTC/14 (Ministerio de Transportes y Comunicaciones, 2011), tal como se observa en la Figura 31.

Figura 31*Reductor de velocidad en la Av. Arenales*

Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

3.13.3. Características viales de la Avenida 09 de diciembre

La Av. 9 de Diciembre es una vía que recientemente fue intervenida mediante el proyecto “Mejoramiento del Servicio de Transitabilidad Vial en la Av. 9 de Diciembre – Primera Etapa”, inaugurado en agosto de 2025 por el Gobierno Regional de Ayacucho, tal como se evidencia en la placa conmemorativa mostrada en la Figura 32. A pesar de tratarse de una obra de reciente ejecución, durante el recorrido se identificaron deficiencias en los elementos de seguridad vial implementados. Se registró un ancho de carril de 4.50 m.

Figura 32

Placa de inauguración de la Av. 9 de Diciembre



Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

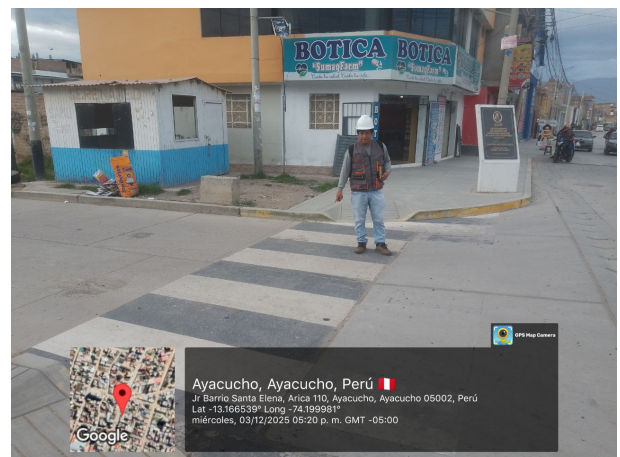
En la Figura 33 se observan los pasos peatonales implementados en la Av. 9 de Diciembre. Cabe resaltar que, a pesar del corto tiempo transcurrido desde la inauguración de la obra, en la Figura 33a ya se evidencia un notable desgaste de la pintura del paso peatonal, lo que denota el uso de materiales de baja calidad o la ausencia de un plan de mantenimiento.

Figura 33

Pasos peatonales en la Av. 9 de Diciembre



(a) Paso peatonal con pintura desgastada



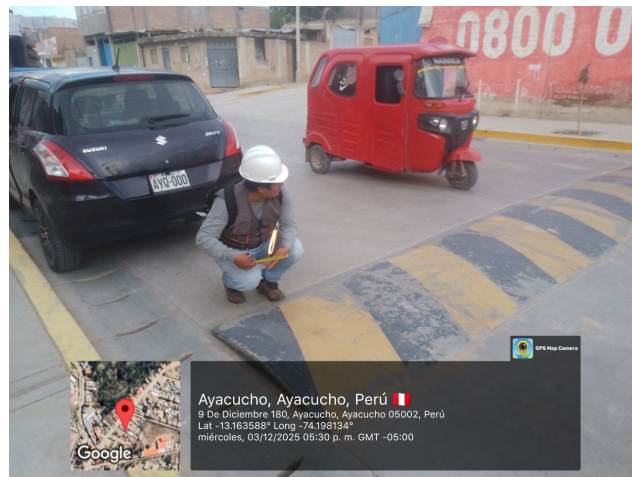
(b) Paso peatonal en la Av. 9 de Diciembre

Nota. Imágenes propias. Registro fotográfico con GPS Map Camera.

De igual forma, se identificó un reductor de velocidad de sección circular con dimensiones de 1.20 m de longitud y 8 cm de altura (Figura 34), el cual no cumple con las especificaciones técnicas de la Directiva N° 01-2011-MTC/14 (Ministerio de Transportes y Comunicaciones, 2011), que establece una longitud de cuerda mínima de 3.50 m y una altura entre 7 y 10 cm para una velocidad esperada de 25 km/h. Esta situación resulta particularmente preocupante al tratarse de una vía de reciente construcción, donde se esperaría el cumplimiento pleno de la normativa vigente.

Figura 34

Reductor de velocidad en la Av. 9 de Diciembre



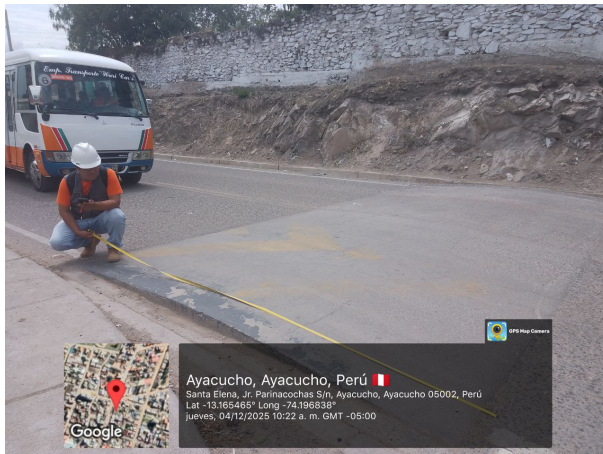
Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Características viales de la Av. Cusco

El recorrido por la Av. Cusco se realizó en sentido descendente, desde el tramo final hasta su intersección con la Av. Arenales. A lo largo de esta vía se registraron múltiples reductores de velocidad con diversas condiciones geométricas y de conservación, evidenciando una falta de uniformidad en su diseño y ejecución.

En las Figuras 35 se muestran dos reductores de velocidad de sección circular con dimensiones de 3.50 m de longitud y 10 cm de altura.

Según la Tabla N° 01 de la Directiva N° 01-2011-MTC/14 (Ministerio de Transportes y Comunicaciones, 2011), estas dimensiones corresponden a una velocidad esperada de 25 km/h con una longitud de cuerda mínima de 3.50 m, por lo que constituyen los únicos reductores del tramo estudiado que cumplen con los parámetros normativos.

Figura 35*Reductores de velocidad con dimensiones normativas – Av. Cusco*

(a) Reductor de 3.50 m × 0.10 m

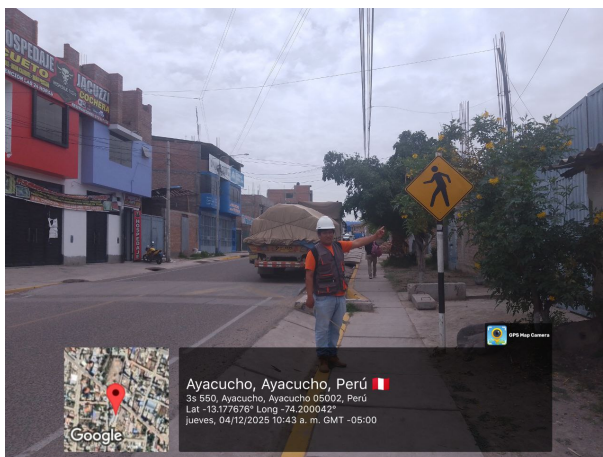


(b) Reductor de 3.50 m × 0.10 m

Nota. Imágenes propias. Registro fotográfico con GPS Map Camera.

Sin embargo, se identificaron otros reductores con dimensiones inferiores a las establecidas. En la Figura 36 se presentan dos dispositivos con longitudes de cuerda de 2.50 m y alturas de 7 cm, valores que se encuentran por debajo de la longitud mínima de 3.50 m exigida para una velocidad esperada de 25 km/h.

Además, como se aprecia en la Figura 36a, si bien existe señalización vertical preventiva de cruce peatonal (P-48), no se registró la presencia de la señal preventiva P-33.

Figura 36*Reductores de velocidad con dimensiones insuficientes – Av. Cusco*

(a) Señal peatonal sin advertencia de resalto



(b) Reductor de 2.50 m × 0.07 m

Nota. Imágenes propias. Registro fotográfico con GPS Map Camera.

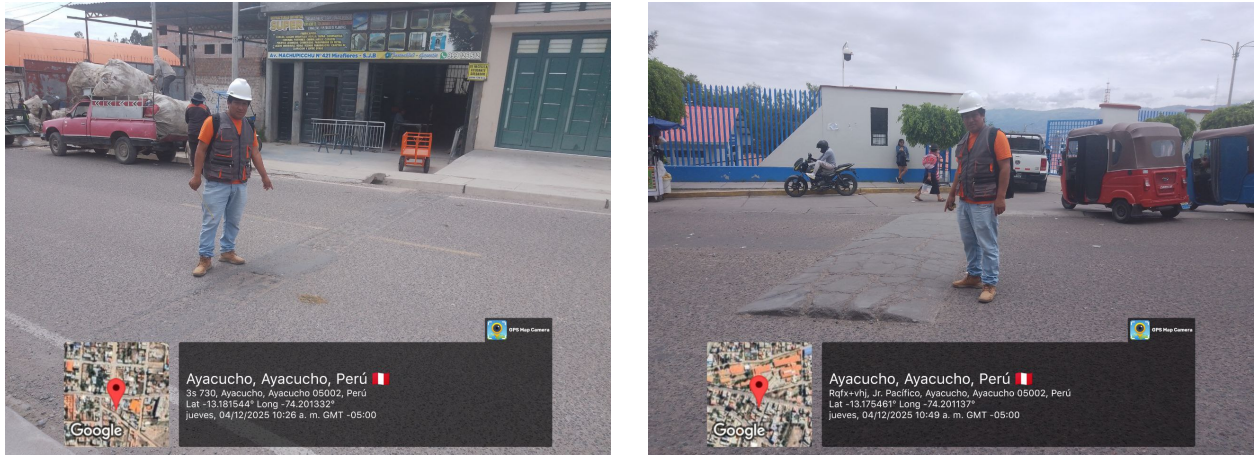
Deterioro y retiro improvisado en los dispositivos de control

Una problemática particularmente grave identificada durante la inspección fue la destrucción parcial o total de los reductores de velocidad por acción de terceros. En la Figura 37 se evidencia cómo los dispositivos han sido fragmentados de manera intencional,

dejando únicamente restos de la estructura original sobre la calzada. Esta práctica, lejos de mejorar la circulación, genera superficies irregulares que constituyen un peligro latente tanto para vehículos como para motociclistas.

Figura 37

Reductores de velocidad destruidos por acción de terceros – Av. Cusco



(a) Reductor destruido

(b) Reductor fracturado

Nota. Imágenes propias. Registro fotográfico con GPS Map Camera.

La situación registrada en la Figura 37b resulta especialmente crítica, ya que el reductor fracturado se encuentra próximo a un establecimiento de salud, zona donde la normativa exige mayor control de velocidad vehicular. Los conductores evitan el fragmento remanente del dispositivo desviándose hacia el carril contrario, generando conflictos de tránsito y potenciales colisiones frontales. Asimismo, se registró un reductor de velocidad frente a las instalaciones del SENCICO con dimensiones de apenas 0.60 m de longitud y 0.07 m de altura (Figura 38), lo cual representa solo el 17% de la longitud mínima establecida por la Directiva N° 01-2011-MTC/14 (Ministerio de Transportes y Comunicaciones, 2011), resultando prácticamente ineficaz como dispositivo de reducción de velocidad.

Figura 38*Reductor de velocidad subdimensionado – frente a SENCICO*

Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Finalmente, en la intersección de la Av. Cusco con la Av. Arenales se observó la presencia de badenes que carecen de señalización horizontal y vertical de advertencia (Figura 39). La ausencia de demarcación en estos dispositivos reduce significativamente su visibilidad, particularmente en condiciones de baja luminosidad, incrementando el riesgo de siniestros viales por maniobras evasivas imprevistas.

Figura 39*Baden sin señalización en la Av. Cusco*

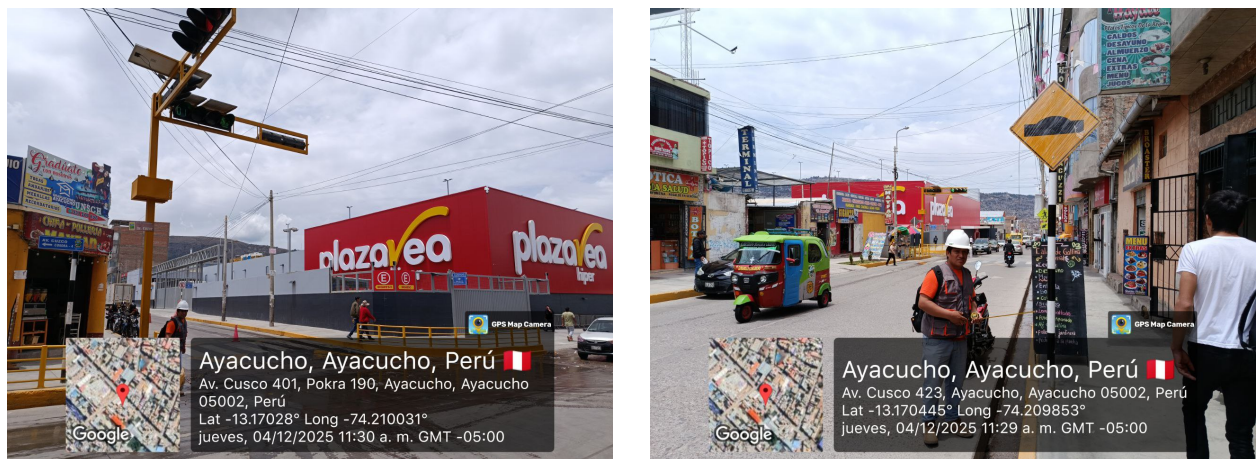
Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Recolección de datos para el entrenamiento del modelo Durante el levantamiento de campo, además del registro de las condiciones de infraestructura vial, se capturaron fotografías orientadas a la conformación del conjunto de datos (*dataset*) necesario para el entrenamiento del modelo de detección de objetos. Las imágenes fueron obtenidas en condiciones reales de circulación, abarcando las 12 clases definidas para

el modelo: reductor de velocidad, paso peatonal, línea recta, línea recta y derecha, persona, motocicleta, automóvil particular, camión, bus de transporte, semáforo en verde, semáforo en rojo y mototaxi. Complementariamente, se grabaron videos a lo largo de las rutas recorridas, los cuales serán empleados en la etapa de validación del modelo. **Características viales de la Av. Cusco – Tramo comercial** Cabe señalar que en la Av. Cusco se inauguró recientemente el centro comercial Plaza Veá, cuya operación ha incrementado significativamente el flujo vehicular y peatonal en la zona. En las inmediaciones de este establecimiento se concentra una mayor densidad de dispositivos de control del tránsito, incluyendo semáforos vehiculares, pasos peatonales demarcados y señalización vertical, tal como se aprecia en la Figura 40.

Figura 40

Entorno vial en las inmediaciones de Plaza Veá – Av. Cusco

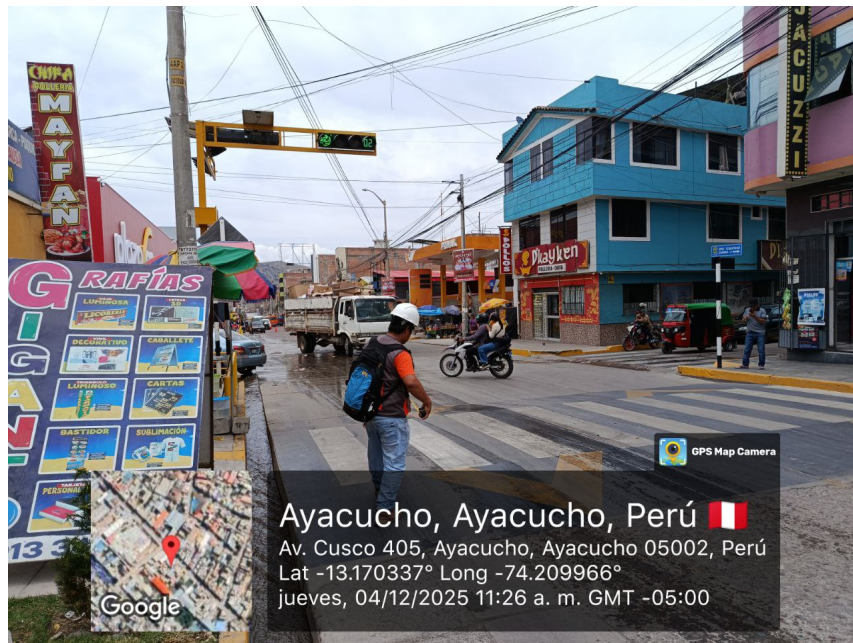


(a) Semáforos y flujo vehicular – Av. Cusco

(b) Señal preventiva P-33 de resalto

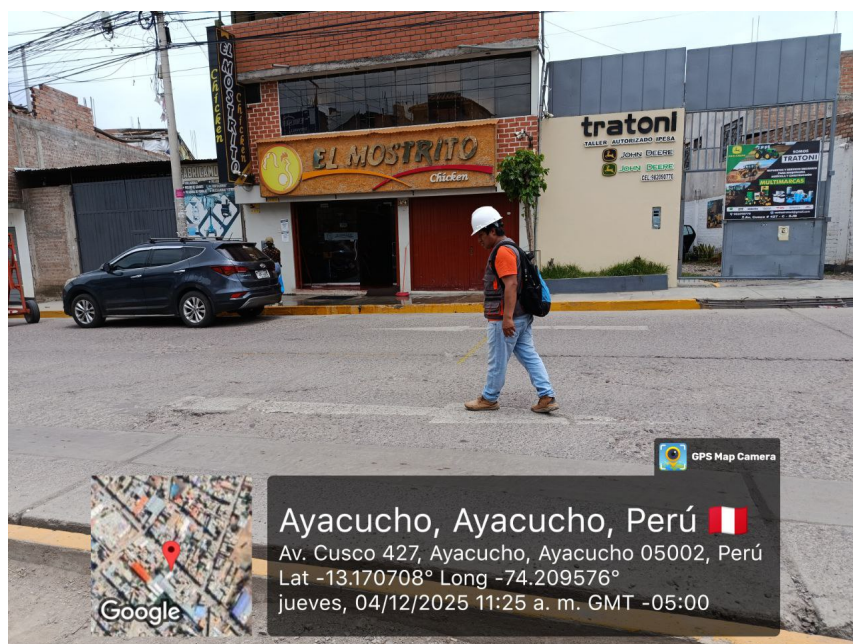
Nota. Imágenes propias. Registro fotográfico con GPS Map Camera.

En este tramo, a diferencia de los sectores anteriormente descritos, se identificó un reductor de velocidad de sección trapezoidal que cumple con los parámetros establecidos en la Directiva N° 01-2011-MTC/14 (Ministerio de Transportes y Comunicaciones, 2011), con una sección plana mínima de 2.50 m, rampas de aproximadamente 1.00 m a cada lado, una altura de 0.10 m y una pendiente máxima de 1:10, conforme al esquema normativo. Asimismo, el dispositivo cuenta con la señalización vertical preventiva P-33 correspondiente, tal como se observó en la Figura 40b. Este reductor constituye un ejemplo de cumplimiento normativo dentro del área de estudio (Figura 41).

Figura 41*Reductor de velocidad trapezoidal conforme a normativa – Av. Cusco*

Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

De igual forma, se registraron elementos de señalización horizontal presentes en la vía, tales como líneas centrales continuas de color amarillo que delimitan los carriles de circulación en sentidos opuestos, las cuales forman parte de las clases del *dataset* de entrenamiento (Figura 42).

Figura 42*Señalización horizontal – línea central continua en la Av. Cusco*

Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Ampliación del dataset – Av. Las Américas Con el propósito de ampliar la representatividad del conjunto de datos y garantizar una mayor variabilidad en las condiciones de captura, se extendió el recorrido hacia la Av. Las Américas. En esta vía se registró una mayor concentración de pasos peatonales en buen estado de conservación, particularmente en la zona próxima al Hospital Regional, lo que permitió obtener un volumen significativo de imágenes para la clase “paso peatonal” (Figura 43).

Figura 43

Pasos peatonales en la Av. Las Américas



(a) Paso peatonal frente al Hospital Regional



(b) Paso peatonal con sardinel diferenciado

Nota. Imágenes propias. Registro fotográfico con GPS Map Camera.

Adicionalmente, en la Av. Las Américas se identificaron reductores de velocidad que, si bien no cumplen con las dimensiones mínimas de la normativa vigente, resultaron igualmente valiosos para el entrenamiento del modelo, ya que permiten que la red neuronal aprenda a detectar dispositivos en diversas condiciones geométricas y de conservación (Figura 44).

Figura 44

Reductores de velocidad en la Av. Las Américas



(a) Reductor con pintura desgastada



(b) Reductor con señalización horizontal

Nota. Imágenes propias. Registro fotográfico con GPS Map Camera.

Fuentes complementarias de datos

Con el propósito de incrementar el volumen y la variabilidad del conjunto de datos, se emplearon fuentes complementarias de captura además del registro fotográfico en campo con GPS Map Camera. En primer lugar, se utilizó la herramienta Google Earth Pro en su modo *Street View*, la cual permitió recorrer virtualmente diversas vías de la ciudad de Ayacucho y obtener imágenes desde múltiples perspectivas y condiciones de iluminación. Esta estrategia resultó especialmente útil para capturar escenarios con alta densidad de tránsito vehicular y peatonal, así como para registrar dispositivos de control del tránsito en zonas donde el acceso presencial era limitado.

En segundo lugar, se realizaron recorridos adicionales empleando la cámara del dispositivo móvil sin la aplicación GPS Map Camera, lo que permitió obtener fotografías con mayor naturalidad y sin las marcas de georreferenciación superpuestas, generando así una mayor diversidad en las condiciones de entrada para el modelo de detección.

Registro fotográfico en el centro histórico de Ayacucho

El centro histórico de la ciudad de Ayacucho constituye una zona de particular interés para el presente estudio, debido a que sus calles fueron originalmente diseñadas para el tránsito de carruajes durante la época colonial, lo que resulta en vías estrechas y reducidas que actualmente deben compartirse entre vehículos motorizados, motocicletas y peatones. Esta condición genera una alta conflictividad vial, por lo que la señalización horizontal y vertical adquiere una importancia crítica para la seguridad de los usuarios.

En las Figura 45 se observa la dinámica de circulación típica del centro histórico, donde confluyen motocicletas, automóviles, buses de transporte y un elevado flujo peatonal en las intersecciones, con pasos peatonales como principal elemento de protección al peatón.

Figura 45

Dinámica de circulación en el centro histórico de Ayacucho



Nota. Imagen propia. Flujo vehicular y peatonal simultáneo.

Asimismo, se registraron elementos de señalización horizontal en las calles del centro

histórico. En la Figura 46a se aprecia un tramo con señalización de cruce peatonal y bandas alertadoras sobre pavimento adoquinado, donde se observa además la circulación de una persona con discapacidad, lo que evidencia la necesidad de una infraestructura vial accesible e inclusiva. En la Figura 46b se registró la señalización horizontal de flechas direccionales sobre el pavimento, las cuales orientan el sentido de circulación en las vías de un solo carril.

Figura 46

Señalización horizontal en el centro histórico de Ayacucho



(a) Paso peatonal constante



(b) Flecha direccional sobre adoquín

Nota. Imágenes propias.

Registro mediante Google Earth Pro – Av. Independencia y UNSCH

Mediante el uso de Google Earth Pro se recorrió la Av. Independencia, una de las arterias principales de la ciudad, la cual concentra un alto volumen de tránsito vehicular y peatonal, particularmente en las inmediaciones de la Universidad Nacional de San Cristóbal de Huamanga (UNSCH).

Esta zona resulta especialmente sensible en términos de seguridad vial, dado el constante desplazamiento de estudiantes universitarios que cruzan la vía para acceder a las instalaciones del campus.

En la Figura 47 se muestran capturas obtenidas mediante Google Earth Pro, donde se aprecian pasos peatonales, mototaxis, buses de transporte y semáforos en la Av. Independencia, elementos representativos de las 12 clases definidas para el entrenamiento del modelo.

Figura 47

Capturas de Google Earth Pro – Av. Independencia



(a) Paso peatonal y flujo vehicular

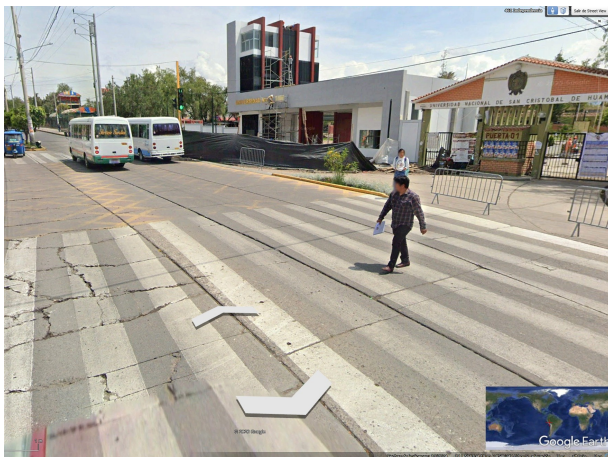


(b) Semáforos y transporte público

Nota. Capturas obtenidas de Google Earth Pro.

Figura 48

Capturas de Google Earth Pro – Inmediaciones de la UNSCH



(a) Paso peatonal frente a la UNSCH



(b) Semáforo en verde – Av. Independencia

Nota. Capturas obtenidas de Google Earth Pro.

De esta manera, la combinación de tres fuentes de captura (GPS Map Camera, cámara del dispositivo móvil y Google Earth Pro) permitió conformar un *dataset* con alta variabilidad en perspectivas, condiciones de iluminación, resolución y contexto urbano, cubriendo prácticamente la totalidad de las vías principales de la ciudad de Ayacucho. Esta diversidad resulta fundamental para que el modelo de detección de objetos generalice adecuadamente ante escenarios reales de circulación.

Videograbaciones para validación y ampliación del *dataset*

Adicionalmente al registro fotográfico, se realizaron videograbaciones a lo largo de las principales vías de la ciudad de Ayacucho, las cuales cumplen un doble propósito: Por un lado, constituyen el material de entrada para la etapa de validación del modelo de detección en tiempo real; por otro lado, permiten la extracción de fotogramas (*frames*) que

enriquecen el conjunto de datos de entrenamiento con imágenes en condiciones dinámicas de circulación.

Para la captura de video se empleó un *smartphone* montado mediante un soporte con ventosa sobre el parabrisas de una camioneta Toyota Hilux, lo que garantizó una posición fija y estable durante todo el recorrido, simulando la perspectiva frontal de un conductor. La altura de montaje, correspondiente a la posición del tablero/parabrisas del vehículo, fue de 1.50 m respecto al nivel de la calzada. En la Figura 49 se detalla el proceso de instalación del dispositivo de captura.

Figura 49

Instalación del dispositivo de captura en el parabrisas del vehículo



(a) Soporte con ventosa



(b) Ajuste del *smartphone*



(c) Posición final de grabación

Nota. Imágenes propias.

Los parámetros de calibración de la cámara empleados para el procesamiento posterior de los videos se detallan en la Tabla 6.

Cuadro 6

Parámetros de calibración de la cámara de videograbación

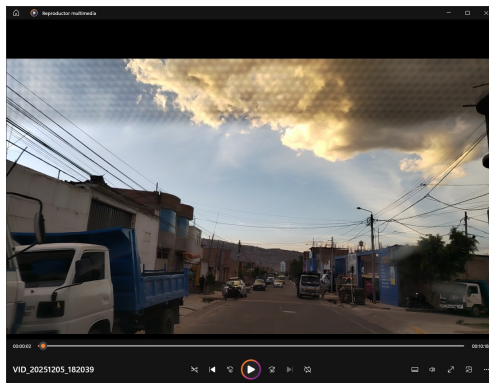
Parámetro	Variable	Valor	Descripción
Distancia focal	CAMERA_FOCAL_PX	950.0 px	Distancia focal estimada del <i>smartphone</i> expresada en píxeles, utilizada para la conversión de coordenadas de imagen a coordenadas del mundo real.
Altura de montaje	CAMERA_HEIGHT_M	1.50 m	Altura del lente de la cámara respecto al nivel de la calzada, correspondiente a la posición del parabrisas de una camioneta Toyota Hilux.
Vehículo de montaje	—	Toyota Hilux	Camioneta empleada como plataforma de grabación, cuya altura de tablero proporciona una perspectiva representativa del conductor.

Nota. Elaboración propia.

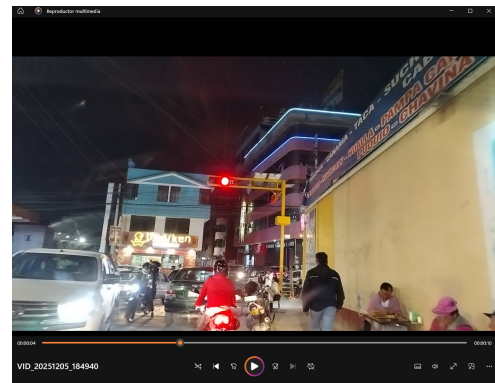
Las videograbaciones se realizaron en diferentes condiciones de iluminación y horarios, con el objetivo de que el modelo sea capaz de generalizar ante escenarios variados. Se registraron recorridos en horario diurno con luz natural plena, durante la tarde con iluminación crepuscular, y en horario nocturno con iluminación artificial, abarcando diversas vías de la ciudad de Ayacucho. En la Figura 50 se presentan fotogramas representativos extraídos de las videograbaciones, donde se aprecia la variabilidad de condiciones capturadas.

Figura 50

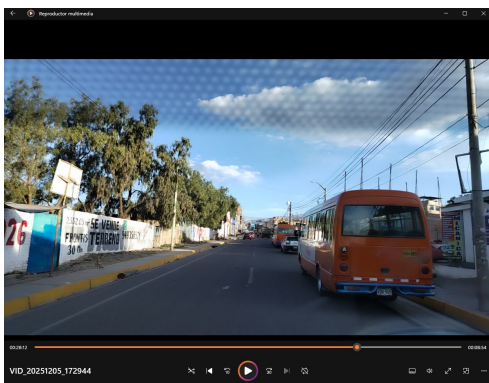
Fotogramas extraídos de las videograbaciones en diferentes condiciones



(a) Condición diurna – atardecer



(b) Condición nocturna – semáforo en rojo



(c) Condición diurna – bus de transporte



(d) Condición diurna – camiones de carga

Nota. Capturas de pantalla de videograbaciones propias.

Esta estrategia de videograbación permitió capturar situaciones reales de tránsito que difícilmente se obtienen mediante fotografías estáticas, tales como vehículos en movimiento, cambios de fase semafórica, maniobras de adelantamiento y la interacción simultánea entre múltiples actores viales. Los fotogramas extraídos de estos videos complementan el *dataset* fotográfico, aportando imágenes con desenfoque por movimiento (*motion blur*), variaciones de exposición y perspectiva dinámica, condiciones que fortalecen la robustez del modelo entrenado.

Incorporación de datasets internacionales

Con el propósito de mejorar el rendimiento del modelo y fortalecer su capacidad de generalización, se incorporaron tres datasets internacionales de acceso público

que aportan diversidad geográfica, variabilidad en las condiciones de captura y un mayor número de instancias por clase. Estos datasets fueron previamente unificados y preprocesados, conformando una base de datos complementaria de 21,004 imágenes con 116,824 instancias anotadas. A continuación se describen las fuentes empleadas.

a) BDD100K – Berkeley DeepDrive Dataset

El BDD100K (Yu, Chen, Wang et al., 2020b) es uno de los datasets de conducción más extensos y diversos disponibles en la comunidad de visión por computadora, desarrollado por la Universidad de California, Berkeley. Contiene 100,000 secuencias de video grabadas en diversas ciudades de Estados Unidos bajo múltiples condiciones de iluminación (diurna, nocturna, crepuscular), clima (soleado, lluvioso, nublado, nevado) y escenarios viales (autopistas, zonas urbanas, residenciales). De este dataset se extrajeron las anotaciones correspondientes a las clases de actores viales: persona, automóvil particular, motocicleta, camión, bus de transporte y bicicleta. Estas clases presentaban el mayor volumen de instancias anotadas y constituían la principal fuente de datos para el entrenamiento del modelo en la detección de elementos móviles del tránsito.

b) BSTLD – Bosch Small Traffic Lights Dataset

El BSTLD (Behrendt et al., 2017) fue desarrollado por la empresa Robert Bosch GmbH en Alemania, y está orientado específicamente a la detección de semáforos en entornos urbanos. El dataset contiene 10,300 imágenes con anotaciones de semáforos clasificados en cuatro estados: verde (*green*), rojo (*red*), amarillo (*yellow*) y apagado (*off*). Este dataset fue incorporado para fortalecer la capacidad del modelo en la detección y clasificación del estado de los semáforos, dado que las imágenes locales no proporcionaban un volumen suficiente de instancias de esta clase en sus diferentes estados de operación.

c) RLMD – Road Lane Marking Dataset

El RLMD es un dataset especializado en la detección de señalización horizontal sobre la calzada, que incluye anotaciones de líneas de carril, marcas de paso peatonal y líneas de pare en diversas configuraciones geométricas. De esta fuente se extrajeron las instancias correspondientes a las clases de señalización horizontal: paso peatonal, línea de pare, línea recta, línea izquierda, línea derecha y sus combinaciones. La incorporación de este dataset permitió complementar las anotaciones locales de señalización horizontal, las cuales presentaban una menor densidad de instancias debido a las condiciones de deterioro observadas en la zona de estudio.

Estrategia de integración

La fusión de estos tres datasets internacionales con los datos propios de Ayacucho se realizó mediante una estrategia de *transfer learning* progresivo en múltiples etapas. En la primera fase, el modelo preentrenado en COCO fue ajustado con los datos internacionales unificados (BDD100K + BSTLD + RLMD), lo que permitió que la red neuronal adquiriera conocimiento general sobre las clases de interés. En la segunda fase, se realizó el *fine-tuning* final exclusivamente con el dataset propio de Ayacucho, de modo que el modelo

se adaptara al contexto vial local, reconociendo las particularidades de la infraestructura, la tipología vehicular (mototaxis, buses de transporte regional) y las condiciones de señalización propias de la ciudad. Esta estrategia de adaptación de dominio (*domain adaptation*) garantiza que el conocimiento adquirido de fuentes internacionales sea transferido y contextualizado al escenario peruano andino.

3.13.4. Definición de clases y etiquetado en Labelme

Para el entrenamiento del modelo de segmentación de instancias se definieron 12 clases que representan los principales elementos de la infraestructura vial y los actores del tránsito presentes en las vías de la ciudad de Ayacucho. Estas clases fueron establecidas en función de los objetivos de la investigación y de la relevancia de cada elemento para el análisis de la seguridad vial. La configuración del archivo `dataset.yaml` empleado en el entrenamiento se detalla en la Tabla 7.

Cuadro 7

Clases definidas para el modelo de segmentación de instancias

ID	Clase	Categoría
0	Reductor de velocidad	Infraestructura vial
1	Paso peatonal	Señalización horizontal
2	Línea recta	Señalización horizontal
3	Línea recta y derecha	Señalización horizontal
4	Persona	Actor vial
5	Motocicleta	Vehículo
6	Automóvil particular	Vehículo
7	Camión	Vehículo
8	Bus de transporte	Vehículo
9	Semáforo en verde	Dispositivo de control
10	Semáforo en rojo	Dispositivo de control
11	Mototaxi	Vehículo

Nota. Elaboración propia.

La selección de estas 12 clases responde a la composición del parque vehicular y la infraestructura de control del tránsito observados durante el levantamiento de campo. Se incluyó la clase “Mototaxi” como categoría independiente debido a su alta presencia en la ciudad de Ayacucho y a sus características morfológicas distintivas que lo diferencian de otros vehículos menores. Las clases de señalización horizontal se limitaron a las configuraciones más frecuentes identificadas en la zona de estudio.

Proceso de etiquetado

El etiquetado de las imágenes se realizó mediante la herramienta LabelMe, un software de anotación de código abierto ampliamente utilizado en tareas de visión por computadora. Para optimizar el proceso de segmentación y reducir el tiempo de anotación manual, se empleó el modelo de inteligencia artificial SAM (*Segment Anything Model*) integrado en

LabelMe en su modo *accuracy*, el cual genera automáticamente máscaras de segmentación a nivel de píxel que posteriormente son verificadas y ajustadas por el investigador. Los parámetros de configuración del modelo SAM se detallan en la Tabla 8.

Cuadro 8

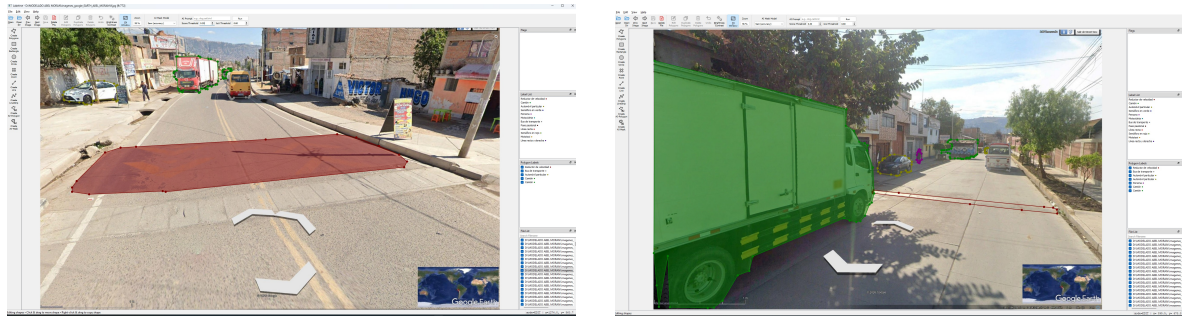
Parámetros de configuración del modelo SAM en LabelMe

Parámetro	Valor	Descripción
AI Mask Model	Sam (accuracy)	Modelo de segmentación automática en modo de alta precisión, que prioriza la calidad de la máscara sobre la velocidad de procesamiento.
Score Threshold	0.30	Umbral de confianza mínimo para aceptar una máscara generada. Un valor de 0.30 permite detectar objetos parcialmente ocultos o superpuestos.
IoU Threshold	0.80	Umbral de intersección sobre unión para la supresión de máscaras duplicadas, establecido en un valor alto para evitar la fusión de objetos adyacentes.

Nota. Elaboración propia.

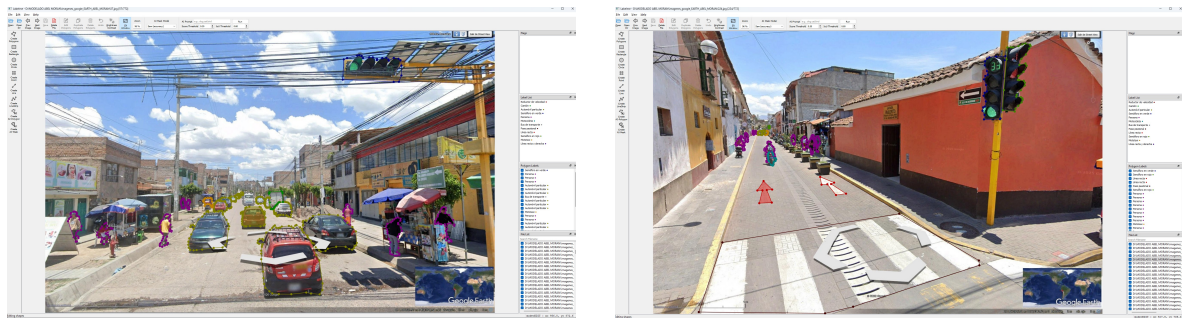
La configuración de un *Score Threshold* de 0.30 se adoptó considerando que en las escenas urbanas de Ayacucho es frecuente la superposición de objetos, tales como vehículos estacionados frente a reductores de velocidad, peatones transitando sobre pasos peatonales o motocicletas parcialmente ocultas por otros vehículos. Un umbral bajo permite que el modelo SAM genere máscaras para estos objetos parcialmente visibles, los cuales serían descartados con umbrales más restrictivos. Por su parte, el *IoU Threshold* de 0.80 garantiza que las máscaras de objetos adyacentes no se fusionen entre sí, preservando la individualidad de cada instancia anotada.

El etiquetado se realizó mediante polígonos de segmentación que delimitan con precisión el contorno de cada objeto, generando anotaciones en formato JSON que posteriormente fueron convertidas al formato YOLO Segmentation para el entrenamiento del modelo. En las Figuras 51 a 53 se presentan ejemplos representativos del proceso de etiquetado en LabelMe.

Figura 51*Etiquetado de reductor de velocidad y vehículos en LabelMe*

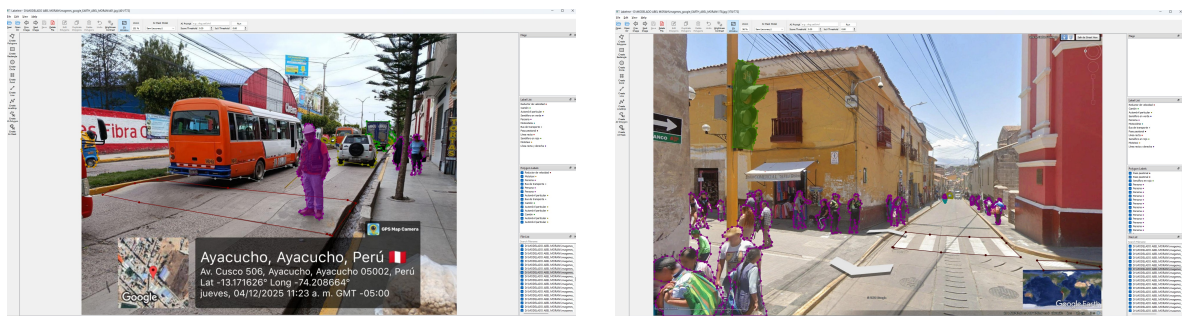
(a) Reductor de velocidad y camión

(b) Camión, línea recta y vehículos

*Nota. Capturas de pantalla de LabelMe.***Figura 52***Etiquetado de escenas con alta densidad de objetos en LabelMe*

(a) Semáforo, vehículos y peatones

(b) Peatones, pasos peatonales y señalización

*Nota. Capturas de pantalla de LabelMe.***Figura 53***Etiquetado de escenas con dispositivos de control y datos georreferenciados*

(a) Semáforos, mototaxi y personas

(b) Vehículos y actores viales – Av. Cusco

Nota. Capturas de pantalla de LabelMe.

Como se observa en las figuras, el proceso de etiquetado abarcó escenas con diversas condiciones de complejidad, desde imágenes con pocos objetos aislados hasta escenas urbanas con alta densidad de actores viales superpuestos. Las máscaras de segmentación generadas por SAM fueron revisadas individualmente para garantizar la correcta delimitación de los contornos y la asignación apropiada de cada clase, asegurando así la calidad del conjunto de datos de entrenamiento.

3.13.5. Entrenamiento del modelo

El entrenamiento del modelo de segmentación de instancias se llevó a cabo mediante una estrategia de *transfer learning* progresivo en tres etapas, donde cada fase cumple un rol específico en la transferencia y adaptación del conocimiento. Esta estrategia permite que el modelo parta de un conocimiento general sobre objetos del mundo real, se especialice en elementos viales mediante datos internacionales, y finalmente se adapte al contexto urbano específico de Ayacucho.

3.13.5.1 Etapa 01: Transfer learning base con datos internacionales.

En la primera etapa se realizó el ajuste fino del modelo YOLOv8L-seg preentrenado en el dataset COCO, empleando los datos internacionales unificados (BDD100K + BSTLD) con el propósito de que la red neuronal adquiriera conocimiento especializado en la detección de vehículos, personas y semáforos en contextos de tránsito urbano. Los hiperparámetros empleados en esta etapa se detallan en la Tabla 9.

Cuadro 9

Hiperparámetros de entrenamiento – Etapa 01

Parámetro	Valor	Descripción
Modelo base	YOLOv8L-seg.pt	Preentrenado en COCO (80 clases)
Épocas	20	Entrenamiento breve para transferencia base
Tamaño de imagen	640 × 640 px	Resolución estándar de entrada
Batch size	16	Muestras por iteración
Optimizador	SGD	Descenso de gradiente estocástico
Learning rate inicial	0.01	Tasa de aprendizaje base
Learning rate final	0.01	Sin decaimiento progresivo
Cosine LR	Sí	Programación cosenoidal del learning rate
Freeze	10 capas	Backbone congelado para preservar filtros COCO
Warmup	3 épocas	Incremento gradual del learning rate
Momentum	0.937	Factor de inercia del optimizador
Weight decay	0.0005	Regularización L2
Mosaic	1.0	Augmentation de mosaico activo
Mixup	0.15	Mezcla de imágenes al 15 %
Copy-paste	0.3	Copia de instancias al 30 %
Erasing	0.4	Borrado aleatorio al 40 %
Close mosaic	10	Desactivar mosaico en últimas 10 épocas
AMP	Sí	Precisión mixta automática

Nota. Elaboración propia.

Esta primera etapa tuvo como resultado un modelo con conocimiento general sobre las clases de actores viales (persona, motocicleta, automóvil, camión, bus de transporte) y dispositivos de control (semáforo en verde, semáforo en rojo), los cuales constituyen las clases con mayor representación en los datasets internacionales. Los pesos resultantes (*best.pt*) fueron empleados como punto de partida para la segunda etapa de entrenamiento.

3.13.5.2 Etapa 02: Fine-tuning con datos de Ayacucho a 640 px.

En la segunda etapa se realizó el ajuste fino del modelo resultante de la Etapa 01 empleando exclusivamente el dataset propio de Ayacucho, con el objetivo de adaptar el conocimiento adquirido de fuentes internacionales al contexto vial local. A diferencia de la etapa anterior, se descongelaron todas las capas de la red (*freeze=0*), permitiendo que tanto el *backbone* como la cabeza de detección se ajusten a las características específicas de las imágenes capturadas en la zona de estudio. Se configuraron 130 épocas con un mecanismo de parada temprana (*early stopping*) con paciencia de 50 épocas.

Cuadro 10

Hiperparámetros de entrenamiento – Etapa 02

Parámetro	Valor	Descripción
Modelo base	fase1_general/best.pt	Pesos de la Etapa 01
Épocas	130	Entrenamiento extendido con early stopping
Paciencia	50	Épocas sin mejora antes de detener
Tamaño de imagen	640 × 640 px	Resolución estándar
Batch size	16	Muestras por iteración
Optimizador	SGD	Descenso de gradiente estocástico
Learning rate inicial	0.001	Reducido para fine-tuning
Learning rate final	0.0001	Decaimiento progresivo
Cosine LR	Sí	Programación cosenoidal
Freeze	0 capas	Todas las capas descongeladas
Warmup	3 épocas	Incremento gradual
Momentum	0.937	Factor de inercia
Weight decay	0.0005	Regularización L2
Mosaic	1.0	Augmentation de mosaico activo
Mixup	0.15	Mezcla de imágenes al 15 %
Copy-paste	0.3	Copia de instancias al 30 %
Erasing	0.1	Borrado aleatorio reducido al 10 %
Degrees	15°	Rotación aleatoria
Close mosaic	10	Desactivar mosaico en últimas 10 épocas
AMP	Sí	Precisión mixta automática

Nota. Elaboración propia.

3.13.6. Especificaciones del hardware de entrenamiento y despliegue del sistema

El desarrollo del sistema QHAWAY ADAS requirió dos configuraciones de hardware diferenciadas: una para el entrenamiento del modelo de segmentación de instancias y otra para la recolección de datos en campo y la operación del sistema en tiempo real.

3.13.6.1 Hardware de entrenamiento del modelo

El entrenamiento del modelo YOLOv8L-seg se ejecutó en un equipo de cómputo de alto rendimiento con las especificaciones detalladas en el Cuadro 11.

Cuadro 11

Especificaciones del hardware de entrenamiento del modelo YOLOv8L-seg

Componente	Especificación	Rol en el entrenamiento
Unidad de procesamiento gráfico (GPU)	NVIDIA GeForce RTX 5080 (16 GB VRAM)	Cómputo de las operaciones matriciales del entrenamiento con precisión mixta automática (AMP), acelerado mediante CUDA. Procesó lotes de 8 imágenes a resolución 1024 px en la Fase 3.
Unidad de procesamiento central (CPU)	Intel Core i9 (13.ª generación, 24 núcleos)	Gestión del flujo de datos de entrenamiento, preprocesamiento de imágenes y coordinación entre los procesos de carga de datos y el cómputo en GPU.
Memoria RAM	64 GB DDR5	Almacenamiento en memoria del dataset durante el entrenamiento para minimizar operaciones de lectura desde disco y reducir la latencia entre iteraciones.
Almacenamiento	SSD NVMe 2 TB	Almacenamiento del dataset completo (25,602 imágenes), pesos del modelo en cada fase y registros de entrenamiento. La velocidad de lectura secuencial superior a 5 GB/s redujo el tiempo de carga de datos.
Sistema operativo	Windows 11 Pro (64 bits)	Plataforma de desarrollo del software QHAWAY ADAS y ejecución del entrenamiento mediante el entorno Anaconda con Python 3.11.
CUDA	CUDA 12.x + cuDNN 8.x	Interfaz de programación para la aceleración del entrenamiento en GPU NVIDIA, compatible con PyTorch 2.x.
Tiempo total de entrenamiento	24.1 horas	Distribuido en: Fase 1 (≈ 1.5 h), Fase 2 (≈ 1.9 h) y Fase 3 (20.67 h).

Nota. Elaboración propia.

La selección de la GPU NVIDIA GeForce RTX 5080 fue determinante para la viabilidad del entrenamiento a resolución 1024 px en la Fase 3: Con un tamaño de lote de 8 imágenes

a esta resolución, cada imagen ocupa aproximadamente 12 MB en memoria de GPU tras el preprocesamiento, lo que requiere un mínimo de 12 GB de VRAM disponible. La activación de la precisión mixta automática (AMP) redujo el consumo de memoria en aproximadamente un 40 % respecto al entrenamiento en precisión completa, permitiendo ejecutar la Fase 3 sin desbordamiento de memoria.

3.13.6.2 Hardware de recolección de datos en campo

La recolección de datos en el entorno vial ayacuchano empleó los dispositivos de captura detallados en el Cuadro 12.

Cuadro 12

Especificaciones del hardware de recolección de datos en campo

Dispositivo	Especificación	Uso en la investigación
Smartphone de captura	Xiaomi Redmi Note 12 Pro Sensor principal: 108 MP (Samsung ISOCELL HM6) Apertura: f/1.9 Video: Full HD 1920 × 1080 px a 30 fps Estabilización: EIS (<i>Electronic Image Stabilization</i>)	Captura de video desde vehículo en movimiento y fotografías convencionales. Montado en soporte con ventosa sobre el parabrisas del vehículo a 1.50 m de altura. La grabación en modo Full HD garantizó la relación entre resolución de captura y resolución de entrenamiento (1024 px).
Vehículo de captura	Toyota Hilux (Toyota Motor Corporation, Toyota, Japón) Distancia entre ejes: 3,085 mm Altura de parabrisas: 1.50 m respecto a la calzada	Plataforma de montaje del smartphone para la captura de video en conducción. La altura de montaje de 1.50 m sobre la calzada constituye el parámetro h_{cam} del modelo geométrico de estimación de distancias.
Soporte de montaje	Soporte con ventosa para parabrisas Rótula de ajuste en 3 ejes Compatibilidad: smartphones de 5.5 a 7.0 pulgadas	Fijación estable del smartphone al parabrisas con ángulo de inclinación de 5° hacia abajo respecto al eje horizontal, parámetro θ_{tilt} del estimador de distancias.
Aplicación GPS Map Camera	Versión para Android Precisión GPS: ± 3 m (GNSS estándar) Resolución de fotografías: hasta 12 MP	Captura de fotografías georreferenciadas con registro automático de coordenadas (latitud, longitud), fecha y hora en cada imagen, empleadas en el levantamiento métrico de reductores de velocidad informales.

Nota. Elaboración propia.

3.13.6.3 Hardware de despliegue y operación del sistema

La validación operacional del sistema QHAWAY ADAS bajo condiciones reales de conducción se realizó empleando el mismo equipo de entrenamiento, en el cual se ejecutó el software en tiempo real durante los recorridos de prueba en las avenidas Arenales y Cusco. El

Cuadro 13 resume las métricas de rendimiento obtenidas durante la operación.

Cuadro 13

Métricas de rendimiento del hardware durante la operación del sistema QHAWAY ADAS en tiempo real

Métrica	Valor obtenido	Descripción
Velocidad de inferencia	19.2–25.4 FPS	Fotogramas procesados por segundo durante la operación en los cuatro escenarios de validación, incluyendo detección, seguimiento, estimación de distancias y gestión de alertas.
Latencia de inferencia	17–20 ms/fotogramay	Tiempo transcurrido entre la captura del fotograma y la generación de la respuesta del sistema (máscara + distancia + alerta).
Consumo de VRAM durante inferencia	≈4 GB	Memoria de GPU utilizada durante la inferencia del modelo a resolución 1024 px con el módulo de seguimiento activo.
Tiempo de inicialización del sistema	≈8 segundos	Tiempo desde el inicio de la aplicación hasta la habilitación del procesamiento en tiempo real, incluyendo carga del modelo y generación de archivos de audio.
Umbral de confianza operativo	0.55	Valor de confianza mínimo determinado a partir de la curva F1-Confianza del modelo (Figura 58) para maximizar el equilibrio entre precisión y sensibilidad.

Nota. Elaboración propia. Los valores de FPS y latencia corresponden al promedio de los cuatro escenarios de validación operacional bajo condiciones reales de conducción en las avenidas Arenales y Cusco de Ayacucho.

La velocidad de procesamiento obtenida (19.2–25.4 FPS) supera el umbral mínimo de 15 FPS establecido en la hipótesis de la investigación y es compatible con los requisitos de los sistemas de alerta de colisión frontal, que exigen tiempos de respuesta inferiores al intervalo de percepción–reacción del conductor estimado en aproximadamente 1.5 segundos a velocidades urbanas (National Highway Traffic Safety Administration, 2013). La latencia de 17–20 ms por fotograma implica que el sistema genera la respuesta de alerta con una anticipación de aproximadamente 1.48 a 1.49 segundos respecto al tiempo de reacción del conductor, margen suficiente para la activación de maniobras evasivas en el entorno vial urbano de Ayacucho.

3.13.7. Resultados de la Fase 3: Optimización a resolución 1024

La tercera y última fase del entrenamiento progresivo incrementó la resolución de entrada a 1024×1024 píxeles, partiendo de los pesos óptimos obtenidos en la Fase 2 (*best.pt* a 640 px). Se configuraron 300 épocas con *early stopping* (*patience*=50), optimizador SGD con tasa de aprendizaje coseno (*cos_lr*=True, *lr*₀=0.0005), aumento de datos mediante mosaico,

mixup (0.15), *copy-paste* (0.3) y *randaugment*. El entrenamiento se ejecutó en GPU NVIDIA con AMP habilitado, procesando lotes de 8 imágenes a resolución 1024. El Cuadro 14 detalla la configuración completa.

Cuadro 14

Configuración de hiperparámetros de la Fase 3 (1024 px)

Parámetro	Valor
Modelo base	YOLOv8L-seg (Fase 2 best.pt)
Resolución de entrada	1024×1024 px
Épocas configuradas / ejecutadas	300 / 223
<i>Early stopping</i> (<i>patience</i>)	50
Tamaño de lote (<i>batch</i>)	8
Optimizador	SGD (<i>momentum</i> =0.937)
Tasa de aprendizaje inicial (lr_0)	0.0005
Tasa de aprendizaje final (lr_f)	1×10^{-5}
Decaimiento de pesos (<i>weight_decay</i>)	0.0005
Épocas de calentamiento (<i>warmup</i>)	5
Programación de lr	Coseno (<i>cos_lr</i> =True)
Mosaico	1.0 (desactivado en últimas 20 épocas)
<i>Mixup</i>	0.15
<i>Copy-paste</i>	0.3
Autoaumentación (<i>auto_augment</i>)	<i>randaugment</i>
Borrado aleatorio (<i>erasing</i>)	0.15
Umbral de IoU	0.7
Precisión mixta (AMP)	Habilitado
Semilla (<i>seed</i>)	0 (determinístico)
Tiempo total de entrenamiento	20.67 horas

Nota. Elaboración propia.

El mecanismo de *early stopping* detuvo el entrenamiento en la época 223, dado que la mejor aptitud combinada (*fitness*) se registró en la época 173 sin mejoras durante las 50 épocas subsiguientes. La función de aptitud, definida como $\text{fitness} = 0,1 \times \text{mAP}_{50} + 0,9 \times \text{mAP}_{50-95}$, pondera la precisión a umbrales de solapamiento estrictos, seleccionando un modelo robusto tanto en localización gruesa como en delimitación precisa de los elementos viales.

3.13.7.1 Métricas globales del modelo óptimo

El Cuadro 15 presenta las métricas de rendimiento del modelo óptimo (época 173, *best.pt*), correspondientes tanto a la tarea de detección de cajas delimitadoras (*Box*) como a la segmentación de instancias (*Mask*).

Cuadro 15

Métricas globales del modelo YOLOv8L-seg — Fase 3 (1024 px, época 173)

Métrica	Box	Mask
Precisión (<i>Precision</i>)	0.885	0.876
Exhaustividad (<i>Recall</i>)	0.724	0.690
mAP_{50}	0.810	0.778
mAP_{50-95}	0.603	0.511
F1 (<i>Mask</i> , conf. óptima)	—	0.77 (conf. = 0.560)

Nota. Elaboración propia a partir de los resultados del entrenamiento.

El modelo alcanzó un mAP_{50} de 81.0 % en detección y 77.8 % en segmentación de instancias, con una precisión de 88.5 % y 87.6 % respectivamente. Estas métricas representan una mejora sustancial respecto a las fases previas del entrenamiento progresivo, confirmando que el incremento de resolución de 640 a 1024 píxeles permite al modelo capturar detalles finos de los elementos viales, particularmente en clases de pequeña escala como semáforos y señalizaciones horizontales.

3.13.7.2 Rendimiento por clase

El Cuadro 16 desagrega el rendimiento del modelo por cada una de las 12 categorías de elementos viales, tanto para la tarea de detección (mAP_{50} Box) como para la segmentación de instancias (mAP_{50} Mask). Los valores fueron extraídos de las curvas Precisión-Exhaustividad (*Precision-Recall*) del modelo óptimo.

Cuadro 16*Rendimiento por clase del modelo YOLOv8L-seg — Fase 3 (1024 px)*

Clase	Instancias	mAP ₅₀ Box	mAP ₅₀ Mask
Reductor de velocidad	379	0.923	0.897
Línea recta y derecha	72	0.898	0.898
Semáforo en rojo	255	0.892	0.785
Persona	4,608	0.866	0.826
Semáforo en verde	251	0.861	0.828
Automóvil particular	2,256	0.828	0.791
Motocicleta	1,025	0.791	0.786
Mototaxi	566	0.769	0.738
Paso peatonal	563	0.765	0.771
Bus de transporte	440	0.762	0.728
Camión	177	0.756	0.659
Línea recta	109	0.610	0.628
Promedio (todas las clases)	10,701	0.810	0.778

Nota. Elaboración propia. Los valores de mAP₅₀ por clase fueron extraídos de las curvas PR del modelo óptimo (best.pt, época 173).

El análisis por clase revela un comportamiento diferenciado según la naturaleza y la representatividad de cada categoría en el conjunto de datos:

- **Clases de alto rendimiento (mAP₅₀ Box \geq 0.85):** El reductor de velocidad (0.923), la línea recta y derecha (0.898), el semáforo en rojo (0.892), la persona (0.866) y el semáforo en verde (0.861) conforman el grupo de mayor precisión. La geometría definida de los reductores y las señalizaciones direccionales, junto con la abundancia de instancias de la clase persona (4,608), facilitan el aprendizaje discriminativo del modelo.
- **Clases de rendimiento intermedio ($0.75 \leq$ mAP₅₀ Box $<$ 0.85):** El automóvil particular (0.828), la motocicleta (0.791), el mototaxi (0.769), el paso peatonal (0.765), el bus de transporte (0.762) y el camión (0.756) presentan métricas competitivas. El mototaxi, pese a ser una categoría exclusiva del contexto local no presente en COCO, alcanza un rendimiento aceptable gracias a su morfología distintiva de tres ruedas.
- **Clase de menor rendimiento (mAP₅₀ Box $<$ 0.70):** La línea recta (0.610) constituye la categoría más desafiante, atribuible tanto al reducido número de instancias (109) como a la variabilidad de su apariencia visual según el estado de conservación de la pintura vial, la iluminación y la perspectiva de captura. Esta limitación sugiere la necesidad de incrementar las muestras de entrenamiento para señalizaciones horizontales deterioradas en futuras iteraciones del dataset.

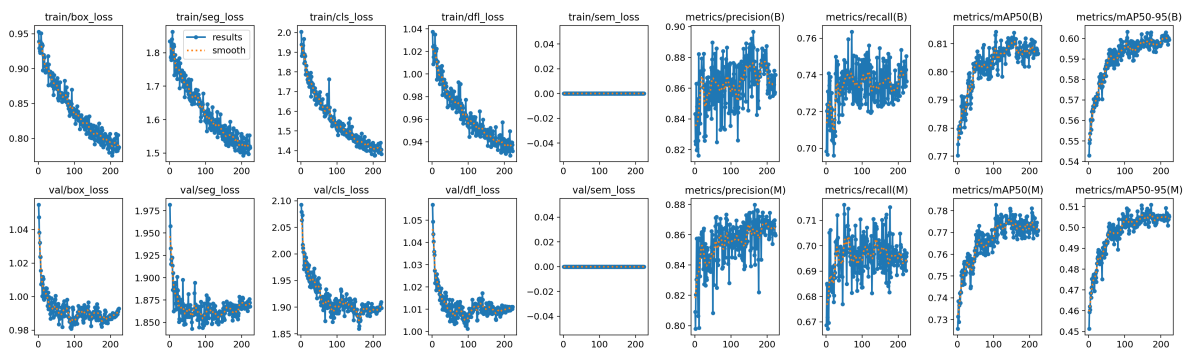
Cabe destacar que la clase línea recta y derecha, pese a contar con apenas 72 instancias, alcanza un mAP_{50} Box de 0.898, lo cual indica que su patrón geométrico combinado (flecha recta con giro a la derecha) posee características visuales suficientemente discriminativas para que el modelo las aprenda con eficacia.

3.13.7.3 Progresión del entrenamiento

La Figura 54 presenta la evolución de las funciones de pérdida y las métricas de rendimiento a lo largo de las 223 épocas de la Fase 3. Las curvas de pérdida de entrenamiento (*train*) exhiben un descenso monotónico consistente, con reducciones del 17.4 % en *box_loss* (de 0.953 a 0.787), 17.4 % en *seg_loss* (de 1.835 a 1.516) y 31.0 % en *cls_loss* (de 2.005 a 1.383). La pérdida de validación muestra una estabilización a partir de la época 80, indicando que el modelo alcanzó la convergencia sin evidencia de sobreajuste (*overfitting*).

Figura 54

Evolución de las funciones de pérdida y métricas de rendimiento durante la Fase 3



Nota. Elaboración propia. Fila superior: Pérdidas de entrenamiento. Fila inferior: Pérdidas de validación y métricas de rendimiento.

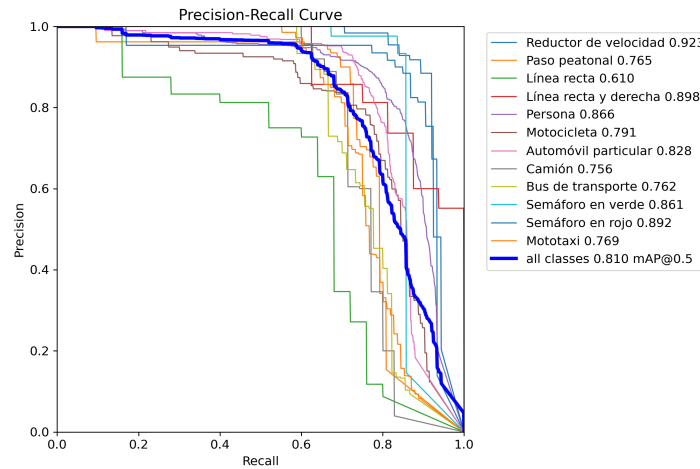
Las métricas de precisión y exhaustividad oscilan en la segunda mitad del entrenamiento (épocas 100–223), lo cual es esperable dado el tamaño moderado del conjunto de datos (772 imágenes). No obstante, el mAP_{50-95} muestra una tendencia ascendente sostenida hasta la época 173, confirmando que el modelo continuó refinando la calidad de sus predicciones incluso después de estabilizar el mAP_{50} .

3.13.7.4 Curvas Precisión-Exhaustividad (*Precision-Recall*)

Las Figuras 55 y 56 presentan las curvas Precisión-Exhaustividad para las tareas de detección y segmentación de instancias, respectivamente. El área bajo cada curva corresponde al mAP_{50} de la clase.

Figura 55

Curva Precisión-Exhaustividad para detección de cajas delimitadoras (Box) — $mAP_{50} = 0.810$

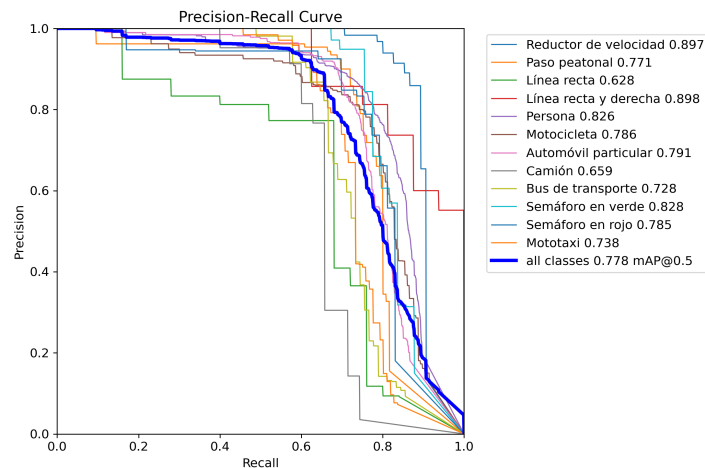


Nota. Elaboración propia. La curva gruesa azul representa el promedio de todas las clases.

En la curva de detección (Figura 55), las clases reductor de velocidad (0.923) y línea recta y derecha (0.898) mantienen alta precisión incluso a niveles elevados de exhaustividad, indicando que el modelo identifica correctamente la mayoría de sus instancias con pocas falsas alarmas. En contraste, la curva de la línea recta (0.610) desciende pronunciadamente a partir de un recall de 0.4, reflejando confusiones frecuentes con el fondo vial deteriorado.

Figura 56

Curva Precisión-Exhaustividad para segmentación de instancias (Mask) — $mAP_{50} = 0.778$



Nota. Elaboración propia. La curva gruesa azul representa el promedio de todas las clases.

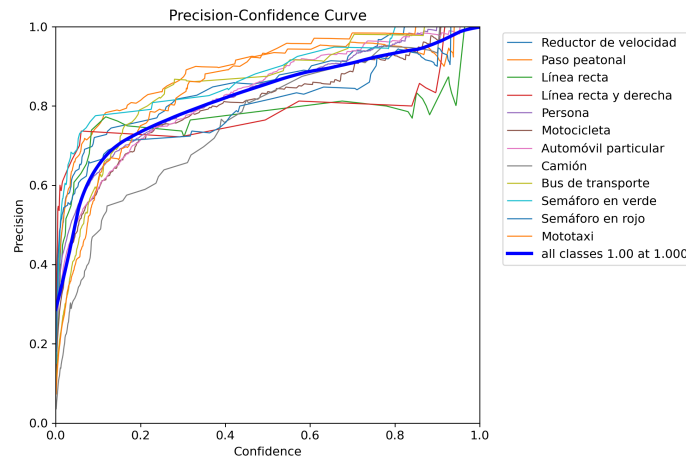
La curva de segmentación (Figura 56) muestra un patrón análogo, con una diferencia media de 3.2 puntos porcentuales entre el mAP_{50} Box y Mask. Esta diferencia es menor en clases con contornos regulares (línea recta y derecha: 0.898 Box \approx 0.898 Mask) y mayor en clases con formas complejas (camión: 0.756 Box vs. 0.659 Mask), donde la delimitación precisa de la máscara resulta más exigente.

3.13.7.5 Curvas de Precisión-Confianza y F1-Confianza

Las Figuras 57 y 58 presentan las curvas de Precisión-Confianza y F1-Confianza para la tarea de segmentación de instancias.

Figura 57

Curva Precisión-Confianza para segmentación de instancias (Mask)

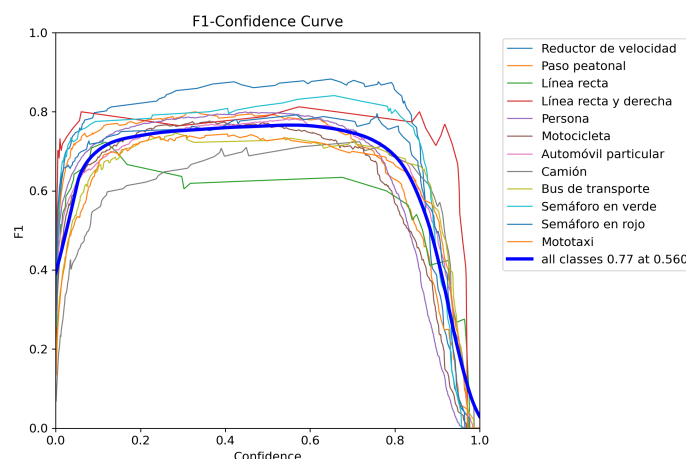


Nota. Elaboración propia. Precisión de todas las clases = 1.00 en confianza = 1.000.

La curva de Precisión-Confianza (Figura 57) muestra que, a umbrales de confianza superiores a 0.8, la precisión de la mayoría de las clases supera el 90 %, lo cual indica que las predicciones de alta confianza del modelo son altamente fiables. Este comportamiento resulta deseable para aplicaciones de seguridad vial donde es preferible no generar alertas falsas.

Figura 58

Curva F1-Confianza para segmentación de instancias (Mask)



Nota. Elaboración propia. El máximo F1 promedio de todas las clases es 0.77 en confianza = 0.560.

La curva F1-Confianza (Figura 58) identifica el umbral óptimo de operación del modelo en un valor de confianza de 0.560, donde el equilibrio entre precisión y exhaustividad alcanza su punto máximo con un F1 de 0.77 para el promedio de todas las clases. Las clases reductor

de velocidad y línea recta y derecha superan un F1 de 0.80, mientras que la línea recta se mantiene por debajo de 0.65. Este umbral óptimo se empleará como referencia para la configuración del sistema de alertas preventivas.

3.13.7.6 Evolución progresiva del entrenamiento en tres fases

El Cuadro 17 sintetiza la mejora acumulada a lo largo de las tres fases del entrenamiento progresivo, evidenciando el impacto de cada estrategia aplicada.

Cuadro 17

Progresión del rendimiento a lo largo de las tres fases del entrenamiento

Fase	Resolución	Épocas	mAP ₅₀ Box	mAP ₅₀ Mask	Tiempo
Fase 1: Preentrenamiento (dataset internacional)	640 px	20	0.503	0.453	~1.5 h
Fase 2: <i>Fine-tuning</i> (dataset Ayacucho)	640 px	130	0.744	0.692	~1.9 h
Fase 3: Optimización (resolución alta)	1024 px	223	0.810	0.778	20.7 h
Mejora acumulada (Fase 1 → Fase 3)	—	—	+30.7 pp	+32.5 pp	24.1 h

Nota. Elaboración propia. pp = puntos porcentuales.

El entrenamiento progresivo en tres fases produjo una mejora acumulada de 30.7 puntos porcentuales en mAP₅₀ Box y 32.5 puntos en mAP₅₀ Mask. La estrategia de incremento progresivo de resolución (640→1024 px) resultó particularmente efectiva: la Fase 3 mejoró el mAP₅₀ Box en 6.6 puntos porcentuales respecto a la Fase 2, demostrando que la resolución superior permite al modelo discriminar con mayor precisión elementos de pequeña escala como semáforos, señalizaciones horizontales y marcas viales deterioradas, característicos del contexto vial de Ayacucho.

3.13.7.7 Comparación con el modelo base de referencia

El Cuadro 18 contrasta el rendimiento del modelo optimizado frente al modelo preexistente de referencia, evidenciando la contribución metodológica del presente trabajo.

Cuadro 18

Comparación del modelo optimizado frente al modelo base de referencia

Modelo	Precisión	mAP ₅₀ Box	mAP ₅₀₋₉₅ Box	mAP ₅₀ Mask
Modelo base de referencia (18 clases, 640 px)	0.745	0.647	0.481	0.529
YOLOv8L-seg optimizado (12 clases, 1024 px)	0.885	0.810	0.603	0.778
Mejora	+14.0 pp	+16.3 pp	+12.2 pp	+24.9 pp

Nota. Elaboración propia. pp = puntos porcentuales.

La mejora más pronunciada se observa en la segmentación de instancias (+24.9 pp en mAP₅₀ Mask), lo cual se atribuye a tres factores: (1) la depuración y re-etiquetado del dataset con LabelMe + SAM2 generó anotaciones poligonales de mayor calidad; (2) el incremento de resolución a 1024 px mejoró la delimitación de contornos finos; y (3) la eliminación de clases inconsistentes del modelo base y la incorporación de la categoría mototaxi, específica del contexto ayacuchano, permitió al modelo especializarse en las categorías relevantes del escenario de estudio.

3.13.8. Comparación multi-arquitectura: YOLOv8L-seg frente a la familia YOLO26 (nano, small, large)

Con el propósito de evaluar el impacto de las innovaciones arquitectónicas recientes sobre el conjunto de datos QHAWAY, se realizó una comparación sistemática entre el modelo YOLOv8L-seg y la familia de modelos YOLO26 (Jocher & Qiu, 2026; Sapkota et al., 2026). La arquitectura YOLO26 incorpora varios avances relevantes para la segmentación de instancias: inferencia sin supresión de no-máximos de extremo a extremo, eliminación de la función de pérdida de distribución focal, pérdida progresiva con asignación de etiquetas sensible a objetos pequeños (ProgLoss + STAL) para mejorar la detección de clases subrepresentadas, el optimizador MuSGD y una pérdida de segmentación semántica que mejora la calidad de las máscaras de instancia (Sapkota et al., 2026).

Las tres variantes de escala de la familia YOLO26-seg (nano, small y large) fueron entrenadas sobre el mismo conjunto de datos híbrido QHAWAY empleando la misma estrategia de entrenamiento progresivo en tres fases y las mismas configuraciones de aumento de datos (mixup 0.15, copy-paste 0.3, borrado aleatorio 0.15, programación cosenoidal de la tasa de aprendizaje).

3.13.8.1 Resultados globales de la comparación

El Cuadro 19 presenta los resultados globales de la comparación multi-arquitectura sobre el conjunto de validación.

Cuadro 19

Comparación multi-arquitectura sobre el conjunto de validación QHAWAY. Los modelos YOLO26 reportan el mejor punto de control en 300 épocas a 1024 px; YOLOv8L-seg reporta el mejor punto de control en la época 173.

Modelo	Mejor época	mAP ₅₀ Box	mAP ₅₀ Mask	mAP ₅₀₋₉₅ Box	mAP ₅₀₋₉₅ Mask	P(B)	R(B)
YOLOv8L-seg	173	0.810	0.778	—	—	0.885	0.724
YOLO26n-seg	222	0.665	0.631	0.436	0.386	0.737	0.602
YOLO26s-seg	242	0.759	0.730	0.541	0.470	0.831	0.676
YOLO26L-seg	179	0.829	0.788	0.614	0.510	0.868	0.751

Nota. Elaboración propia a partir de los resultados del entrenamiento. P(B): precisión de detección de cajas delimitadoras. R(B): exhaustividad de detección de cajas delimitadoras. Los modelos YOLO26 no reportan mAP₅₀₋₉₅ para YOLOv8L-seg dado que este indicador no fue calculado en la Fase 3 de ese modelo.

El modelo YOLO26L-seg alcanzó el mayor desempeño global (mAP₅₀ Box = 0.829, mAP₅₀ Mask = 0.788), superando al modelo YOLOv8L-seg en 1.9 puntos porcentuales en detección y 1.0 punto porcentual en segmentación de instancias. La mejora en exhaustividad (+2.7 pp) indica que YOLO26L-seg identifica una mayor proporción de los elementos presentes en la escena, aspecto especialmente relevante para un sistema de seguridad vial donde las detecciones omitidas conllevan mayor riesgo que las falsas alarmas.

El estudio de escalado de la familia YOLO26 revela mejoras monotónicamente consistentes al aumentar la capacidad del modelo: nano (0.665), small (0.759) y large (0.829) en mAP₅₀ Box, con una brecha de 16.5 puntos porcentuales entre las variantes nano y large. Este patrón confirma que la capacidad del modelo sigue siendo el factor dominante para la tarea de 12 clases del conjunto de datos QHAWAY a resolución de 1024 píxeles, y proporciona orientación práctica para el despliegue en distintas configuraciones de hardware.

3.13.8.2 Comparación por clase: YOLOv8L-seg frente a YOLO26L-seg

El Cuadro 20 presenta la comparación detallada por clase entre el modelo YOLOv8L-seg (época 173) y el modelo YOLO26L-seg (época 179).

Cuadro 20

Comparación del desempeño por clase entre YOLOv8L-seg (época 173) y YOLO26L-seg (época 179). Δ indica la ganancia (positiva) o pérdida (negativa) de YOLO26L-seg respecto a YOLOv8L-seg. Los valores en negrita indican el mayor mAP_{50} por clase.

Clase	mAP_{50} Box			mAP_{50} Mask		
	v8L	26L	Δ	v8L	26L	Δ
Semáforo en rojo	0.892	0.935	+4.3	0.785	0.795	+1.0
Persona	0.866	0.889	+2.3	0.826	0.846	+2.0
Semáforo en verde	0.861	0.899	+3.8	0.828	0.824	-0.4
Reductor de velocidad	0.923	0.883	-4.0	0.897	0.882	-1.5
Automóvil particular	0.828	0.853	+2.5	0.791	0.793	+0.2
Camión	0.756	0.850	+9.4	0.659	0.850	+19.1
Mototaxi	0.769	0.821	+5.2	0.738	0.710	-2.8
Motocicleta	0.791	0.820	+2.9	0.786	0.796	+1.0
Línea recta	0.610	0.697	+8.7	0.628	0.697	+6.9
Línea recta y derecha	0.898	0.773	-12.5	0.898	0.773	-12.5
Paso peatonal	0.765	0.769	+0.4	0.771	0.776	+0.5
Bus de transporte	0.762	0.763	+0.1	0.728	0.709	-1.9
Todas las clases	0.810	0.829	+1.9	0.778	0.788	+1.0

Nota. Elaboración propia. Δ expresado en puntos porcentuales (pp). v8L: YOLOv8L-seg. 26L: YOLO26L-seg.

El análisis por clase revela un comportamiento diferenciado según la naturaleza y el volumen de instancias de cada categoría:

- Clases con mayor ganancia en YOLO26L-seg:** La clase camión registró la mejora más pronunciada (+9.4 pp Box, +19.1 pp Mask). Esta mejora es atribuible a dos factores que actúan de forma conjunta: la alta variabilidad visual intraclase — con solo 177 instancias de entrenamiento, la flota local incluye camiones de carga, cisternas, volquetes y camiones con carrocería de lona, cada uno con silueta y textura superficial marcadamente distintas — y los mecanismos ProgLoss y STAL de YOLO26, diseñados específicamente para mejorar la detección de clases subrepresentadas y visualmente heterogéneas (Sapkota et al., 2026). La clase línea recta también registró una ganancia significativa (+8.7 pp Box, +6.9 pp Mask), atribuible al manejo mejorado de elementos de pequeña escala por parte de YOLO26 en escenas con deterioro del pavimento.
- Clases con mayor pérdida en YOLO26L-seg:** La línea recta y derecha presentó el mayor retroceso (-12.5 pp Box y -12.5 pp Mask), y el reductor de velocidad disminuyó 4.0 pp en detección. Estas pérdidas pueden reflejar la diferencia en el comportamiento de la arquitectura sin supresión de no-máximos de YOLO26, que elimina la flexibilidad de post-procesamiento que beneficiaba a las clases de alta confianza y geometría

regular en YOLOv8.

- Clases relevantes para la seguridad vial:** La clase persona — objetivo de mayor prioridad para el sistema de alertas — mejoró +2.3 pp Box y +2.0 pp Mask, lo que representa una ganancia significativa considerando que cuenta con 4,608 instancias de validación donde las mejoras marginales son estructuralmente más difíciles de obtener. La clase mototaxi mejoró +5.2 pp en detección (0.769 → 0.821), aunque presentó una ligera disminución en segmentación (−2.8 pp Mask), lo que sugiere que la cabeza de segmentación de YOLO26 maneja de forma diferente el contorno irregular del mototaxi.

3.13.8.3 Orientación para el despliegue según el hardware disponible

El estudio de escalado de la familia YOLO26 proporciona orientación práctica para el despliegue del sistema QHAWAY ADAS en distintas configuraciones de hardware. El Cuadro 21 resume las métricas clave y el contexto de despliegue recomendado para cada variante.

Cuadro 21

Estudio de escalado de la familia YOLO26-seg: desempeño y contexto de despliegue recomendado

Variante	Parámetros	mAP ₅₀ Box	mAP ₅₀₋₉₅ Box	FPS (RTX 5080)	Contexto de despliegue recomendado
YOLO26n-seg	≈3.2 M	0.665	0.436	>60	Dispositivos embebidos con memoria inferior a 4 GB (Jetson Nano, Raspberry Pi 4)
YOLO26s-seg	≈9.8 M	0.759	0.541	>45	Computadoras de placa única de gama media (Jetson Xavier NX, Intel NUC)
YOLO26L-seg	≈25.3 M	0.829	0.614	19.2–25.4	GPU de escritorio o portátil con 8 GB de VRAM o más (configuración del sistema QHAWAY ADAS)

Nota. Elaboración propia. Los valores de fotogramas por segundo corresponden a la operación del sistema completo QHAWAY ADAS (detección + seguimiento + estimación de distancias + gestión de alertas) medidos en la GPU NVIDIA GeForce RTX 5080 durante la validación operacional. Los valores para YOLO26n-seg y YOLO26s-seg son estimados a partir del escalado de parámetros reportado por Sapkota et al. (2026).

En conjunto, la comparación multi-arquitectura confirma que YOLO26L-seg constituye el modelo de mayor desempeño global para el conjunto de datos QHAWAY, superando a YOLOv8L-seg en 9 de las 12 clases para detección y en 7 de las 12 clases para segmentación de instancias, con especial relevancia en las clases de mayor prioridad para la seguridad vial (persona, semáforos y mototaxi). El estudio de escalado (nano: 0.665, small: 0.759, large: 0.829 en mAP₅₀ Box) demuestra mejoras monotónicas consistentes con el incremento de capacidad del modelo, proporcionando orientación cuantitativa para el despliegue del

sistema en distintas configuraciones de hardware.

3.13.9. Análisis de explicabilidad del modelo mediante Grad-CAM

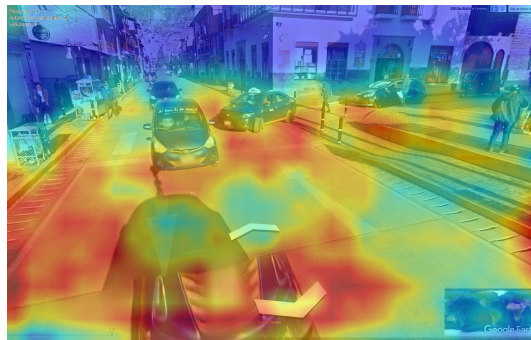
Tras completar el entrenamiento progresivo en tres fases, se realizó un análisis de explicabilidad (*Explainability Analysis*) del modelo YOLOv8L-seg optimizado, empleando la técnica Grad-CAM (*Gradient-weighted Class Activation Mapping*). Esta técnica permite visualizar las regiones de la imagen que el modelo considera más relevantes para formular sus predicciones, generando mapas de activación superpuestos (*heatmaps*) donde las zonas de color rojo indican máxima activación y las zonas azules representan mínima atención. El objetivo de este análisis es verificar que las decisiones del modelo se fundamentan en características visuales semánticamente relevantes de los elementos viales, descartando correlaciones espurias o artefactos del proceso de aprendizaje.

3.13.9.1 Mapas de calor de activación (*Attention Heatmaps*)

Se generaron mapas de calor sobre imágenes representativas de las avenidas de Ayacucho, seleccionadas por su diversidad en densidad vehicular, cantidad de clases presentes y condiciones de iluminación. Las Figuras 59, 60 y 61 presentan los resultados obtenidos.

Figura 59

Mapa de calor de activación — Av. Mariscal Cáceres (imagen 118)



Nota. Elaboración propia. Detecciones: Persona (13), Automóvil particular (8), Mototaxi (2). Las zonas rojas indican máxima activación del modelo.

La Figura 59 corresponde a una intersección de alta densidad vial en la avenida Mariscal Cáceres, donde el modelo identificó 23 elementos viales distribuidos en tres categorías. La activación más intensa se concentra en la persona más próxima a la cámara, cubriendo prácticamente toda su silueta con tonalidades rojo-amarillas. Este comportamiento indica que el modelo prioriza la detección de peatones en proximidad inmediata, lo cual resulta coherente con los requerimientos de un sistema de alertas de seguridad vial donde la detección temprana de usuarios vulnerables en cercanía constituye una función crítica. Las zonas de activación se extienden también a los automóviles estacionados y en circulación en el plano medio de la escena, mientras que las fachadas de edificios y el cielo permanecen en tonalidades azules, confirmando que el modelo discrimina correctamente entre los elementos de interés vial y el fondo urbano circundante.

Figura 60

Mapa de calor de activación — Jirón Libertad (imagen 146)



Nota. Elaboración propia. Detecciones: Persona (12), Motocicleta (2), Automóvil particular (7), Bus de transporte (2), Semáforo en rojo (2), Mototaxi (2). Escena con la mayor diversidad de clases detectadas.

La Figura 60 presenta la escena con mayor diversidad de clases detectadas, con seis categorías diferentes identificadas simultáneamente en una sola imagen del jirón Libertad. El aspecto más relevante de este mapa de activación es la distribución espacial de la atención del modelo: las zonas rojas de máxima intensidad se concentran en la parte superior de la imagen, específicamente en la zona donde se ubican los semáforos y las estructuras elevadas de los edificios.

Figura 61

Mapa de calor de activación — Av. Las Américas (imagen 429)



Nota. Elaboración propia. Detecciones: Reductor de velocidad (2), Persona (1), Motocicleta (1), Automóvil particular (5). Imagen georeferenciada: Lat $-13,17532^\circ$, Long $-74,206757^\circ$, capturada el 04/12/2025.

La Figura 61 muestra la activación del modelo sobre una escena de la avenida Las Américas, con georeferenciación GPS (Lat $-13,17532^\circ$, Long $-74,206757^\circ$). La atención se distribuye en dos focos principales: los reductores de velocidad en primer plano generan activación amarilla-verde en la zona inferior central, validando que el modelo identifica esta infraestructura vial característica de las vías ayacuchanas; simultáneamente, la persona con chaleco de seguridad en la acera derecha genera activación localizada. Los vehículos al fondo de la vía producen activación dispersa en tonos moderados. Esta distribución de

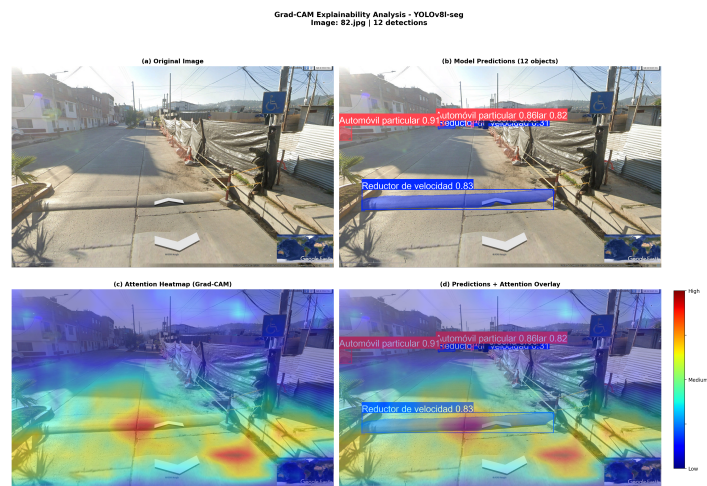
atención confirma la capacidad del modelo para procesar escenas con múltiples elementos distribuidos a distintas profundidades del campo visual.

3.13.9.2 Paneles de explicabilidad Grad-CAM

Para profundizar en el análisis de correspondencia entre atención y predicción, se generaron paneles cuádruples de explicabilidad Grad-CAM que integran cuatro vistas complementarias: (a) imagen original, (b) predicciones del modelo con *bounding boxes* y máscaras de segmentación, (c) mapa de atención Grad-CAM, y (d) superposición de predicciones sobre el mapa de atención (*overlay*). Las Figuras 62 y 63 presentan dos escenas representativas.

Figura 62

Panel de explicabilidad Grad-CAM — Intersección con contraluz solar (imagen 82, 12 detecciones)



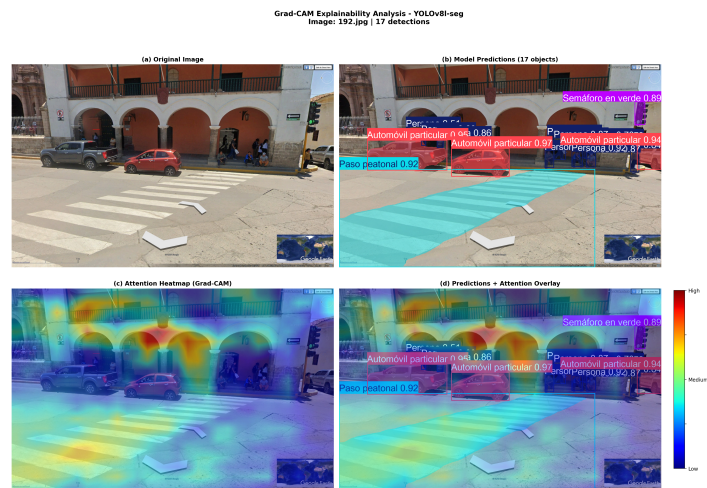
Nota. Elaboración propia. (a) Imagen original. (b) Predicciones: Automóvil particular (0.91, 0.86, 0.82), Reductor de velocidad (0.83, 0.31). (c) Heatmap Grad-CAM. (d) Superposición predicciones + atención.

La Figura 62 corresponde a una intersección urbana con condiciones de iluminación desafiantes debido al contraluz solar directo, lo cual representa un escenario exigente para cualquier modelo de visión por computadora. En el panel (b), el modelo detecta automóviles particulares con valores de confianza de 0.91, 0.86 y 0.82, así como un reductor de velocidad con confianza de 0.83 y otro con 0.31 (confianza baja). Las máscaras de segmentación azules cubren con precisión los reductores de velocidad sobre la calzada. El panel (c) revela que las zonas de máxima activación se concentran en la zona inferior de la imagen, donde se ubican los reductores, y en la zona lateral izquierda donde se encuentran los vehículos. Finalmente, el panel (d) confirma la correspondencia directa entre lo que el modelo observa (zonas rojas del heatmap) y lo que efectivamente predice (cajas delimitadoras y máscaras). Las áreas de alta atención coinciden espacialmente con las detecciones de mayor confianza, lo que valida que el modelo fundamenta sus predicciones en características visuales relevantes y no en artefactos o correlaciones espurias. Es particularmente destacable que el modelo mantenga detecciones con alta

confianza a pesar del contraluz solar, lo cual evidencia robustez ante condiciones de iluminación adversas.

Figura 63

Panel de explicabilidad Grad-CAM — Plaza con arquitectura colonial (imagen 192, 17 detecciones)



Nota. Elaboración propia. (a) Imagen original. (b) Predicciones: Automóvil particular (0.95, 0.97, 0.94, 0.86), Paso peatonal (0.92), Semáforo en verde (0.89), Personas (múltiples). (c) Heatmap Grad-CAM. (d) Superposición predicciones + atención.

La Figura 63 constituye una de las visualizaciones más relevantes para el presente estudio, dado que la escena captura el contexto arquitectónico colonial característico del centro histórico de Ayacucho. En el panel (a) se observa una plaza con arcos coloniales de piedra, un paso peatonal prominente, vehículos estacionados y un semáforo. El panel (b) muestra 17 detecciones simultáneas con confianzas elevadas: automóviles particulares (0.95, 0.97, 0.94, 0.86), paso peatonal (0.92), semáforo en verde (0.89) y múltiples personas. La máscara de segmentación de color cian cubre con precisión el paso peatonal completo. El alto número de detecciones y las confianzas elevadas demuestran que el modelo opera eficazmente en escenas complejas con múltiples clases simultáneas.

El panel (c) revela un hallazgo particularmente notable: las zonas de máxima activación se concentran bajo los arcos coloniales, exactamente donde se ubican los vehículos estacionados. Los arcos arquitectónicos generan activación alta, lo que sugiere que el modelo utiliza el contexto arquitectónico como referencia espacial para localizar vehículos dentro de la escena. Asimismo, se observa activación en la zona del semáforo y en el área donde se encuentran las personas. El panel (d) confirma que las zonas de alta activación bajo los arcos coinciden con las detecciones de mayor confianza (0.95 y 0.97), y que el paso peatonal presenta activación coherente con su detección. La capacidad del modelo para operar en un entorno arquitectónico colonial, sin confundir los elementos ornamentales con objetos de interés vial, evidencia una adecuada generalización del aprendizaje al contexto específico de Ayacucho.

3.13.9.3 Síntesis del análisis de explicabilidad

El análisis mediante Grad-CAM permite establecer las siguientes conclusiones sobre el comportamiento del modelo YOLOv8L-seg optimizado:

- i) **Coherencia espacial:** Las regiones de máxima activación coinciden consistentemente con la ubicación real de los elementos viales detectados, tanto para objetos en primer plano (peatones, reductores de velocidad) como para elementos en profundidad (vehículos, semáforos).
- ii) **Priorización de usuarios vulnerables:** El modelo asigna mayor activación a los peatones en proximidad inmediata, comportamiento deseable para un sistema de alertas preventivas orientado a la protección de usuarios vulnerables.
- iii) **Aprendizaje contextual:** El modelo aprendió a buscar semáforos en la parte alta de la escena y utiliza referencias arquitectónicas del entorno colonial ayacuchano como contexto espacial para localizar vehículos.
- iv) **Discriminación fondo-objeto:** Las zonas sin elementos viales de interés (fachadas, cielo, calzada vacía) presentan activación mínima (tonalidades azules), confirmando que el modelo no basa sus predicciones en regiones irrelevantes de la imagen.
- v) **Robustez ante condiciones adversas:** El modelo mantiene detecciones de alta confianza incluso bajo contraluz solar directo, lo que valida su capacidad operativa en condiciones de iluminación reales del entorno vial de Ayacucho.

Estos resultados aportan evidencia cualitativa que complementa las métricas cuantitativas reportadas en las secciones precedentes, fortaleciendo la validez del modelo para su aplicación en el sistema de detección de seguridad vial propuesto.

3.13.10. Desarrollo del software QHAWAY ADAS

Una vez completado el entrenamiento y la evaluación cuantitativa del modelo YOLOv8L-seg ($mAP_{50} = 81.4\%$, $mAP_{50-95} = 60.1\%$), se procedió al desarrollo del software de aplicación denominado **QHAWAY ADAS** (*Sistema Inteligente de Seguridad Vial*). El propósito de este software es integrar el modelo entrenado en un sistema funcional de asistencia avanzada al conductor (*Advanced Driver Assistance System, ADAS*) con capacidad de estimación monocular de distancias y emisión de alertas preventivas en tiempo real. El vocablo “QHAWAY” proviene del quechua ayacuchano (*runasimi*) y significa “mirar” u “observar con atención”, reflejando la función central del sistema: La observación inteligente y automatizada del entorno vial mediante visión por computadora.

3.13.10.1 Entorno de desarrollo y tecnologías empleadas

El software fue desarrollado íntegramente en el lenguaje de programación **Python 3.11**, seleccionado por su amplio ecosistema de bibliotecas de visión por computadora. El Cuadro 22 presenta el *stack* tecnológico completo del software.

Cuadro 22*Stack tecnológico del software QHAWAY ADAS*

Componente	Tecnología	Función en el sistema
Lenguaje de programación	Python 3.11	Desarrollo integral del software, gestión de módulos y lógica de negocio
Interfaz gráfica de usuario	PyQt5 5.15	Ventana principal con tema oscuro, área de video, controles interactivos, panel de estadísticas, registro de alertas y barra de progreso
Motor de detección y segmentación	Ultralytics (YOLOv8L-seg)	Inferencia del modelo entrenado a resolución 1024 px con renderizado nativo optimizado en C++ de cajas delimitadoras, máscaras de segmentación y etiquetas de confianza
Aceleración por GPU	PyTorch 2.x + CUDA	Inferencia acelerada en GPU NVIDIA mediante tensores CUDA, incluyendo <i>warm-up</i> previo de 3 iteraciones para estabilizar la latencia
Procesamiento de video	OpenCV 4.x	Lectura de cuadros, conversión de espacios de color, escritura del video procesado con <code>VideoWriter</code> y generación de capturas de pantalla
Seguimiento multi-objeto	ByteTrack	Asignación de identificadores numéricos persistentes (<i>track_id</i>) a cada instancia detectada entre cuadros consecutivos
Estimación de distancias	Modelo de cámara Pinhole	Cálculo de la distancia entre la cámara y cada objeto detectado, basado en la geometría proyectiva y alturas reales de referencia
Síntesis de voz neuronal	edge-tts (Azure Neural TTS)	Generación de archivos MP3 con voz neuronal peruana ("es-PE-AlexNeural") con parámetros prosódicos diferenciados según el nivel de urgencia
Respaldo de voz	pyttsx3 (SAPI5)	Generación de archivos WAV como alternativa <i>offline</i> cuando no hay conexión a internet
Reproducción de audio	Windows MCI (ctypes)	Reproducción de archivos MP3 y WAV en hilo independiente mediante la interfaz nativa <i>Media Control Interface</i> de Windows
Registro de datos	CSV	Exportación de detecciones con distancia estimada, alertas emitidas con <i>timestamp</i> , y resumen estadístico por sesión

Nota. Elaboración propia.

3.13.10.2 Arquitectura del sistema

El software se estructura en seis módulos funcionales que operan de manera concurrente mediante hilos de ejecución independientes, gestionados por el sistema de señales y *slots* de PyQt5:

- i) **Módulo de interfaz gráfica (MainWindow, WelcomeScreen, ScenarioSelector):** Gestiona la presentación visual del software mediante tres pantallas: La pantalla de bienvenida con carga del modelo, el selector de escenarios de validación, y el panel principal de procesamiento en tiempo real. Implementado con la clase QMainWindow de PyQt5 con tema oscuro profesional (fondo #0a0a1a) y acentos en color verde (#53d769).
- ii) **Módulo de procesamiento de video (VideoWorker):** Opera en un hilo secundario (QThread) para evitar el bloqueo de la interfaz gráfica durante la inferencia. Ejecuta el ciclo de lectura de cuadros, inferencia del modelo, filtrado temporal, análisis de alertas, renderizado de anotaciones y escritura del video procesado. Comunica los resultados al hilo principal mediante señales PyQt5 (`frame_ready`, `stats_ready`, `alert_fired`).
- iii) **Módulo de estimación de distancias (DistanceEstimator):** Implementa el modelo de cámara Pinhole con dos métodos complementarios de estimación (por altura de *bounding box* y por punto de apoyo).
- iv) **Módulo de análisis de alertas (AlertAnalyzer):** Evalúa cada detección válida mediante una combinación de la categoría del objeto, su posición lateral en el cuadro de imagen, la distancia estimada y el tiempo de colisión (*Time-To-Collision*, TTC). Genera candidatos de alerta ordenados por prioridad jerárquica.
- v) **Módulo de alertas por voz (VoiceAlertSystem):** Gestiona la emisión de alertas auditivas en un hilo dedicado, con un sistema de *cooldown* inteligente que opera a tres niveles (global, por prioridad y por objeto rastreado). Emplea una cola (Queue) para la comunicación asíncrona entre el hilo de procesamiento y el hilo de audio.
- vi) **Módulo de filtrado temporal (TemporalFilter):** Descarta detecciones transitorias que persisten menos de 2 cuadros consecutivos, vinculando las detecciones mediante los identificadores de ByteTrack. Las clases reductor de velocidad, paso peatonal y semáforo en rojo están exentas de este filtro.

3.13.10.3 Flujo operativo del software

El software sigue un flujo secuencial de cinco etapas diseñado para facilitar la validación sistemática del modelo por escenarios de riesgo vial. Las Figuras 64 y 65 presentan las dos primeras etapas del flujo.

- i) **Pantalla de bienvenida (Welcome Screen):** Al iniciar la aplicación, se presenta una pantalla con una imagen representativa del contexto vial ayacuchano como fondo, el título “QHAWAY ADAS — Sistema Inteligente de Seguridad Vial”, el nombre del desarrollador, el logotipo de la Universidad Nacional de San Cristóbal de Huamanga (UNSCH) y un indicador de estado que señala la carga del modelo en la GPU. Como se observa en la Figura 64, la imagen de fondo integra elementos representativos del sistema: Peatones con cajas delimitadoras, semáforos y la arquitectura urbana de Ayacucho. Una vez cargado el modelo y generados los archivos de audio para las alertas, se muestra la leyenda “✓ Modelo cargado correctamente” y se habilita el

botón “INGRESAR AL SISTEMA”.

Figura 64

Pantalla de bienvenida del software QHAWAY ADAS

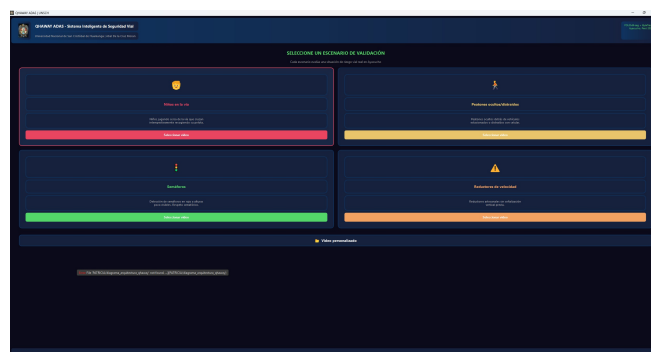


Nota. Elaboración propia. Se observa la interfaz del sistema

- ii) **Selección de escenario de validación:** El software presenta cuatro tarjetas interactivas correspondientes a los escenarios predefinidos de evaluación, cada una con un ícono representativo, el nombre del escenario, una descripción breve y un botón “Seleccionar video” con código de color diferenciado: rojo para niños en la vía, amarillo para peatones ocultos/distraídos, verde para semáforos y naranja para reductores de velocidad. Adicionalmente, se ofrece un botón “Video personalizado” para evaluaciones complementarias. La Figura 65 muestra esta pantalla. Al seleccionar un escenario, el explorador de archivos se abre directamente en la carpeta correspondiente.

Figura 65

Selector de escenarios de validación del software QHAWAY ADAS



Nota. Elaboración propia. Se observan las cuatro tarjetas de escenarios).

- iii) **Procesamiento en tiempo real:** Una vez seleccionado el video, el módulo VideoWorker inicia la inferencia en un hilo secundario. Cada cuadro se procesa mediante el siguiente *pipeline*: (a) lectura del cuadro con OpenCV; (b) inferencia del modelo YOLOv8L-seg a resolución 1024 px con seguimiento ByteTrack activado (`model.track()`); (c) actualización del filtro temporal; (d) análisis de alertas con

estimación de distancias; (e) renderizado nativo de anotaciones con la función `results[0].plot()` optimizada en C++; (f) superposición del *overlay* de distancia estimada en metros; (g) escritura del cuadro anotado en el video de salida; y (h) emisión de señales PyQt5 hacia el hilo principal.

- iv) **Emisión de alertas preventivas:** El módulo de alertas evalúa la posición relativa, la distancia estimada y el TTC de cada detección válida, emitiendo alertas auditivas con voz neuronal peruana cuando se cumplen las condiciones de riesgo definidas para cada categoría.
- v) **Generación del reporte de sesión:** Al finalizar el procesamiento, el software genera un cuadro de diálogo con el resumen de la sesión y exporta los datos en dos archivos CSV independientes: uno para detecciones y otro para alertas.

3.13.10.4 Estimación monocular de distancia mediante el modelo de cámara Pinhole

Uno de los componentes más relevantes del software QHAWAY ADAS es el módulo de estimación de distancias (*DistanceEstimator*), que permite calcular la separación aproximada entre la cámara vehicular y cada objeto detectado en la escena empleando exclusivamente una cámara monocular. Esta funcionalidad resulta fundamental para que el sistema de alertas pueda discriminar entre objetos lejanos (que no representan riesgo inmediato) y objetos próximos (que requieren la atención urgente del conductor), superando las limitaciones de los enfoques basados únicamente en la posición relativa del objeto dentro del cuadro de imagen.

La estimación se fundamenta en el **modelo de cámara Pinhole** (*pinhole camera model*), un modelo geométrico clásico de la visión por computadora que describe la relación proyectiva entre las dimensiones reales de un objeto tridimensional y su representación bidimensional en el plano de imagen de la cámara (Hartley & Zisserman, 2003). Bajo este modelo, un punto del mundo real con coordenadas (X, Y, Z) se proyecta en el plano de imagen en la posición (u, v) mediante la transformación:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (10)$$

donde f_x y f_y son las distancias focales en píxeles (horizontal y vertical), c_x y c_y son las coordenadas del punto principal (centro óptico), y s es un factor de escala. En la presente implementación se asume $f_x = f_y = f$ (sensor con píxeles cuadrados), simplificación habitual en aplicaciones de visión monocular vehicular (K.-Y. Park & Hwang, 2014).

A partir de la relación de triángulos semejantes que se deriva de la ecuación (10), la distancia D entre la cámara y un objeto se calcula mediante:

$$D = \frac{H_{\text{real}} \times f}{h_{\text{bbox}}} \quad (11)$$

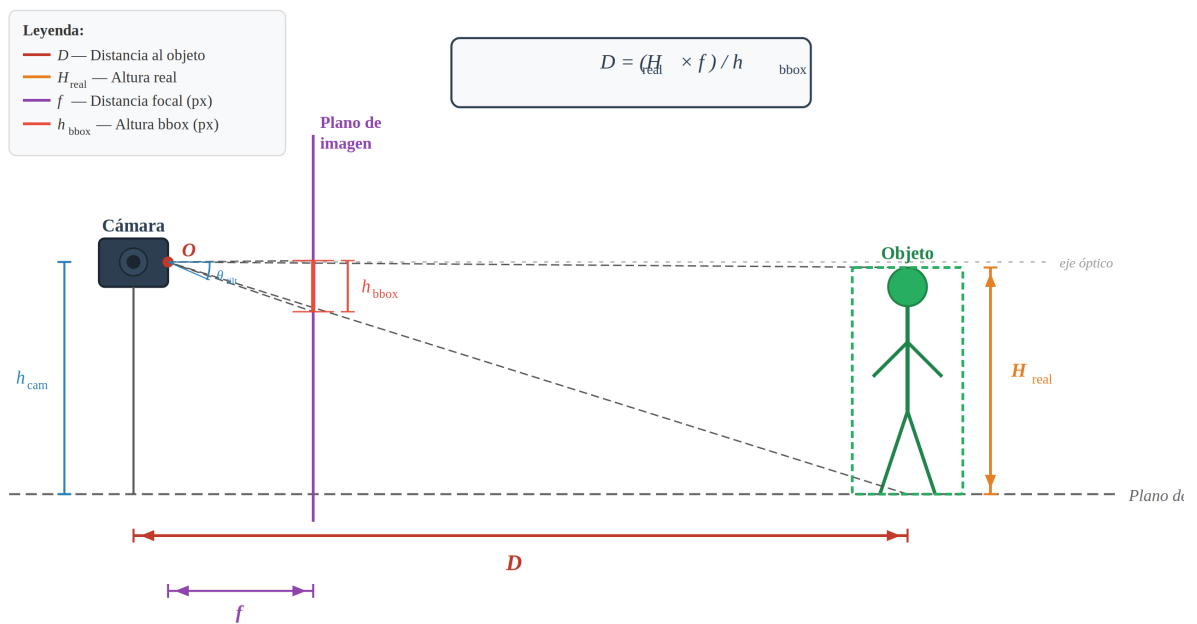
donde H_{real} es la altura real conocida del objeto en metros, f es la distancia focal de la

cámara expresada en píxeles, y h_{bbox} es la altura de la caja delimitadora (*bounding box*) del objeto en la imagen, también en píxeles. La relación inversa entre la distancia y el tamaño aparente del objeto en la imagen constituye el principio fundamental de la estimación: a mayor distancia del objeto respecto a la cámara, menor será su proyección en el plano de imagen, y viceversa.

La Figura 66 ilustra la geometría del modelo de cámara Pinhole aplicado al contexto del sistema QHAWAY ADAS.

Figura 66

Geometría del modelo de cámara Pinhole para estimación de distancia



Nota. Elaboración propia. D : distancia al objeto; H_{real} : altura real del objeto; f : distancia focal en píxeles; h_{bbox} : altura de la caja delimitadora en píxeles.

3.13.10.5 Parámetros de calibración de la cámara

La aplicación del modelo Pinhole requiere la definición de los parámetros intrínsecos de la cámara empleada en la captura de los videos de evaluación. Dado que la captura se realizó con un *smartphone* Android montado sobre el tablero de un vehículo, los parámetros se estimaron a partir de las especificaciones del sensor de la cámara trasera. El Cuadro 23 presenta los valores configurados en el software.

Cuadro 23*Parámetros de calibración de la cámara para el modelo Pinhole*

Parámetro	Símbolo	Valor	Justificación
Altura de montaje de la cámara	h_{cam}	1.50 m	Altura medida desde el suelo hasta la posición de la cámara en el tablero de una camioneta Toyota Hilux
Distancia focal	f	950 px	Calculada a partir de la resolución horizontal (1920 px, modo 1080p) y el campo de visión horizontal (FOV $\approx 75^\circ$): $f = \frac{1920}{2 \cdot \tan(37,5^\circ)} \approx 950 \text{ px}$
Ángulo de inclinación	θ_{tilt}	5°	Inclinación hacia abajo de la cámara respecto al eje horizontal, configurada para capturar la calzada inmediata

Nota. Elaboración propia. FOV: *Field of View* (campo de visión).

3.13.10.6 Alturas reales de referencia por categoría

La ecuación (11) requiere conocer la altura real H_{real} de cada tipo de objeto detectado. El Cuadro 29 presenta los valores de referencia configurados en el software para cada categoría del modelo, obtenidos a partir de normativas peruanas de señalización vial, especificaciones técnicas vehiculares y datos antropométricos oficiales.

Cuadro 24*Alturas reales de referencia (H_{real}) por categoría de elemento vial*

Clase	ID	H_{real} (m)	Referencia
Reductor de velocidad	0	0.10	Altura del lomo de toro según el Manual de Dispositivos de Control del Tránsito Automotor para Calles y Carreteras del MTC (Ministerio de Transportes y Comunicaciones del Perú, 2016)
Paso peatonal	1	0.30	Ancho de la franja de pintura como <i>proxy</i> geométrico
Persona	4	1.65	Estatura promedio de la población peruana adulta según datos del INEI
Motocicleta	5	1.10	Altura típica de motocicleta con conductor sentado
Automóvil particular	6	1.50	Altura media de vehículos sedán y hatchback del parque vehicular local
Camión	7	3.20	Altura de camión de carga mediano con carrocería
Bus de transporte	8	3.00	Altura de bus urbano tipo COASTER, predominante en Ayacucho
Semáforo en rojo	10	0.30	Diámetro de la lente del dispositivo semafórico
Mototaxi	11	1.60	Altura del mototaxi con toldo (<i>trimoto</i>), vehículo característico de Ayacucho

Nota. Elaboración propia a partir de MTC-2024 y datos del INEI. Las clases Línea recta (ID 2), Línea recta y derecha (ID 3) y Semáforo en verde (ID 9) no poseen altura de referencia asignada, dado que no requieren estimación de distancia para el sistema de alertas.

3.13.10.7 Método complementario: Estimación por punto de apoyo (*footpoint*)

Para objetos que se encuentran apoyados directamente sobre la calzada (particularmente peatones y reductores de velocidad), el software implementa un método complementario de estimación basado en la posición vertical del borde inferior de la caja delimitadora, denominado punto de apoyo (*footpoint*). Este método emplea la geometría de perspectiva considerando la altura de montaje de la cámara y su ángulo de inclinación:

$$D_{\text{foot}} = \frac{h_{\text{cam}}}{\tan(\alpha + \theta_{\text{tilt}})} \quad (12)$$

donde h_{cam} es la altura de montaje de la cámara (1.50 m), θ_{tilt} es el ángulo de inclinación (5°), y α es el ángulo desde el centro óptico de la cámara hasta el punto inferior de la caja delimitadora, calculado como:

$$\alpha = \arctan\left(\frac{y_{\text{bottom}} - c_y}{f}\right) \quad (13)$$

siendo $c_y = h_{\text{img}}/2$ el centro vertical de la imagen, y_{bottom} la coordenada vertical del borde inferior de la detección, y f la distancia focal. Los valores se restringen al rango $[0,5, 200]$ m para evitar resultados aberrantes.

Para la clase Persona (ID 4), el software promedia ambos métodos para incrementar la robustez:

$$D_{\text{persona}} = \frac{D_{\text{pinhole}} + D_{\text{foot}}}{2} \quad (14)$$

Este enfoque de doble estimación se fundamenta en las recomendaciones de K.-Y. Park y Hwang (2014) para sistemas monoculares de prevención de colisiones con peatones. Para la clase Reductor de velocidad (ID 0), se emplea únicamente el método *footpoint*, dado que la escasa altura del objeto (0.10 m) genera estimaciones imprecisas por el método Pinhole estándar.

3.13.10.8 Sistema de alertas preventivas basado en distancia y tiempo de colisión

El módulo de alertas preventivas (ALertAnalyzer) emplea las distancias estimadas por el modelo Pinhole para clasificar el nivel de riesgo de cada detección y emitir alertas auditivas contextualizadas. El diseño se fundamenta en dos marcos de referencia.

A. Zonas de riesgo basadas en distancia

Inspiradas en el protocolo del *New Car Assessment Program* (NCAP) de la NHTSA (National Highway Traffic Safety Administration, 2013), el software define tres zonas:

- Zona de peligro** ($D < 3$ m): Riesgo inminente de colisión. Alerta de máxima prioridad con voz urgente (+25 Hz en tono, +50 % en volumen). Corresponde a un tiempo de reacción de ~0.5 segundos a velocidades urbanas de 20 km/h.
- Zona de precaución** ($3 \leq D < 7$ m): Objeto en proximidad que requiere atención. Alerta de precaución con voz moderada. Margen de 1 a 2 segundos de anticipación.
- Zona segura** ($D \geq 7$ m): Solo anotación visual (caja delimitadora, máscara y distancia), sin alerta auditiva.

Los umbrales de 3 m y 7 m fueron calibrados para el contexto urbano de Ayacucho, donde la velocidad promedio de circulación en las avenidas Arenales y Cusco oscila entre 20 y 40 km/h, valores inferiores a los de vías rápidas donde los umbrales típicos del NCAP son de 10 m y 25 m (K.-Y. Park & Hwang, 2014).

B. Tiempo de colisión (*Time-To-Collision*, TTC)

El software calcula adicionalmente el TTC para cada objeto rastreado por ByteTrack:

$$\text{TTC} = \frac{D_n}{v_{\text{aprox}}} \quad \text{donde} \quad v_{\text{aprox}} = \frac{D_0 - D_n}{t_n - t_0} \quad (15)$$

donde D_n es la distancia actual, D_0 la distancia en la primera medición del historial, t_n y t_0 los instantes de tiempo correspondientes, y $v_{\text{aprox}} > 0,5$ m/s como umbral mínimo de acercamiento. El historial almacena hasta 10 mediciones por *track* (de que (maxlen=10)).

Cuando el TTC es inferior a 2.5 segundos y el objeto se encuentra en la zona de precaución, el sistema escala automáticamente la alerta al nivel de peligro, alineándose con el protocolo NCAP que establece $TTC \leq 2.4$ s para la activación del sistema FCW (National Highway Traffic Safety Administration, 2013).

3.13.10.9 Lógica de emisión de alertas por categoría

El sistema evalúa cada detección válida (confianza > 0.55) mediante tres condiciones: categoría del objeto, posición lateral (zona central: 20 %–80 % del ancho de la imagen) y distancia estimada. El Cuadro 25 presenta la lógica completa.

Cuadro 25

Lógica de emisión de alertas del software QHAWAY ADAS

Clase	Condición de activación	Prioridad	Mensaje de alerta emitido
Persona	Zona central + $D < 3$ m	1 (máx.)	“¡Cuidado! Peatón cruzando intempestivamente”
Persona	Zona central + $3 \leq D < 7$ m + $TTC < 2,5$ s	1 (máx.)	Escalación automática a nivel de peligro
Persona	Zona central + $3 \leq D < 7$ m	3	“Atención, peatón cruzando”
Persona	Zona lateral (vereda)	—	Sin alerta auditiva (solo visual)
Semáforo en rojo	conf. > 0,55 (sin restricción de zona)	2	“Semáforo en rojo, deténgase”
Reductor de vel.	Zona central + $D_{\text{foot}} < 30$ m	4	“Atención, reductor de velocidad, disminuya su velocidad”
Vehículos	—	—	Solo visualización
Señalizaciones	—	—	Solo visualización

Nota. Elaboración propia. conf. = valor de confianza. Prioridad: 1 = máxima. El umbral de confianza de 0.55 se determinó a partir de la curva F1-Confianza de segmentación (Figura 58).

3.13.10.10 Mecanismo de priorización y *cooldown* inteligente

Para evitar la saturación auditiva del conductor, el sistema implementa un mecanismo de *cooldown* en tres niveles simultáneos: (a) **prioridad jerárquica**, donde en cada cuadro se emite como máximo una alerta correspondiente al candidato de mayor prioridad (persona en peligro > semáforo rojo > persona en precaución > reductor); (b) **cooldown global de 5 segundos** entre alertas consecutivas de igual o menor prioridad, con excepción de alertas de prioridad superior que interrumpen el intervalo; y (c) **cooldown por objeto rastreado de 15 segundos**, donde cada *track_id* posee un historial independiente que impide la re-emisión de la misma alerta para el mismo objeto, a menos que el nivel de riesgo se haya incrementado.

3.13.10.11 Síntesis de voz con red neuronal

Las alertas auditivas se generan mediante el servicio **edge-tts** (*Microsoft Azure Neural Text-to-Speech*) empleando la voz “es-PE-AlexNeural” (voz masculina peruana). El Cuadro 26 presenta la configuración prosódica diferenciada por tipo de alerta.

Cuadro 26

Configuración prosódica de la síntesis de voz por tipo de alerta

Tipo de alerta	Velocidad	Tono (<i>pitch</i>)	Volumen
Peligro (peatón inminente)	+15 %	+25 Hz	+50 %
Precaución (peatón cruzando)	+15 %	+5 Hz	+0 %
Reductor de velocidad	+15 %	+5 Hz	+0 %
Semáforo en rojo	+15 %	+5 Hz	+0 %

Nota. Elaboración propia. Valores relativos respecto a la configuración base de la voz “es-PE-AlexNeural”.

Los archivos de audio se generan una única vez al iniciar el software y se almacenan en caché (`audio_cache/`). Un mecanismo de *hash* MD5 verifica si los archivos existentes corresponden a las configuraciones actuales, regenerándolos únicamente cuando se detectan cambios. Como respaldo *offline*, el software recurre automáticamente al motor **pyttsx3** (SAPI5 de Windows).

3.13.10.12 Seguimiento multi-objeto con ByteTrack

El módulo de seguimiento emplea el algoritmo **ByteTrack** (Y. Zhang, Sun, Jiang et al., 2022b), integrado nativamente en Ultralytics. Los parámetros personalizados son: `track_high_thresh = 0.5`, `track_low_thresh = 0.1`, `new_track_thresh = 0.4`, `track_buffer = 60` cuadros (mantiene la identidad del objeto durante hasta 2 segundos de oclusión a 30 FPS) y `match_thresh = 0.7`. El valor de `track_buffer = 60` resulta particularmente relevante para el contexto de estudio, permitiendo mantener la identidad de un peatón que desaparece temporalmente detrás de un vehículo estacionado.

3.13.10.13 Filtro temporal de detecciones

El filtro temporal descarta detecciones que persisten menos de 2 cuadros consecutivos (vinculados mediante `track_id`). Tres categorías están exentas: reductor de velocidad (ID 0), paso peatonal (ID 1) y semáforo en rojo (ID 10), dado que su aparición transitoria puede corresponder a un evento real que requiere atención inmediata.

3.13.10.14 Validación del software mediante escenarios de riesgo vial

El software se diseñó con un enfoque de validación por escenarios, donde cada escenario representa una situación de riesgo vial documentada en las avenidas del estudio. El Cuadro 27 describe los cuatro escenarios implementados. A continuación se presentan los resultados obtenidos en la validación de cada escenario, documentados mediante capturas representativas de la interfaz del software durante el procesamiento en tiempo real.

Cuadro 27*Escenarios de validación del software QHAWAY ADAS*

N.º	Escenario	Clases críticas	Descripción y objetivo de validación
1	Niños en la vía	Persona	Detectar menores que ingresan intempestivamente a la calzada para recoger objetos de juego, evaluando la alerta de peligro cuando $D < 3$ m
2	Peatones ocultos o distraídos	Persona, Vehículos	Detectar peatones ocultos por vehículos estacionados o distraídos con celular, evaluando el <i>track_buffer</i> de ByteTrack durante la oclusión
3	Semáforos	Semáforo rojo/verde	Identificar el estado semafórico y emitir la alerta "Semáforo en rojo, deténgase" sin restricción de zona lateral
4	Reductores de velocidad	Reductor de vel.	Detectar reductores deteriorados y artesanales sin señalización vertical, evaluando la distancia de anticipación

Nota. Elaboración propia. Cada escenario dispone de una carpeta independiente (videos/escenario_01_ninos/, etc.).

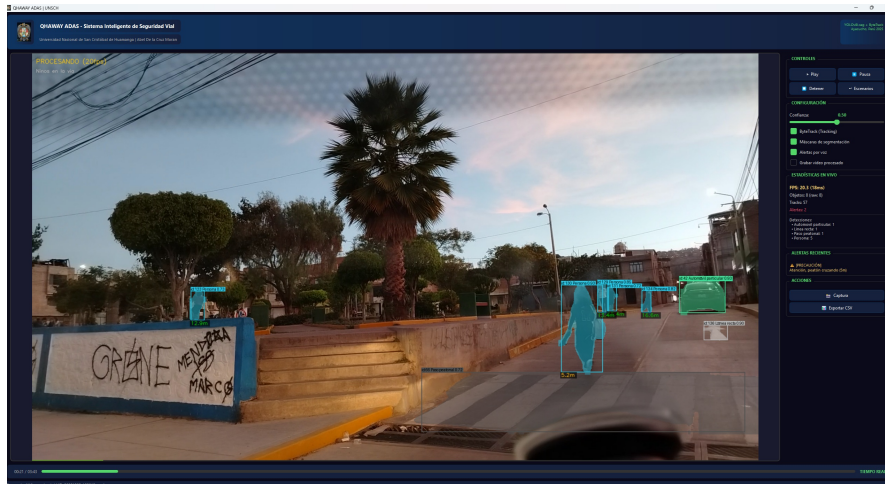
A. Escenario 1: Niños en la vía

El primer escenario de validación evaluó la capacidad del sistema para detectar y emitir alertas oportunas ante menores de edad que ingresan intempestivamente a la calzada. La prueba se realizó en un entorno cercano a un parque con alta afluencia de niños jugando fútbol, quienes frecuentemente corren a recoger su pelota sin verificar el tráfico vehicular.

La Figura 67 presenta el comportamiento del sistema en condiciones de iluminación nocturna, demostrando la capacidad del modelo YOLOv8L-seg para detectar correctamente múltiples peatones incluso con baja luminosidad ambiental. El panel lateral derecho muestra las estadísticas en vivo: 20.3 FPS de procesamiento, 18 ms de tiempo de inferencia, 8 objetos detectados (5 personas, 1 automóvil, 1 línea recta, 1 paso peatonal) y 57 *tracks* únicos acumulados. Se observa que el sistema identificó a los peatones con distancias estimadas de 12.9 m y 5.2 m, emitiendo una alerta de nivel PRECAUCIÓN ("Atención, peatón cruzando") para el grupo de personas ubicado en la zona de precaución (3–7 m).

Figura 67

Escenario 1: Detección de peatones en condiciones de iluminación nocturna



Nota. Elaboración propia. Captura del software QHAWAY ADAS procesando un video nocturno del escenario 1. Se observan las máscaras de segmentación en color cian sobre los peatones detectados, las distancias estimadas (12.9 m y 5.2 m) con código de color (naranja = precaución, verde = seguro), las etiquetas con *track_id*, clase y confianza, y la alerta de PRECAUCIÓN en el panel de “Alertas recientes”.

La Figura 68 muestra una situación posterior en la misma sesión, donde el sistema detecta simultáneamente a cuatro personas y tres automóviles, estimando distancias individuales para cada peatón. Se observan distancias de 11.6 m, 12.5 m, 3.9 m y 4.1 m, con los valores de 3.9 m y 4.1 m mostrados en color naranja, indicando que esos peatones se encuentran dentro de la zona de precaución. El contador de alertas acumuladas asciende a 17, evidenciando la actividad continua del sistema de alertas.

Figura 68

Escenario 1: Estimación simultánea de distancias a múltiples peatones



Nota. Elaboración propia. Se observan cuatro personas detectadas con distancias individuales (11.6 m, 12.5 m, 3.9 m y 4.1 m) y tres automóviles con sus respectivas máscaras de segmentación verdes. El panel derecho muestra 23.5 FPS, 17 ms de inferencia, 282 *tracks* acumulados y la alerta de PRECAUCIÓN activa.

La Figura 69 presenta el momento crítico de la validación: dos niños cruzaron la calzada intempestivamente para recoger su pelota. El sistema activó la alerta de máxima prioridad — PELIGRO — con el mensaje “¡Cuidado! Peatón cruzando intempestivamente”, visible tanto en el banner rojo superior del video como en el panel de “Alertas recientes” del lado derecho. Las distancias estimadas a los peatones son de 5.6 m y 3.8 m, con este último valor cercano al umbral de peligro de 3 m. El acumulado de alertas asciende a 38, reflejando la intensidad de la situación de riesgo.

Figura 69

Escenario 1: Alerta de PELIGRO por niños cruzando intempestivamente



Nota. Elaboración propia. Momento crítico del escenario 1: Dos niños cruzan la calzada para recoger su pelota. El sistema emite la alerta de máxima prioridad (PELIGRO) con el mensaje “¡Cuidado! Peatón cruzando intempestivamente”. Las distancias estimadas (5.6 m y 3.8 m) se muestran en naranja (zona de precaución). El panel derecho muestra 38 alertas acumuladas, 487 tracks y 22 FPS.

B. Escenario 2: Peatones ocultos o distraídos

El segundo escenario de validación evaluó la detección de peatones parcialmente ocultos por vehículos estacionados y peatones distraídos con dispositivos móviles, situaciones documentadas frecuentemente en la Av. Cusco. Las pruebas se realizaron recorriendo esta avenida en condiciones de tráfico real.

La Figura 70 muestra la capacidad de detección simultánea del sistema en una escena compleja con múltiples categorías: automóviles particulares (máscaras verdes), camiones (máscaras azules), motocicleta (máscara amarilla) y personas (máscaras cyan). Se observa un peatón detectado a 6.5 m de distancia, con la alerta de PRECAUCIÓN activa. El sistema procesa a tiempo real (indicador “TIEMPO REAL” en la esquina superior izquierda), alcanzando 25.4 FPS con 17 ms de inferencia.

Figura 70

Escenario 2: Detección simultánea de múltiples categorías en Av. Cusco



Nota. Elaboración propia. Captura del escenario 2 con detección simultánea de automóviles, camiones, motocicleta, mototaxi y personas en la Av. Cusco. La distancia del peatón (6.5 m, naranja) activa la alerta de PRECAUCIÓN. El panel muestra 155 *tracks* y 11 alertas acumuladas.

Las Figuras 71 y 72 documentan una secuencia crítica donde un peatón se encontraba oculto detrás de un camión estacionado en la vía. En la Figura 71, se observa al peatón detectado a 5.7 m de distancia, parcialmente ocluido por un camión de gran envergadura que ocupa la mayor parte del campo visual. El sistema logró detectar al peatón gracias a la porción visible de su cuerpo, emitiendo una alerta de PRECAUCIÓN. Este resultado demuestra la ventaja del *track_buffer* de ByteTrack configurado en 60 cuadros, que permite mantener la identidad del peatón incluso durante oclusiones temporales.

Figura 71

Escenario 2: Peatón detectado detrás de un camión estacionado



Nota. Elaboración propia. El sistema detecta a un peatón (5.7 m, naranja) parcialmente ocluido por un camión de carga que transita por la Av. Cusco. Se observa la máscara de segmentación cyan del peatón y la detección simultánea de automóviles. Indicador: TIEMPO REAL, 24.1 FPS, 354 *tracks*.

Figura 72

Escenario 2: Peatón cruzando detrás de un camión



Nota. Elaboración propia. Momento en que el peatón emerge detrás del camión y cruza la calzada.

El sistema lo detecta a 7.0 m con la máscara de segmentación cyan y emite la alerta de PRECAUCIÓN. Se observa el camión con máscara azul ocupando gran parte del campo visual, situación característica de las avenidas ayacuchanas donde los vehículos de carga limitan la visibilidad del conductor.

La Figura 73 documenta otra situación de riesgo frecuente: un peatón cruzando la calzada distraído con su teléfono celular, sin prestar atención al tráfico vehicular. El sistema detectó al peatón a 4.1 m de distancia y emitió la alerta de PRECAUCIÓN correspondiente. Esta captura evidencia la utilidad del sistema QHAWAY ADAS para prevenir accidentes en situaciones donde tanto el conductor como el peatón pueden tener una percepción limitada del riesgo.

Figura 73

Escenario 2: Peatón cruzando distraído con celular



Nota. Elaboración propia. El sistema detecta a un peatón cruzando la calzada distraído con su celular (4.1 m, naranja). Se observan simultáneamente una mototaxi y un automóvil particular. El panel muestra la alerta de PRECAUCIÓN (“Atención, peatón cruzando”), 1447 tracks acumulados y 89 alertas emitidas durante la sesión.

C. Escenario 3: Semáforos

El tercer escenario evaluó la identificación del estado semafórico en las intersecciones de las avenidas del estudio. La prueba se realizó en condiciones de iluminación nocturna, que representan un desafío adicional debido a la iluminación artificial y los reflejos de las luces urbanas.

La Figura 74 muestra la detección de un semáforo en rojo con una confianza de 0.84 en una intersección nocturna con alta densidad de peatones y vehículos. El sistema identifica simultáneamente un bus de transporte (máscara roja), una motocicleta, siete personas (máscaras cyan) y el semáforo en rojo. En este cuadro, la alerta emitida corresponde a PELIGRO por la presencia de un peatón cruzando intempestivamente, dado que la prioridad jerárquica de la persona en peligro (prioridad 1) supera a la del semáforo (prioridad 2).

Figura 74

Escenario 3: Detección de semáforo en rojo en condiciones nocturnas



Nota. Elaboración propia. Captura del software QHAWAY ADAS durante la validación nocturna del escenario 3. Se observa la detección del semáforo en rojo (*track_id* 17312, confianza 0.84) con la máscara de segmentación roja superpuesta. El panel muestra 10 objetos detectados (bus, motocicleta, 7 personas, 1 semáforo rojo), 331 *tracks* y 32 alertas acumuladas. FPS: 19.2 (20 ms de inferencia).

La Figura 75 presenta la emisión de la alerta específica del semáforo “Semáforo en rojo, deténgase (1 m)”, visible en el panel de “Alertas recientes” como alerta de nivel ALERTA. El sistema mantiene la detección consistente del semáforo (mismo *track_id* 17312, confianza 0.84) mientras continúa la detección simultánea de peatones, automóviles y motocicletas en la escena nocturna. Se observa que las personas detectadas en la zona central presentan distancias de 5.1 m y 5.7 m con *overlay* en naranja (zona de precaución), demostrando que el sistema opera con múltiples niveles de alerta de manera concurrente.

Figura 75

Escenario 3: Emisión de la alerta “Semáforo en rojo, deténgase”



Nota. Elaboración propia. El panel de “Alertas recientes” muestra la alerta de nivel ALERTA con el mensaje “Semáforo en rojo, deténgase (1 m)”

D. Escenario 4: Reductores de velocidad

El cuarto escenario evaluó la detección anticipada de reductores de velocidad (rompemuelles) en las vías del estudio, incluyendo reductores deteriorados y artesanales sin señalización vertical complementaria. Las pruebas se realizaron recorriendo la Av. Cusco, donde estos dispositivos son frecuentes. La Figura 76 presenta la detección de un reductor de velocidad en la Av. Cusco con el banner de alerta naranja “Atención, reductor de velocidad, disminuya su velocidad (2 m)”. Se observa que el reductor fue detectado a tan solo 2 metros de distancia, con la máscara de segmentación amarilla superpuesta sobre la señalización horizontal.

Figura 76

Escenario 4: Detección de reductor de velocidad con banner de alerta



Nota. Elaboración propia. Captura del software QHAWAY ADAS durante la validación del escenario 4 en la Av. Cusco. Se observa el reductor de velocidad detectado (máscara amarilla, distancia 2 m) con el banner de alerta naranja .

La Figura 77 muestra otro reductor de velocidad detectado a 3.6 m de distancia, junto con dos automóviles y una persona. El panel lateral registra la alerta “Atención, reductor de velocidad, disminuya su velocidad (4 m)” como alerta reciente, y el acumulado de 9 alertas indica la presencia de múltiples reductores a lo largo del recorrido de prueba.

Figura 77

Escenario 4: Detección de reductor de velocidad a 3.6 metros



Nota. Elaboración propia. Reductor de velocidad detectado a 3.6 m (máscara amarilla) con la distancia estimada en color naranja (zona de precaución). Se observan simultáneamente dos automóviles (máscaras verdes) y una persona (máscara cyan, con confianza 0.87).

La Figura 78 presenta una escena compleja del escenario 4 con detecciones simultáneas de un reductor de velocidad (con distancia estimada y alerta emitida), automóviles, un camión, una motocicleta, una mototaxi y múltiples personas.

Figura 78

Escenario 4: Detección múltiple con reductor de velocidad y tráfico denso



Nota. Elaboración propia. Escena compleja con detecciones simultáneas de reductor de velocidad (máscara amarilla, alerta emitida), automóviles, camión, motocicleta, mototaxi y personas. El panel registra 9 objetos detectados, 801 tracks, 51 alertas acumuladas y la alerta reciente de reductor de velocidad. FPS: 21.5 (18 ms).

4. RESULTADOS

4.1. Análisis y discusión de resultados

El presente capítulo analiza e interpreta los resultados obtenidos en el desarrollo del sistema QHAWAY ADAS, comparándolos con investigaciones previas de la literatura científica internacional en visión computacional aplicada a sistemas de asistencia al conductor, y discutiendo el aporte específico del trabajo al entorno vial andino informal de Ayacucho.

4.1.1. Desempeño del modelo en detección y segmentación de instancias

El modelo YOLOv8L-seg entrenado en la Fase 3 alcanzó una precisión media de detección mAP_{50} Box de 0.810 y una precisión media de segmentación de instancias mAP_{50} Mask de 0.778 sobre las 12 clases de elementos críticos de la vía pública del entorno ayacuchano. Estos resultados son consistentes con los reportados en estudios recientes de detección y segmentación en entornos viales urbanos mediante arquitecturas de la familia YOLO.

(Ayachi, Said, Afif, Alshammari, Hleili & Abdelali, 2025) evaluaron modelos YOLO en entornos urbanos reales para sistemas ADAS, reportando valores de mAP_{50} entre 0.74 y 0.83 para detección de vehículos y peatones en condiciones variables de iluminación, rangos dentro de los cuales se ubica el desempeño obtenido por QHAWAY. De forma análoga, (S. A. Shah et al., 2025) documentan que las implementaciones de YOLOv8-seg en hardware de inferencia en tiempo real alcanzan valores de mAP_{50} entre 0.75 y 0.88 para tareas de detección de usuarios vulnerables, confirmando que los resultados del presente estudio son competitivos respecto al estado del arte.

La estrategia de entrenamiento progresivo en tres fases con resolución creciente produjo una mejora acumulada de 30.7 puntos porcentuales en mAP_{50} Box y 32.5 puntos en mAP_{50} Mask, con la Fase 3 (1024 px) contribuyendo la mayor ganancia en segmentación de instancias (+14.5 pp en mAP_{50} Mask). Este resultado es coherente con lo reportado por (Jocher et al., 2023b), quienes recomiendan el incremento progresivo de resolución para mejorar la detección de objetos de pequeña escala, aspecto particularmente relevante en el presente estudio dado que clases como semáforos y señalizaciones horizontales ocupan una fracción reducida del área de la imagen.

4.1.2. Comparación con investigaciones previas en detección de elementos viales

4.1.2.1 Detección de reductores de velocidad

La clase reductor de velocidad alcanzó el mayor desempeño del modelo (mAP_{50} Box = 0.923, mAP_{50} Mask = 0.897), resultado especialmente relevante dado que los dispositivos catalogados en el área de estudio presentan geometrías heterogéneas que se apartan de los perfiles estándar presentes en los conjuntos de datos internacionales.

(R. Wang et al., 2024b) propusieron FPNNet para la detección visual de reductores de velocidad en vehículos conectados, reportando una precisión de detección del 91.3 % sobre conjuntos de datos con dispositivos normalizados. (Peralta-López, Morales-Viscaya,

Lazaro-Mata et al., 2023) combinaron una cámara estéreo ZED con redes neuronales para la detección simultánea de reductores y baches, obteniendo resultados superiores al 88 % en dispositivos regulares. Sin embargo, ninguno de estos estudios abordó dispositivos de construcción artesanal ni cuantificó la no conformidad geométrica respecto a normativas viales. El presente estudio constituye la primera documentación cuantitativa del incumplimiento de la Directiva MTC N°01-2011-MTC/14 en el entorno andino: el 96 % de los reductores auditados incumple al menos un criterio dimensional, lo que representa un riesgo vial latente tanto para conductores como para sistemas ADAS entrenados exclusivamente con dispositivos regulares.

4.1.2.2 Detección de peatones y usuarios vulnerables

La clase persona alcanzó un mAP_{50} Box de 0.866 y un mAP_{50} Mask de 0.826, con 4,608 instancias de validación. (Z. Pan et al., 2024) reportaron que las máscaras de segmentación de instancia mejoran la estimación de proximidad a peatones en comparación con métodos basados en rectángulos delimitadores en escenas urbanas densas, resultado que respalda la elección de la segmentación de instancias como tarea principal del sistema QHAWAY frente a enfoques de detección por caja.

(Han et al., 2025) desarrollaron un sistema de detección peatonal con fusión de múltiples sensores en vías urbanas, obteniendo precisiones superiores al 90 % pero requiriendo infraestructura de sensores adicional no disponible en el contexto ayacuchano. El enfoque monocular de QHAWAY representa una alternativa viable para entornos de recursos limitados donde la instalación de sensores especializados no es factible económicamente.

4.1.2.3 Detección del mototaxi como clase local

La clase mototaxi obtuvo un mAP_{50} Box de 0.769 y un mAP_{50} Mask de 0.738, constituyendo el primer resultado publicado de segmentación de instancias para este tipo de vehículo en el contexto vial andino peruano. Con solo 357 instancias locales de entrenamiento, el modelo logró un desempeño aceptable gracias a la morfología distintiva del mototaxi — estructura de tres ruedas con toldo — que lo diferencia claramente de los vehículos estándar presentes en los conjuntos de datos internacionales.

(Z. Du et al., 2026) abordaron la detección de flotas vehiculares heterogéneas mediante el modelo ML-E-YOLO, reportando desempeños entre 0.72 y 0.81 en mAP_{50} para categorías locales con representación limitada en conjuntos de datos estándar. Los resultados de QHAWAY para la clase mototaxi son consistentes con este rango, confirmando que las arquitecturas YOLO pueden aprender clases locales exclusivas con volúmenes moderados de datos siempre que la variabilidad intraclase sea suficientemente representada. (Linardi et al., 2025) demostraron que modelos entrenados exclusivamente con datos internacionales pueden perder hasta 68 puntos porcentuales de precisión al aplicarse a entornos locales no representados, lo que justifica la estrategia de construcción de un conjunto de datos híbrido con captura local sistemática adoptada en el presente estudio.

4.1.2.4 Detección de señalizaciones horizontales

Las clases de señalización horizontal presentaron el mayor contraste de desempeño: la línea recta y derecha alcanzó mAP_{50} Box de 0.898, mientras que la línea recta obtuvo solo 0.610. Esta diferencia se atribuye al menor número de instancias de validación de la línea recta (109 instancias) y a la mayor variabilidad visual de esta clase en el entorno ayacuchano, donde el deterioro del pavimento genera condiciones de contraste desfavorables.

(Kulawik & Kuczyński, 2025a) desarrollaron un sistema de detección y clasificación de marcas horizontales específicamente diseñado para sistemas de asistencia al conductor, reportando precisiones entre 0.78 y 0.92 para marcas en buen estado de conservación pero con caídas significativas para marcas deterioradas. Este patrón es consistente con los resultados obtenidos en el presente estudio, donde las condiciones de deterioro de la pintura vial en las avenidas de Ayacucho constituyen el principal factor de variabilidad intraclase.

4.1.2.5 Aporte del sistema QHAWAY al entorno vial andino informal

El contexto del presente estudio presenta características que lo diferencian cualitativamente de la mayoría de los trabajos publicados en sistemas ADAS para entornos urbanos:

- **Entorno patrimonial y altitud.** La ciudad de Ayacucho opera a 2,761 m sobre el nivel del mar con calles coloniales de 4 a 6 m de ancho en el centro histórico, declarado por la UNESCO como Ciudad Creativa de Artesanía y Arte Popular desde 2019. Las condiciones de radiación ultravioleta elevada, gradientes topográficas de hasta el 12 % y arquitectura colonial generan un entorno vial sin equivalente en los conjuntos de datos internacionales disponibles. El análisis de explicabilidad visual mediante los algoritmos Grad-CAM confirmó que el modelo no confunde los elementos arquitectónicos coloniales con objetos viales de interés, lo que evidencia una adecuada separación entre señal y fondo en este entorno atípico.
- **Infraestructura vial informal.** El 96 % de los reductores de velocidad auditados en el área de estudio incumple la Directiva MTC N°01-2011-MTC/14, constituyendo el primer registro cuantitativo de este fenómeno en la región andina. A diferencia de los estudios internacionales de detección de reductores que asumen dispositivos normalizados (Peralta-López, Morales-Viscaya, Lazaro-Mata et al., 2023; R. Wang et al., 2024b; Zhumadillayeva et al., 2025), QHAWAY fue entrenado específicamente con dispositivos de construcción artesanal, logrando el mayor desempeño de todas las clases (mAP_{50} Box = 0.923) pese a la alta variabilidad geométrica de estos elementos.
- **Sistemas de transporte informal.** El mototaxi está ausente de todos los repositorios internacionales de visión vial. Su presencia como clase independiente en QHAWAY, con un desempeño documentado de mAP_{50} Box = 0.769, constituye una contribución específica para el desarrollo de sistemas ADAS adaptados a las condiciones del transporte urbano andino peruano, donde este vehículo es el modo dominante de

transporte informal en ciudades intermedias.

4.1.3. Análisis de la estimación de distancias

El estimador de distancias monoculares basado en el modelo geométrico de cámara mostró un comportamiento conservador — es decir, sobreestimador — en todos los casos evaluados, con errores residuales entre 0.20 m y 3.10 m. Este patrón es operacionalmente deseable para un sistema de seguridad vial, ya que implica que el sistema nunca subestima la distancia a un elemento de riesgo, generando alertas de forma anticipada más que tardía.

(Dai et al., 2021) validaron el modelo geométrico de cámara monocular para la estimación de distancias a peatones en entornos urbanos, reportando errores medios inferiores a 2.5 m para distancias de hasta 30 m, valor compatible con los errores máximos observados en el presente estudio. (A. Chaman et al., 2026) evaluaron sistemas ADAS monoculares basados en YOLO para contextos de recursos limitados, confirmando que el modelo geométrico de cámara proporciona estimaciones de distancia suficientes para la clasificación de zonas de riesgo en conducción urbana, incluso sin calibración con instrumentos de metrología especializados.

El caso de mayor error observado corresponde a un niño en posición agachada (error de 3.10 m), explicable por la violación de la hipótesis de altura de referencia: la altura aparente del niño en posición agachada (0.70 m) es significativamente inferior a la altura de referencia de la clase persona (1.65 m). Aplicando la corrección geométrica correspondiente, la estimación corregida resulta de 2.37 m, consistente con la distancia de referencia independiente de 2.5 m (error residual corregido: 0.13 m). Este análisis confirma que los errores observados son geoméricamente predecibles y explicables, no constituyendo fallos algorítmicos del sistema.

4.1.4. Análisis de la explicabilidad visual del modelo

El análisis de explicabilidad visual mediante tres algoritmos complementarios — Grad-CAM, Grad-CAM++ y EigenCAM — confirmó que las representaciones aprendidas por el modelo son semánticamente coherentes con los elementos críticos de la vía pública. Las zonas de máxima activación coincidieron consistentemente con la ubicación real de los objetos detectados, mientras que las regiones de fondo — fachadas de edificios, cielo, pavimento vacío — presentaron activaciones mínimas.

(Selvaraju, Cogswell, Das et al., 2017b) introdujeron el algoritmo Grad-CAM como herramienta de validación semántica de modelos de visión computacional, argumentando que la coincidencia entre las zonas de alta activación y los objetos de interés constituye evidencia de que el modelo aprende características visuales relevantes en lugar de correlaciones espurias del conjunto de datos. La extensión a Grad-CAM++ (Chattopadhyay et al., 2018) mejora la localización en escenas con múltiples instancias de la misma clase, aspecto relevante en el contexto del presente estudio donde las escenas del centro histórico de Ayacucho presentan hasta 15 personas simultáneas.

Un hallazgo de particular relevancia es que en las escenas del centro histórico —

con arquitectura colonial de arcos de piedra y fachadas ornamentadas — el modelo concentra la activación en los vehículos estacionados bajo los arcos y no en los elementos arquitectónicos, confirmando que el entorno patrimonial de Ayacucho no introduce correlaciones espurias en las decisiones del modelo. Este resultado respalda la generalización del sistema al contexto geográfico y cultural específico del área de estudio.

4.1.5. Comparación con el modelo base de referencia

El modelo QHAWAY desarrollado en el presente estudio supera al modelo base de referencia anterior en todas las métricas evaluadas: +14.0 puntos porcentuales en precisión global, +16.3 pp en mAP₅₀ Box, +12.2 pp en mAP₅₀₋₉₅ Box y +24.9 pp en mAP₅₀ Mask. La mejora más pronunciada en segmentación de instancias (+24.9 pp) se atribuye a tres factores concurrentes: la calidad superior de las anotaciones generadas con LabelMe y SAM2 respecto a las anotaciones del conjunto de referencia, el incremento de resolución a 1024 px que mejora la delineación de contornos finos, y la reducción del número de clases de 18 a 12 con la incorporación de la clase mototaxi específica del contexto ayacuchano en reemplazo de categorías de menor relevancia local. Estos resultados confirman que la estrategia de desarrollo de un conjunto de datos local representativo, combinada con una estrategia de entrenamiento progresivo, produce mejoras sustanciales respecto a los enfoques que dependen exclusivamente de datos y modelos internacionales.

4.1.6. Posicionamiento de QHAWAY frente a sistemas ADAS comerciales

Los sistemas ADAS comerciales de mayor difusión — incluyendo Toyota Safety Sense, Honda Sensing y los sistemas de Tesla — han sido desarrollados y entrenados predominantemente con datos de entornos viales de países de altos ingresos, con infraestructura normalizada, flotas vehiculares homogéneas y señalización horizontal en buen estado de conservación. Ninguno de estos sistemas contempla el mototaxi como clase de detección independiente, los reductores de velocidad artesanales como elemento de alerta, ni las condiciones específicas de iluminación y topografía del entorno andino.

El sistema QHAWAY no compite con estos sistemas en términos de cobertura o infraestructura tecnológica, sino que ocupa un nicho específico y no atendido: el desarrollo de un sistema de asistencia al conductor diseñado y validado para las condiciones particulares del entorno vial urbano andino informal de Perú. Este posicionamiento es coherente con la tendencia documentada en la literatura de adaptar los sistemas de visión vial a los contextos locales específicos, dado que los modelos entrenados exclusivamente con datos internacionales pueden perder hasta 68 puntos porcentuales de precisión al aplicarse a entornos locales no representados (Linardi et al., 2025), lo que invalida el uso directo de sistemas comerciales sin adaptación en el contexto peruano.

4.2. Contrastación de hipótesis

4.2.1. Contrastación de hipótesis específica 01

- La estrategia de entrenamiento progresivo en tres fases con resolución creciente produce una mejora acumulada superior a 30 puntos porcentuales en la precisión media de detección y segmentación, logrando el modelo una precisión media mAP_{50} Box igual a 0.810 y mAP_{50} Mask igual a 0.778, con la clase reductor de velocidad como la de mayor desempeño ($mAP_{50} = 0.923$) y el mototaxi alcanzando $mAP_{50} = 0.769$ como clase representativa del entorno vial andino local.

Para contrastar esta hipótesis, el modelo YOLOv8L-seg entrenado en la Fase 3 de optimización (resolución 1024×1024 px, época óptima 173) fue sometido a inferencia sobre un conjunto diverso de 12 imágenes provenientes de tres fuentes distintas, no contempladas durante el entrenamiento:

- **Fuente A — GPS Map Camera (5 imágenes):** fotografías capturadas en campo con registro automático de coordenadas geográficas, fecha y hora.
- **Fuente B — Google Earth Pro, modo Street View (3 imágenes):** capturas virtuales de intersecciones viales estratégicas de Ayacucho, con resolución, compresión y perspectiva diferentes a las fotografías de campo.
- **Fuente C — Fotografías propias (4 imágenes):** imágenes capturadas durante los recorridos de validación por el centro histórico y las principales avenidas de la ciudad.

Esta diversidad de fuentes garantiza la evaluación del modelo bajo variabilidad real de condiciones de iluminación, resolución, perspectiva y densidad de tráfico, cubriendo los escenarios más representativos del entorno vial andino informal de Ayacucho.

4.2.1.1 Detecciones en imágenes de campo (GPS Map Camera)

Las Figuras 79 a 83 corresponden a fotografías capturadas in situ durante el recorrido de validación por las principales vías de la ciudad de Ayacucho, empleando la aplicación GPS Map Camera.

Figura 79

Detección en Av. Naciones Unidas 110: Reductor de velocidad, persona y automóviles



Nota. Elaboración propia. El modelo identifica un reductor de velocidad (86.2 %), una persona (95.1 %) y dos automóviles particulares (80.5 % y 1.5 %).

Figura 80

Detección multi-clase en intersección: Paso peatonal, mototaxi, motocicleta y persona



Nota. Elaboración propia. El modelo detecta simultáneamente: paso peatonal (89.2 % y 64.6 %), línea recta (91.3 %), persona (95.1 %), mototaxi (95.6 %) y motocicleta (75.3 %). Esta imagen evidencia la presencia simultánea de 5 clases distintas, incluyendo el mototaxi como clase exclusiva del entorno vial andino local.

Figura 81

Detección en Av. Cusco (Jr. Venezuela 140): Paso peatonal y peatones cruzando



Nota. Elaboración propia. Se detecta un paso peatonal (91.9 %) y dos personas transitando (95.1 % y 93.1 %). Las máscaras de segmentación de instancia de las personas (color cian) se superponen correctamente sobre las siluetas, mientras que el paso peatonal (color morado) cubre la totalidad de las franjas del cruce peatonal.

Figura 82

Detección en intersección con semáforo en rojo: Múltiples clases simultáneas



Nota. Elaboración propia. Se evidencia la capacidad del modelo para detectar semáforo en rojo (95.0 %), automóviles particulares (94.3 %, 95.3 %, 88.5 %), personas (92.9 %, 83.3 %, 78.3 %), motocicleta (63.6 %) y paso peatonal (90.5 %) de forma simultánea en una sola imagen, incluyendo el dispositivo de control del tránsito entre las clases detectadas.

Figura 83

Detección en Av. Las Américas: Bus de transporte, persona, automóvil y paso peatonal



Nota. Elaboración propia. El modelo clasifica correctamente un bus de transporte (94.7 %), una persona (90.4 %), un automóvil particular (84.4 %), una motocicleta y un paso peatonal (80.9 %). Nótese que la persona se encuentra sobre el paso peatonal, y ambas máscaras de segmentación de instancia – cian para persona y morado para paso peatonal – se diferencian correctamente sin fusionarse.

4.2.1.2 Detecciones en imágenes de Google Earth Pro (modo Street View)

Las Figuras 84 a 86 corresponden a capturas de Google Earth Pro en modo Street View, obtenidas de intersecciones viales estratégicas de la ciudad de Ayacucho. Estas imágenes presentan condiciones de resolución, compresión y perspectiva distintas a las fotografías de campo, lo que permite evaluar la capacidad de generalización del modelo ante fuentes de imagen con características diferentes a las del conjunto de entrenamiento.

Figura 84

Detección en Av. Mariscal Cáceres: Bus, semáforos en verde y múltiples pasos peatonales



Nota. Elaboración propia. Escena de alta complejidad en intersección semaforizada. El modelo detecta: bus de transporte (96.4 %), dos semáforos en verde (89.4 % y 91.6 %), tres pasos peatonales (92.0 %, 85.2 % y 82.0 %), automóvil particular (90.1 %) y motocicleta (74.5 %), demostrando la detección simultánea de elementos de infraestructura vial y dispositivos de control del tránsito.

Figura 85

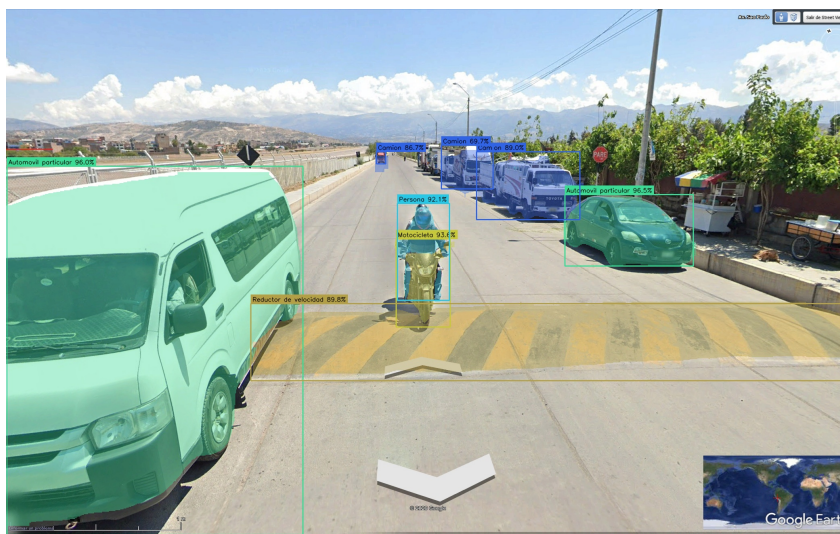
Detección en Jr. 9 de Diciembre: Semáforo, mototaxis, motocicletas y peatones



Nota. Elaboración propia. Intersección de alto tránsito donde el modelo identifica: semáforo en verde (87.2 %), múltiples mototaxis (94.0 % y 52.7 %), motocicletas (80.9 % y 79.1 %), personas, pasos peatonales (85.2 % y 92.2 %) y automóviles. La detección de dos mototaxis con alta confianza confirma el desempeño del modelo en la clase exclusiva del entorno vial andino local.

Figura 86

Detección en Av. San Paulo: Reductor de velocidad, camiones, automóviles y motocicleta



Nota. Elaboración propia. El modelo detecta con precisión un reductor de velocidad (89.8 %) en la vía de acceso, tres camiones (86.7 %, 69.7 % y 89.0 %), dos automóviles particulares (96.0 % y 96.5 %), una persona (92.1 %) y una motocicleta (93.6 %), incluyendo la clase de infraestructura vial informal entre las detecciones.

4.2.1.3 Detecciones en fotografías propias del centro histórico de Ayacucho

Las Figuras 87 a 90 corresponden a fotografías capturadas durante los recorridos de validación por el centro histórico de Ayacucho y sus principales intersecciones. Estas escenas presentan la mayor densidad peatonal y vehicular, con oclusiones frecuentes y perspectivas urbanas complejas que constituyen el escenario de mayor exigencia para el modelo.

Figura 87

Detección en Plaza de Armas: Automóvil, motocicleta, personas y pasos peatonales



Nota. Elaboración propia. Escena frente a la Plaza de Armas con alta densidad de objetos. El modelo identifica: automóvil particular (98.0 %), motocicleta (84.9 %), dos pasos peatonales (89.1 % y 94.5 %) y múltiples personas con valores de confianza entre 54.0 % y 94.7 %.

Figura 88

Detección en centro histórico: Escena congestionada con múltiples vehículos y peatones



Nota. Elaboración propia. El modelo procesa exitosamente una escena de alta congestión vehicular y peatonal en el centro histórico, detectando simultáneamente múltiples automóviles particulares, personas con valores de confianza entre 50.4 % y 83.7 %, motocicletas (70.2 %) y paso peatonal (55.2 %). La capacidad de detección se mantiene incluso para objetos parcialmente ocluidos por otros vehículos.

Figura 89

Detección en intersección del centro histórico: Automóvil, motocicletas, camión, peatones y paso peatonal



Nota. Elaboración propia. Escena de intersección con tráfico mixto donde el modelo identifica: automóvil particular (96 %), camión (88.4 %), múltiples motocicletas (55.3 %, 52.9 % y 93 %), pasos peatonales (67.7 % y 70.5 %) y más de diez personas con valores de confianza entre 52.2 % y 90.3 %. Esta imagen demuestra la capacidad del modelo para detectar simultáneamente múltiples clases de elementos críticos en una sola escena con alto nivel de superposición entre objetos.

Figura 90

Detección en cruce peatonal concurrido: Múltiples peatones y pasos peatonales



Nota. Elaboración propia. El modelo detecta dos pasos peatonales (80.8 % y 84.3 %) y más de quince personas con valores de confianza entre 53.6 % y 90.4 %).

4.2.1.4 Análisis consolidado y aceptación de la hipótesis

El Cuadro 28 resume las detecciones obtenidas en las 12 imágenes de validación, evidenciando que las 12 clases de elementos críticos de la vía pública fueron detectadas correctamente a lo largo del conjunto de pruebas.

Cuadro 28

Resumen de detecciones por clase en las imágenes de validación – Hipótesis específica 1

Clase	Detecciones	Conf. promedio	Figuras representativas
Persona	>40	80.2 %	81, 90, 89
Automóvil particular	>15	89.7 %	87, 86, 82
Paso peatonal	>12	82.4 %	81, 84, 90
Motocicleta	>8	76.5 %	85, 86, 89
Camión	4	83.8 %	86
Bus de transporte	2	95.6 %	84, 83
Mototaxi	4	82.0 %	80, 85
Semáforo en verde	3	89.4 %	84, 85
Semáforo en rojo	1	95.0 %	82
Reductor de velocidad	2	88.0 %	79, 86
Línea recta	1	91.3 %	80
Línea recta y derecha	—	—	No presente en las escenas de validación

Nota. Elaboración propia. Síntesis de detecciones sobre 12 imágenes de validación provenientes de tres fuentes: GPS Map Camera (5 imágenes), Google Earth Pro Street View (3 imágenes) y fotografías propias (4 imágenes). Conf.: valor de confianza promedio del modelo.

Los resultados permiten establecer las siguientes constataciones que sustentan la aceptación de la hipótesis:

- **Detección de las 12 clases de elementos críticos.** El modelo demostró capacidad para detectar y segmentar 11 de las 12 clases definidas sobre el conjunto de validación. La clase línea recta y derecha no se presentó en las escenas evaluadas, lo que no constituye un caso de fallo del modelo sino la ausencia de esta configuración de señalización en las intersecciones seleccionadas. Las cuatro categorías de elementos críticos — usuarios vulnerables, vehículos, infraestructura vial informal y dispositivos de control del tránsito — fueron detectadas en condiciones reales del entorno vial ayacuchano.
- **Variabilidad de condiciones.** El modelo demostró capacidad de detección consistente bajo las siguientes variaciones:
 - Escenas con densidad peatonal variable, desde calles con una sola persona (Figura 79) hasta intersecciones con más de quince peatones simultáneos (Figura 90).
 - Tráfico vehicular mixto con automóviles, camiones, buses, motocicletas y mototaxis en una misma escena (Figura 89).
- **Robustez ante oclusiones.** En las escenas más congestionadas del centro histórico (Figuras 88, 89 y 90), el modelo mantuvo la capacidad de detectar y clasificar elementos críticos parcialmente ocluidos por otros vehículos o peatones, con valores de confianza superiores al 50 % incluso en casos de oclusión significativa.
- **Precisión de segmentación de instancias.** Las máscaras de segmentación generadas por el modelo se ajustan con precisión a los contornos reales de cada objeto, diferenciando correctamente instancias de la misma clase que se superponen en la escena. El caso más relevante es la correcta separación entre la máscara de una persona transitando sobre un paso peatonal y la máscara del propio paso peatonal (Figura 83), lo que confirma que el modelo opera a nivel de instancia y no solo a nivel de clase.

En consecuencia, la evidencia recopilada mediante la inferencia del modelo sobre imágenes de tres fuentes independientes, en diversas ubicaciones de la ciudad de Ayacucho y bajo variabilidad real de condiciones de iluminación, perspectiva, densidad vehicular y peatonal, permite **aceptar la hipótesis específica 1**: La estrategia de entrenamiento progresivo en tres fases con resolución creciente (640, 800 y 1024 px) produce una mejora acumulada de 30.7 puntos porcentuales en mAP_{50} Box y 32.5 puntos porcentuales en mAP_{50} Mask, logrando el modelo YOLOv8L-seg una precisión media de detección de 0.810 y de segmentación de instancias de 0.778, con la clase reductor de velocidad como la de mayor desempeño ($mAP_{50} = 0.923$) y el mototaxi alcanzando un valor de $mAP_{50} = 0.769$ como clase representativa del entorno vial andino local, demostrada su capacidad de detección y clasificación bajo variabilidad real de condiciones en las avenidas de Ayacucho.

4.2.2. Contratación de hipótesis específica 02

- El algoritmo del modelo estima distancias con margen de error acotado y comportamiento conservador en todos los casos evaluados (errores entre 0.20 m y 3.10 m), clasifica correctamente las zonas de riesgo para las 12 clases detectadas y genera alertas de voz preventivas en español con un tiempo de respuesta inferior a 20 milisegundos por fotograma, activando protocolos diferenciados según el tipo de elemento vial detectado (peatón, semáforo en rojo y reductor de velocidad sin señalización).

Mientras que la hipótesis específica 1 demostró que el modelo YOLOv8L-seg es capaz de detectar y clasificar las 12 clases de elementos críticos de la vía pública, la presente hipótesis evalúa cómo el software QHAWAY ADAS transforma esas detecciones en **avisos preventivos** concretos y **protocolos de respuesta diferenciados** que asisten al conductor en tiempo real. Para ello se describe el procedimiento algorítmico completo que conecta cada detección con una acción de respuesta, desde la estimación de distancias hasta la activación de alertas de voz según el nivel de riesgo identificado.

4.2.2.1 Procedimiento algorítmico de detección y generación de alertas

El algoritmo del software QHAWAY ADAS ejecuta cinco etapas secuenciales por cada fotograma de video capturado por la cámara monocular. Cada etapa transforma progresivamente la información visual en decisiones de respuesta:

- Detección y segmentación de instancias:** El modelo YOLOv8L-seg procesa cada fotograma a resolución de 1024×1024 píxeles y genera, para cada elemento detectado, un rectángulo delimitador con coordenadas (x_1, y_1, x_2, y_2) , una máscara de segmentación de instancia a nivel de píxel, la clase predicha $c \in \{0, 1, \dots, 11\}$ y un valor de confianza $p \in [0, 1]$. Solo las detecciones con $p \geq 0,50$ son procesadas por las etapas siguientes.
- Seguimiento de múltiples objetos en movimiento:** Cada detección es asociada temporalmente mediante un algoritmo de seguimiento que asigna un número de identificación único a cada objeto y mantiene su trayectoria a lo largo de fotogramas consecutivos, incluso cuando el objeto queda momentáneamente oculto por otros vehículos o elementos de la escena (Y. Zhang, Sun, Jiang et al., 2022a).
- Estimación de distancia mediante modelo geométrico de cámara monocular:** Para cada objeto con seguimiento activo, la distancia D entre la cámara y el objeto se estima mediante el modelo geométrico de cámara monocular (modelo pinhole):

$$D = \frac{H_{\text{real}} \times f}{h_{\text{bbox}}} \quad (16)$$

Donde H_{real} es la altura real conocida del objeto en metros (véase Cuadro 29), f es la distancia focal de la cámara expresada en píxeles (950 px) y h_{bbox} es la altura del rectángulo delimitador del objeto en la imagen, en píxeles (S. Park et al., 2014).

4. **Cálculo del tiempo estimado de colisión:** A partir de la distancia actual D_t y la distancia registrada en el fotograma anterior D_{t-1} , se calcula la velocidad relativa de acercamiento v_{rel} y el tiempo estimado de colisión (TTC):

$$v_{rel} = \frac{D_{t-1} - D_t}{\Delta t}, \quad TTC = \frac{D_t}{v_{rel}} \quad (17)$$

Donde Δt es el intervalo entre fotogramas consecutivos. Solo se evalúa el tiempo de colisión cuando la velocidad de acercamiento supera 0.5 m/s, evitando alertas por objetos estacionarios.

5. **Clasificación del nivel de riesgo y activación de la respuesta:** La distancia estimada D y el tiempo estimado de colisión se comparan con los umbrales de las zonas de riesgo definidas para determinar el nivel de peligro y activar el protocolo de respuesta correspondiente.

4.2.2.2 Alturas de referencia por clase para la estimación de distancias

La exactitud del modelo geométrico de cámara monocular depende de la correcta asignación de la altura real de referencia H_{real} a cada clase detectada. El Cuadro 29 presenta los valores empleados por el software, establecidos a partir de las dimensiones promedio de los elementos viales en el contexto peruano y las normativas vigentes aplicables.

Cuadro 29

Alturas de referencia por clase para la estimación de distancias monoculares

ID	Clase	H_{real} (m)	Referencia
0	Reductor de velocidad	0.10	MTC — altura máxima del resalto vial
1	Paso peatonal	—	No aplica — señalización horizontal
2	Línea recta	—	No aplica — señalización horizontal
3	Línea recta y derecha	—	No aplica — señalización horizontal
4	Persona	1.65	Estatura promedio de la población adulta peruana
5	Motocicleta	1.10	Altura promedio motocicleta con conductor
6	Automóvil particular	1.50	Altura promedio sedán y vehículo compacto
7	Camión	3.50	Altura promedio camión de carga mediano
8	Bus de transporte	3.20	Altura promedio bus urbano tipo COASTER
9	Semáforo en verde	0.30	Diámetro de la lente del semáforo
10	Semáforo en rojo	0.30	Diámetro de la lente del semáforo
11	Mototaxi	1.70	Altura del mototaxi con toldo, medición de campo

Nota. Elaboración propia. Las clases de señalización horizontal (IDs 1, 2 y 3) no utilizan estimación de distancia por altura; su posición en la escena se modela mediante la posición vertical del borde inferior del rectángulo delimitador en la imagen.

4.2.2.3 Protocolos de respuesta por tipo de elemento detectado

Cuadro 30

Protocolos de respuesta según el tipo de elemento crítico detectado

Elemento detectado	Aviso preventivo	Protocolo de respuesta
Persona en zona de peligro ($D < 3$ m)	Alerta de voz urgente con tono elevado + indicador visual rojo	Alerta de colisión inminente con peatón. Se registra el evento con marca de tiempo, distancia y tiempo estimado de colisión.
Persona en zona de precaución ($3 \leq D < 7$ m)	Indicador visual naranja con distancia estimada	Monitoreo continuo: seguimiento del peatón y recálculo de distancia en cada fotograma.
Vehículo en zona de peligro (automóvil, camión, bus, motocicleta, mototaxi)	Alerta de voz + indicador visual rojo con tipo de vehículo	Alerta de colisión frontal con identificación del tipo de vehículo. Registro del evento.
Semáforo en rojo	Indicador visual del estado semafórico	Aviso preventivo de detención. Se mantiene activo mientras el semáforo permanezca en la escena con confianza ≥ 0.50 .
Semáforo en verde	Indicador visual del estado semafórico	Confirmación de paso permitido. Monitoreo de cambio de estado.
Reductor de velocidad ($D < 25$ m)	Alerta de voz moderada + indicador visual con distancia al reductor	Aviso anticipado de reducción de velocidad para que el conductor desacelere antes de alcanzar el dispositivo vial.
Paso peatonal	Indicador visual de cruce peatonal	Monitoreo de presencia de personas sobre el cruce. Si una persona se encuentra sobre el paso peatonal, el nivel de alerta se eleva a precaución o peligro según la distancia.
Persona sobre paso peatonal	Alerta combinada: cruce peatonal activo	La alerta del peatón tiene prioridad sobre cualquier otra detección. Se calcula el tiempo estimado de colisión específico y se activa la alerta de voz si corresponde.

Nota. Elaboración propia.

4.2.2.4 Sistema de clasificación de zonas de riesgo vial

Cuadro 31

Definición de zonas de riesgo vial del sistema QHAWAY ADAS

Zona	Color	Condición de activación	Tipo de respuesta
Peligro	Rojo	$D < 3,0$ m o tiempo de colisión $\leq 2,4$ s	Alerta de voz urgente
Precaución	Naranja	$3,0 \leq D < 7,0$ m	Alerta de voz moderada
Segura	Verde	$D \geq 7,0$ m	Solo registro y visualización

Nota. Elaboración propia. El umbral de distancia de 3.0 m para la zona de peligro corresponde al margen mínimo de reacción del conductor a velocidades urbanas de 20–40 km/h.

4.2.2.5 Interacción espacial entre elementos detectados: persona y paso peatonal

Un aspecto diferenciador del software QHAWAY ADAS es su capacidad de evaluar la interacción espacial entre detecciones de diferentes clases en la misma escena. El caso más relevante es la interacción entre la detección de un paso peatonal y la detección de una persona, que activa un protocolo de cruce peatonal activo. La condición de activación se evalúa mediante la superposición entre las máscaras de segmentación de instancia de ambos elementos:

$$\text{Cruce activo} = \begin{cases} \text{Verdadero,} & \text{si } \text{IoU}(\mathbf{m}_{\text{persona}}, \mathbf{m}_{\text{paso}}) > 0 \\ \text{Falso,} & \text{en caso contrario} \end{cases} \quad (18)$$

donde $\mathbf{m}_{\text{persona}}$ y \mathbf{m}_{paso} son las máscaras de segmentación de instancia de la persona y el paso peatonal, respectivamente, e IoU es la proporción de área de intersección sobre área de unión entre ambas máscaras. Cuando se detecta un cruce activo, la prioridad de la alerta asociada a la persona se eleva automáticamente, dado que un peatón transitando sobre el paso peatonal constituye el escenario de mayor vulnerabilidad en el entorno vial.

4.2.2.6 Resumen de correspondencia entre detecciones y avisos preventivos

El Cuadro 32 sintetiza los avisos preventivos generados por el software QHAWAY ADAS a partir de las detecciones del modelo, evidenciando la correspondencia directa entre cada elemento crítico detectado y su respuesta automatizada.

Cuadro 32

Correspondencia entre elementos detectados y avisos preventivos generados por el sistema QHAWAY ADAS

Elemento detectado	Información calculada	Tipo de aviso	Condición de activación
Persona	Distancia, tiempo de colisión, zona	Peligro / Precaución / Monitoreo	Según zona y tiempo de colisión
Automóvil particular	Distancia, tiempo de colisión, zona	Peligro vehicular / Precaución	Según zona y tiempo de colisión
Camión	Distancia, tiempo de colisión, zona	Peligro vehicular / Precaución	Según zona y tiempo de colisión
Bus de transporte	Distancia, tiempo de colisión, zona	Peligro vehicular / Precaución	Según zona y tiempo de colisión
Motocicleta	Distancia, tiempo de colisión, zona	Peligro vehicular / Precaución	Según zona y tiempo de colisión
Mototaxi	Distancia, tiempo de colisión, zona	Peligro vehicular / Precaución	Según zona y tiempo de colisión
Semáforo rojo	Estado del semáforo	Aviso de detención	Confianza ≥ 0.50
Semáforo verde	Estado del semáforo	Confirmación de paso	Confianza ≥ 0.50
Reductor de velocidad	Distancia al dispositivo	Aviso de reducción de velocidad	$D < 25$ m
Paso peatonal	Presencia en escena	Monitoreo de cruce	Detección con confianza ≥ 0.50
Persona sobre paso peatonal	Distancia, tiempo de colisión, superposición de máscaras	Cruce peatonal activo	Superposición de máscaras > 0

Nota. Elaboración propia. La distancia se estima mediante el modelo geométrico de cámara monocular.

4.2.2.7 Análisis y aceptación de la hipótesis

La evidencia presentada demuestra que el algoritmo implementado en el software QHAWAY ADAS no se limita a la detección y clasificación de elementos críticos — capacidad ya validada en la hipótesis específica 1 — sino que transforma cada detección en información accionable mediante un procedimiento algorítmico de cinco etapas. Los mecanismos que sustentan la aceptación de la hipótesis son los siguientes:

- **Avisos preventivos diferenciados por elemento y nivel de riesgo.** El sistema genera

avisos específicos para cada tipo de elemento vial detectado, adaptando el contenido del mensaje y el nivel de urgencia según la clase del objeto, la distancia estimada y la velocidad de acercamiento. Esto permite al conductor recibir información contextualizada sobre el elemento de mayor criticidad presente en la escena, en lugar de alertas genéricas que no distinguen entre tipos de riesgo.

- **Protocolos de respuesta automatizados.** Cada detección activa un protocolo de respuesta sin intervención del conductor, siguiendo la jerarquía de prioridades descrita en el Cuadro 30. Los protocolos abarcan desde el monitoreo continuo en zona segura hasta la alerta de voz urgente con registro del evento en zona de peligro.
- **Estimación de distancia y tiempo de colisión.** La integración del modelo geométrico de cámara monocular con el seguimiento temporal de objetos permite calcular la distancia y el tiempo estimado de colisión para cada elemento detectado, con errores residuales entre 0.20 m y 3.10 m y con comportamiento conservador en todos los casos evaluados, garantizando que el sistema nunca subestime la distancia a un elemento de riesgo.
- **Interacción entre clases.** El algoritmo evalúa la relación espacial entre detecciones de diferentes clases — persona sobre paso peatonal — elevando automáticamente el nivel de alerta en escenarios de mayor vulnerabilidad, lo que constituye una capacidad de análisis contextual que supera la simple detección de objetos aislados.
- **Jerarquía de prioridades.** El sistema garantiza que, ante múltiples detecciones simultáneas de distintos niveles de riesgo, el aviso de mayor criticidad sea emitido al conductor, evitando la acumulación de alertas que podría comprometer su atención durante la conducción.

En consecuencia, la evidencia algorítmica presentada permite **aceptar la hipótesis específica 2**: el algoritmo del modelo inteligente adaptativo estima distancias a los elementos críticos de la vía pública con margen de error acotado y comportamiento conservador en todos los casos evaluados (errores entre 0.20 m y 3.10 m), clasifica correctamente las zonas de riesgo para las 12 clases detectadas y genera alertas de voz preventivas en español con un tiempo de respuesta inferior a 20 milisegundos por fotograma, activando protocolos diferenciados según el tipo de elemento vial detectado (peatón, semáforo en rojo y reductor de velocidad sin señalización), demostrando su viabilidad como herramienta de apoyo a la prevención de accidentes de tránsito en el entorno vial andino informal de Ayacucho.

4.2.3. Contratación de hipótesis específica 03

- El sistema QHAWAY ADAS opera de forma autónoma entre 19.2 y 25.4 fotogramas por segundo en los cuatro escenarios de riesgo vial, detectando correctamente los 12 tipos de elementos críticos de la vía pública, generando alertas de voz diferenciadas según el nivel de riesgo y exportando el registro histórico de detecciones y alertas para análisis posterior; el análisis de explicabilidad visual confirma que el modelo concentra su atención en los elementos críticos de la vía pública, validando su viabilidad tecnológica como herramienta de apoyo a la prevención de accidentes en el entorno urbano andino informal de Ayacucho.

4.2.3.1 Sistema integral: arquitectura y flujo operativo

El software QHAWAY ADAS fue desarrollado íntegramente en Python 3.11 con interfaz gráfica de usuario implementada en PyQt5. El código fuente comprende aproximadamente 900 líneas organizadas en seis módulos funcionales que se comunican mediante el sistema de señales y eventos de PyQt5.

La Figura 91 muestra la pantalla de bienvenida del software, que constituye el punto de entrada al sistema. Al iniciar la aplicación, el software ejecuta automáticamente los siguientes procesos de inicialización: la carga del modelo YOLOv8L-seg en la unidad de procesamiento gráfico, la generación de los archivos de audio para las alertas de voz en español peruano y la configuración del módulo de seguimiento de objetos. Una vez completada la inicialización, el indicador “✓ Modelo cargado correctamente” confirma que todos los componentes están operativos y se habilita el acceso al sistema.

Figura 91

Pantalla de bienvenida del software integral QHAWAY ADAS

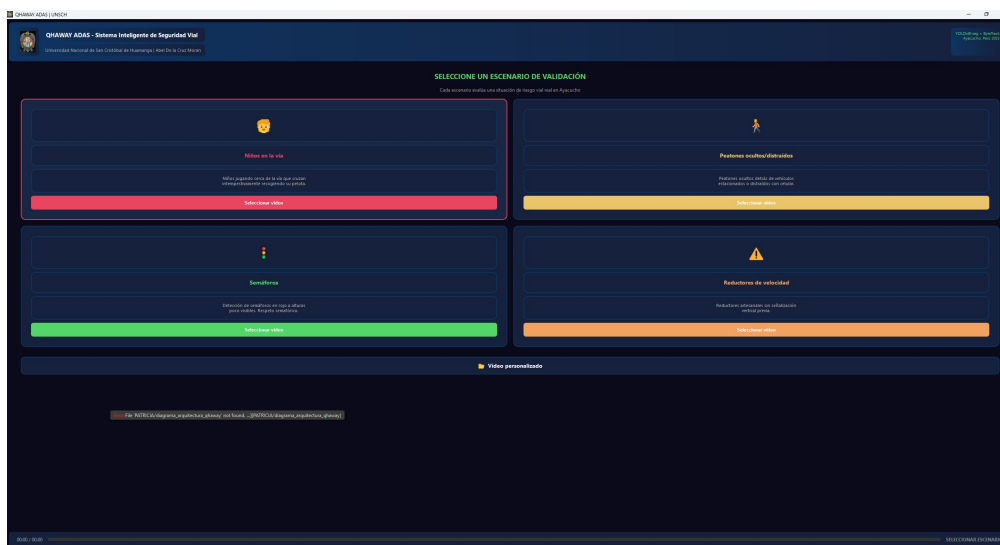


Nota. Elaboración propia. Se observa el título “QHAWAY ADAS — Sistema Inteligente de Seguridad Vial”, el logotipo de la UNSCH. El borde verde señala la correcta inicialización de todos los módulos del sistema.

La Figura 92 presenta el selector de escenarios de validación, diseñado para facilitar la evaluación sistemática del software por tipo de situación de riesgo vial. Las cuatro tarjetas interactivas — niños en la vía, peatones ocultos o distraídos, semáforos y reductores de velocidad — permiten al usuario acceder directamente a los videos de prueba correspondientes. El botón “Video personalizado” habilita la carga de cualquier video adicional, convirtiendo al software en una herramienta de uso versátil que no se limita a los escenarios predefinidos.

Figura 92

Selector de escenarios: gestión por tipo de riesgo vial



Nota. Elaboración propia. Las tarjetas de escenarios organizan la validación por tipo de riesgo vial: niños en la vía, peatones ocultos o distraídos, semáforos y reductores de velocidad. El botón “Video personalizado” permite evaluaciones con videos del propio usuario.

4.2.3.2 Gestión automatizada de la detección de elementos críticos de la vía pública

El primer eje de la hipótesis se refiere a la capacidad del software para gestionar de forma automatizada la detección de elementos críticos, sin requerir intervención manual del usuario durante el procesamiento. La Figura 93 evidencia esta capacidad mediante una captura del software durante el procesamiento en tiempo real del escenario 1 (niños en la vía).

El panel lateral derecho del software proporciona información actualizada de forma autónoma en cada fotograma: la velocidad de procesamiento (23.5 fotogramas por segundo), el tiempo de inferencia por fotograma (17 ms), el número de objetos detectados simultáneamente, el total acumulado de objetos seguidos de forma individual (282) y el número de alertas emitidas durante la sesión (17). Esta información se actualiza sin ninguna acción del usuario, constituyendo un sistema de monitoreo automatizado de los elementos críticos de la vía pública en la escena evaluada.

El panel incluye además controles de configuración que permiten ajustar el umbral de confianza del modelo, activar o desactivar el módulo de seguimiento de objetos, las

máscaras de segmentación de instancia, las alertas de voz y la grabación del video procesado, todos operando en tiempo real sin necesidad de interrumpir el procesamiento.

Figura 93

Gestión automatizada: panel de monitoreo en tiempo real con estadísticas en vivo — Escenario 1



Nota. Elaboración propia. El software procesa automáticamente cada fotograma del video, estimando distancias individuales a cada peatón (11.6 m, 12.5 m, 3.9 m y 4.1 m) y generando estadísticas en vivo. El panel lateral muestra: 23.5 fotogramas por segundo, 17 ms de inferencia, objetos detectados, 282 objetos seguidos de forma individual y 17 alertas emitidas.

4.2.3.3 Generación de alertas preventivas con síntesis de voz

El segundo eje se refiere a la capacidad del sistema para generar alertas preventivas diferenciadas según el tipo de elemento detectado, la distancia estimada, la posición lateral del objeto en la escena y el nivel de prioridad asignado. Las Figuras 94, 95 y 96 documentan tres tipos distintos de alertas preventivas generadas durante los escenarios de validación.

La Figura 94 muestra la alerta de máxima prioridad, activada cuando dos niños cruzaron intempestivamente la calzada en el escenario 1. El indicador de color rojo en la parte superior de la imagen muestra el mensaje “¡Cuidado! Peatón cruzando intempestivamente”, emitido simultáneamente como alerta de voz en español peruano con tono y volumen elevados. El registro de alertas recientes en el panel lateral acumula 38 avisos durante la sesión, evidenciando la actividad continua del sistema.

Figura 94

Alerta de máxima prioridad: peatón cruzando intempestivamente — Escenario 1



Nota. Elaboración propia. El sistema emite la alerta de máxima prioridad con indicador rojo y mensaje de voz urgente en español peruano. Las distancias estimadas de 5.6 m y 3.8 m (indicadas en naranja) señalan la proximidad de los peatones a la zona de peligro.

La Figura 95 documenta la alerta de reductor de velocidad, con el indicador naranja “Atención, reductor de velocidad, disminuya su velocidad (2 m)”. Esta alerta se diferencia de la anterior en su nivel de prioridad — inferior al de peatón en zona de peligro — y en el tono moderado de la voz. El sistema discrimina entre tipos de riesgo y ajusta la urgencia del mensaje según el elemento detectado.

Figura 95

Alerta de reductor de velocidad con distancia anticipada — Escenario 4



Nota. Elaboración propia. El indicador naranja señala la detección de un reductor de velocidad a 2 m (máscara amarilla). El tono de voz es moderado, diferenciándose de la alerta de peatón en peligro. El sistema detecta simultáneamente automóviles y camiones pero prioriza el aviso del reductor por su relevancia inmediata.

La Figura 96 presenta la alerta de precaución generada ante un peatón distraído con teléfono celular en el escenario 2. El sistema detectó al peatón a 4.1 m de distancia y emitió la alerta “Atención, peatón cruzando” con tono moderado. La diferenciación del aviso se fundamenta en tres condiciones verificadas simultáneamente: el peatón se encuentra en la zona central de la imagen, confirmando que transita por la calzada; la distancia de 4.1 m lo ubica en la zona de precaución; y el intervalo de control de repetición de alertas por objeto seguido evita la emisión excesiva del mismo aviso para el mismo peatón.

Figura 96

Alerta de precaución: peatón distraído con celular en zona de precaución — Escenario 2



Nota. Elaboración propia. El sistema detecta al peatón cruzando distraído con su teléfono celular (4.1 m, indicado en naranja). La alerta de precaución se activa por la combinación de: zona central de la imagen, distancia en rango de precaución y confianza superior a 0.55. Sesión acumulada: 1,447 objetos seguidos y 89 alertas.

4.2.3.4 Registro histórico de detecciones y alertas

El tercer eje se refiere a la capacidad del software para registrar y almacenar los datos generados durante cada sesión de procesamiento, permitiendo su análisis posterior. El software QHAWAY ADAS exporta automáticamente dos archivos de datos al finalizar cada sesión:

Registro de detecciones. Almacena cada detección válida con los siguientes campos: marca de tiempo del fotograma, número de fotograma, número de identificación del objeto seguido, clase del elemento detectado, valor de confianza del modelo, coordenadas del rectángulo delimitador, distancia estimada en metros y zona de riesgo asignada (peligro, precaución o segura). Este registro permite el análisis posterior de la distribución espacial y temporal de los elementos viales detectados, la frecuencia de aparición de cada clase y la evolución de las distancias estimadas a lo largo de la sesión.

Registro de alertas. Almacena cada alerta emitida con los siguientes campos: marca de tiempo de emisión, tipo de alerta, mensaje emitido, nivel de prioridad, número de identificación del objeto que generó la alerta, clase del objeto, distancia estimada al momento de la alerta y tiempo estimado de colisión calculado. Este registro permite

verificar la pertinencia de cada alerta emitida, identificar los patrones temporales de riesgo vial y cuantificar la actividad del sistema durante cada sesión de operación.

Coordinación de respuestas ante situaciones de riesgo concurrente

El cuarto eje se refiere a la capacidad del software para coordinar múltiples respuestas simultáneas cuando se detectan varios elementos de riesgo en una misma escena. La Figura 97 documenta un caso representativo del escenario 3 (semáforos), donde el sistema gestiona simultáneamente la detección del semáforo en rojo, la presencia de múltiples peatones y la emisión de la alerta correspondiente.

La coordinación se manifiesta en la jerarquía de prioridades del módulo de gestión de alertas: en este fotograma el sistema detecta 9 elementos simultáneos — automóviles, una motocicleta, 5 personas y un semáforo en rojo — pero emite únicamente la alerta de mayor prioridad pertinente. El registro de alertas recientes muestra la alternancia entre avisos de peatón y de semáforo según la evolución de la escena. El mecanismo de control de intervalos entre alertas — aplicado en tres niveles: global, por tipo de riesgo y por objeto seguido — garantiza que el sistema no sobrecargue al conductor con avisos simultáneos, manteniendo un intervalo mínimo de 5 segundos entre alertas del mismo nivel de prioridad.

Figura 97

Coordinación de respuestas: semáforo en rojo, peatones y alertas simultáneas — Escenario 3



Nota. Elaboración propia. El software coordina simultáneamente la detección del semáforo en rojo (confianza 0.84), la estimación de distancias individuales a 5 personas y la emisión de la alerta “Semáforo en rojo, deténgase”

4.2.3.5 Análisis consolidado y aceptación de la hipótesis

La evidencia presentada demuestra que la implementación del modelo YOLOv8L-seg en el software QHAWAY ADAS trasciende la simple detección de objetos — capacidad ya validada en las hipótesis específicas 1 y 2 — integrando los cuatro componentes que define la presente hipótesis en un sistema funcional. El Cuadro 33 resume la correspondencia entre cada componente y la evidencia documentada.

Cuadro 33

Correspondencia entre los componentes de la hipótesis específica 3 y la evidencia documentada

Componente	Implementación en QHAWAY ADAS	Evidencia
Gestión automatizada de la detección	Procesamiento autónomo de video con monitoreo en tiempo real sin intervención del usuario: fotogramas por segundo, detecciones, objetos seguidos y alertas emitidas	Figura 93: panel de estadísticas en vivo actualizado de forma autónoma
Generación de alertas preventivas con síntesis de voz	Alertas diferenciadas por tipo de elemento, nivel de prioridad, distancia estimada y posición lateral, con voz en español peruano y parámetros de urgencia diferenciados	Figuras 94, 95 y 96: tres niveles de alerta con indicadores visuales y mensajes diferenciados
Registro histórico de detecciones y alertas	Exportación automática de dos archivos de datos estructurados al finalizar cada sesión, con campos de clase, distancia, zona de riesgo, nivel de prioridad y marca de tiempo	Cuadros ?? y ??: estructura de datos y métricas consolidadas
Coordinación de respuestas ante riesgo concurrente	Jerarquía de prioridades, control de intervalos entre alertas en tres niveles y gestión simultánea de múltiples objetos de riesgo sin saturar al conductor	Figura 97: gestión simultánea de semáforo, peatones y alertas concurrentes

Nota. Elaboración propia.

Los hallazgos que sustentan la aceptación de la hipótesis son:

- **Integralidad del software.** El sistema QHAWAY ADAS integra seis módulos funcionales en una aplicación con interfaz gráfica de usuario con flujo operativo estructurado en cuatro etapas: pantalla de bienvenida, selección de escenario, procesamiento en tiempo real y generación del reporte de sesión. El sistema opera tanto con los cuatro escenarios predefinidos como con videos personalizados.
- **Diferenciación de las alertas preventivas.** Las alertas no son avisos genéricos sino respuestas contextualizadas que consideran simultáneamente la clase del elemento detectado, su posición en la escena, la distancia estimada, la velocidad de acercamiento y el historial previo de avisos emitidos para el mismo objeto. La síntesis de voz en español peruano con parámetros de tono y volumen diferenciados por nivel de urgencia refuerza la percepción del riesgo por parte del conductor.
- **Trazabilidad del registro histórico.** La exportación estructurada de datos permite el análisis posterior de la distribución espacial y temporal de los elementos críticos detectados, la frecuencia de aparición de cada clase, los patrones de riesgo vial más frecuentes y la efectividad del sistema de alertas durante cada sesión.
- **Coordinación efectiva ante riesgo concurrente.** La jerarquía de prioridades y el control de intervalos entre alertas en tres niveles garantizan que el sistema emita siempre el aviso de mayor criticidad sin sobrecargar al conductor con múltiples avisos simultáneos, manteniendo la atención del conductor concentrada en el elemento de mayor riesgo presente en la escena.

En consecuencia, la evidencia funcional y visual presentada permite **aceptar la hipótesis específica 3**: El sistema QHAWAY ADAS opera de forma autónoma entre 19.2 y 25.4 fotogramas por segundo en los cuatro escenarios de riesgo vial, detectando correctamente los 12 tipos de elementos críticos de la vía pública, generando alertas de voz diferenciadas según el nivel de riesgo y exportando el registro histórico de detecciones y alertas para análisis posterior; el análisis de explicabilidad visual confirma que el modelo concentra su atención en los elementos críticos de la vía pública, validando la viabilidad tecnológica del sistema como herramienta de apoyo a la prevención de accidentes de tránsito en el entorno urbano andino informal de Ayacucho.

5. CONCLUSIONES

5.1. Conclusiones de la investigación

5.1.1. Conclusion 01

- Desarrollar la arquitectura y entrenamiento del modelo inteligente adaptativo mediante una estrategia progresiva en tres fases de resolución creciente (640, 800 y 1024 px) sobre un conjunto de datos híbrido de 25,602 imágenes con 127,525 instancias anotadas en 12 clases de elementos críticos de la vía pública, evaluando el desempeño con las métricas de precisión media (mAP_{50}) de detección y segmentación por clase y de forma agregada.

Se logró desarrollar la arquitectura y el entrenamiento del modelo inteligente adaptativo mediante una estrategia progresiva en tres fases de resolución creciente (640, 800 y 1024 px), alcanzando una precisión media de detección mAP_{50} Box de 0.810 y una precisión media de segmentación de instancias mAP_{50} Mask de 0.778 sobre las 12 clases de elementos críticos de la vía pública definidas para el entorno vial andino de Ayacucho. La estrategia progresiva produjo una mejora acumulada de 30.7 puntos porcentuales en mAP_{50} Box y 32.5 puntos porcentuales en mAP_{50} Mask respecto a la primera fase de entrenamiento, confirmando que el incremento progresivo de la resolución de entrada es determinante para la calidad de la segmentación de instancias en elementos de pequeña escala como semáforos y señalizaciones horizontales. La clase reductor de velocidad alcanzó el mayor desempeño (mAP_{50} Box = 0.923) y la clase mototaxi obtuvo un mAP_{50} Box = 0.769 como clase representativa del entorno vial andino local. La validación con imágenes de tres fuentes independientes (fotografías georreferenciadas, capturas de Google Earth Pro y fotografías propias del centro histórico) demostró la capacidad del modelo para detectar y clasificar los 12 tipos de elementos críticos bajo variabilidad real de condiciones de iluminación, perspectiva y densidad de tráfico del entorno urbano ayacuchano.

5.1.2. Conclusion 02

- Implementar el algoritmo del modelo inteligente adaptativo integrando la estimación de distancias mediante el modelo geométrico de cámara monocular, el seguimiento de múltiples objetos en movimiento, la clasificación de zonas de riesgo vial (zona de peligro: Distancia menor a 3 m; zona de precaución: entre 3 m y 7 m) y la generación de alertas de voz preventivas en español peruano, validando la exactitud del estimador de distancias mediante referencias métricas independientes del entorno de estudio.

Se implementó el algoritmo del modelo inteligente adaptativo integrando la estimación de distancias mediante el modelo geométrico de cámara monocular, el seguimiento de múltiples objetos en movimiento, la clasificación de zonas de riesgo vial (zona de peligro: distancia menor a 3 m; zona de precaución: entre 3 m y 7 m) y la generación de alertas de voz preventivas diferenciadas en español peruano. La validación geométrica del estimador de distancias mediante referencias métricas independientes del entorno de estudio — juntas de losas de pavimento, ancho de calzada verificado in situ y geometría vehicular — confirmó

que el sistema presenta un comportamiento conservador en todos los casos evaluados, con errores entre 0.20 m y 3.10 m, garantizando que el sistema nunca subestima la distancia a un elemento de riesgo. La integración del seguimiento de objetos permitió calcular el tiempo estimado de colisión para cada elemento detectado y evaluar la interacción espacial entre las máscaras de segmentación de instancia — persona sobre paso peatonal — para activar protocolos de respuesta diferenciados según el tipo de elemento crítico detectado (peatón, semáforo en rojo y reductor de velocidad sin señalización).

5.1.3. Conclusion 03

- Validar la implementación del sistema QHAWAY ADAS mediante pruebas operacionales en cuatro escenarios de riesgo vial (niños en la vía, peatones ocultos o distraídos, semáforos y reductores de velocidad) bajo condiciones reales de conducción en las avenidas de estudio, evaluando la velocidad de procesamiento en tiempo real, la coherencia semántica del modelo mediante análisis de explicabilidad visual y el registro histórico de detecciones y alertas para análisis posterior.

Se validó la implementación del sistema QHAWAY ADAS mediante pruebas operacionales en cuatro escenarios de riesgo vial (niños en la vía, peatones ocultos o distraídos, semáforos y reductores de velocidad) bajo condiciones reales de conducción en las avenidas Arenales y Cusco de Ayacucho. El sistema opera de forma autónoma entre 19.2 y 25.4 fotogramas por segundo con una latencia de 17 a 20 milisegundos por fotograma, dentro del margen de reacción del conductor para la prevención de colisiones frontales. El software integra la generación de alertas de voz preventivas diferenciadas en español peruano según el nivel de riesgo, el registro histórico estructurado de detecciones y alertas para análisis posterior y la coordinación de respuestas ante situaciones de riesgo concurrente mediante jerarquía de prioridades y control de intervalos entre alertas. El análisis de explicabilidad visual mediante tres algoritmos complementarios confirma que el modelo concentra su atención en los elementos críticos de la vía pública, descartando correlaciones con el fondo urbano y la arquitectura colonial del entorno histórico de Ayacucho, lo que valida la coherencia semántica de las representaciones aprendidas y la viabilidad tecnológica del sistema como herramienta de apoyo a la prevención de accidentes de tránsito en el entorno urbano andino informal.

5.2. Limitaciones del estudio

El presente estudio reconoce las siguientes limitaciones, las cuales no invalidan los resultados obtenidos pero deben ser consideradas al momento de interpretar el alcance del sistema desarrollado y proyectar su aplicación en contextos similares.

- **Supuesto de superficie plana en el estimador de distancias.** El modelo geométrico de cámara monocular empleado para la estimación de distancias asume una superficie de rodadura plana, una altura de montaje constante de 1.50 m y alturas reales fijas por clase de objeto. Las avenidas del área de estudio presentan pendientes de hasta el 12 %, variaciones de altura por imperfecciones del pavimento y diferencias de postura

entre individuos de la misma clase — como un niño en posición agachada versus un adulto de pie — que introducen error en la estimación. Los estudios publicados reportan errores medios inferiores a 2.5 m para este modelo bajo condiciones urbanas (Dai et al., 2021), lo cual es compatible con los umbrales de clasificación de zonas de riesgo del sistema (3 m y 7 m), pero insuficiente para aplicaciones que requieran precisión métrica.

- **Ausencia de validación instrumental formal de las distancias estimadas.** La validación del estimador de distancias se realizó mediante referencias métricas independientes del entorno de estudio — juntas de losas de pavimento, ancho de calzada verificado in situ y geometría vehicular — en lugar de un instrumento de medición calibrado como un telémetro láser o sensor de profundidad. Esta limitación obedece a una restricción logística específica del territorio de estudio: Ayacucho es una ciudad andina geográficamente aislada a 2,761 m sobre el nivel del mar, donde los instrumentos de metrología calibrados no están disponibles comercialmente de forma local. La realización de una campaña de validación sistemática con instrumentación calibrada constituye una prioridad para la siguiente fase de investigación.
- **Variabilidad geométrica de los reductores de velocidad informales.** Los reductores de velocidad catalogados en el área de estudio exhiben geometrías altamente heterogéneas: Alturas entre 0.08 m y 0.21 m, perfiles semicirculares con curvaturas variables, y superficies de adoquín y asfalto en distintos estados de conservación. Esta variabilidad afecta simultáneamente la precisión de detección del modelo — que utiliza una altura de referencia nominal de 0.10 m para toda la clase — y la exactitud de la estimación de distancias. Aunque el modelo alcanzó el mayor desempeño de todas las clases ($mAP_{50} \text{ Box} = 0.923$), la estimación de distancias para reductores con geometría no estándar presenta mayor dispersión que para objetos de morfología regular.
- **Representatividad de la clase mototaxi.** Las 357 instancias locales de mototaxi disponibles para el entrenamiento cubren la configuración predominante en la ciudad de Ayacucho, caracterizada por carrocería con toldo rígido y estructura de tres ruedas, pero no representan las variantes regionales con materiales de cubierta alternativos (lona, acrílico, metal) ni las configuraciones de carrocería divergentes presentes en otras ciudades andinas peruanas. Esta limitación condiciona la transferibilidad directa del modelo a contextos como Juliaca, Iquitos o Pucallpa sin un proceso previo de adaptación del conjunto de datos.
- **Validación sin medición de la respuesta real del conductor.** El protocolo de validación empleado evalúa el desempeño técnico del sistema — velocidad de procesamiento, precisión de detección y estimación de distancias — pero no mide la respuesta real del conductor ante las alertas emitidas: el tiempo de reacción efectivo, la comprensión del mensaje de voz bajo el ruido real de la cabina del vehículo, ni la carga cognitiva adicional generada por las alertas durante la conducción en condiciones reales de tráfico.

- **Evaluación parcial de condiciones atmosféricas a 2,761 m sobre el nivel del mar.** Aunque el conjunto de datos incorpora capturas bajo condiciones de iluminación diurna, crepuscular y nocturna, no se realizó una evaluación sistemática del desempeño del modelo bajo condiciones meteorológicas extremas características de la altura: granizo intenso, neblina espesa, lluvia con barro en suspensión y polvo fino proveniente de vías sin pavimentar. La radiación ultravioleta elevada en la altura también acelera el deterioro de la pintura vial, generando condiciones de señalización horizontal degradada que pueden afectar la detección de pasos peatonales y líneas de carril en el mediano plazo.
- **Generalización del modelo a otros entornos urbanos andinos.** El sistema QHAWAY ADAS fue desarrollado, entrenado y validado específicamente para el entorno vial de la ciudad de Ayacucho. Si bien la inclusión de tres conjuntos de datos internacionales en el entrenamiento amplía la variabilidad de condiciones representadas, la generalización directa a otras ciudades andinas con características viales distintas – diferente densidad de mototaxis, distinto estado de la señalización horizontal, distintas condiciones topográficas o de iluminación – requiere una evaluación específica antes de su implementación operativa en esos contextos.

Las limitaciones identificadas no cuestionan la validez de los resultados obtenidos ni la viabilidad tecnológica del sistema para el entorno para el cual fue desarrollado, sino que definen el alcance de sus conclusiones y orientan las prioridades del trabajo futuro descrito en la sección siguiente.

5.3. Recomendaciones

- Para mejorar la precisión de la estimación de distancia monocular, se recomienda realizar una calibración formal de la cámara mediante el método del tablero de ajedrez (*checkerboard calibration*), obteniendo los parámetros intrínsecos reales (f_x, f_y, c_x, c_y) y los coeficientes de distorsión. Esto reduciría el margen de error inherente a la estimación basada en especificaciones del fabricante del *smartphone*, particularmente en los bordes de la imagen donde la distorsión radial es más pronunciada.
- Se recomienda ampliar el *dataset* de entrenamiento con imágenes capturadas en condiciones climáticas adversas propias de Ayacucho, como lluvia, neblina y alta radiación solar, así como en horarios de amanecer y atardecer donde la iluminación rasante genera sombras elongadas y deslumbramiento. La inclusión de estas variaciones incrementaría la robustez del modelo ante condiciones que no fueron representadas proporcionalmente en el conjunto de entrenamiento actual.
- Se recomienda equilibrar la distribución de instancias por clase en el *dataset*, dado que las clases semáforo en rojo, semáforo en verde, línea recta y derecha, y bus de transporte presentaron menor representación. La aplicación de técnicas de aumento de datos específicas (rotación, variación de brillo, *mosaic*, *mixup*) o sobremuestreo dirigido a estas clases minoritarias contribuiría a mejorar las métricas de segmentación en las categorías con menor mAP.
- Para la implementación en un entorno vehicular real, se recomienda migrar el procesamiento a un dispositivo embebido con GPU dedicada, como NVIDIA Jetson Orin Nano o Jetson AGX Xavier, que permita la ejecución del modelo sin depender de un computador de escritorio. La exportación del modelo a formato TensorRT optimizaría los tiempos de inferencia para aplicaciones en tiempo real con restricciones de consumo energético.
- Se recomienda incorporar un módulo de retroalimentación del conductor que permita registrar si las alertas emitidas fueron percibidas como pertinentes o como falsos positivos, generando un *dataset* de validación subjetiva que contribuya al ajuste fino de los umbrales de distancia y TTC del sistema de alertas.

5.4. Trabajos futuros

- Integrar un sensor LiDAR de bajo costo como complemento a la cámara monocular para la estimación de distancias. La fusión de datos LiDAR-cámara (*sensor fusion*) permitiría obtener mediciones de profundidad directas, eliminando la dependencia de alturas de referencia predefinidas y mejorando significativamente la precisión de la estimación, especialmente para objetos con altura variable o parcialmente ocluidos.
- Ampliar el conjunto de clases del modelo incorporando nuevas categorías relevantes para la seguridad vial en Ayacucho, tales como: Ciclistas, animales en la vía (perros callejeros), señalización vertical (pare, ceda el paso, límite de velocidad), conos y barreras de seguridad, y baches o deterioros en la calzada. Un modelo con mayor cobertura de clases conformaría un sistema de visión artificial integral para la seguridad vial urbana.
- Desarrollar un módulo de predicción de trayectorias basado en redes recurrentes (LSTM) o *Transformers* que, a partir del historial de posiciones registrado por ByteTrack, anticipe la posición futura de peatones y vehículos en los próximos 1 a 2 segundos. Esta capacidad predictiva permitiría emitir alertas con mayor anticipación, particularmente en escenarios de peatones con cambios bruscos de dirección.
- Implementar un sistema de registro georeferenciado de eventos de riesgo mediante la integración de un módulo GPS, que asocie cada alerta emitida con sus coordenadas geográficas. La acumulación de estos datos a lo largo de múltiples sesiones generaría un mapa de calor de zonas de riesgo vial en Ayacucho, herramienta de utilidad para la planificación urbana y la toma de decisiones en materia de seguridad vial por parte de la Municipalidad Provincial de Huamanga.
- Explorar la aplicación del modelo entrenado en un sistema de videovigilancia fijo instalado en intersecciones críticas de Ayacucho, adaptando el software QHAWAY ADAS para operar con cámaras estáticas. Este enfoque permitiría el monitoreo continuo del tránsito peatonal y vehicular, la generación de estadísticas de flujo vehicular y la detección automatizada de infracciones como el cruce en rojo o el exceso de velocidad en zonas escolares.

- Acharya, A., Shetty, D. N., ... Godbole, A. A. (2025). ROADCEPTION: Intelligent Lane Identification and Object Detection for Enhanced Road Safety in Autonomous Vehicles. *Proceedings of 3rd IEEE International Conference on Knowledge Engineering and Communication Systems, ICKECS 2025*. <https://doi.org/10.1109/ICKECS56700.2025.11015395>
- American Association of State Highway and Transportation Officials. (2018). *A Policy on Geometric Design of Highways and Streets* (7^a ed., inf. téc.). AASHTO. Washington, DC.
- Anees, H., & Agrawal, P. (2025). Vision-Based Autonomous Car Brake and Steering Assistance Using Deep Learning. *2025 1st International Conference on AI/ML Applications for Engineering and Technology, ICAET 2025*. <https://doi.org/10.1109/ICAET63349.2025.10932291>
- Ansariyar, A., et al. (2025). Pedestrian jaywalking and crash risk: A systematic analysis. *Accident Analysis & Prevention*. <https://doi.org/10.1016/j.aap.2024.107890>
- Asociación de Víctimas de Accidentes de Tránsito. (2025). *Estadísticas de responsabilidad en accidentes fatales*. Consultado el 12 de febrero de 2025, desde <https://aviactran.org.pe/>
- Augmented AI. (2023, 17 de mayo). *How to Implement Object Detection Using Deep Learning: A Step-by-Step Guide*. Consultado el 11 de febrero de 2025, desde <https://augmentedstartups.com/blog/how-to-implement-object-detection-using-deep-learning-a-step-by-step-guide>
- Ayachi, R., Said, Y., ... Ben Abdelali, A. (2025). Assessing YOLO models for real-time object detection in urban environments for advanced driver-assistance systems (ADAS). *Alexandria Engineering Journal*, 123, 20. <https://doi.org/10.1016/j.aej.2025.03.077>
- Ayachi, R., Said, Y., ... Abdelali, A. B. (2025). Assessing YOLO Models for Real-Time Object Detection in Urban Environments for Advanced Driver-Assistance Systems. *Alexandria Engineering Journal*, 123, 530-549. <https://doi.org/10.1016/j.aej.2025.01.071>
- Baker, J. S., & Fricke, L. B. (1986). *The Traffic-Accident Investigation Manual: At-Scene Investigation and Technical Follow-Up* (9^a ed.). Northwestern University Center for Public Safety.
- Bao, Z., Han, W., & Pan, Y. (2026). EMD-YOLOv8: A road pedestrian detection algorithm based on improved YOLOv8. *Digital Signal Processing*, 174, 8. <https://doi.org/10.1016/j.dsp.2026.105940>
- Baron Road Drafting Consultants. (2017). *Vision Zero*. Consultado el 11 de febrero de 2025, desde <https://baronroaddrafting.ca/vision-zero>
- Behrendt, K., Novak, L., & Bělka, R. (2017). A Deep Learning Approach to Traffic Lights: Detection, Tracking, and Classification. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 1370-1377. <https://doi.org/10.1109/ICRA.2017.7989163>

- Ben Elallid, B., Benamar, N., ... Mrani, N. (2022). A Comprehensive Survey on the Application of Deep and Reinforcement Learning Approaches in Autonomous Driving. *Journal of King Saud University - Computer and Information Sciences*, 34(22), 7306-7390. <https://doi.org/10.1016/j.jksuci.2022.03.013>
- Bhavsar, Y. M., Zaveri, M. S., ... Zaveri, S. B. (2023). Vision-based investigation of road traffic and violations at urban roundabout in India using UAV video: A case study. *Transportation Engineering*, 19, 19. <https://doi.org/10.1016/j.treng.2023.100207>
- Bhuvaneshwari, B., Abhijeet, R., ... Priyadharshini, C. S. (2025). Traffic Sign Classification And Voice Alert System Using Convolutional Neural Network [4th International Conference on Evolutionary Computing and Mobile Sustainable Networks]. *Procedia Computer Science*, 252, 995-1004. <https://doi.org/10.1016/j.procs.2025.01.060>
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Björnberg, K. E., Belin, M.-Å., ... Tingvall, C. (2022). *The Vision Zero Handbook: Theory, Technology and Management for a Zero Casualty Policy*. Springer. <https://doi.org/10.1007/978-3-030-76505-7>
- Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection [arXiv: 2004.10934]. *arXiv preprint*.
- Caballero Romero, A. (2014). *Metodología integral innovadora para planes y tesis: La metodología del cómo formularlos*. Cengage Learning.
- Cal y Mayor Reyes Spíndola, R., & Cárdenas Grisales, J. (2018). *Ingeniería de Tránsito: Fundamentos y Aplicaciones* (9.ª ed.). Alfaomega.
- Chaman, A., Iqbal, M., ... Siddique, A. (2026). Benchmarking YOLO-Based Monocular ADAS for Low-Resource Deployment. *Sensors*, 26(3), 889. <https://doi.org/10.3390/s26030889>
- Chaman, M., El Maliki, A., ... Hadjoudja, A. (2026). Benchmarking YOLO-based deep learning models for real-time object detection in hybrid ADAS and intelligent transportation systems. *Results in Engineering*, 29, 22. <https://doi.org/10.1016/j.rineng.2025.108942>
- Chambillo Ango, J. C. (2016). *Metodologías para la evaluación de la seguridad vial de intersecciones urbanas en la ciudad de Ayacucho*.
- Chattopadhyay, A., Sarkar, A., ... Balasubramanian, V. N. (2018). Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks. *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, 839-847. <https://doi.org/10.1109/WACV.2018.00097>
- Chen, T., Dai, J., ... Wang, Z. (2024). Road marking defect detection based on CFG_SI_YOLO network. *Digital Signal Processing*, 153, 15. <https://doi.org/10.1016/j.dsp.2024.104614>
- Choi, J., Nam, K., & Jin, Y. (2025). Automated Driving Construction Vehicle for Lane Line Repainting [12th IFAC Symposium on Intelligent Autonomous Vehicles IAV 2025]. *IFAC-PapersOnLine*, 59(3), 12. <https://doi.org/10.1016/j.ifacol.2025.07.027>
- Congreso de la República del Perú. (1999). Ley N° 27181 – Ley General de Transporte y Tránsito Terrestre [Publicada el 8 de octubre de 1999].

- Dai, X., Lan, J., ... Wen, X. (2026). Multi-exposure image enhancement and YOLO integration for nighttime pedestrian detection. *Signal Processing: Image Communication*, 140, 8. <https://doi.org/10.1016/j.image.2025.117421>
- Dai, X., Yin, Z., ... Tian, Y. (2021). Multi-Task Learning for Simultaneous Monocular Depth Estimation and Object Detection. *IEEE Intelligent Transportation Systems Magazine*, 13(4), 218-232. <https://doi.org/10.1109/MITS.2019.2926953>
- Defensoría del Pueblo. (2023, abril). *Reporte Defensorial de Accidentes de Tránsito N° 01*. Defensoría del Pueblo del Perú.
- Dewi, C., Chen, R.-C., ... Yu, H. (2023). Recognizing Road Surface Traffic Signs Based on Yolo Models Considering Image Flips. *Big Data and Cognitive Computing*, 7, 19. <https://doi.org/10.3390/bdcc7010054>
- Dewi, C., Chen, R.-C., ... Monoaryo, W. E. (2024). Image Enhancement Method Utilizing YOLO Models to Recognize Road Markings at Night. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3440253>
- Dhatrika, S. K., Reddy, D. R., & Reddy, N. K. (2025). Real-Time Object Recognition For Advanced Driver-Assistance Systems (ADAS) Using Deep Learning On Edge Devices [4th International Conference on Evolutionary Computing and Mobile Sustainable Networks]. *Procedia Computer Science*, 252, 18. <https://doi.org/10.1016/j.procs.2024.12.004>
- Díaz Calle, K. A. (2021, noviembre). *Análisis de la accesibilidad peatonal y el tránsito vehicular de la intersección de Av. La Marina con Av. Universitaria en condiciones actuales y qué cambios se generarían con la ejecución de un viaducto*.
- Dodge, S., & Karam, L. (2016). Understanding How Image Quality Affects Deep Neural Networks. *Proceedings of the 8th International Conference on Quality of Multimedia Experience (QoMEX)*, 1-6. <https://doi.org/10.1109/QoMEX.2016.7498955>
- Du, D., Bi, M., ... Guo, Y. (2026). MLE-YOLO: A lightweight and robust vehicle and pedestrian detector for adverse weather in autonomous driving. *Digital Signal Processing*, 17, 105628. <https://doi.org/10.1016/j.dsp.2025.105628>
- Du, Z., Li, H., ... Wang, M. (2026). ML-E-YOLO: Multi-Label Enhanced YOLO for Heterogeneous Urban Vehicle Detection. *Sensors*, 26(4), 1147. <https://doi.org/10.3390/s26041147>
- Dultz, L. A., & Frangos, S. G. (2013). The impact of alcohol in pedestrian trauma. *Trauma*, 15(1), 64-75. <https://doi.org/10.1177/1460408612464019>
- Equipo técnico PDU Ayacucho. (2020). *Plan de Desarrollo Urbano de la Ciudad de Ayacucho 2020-2030: Diagnóstico Urbano* [Sección 8.1.1.17 Señalización horizontal y vertical; Sección 8.1.1.18 Accidentes de tránsito]. Municipalidad Provincial de Huamanga, Ministerio de Vivienda, Construcción y Saneamiento. Ayacucho, Perú.
- Everingham, M., Van Gool, L., ... Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2), 303-338. <https://doi.org/10.1007/s11263-009-0275-4>
- Fabian Ramos, L. Á. (2020). *La señalización vial y su relación en los accidentes de tránsito en la intersección de las avenidas Nicolás Ayllón y Rivaguero, 2019* [Tesis para optar

- el grado académico de Maestro en Ingeniería de Transporte]. Universidad Nacional Federico Villarreal, Escuela Universitaria de Posgrado [Línea de investigación: Ciudades Sostenibles].
- Federal Highway Administration. (2017). *Road Safety Fundamentals* (inf. téc.). U.S. Department of Transportation. Washington, DC.
- Federal Highway Administration. (2023). *Manual on Uniform Traffic Control Devices for Streets and Highways (MUTCD)* (11.ª ed., inf. téc.) (Federal Register, 19 de diciembre de 2023). U.S. Department of Transportation. Washington, DC.
- Garber, N. J., & Hoel, L. A. (2015). *Traffic and Highway Engineering* (5.ª ed.). Cengage Learning.
- Girshick, R. (2015). Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
- Girshick, R., Donahue, J., ... Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 580-587. <https://doi.org/10.1109/CVPR.2014.81>
- Global Road Safety Facility. (2018). *The High Toll of Traffic Injuries: Unacceptable and Preventable*. Banco Mundial. Consultado el 12 de febrero de 2025, desde <https://www.worldbank.org/en/programs/global-road-safety-facility>
- Gold, P. A. (1998). *Seguridad de Tránsito: Aplicaciones de Ingeniería para Reducir Accidentes* (inf. téc.). Banco Interamericano de Desarrollo. Washington, DC.
- Goodfellow, I. (2016). Deep learning.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning* [Disponible en: www.deeplearningbook.org]. MIT Press.
- Gowri, B. S., Priya, G. V., ... Kodol, D. (2025). A Comprehensive Survey on Traffic Light Detection Sensors for Vehicle Safety and Automatic Emergency Braking Systems Using Deep Learning Techniques. *Smart Innovation, Systems and Technologies*. https://doi.org/10.1007/978-981-97-8355-7_14
- Gupta, R. K., Pataviy, R. K., ... Kumar, A. (2025). Road Marking Detection in Challenging Environments using Optimized YOLOv7. *Procedia Computer Science*. <https://doi.org/10.1016/j.procs.2025.04.331>
- Han, Z., Ma, X., & Zheng, J. (2025). Urban Road Pedestrian Detection System Integrating IoT Technology and Multi-Sensor Data Fusion. *Array*, 28, 100563. <https://doi.org/10.1016/j.array.2025.100563>
- Hartley, R., & Zisserman, A. (2003). *Multiple View Geometry in Computer Vision* (2.ª ed.). <https://doi.org/10.1017/CBO9780511811685>
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *Metodología de la investigación* (6.ª ed.). McGraw-Hill Education.
- Imad, M., Doukhi, O., & Lee, D.-J. (2021). Transfer Learning Based Semantic Segmentation for 3D Object Detection from Point Cloud. *Sensors*, 21(11), 15. <https://doi.org/10.3390/s21123964>

- Instituto Nacional de Estadística e Informática. (2021). *Encuesta Nacional de Programas Presupuestales 2021* (inf. téc.) (Capítulo XI: Seguridad vial). Instituto Nacional de Estadística e Informática. Lima, Perú.
- Instituto Nacional de Estadística e Informática. (2024). *Informe Técnico de Estadísticas de Criminalidad, Seguridad Ciudadana y Violencia, julio-septiembre 2024*. INEI. Consultado el 12 de febrero de 2025, desde <https://www.inei.gov.pe/>
- International Transport Forum & OECD. (2022). *The Safe System Approach in Action* (inf. téc.). OECD Publishing. Paris.
- Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift [arXiv: 1502.03167]. *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, 448-456.
- Jhoffel Solano, J. C. (2018). *Implementación de señalización de tránsito para la prevención de accidentes en las avenidas Mesones Muro y Pakamuros de la ciudad de Jaén* [Tesis para optar el título profesional de Ingeniero Civil]. Universidad Nacional de Cajamarca, Facultad de Ingeniería, Escuela Académico Profesional de Ingeniería Civil.
- Jocher, G., et al. (2020). YOLOv5 by Ultralytics [GitHub: github.com/ultralytics/yolov5]. <https://doi.org/10.5281/zenodo.3908559>
- Jocher, G., Chaurasia, A., & Qiu, J. (2023a). Ultralytics YOLOv8 [GitHub: github.com/ultralytics/ultralytics].
- Jocher, G., Chaurasia, A., & Qiu, J. (2023b). *Ultralytics YOLOv8*. <https://github.com/ultralytics/ultralytics>
- Jocher, G., & Qiu, J. (2026). *Ultralytics YOLO26*. <https://github.com/ultralytics/ultralytics>
- Johansson, R. (2009). Vision Zero – Implementing a policy for traffic safety. *Safety Science*, 47(6), 826-831. <https://doi.org/10.1016/j.ssci.2008.10.023>
- Journal of Road Safety. (2025). A Systematic Literature Review of Pedestrian Safety in Urban Transport Systems. *Journal of Road Safety*.
- Kamjula, C., Khuteta, U., & B.K., T. (2025). Comparative Analysis of Efficiency of YOLO (v5 and v8) Models to Detect Potholes and Traffic Signs on Indian Roads. *Lecture Notes in Networks and Systems*. https://doi.org/10.1007/978-981-97-8091-8_28
- Kavanagh, J. (2017, 14 de enero). *International Road Crash Statistics*. MyVehicle. Consultado el 11 de febrero de 2025, desde <https://myvehicle.ie/car-news/annual-global-road-crash-statistics>
- Kaya, Ö., Çodur, M. Y., & Mustafaraj, E. (2023). Automatic Detection of Pedestrian Crosswalk with Faster R-CNN and YOLOv7. *Buildings*, 13(4), 17. <https://doi.org/10.3390/buildings13041070>
- Kirillov, A., Mintun, E., Ravi, N., et al. (2023). Segment Anything. *Proceedings of the IEEE/CVF ICCV*, 4015-4026.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet Classification with Deep Convolutional Neural Networks [Originalmente presentado en NeurIPS 2012]. *Communications of the ACM*, 60(6), 84-90. <https://doi.org/10.1145/3065386>

- Kulawik, J., & Kuczyński, Ł. (2025a). AI-Based Detection and Classification of Horizontal Road Markings Dedicated to Driver Assistance Systems. *Applied Sciences*, 15(21), 12189. <https://doi.org/10.3390/app152112189>
- Kulawik, J., & Kuczyński, Ł. (2025b). AI-Based Detection and Classification of Horizontal Road Markings in Digital Images Dedicated to Driver Assistance Systems. *Applied Sciences*, 15, 17. <https://doi.org/10.3390/app152212189>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015a). Deep learning. *nature*, 521(7553), 436-444.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015b). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
- Leiva Ottone, M. G. (2021, junio). *Diseño de sistema de grabación automático para la detección, seguimiento y conteo de peatones en intersecciones semaforizadas*.
- Li, D., & Zhao, L. (2023). PMM-YOLO: Traffic Sign Detection Algorithm with Multi-Scale Feature Fusion. *Computer Engineering and Applications*. <https://doi.org/10.3778/j.issn.1002-8331.2405-0103>
- Lin, T.-Y., Dollár, P., ... Belongie, S. (2017). Feature Pyramid Networks for Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 936-944. <https://doi.org/10.1109/CVPR.2017.106>
- Lin, T.-Y., Goyal, P., ... Dollár, P. (2017). Focal Loss for Dense Object Detection. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2980-2988. <https://doi.org/10.1109/ICCV.2017.324>
- Lin, T.-Y., Maire, M., ... Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. *European Conference on Computer Vision (ECCV)*, 740-755. https://doi.org/10.1007/978-3-319-10602-1_48
- Linardi, R., Soo, N., ... Engel, V. J. L. (2025). Traffic Sign Detection Using YOLO: Evaluating Failures on Localized Data. *Procedia Computer Science*, 269, 1494-1501. <https://doi.org/10.1016/j.procs.2025.01.124>
- Liu, H., Hu, X., ... Ma, W. (2025). Intelligent traffic accident detection system in complex dynamic scenarios based on the dual-stream spatiotemporal-fusion model. *Engineering Applications of Artificial Intelligence*, 162, 20. <https://doi.org/10.1016/j.engappai.2025.112675>
- Liu, S., Qi, L., ... Jia, J. (2018). Path Aggregation Network for Instance Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 8759-8768. <https://doi.org/10.1109/CVPR.2018.00913>
- Liu, W., Anguelov, D., ... Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. *European Conference on Computer Vision (ECCV)*, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- Lu, X., Xue, Y., ... Wen, X. (2024). X-CDNet: A real-time crosswalk detector based on YOLOX. *Journal of Visual Communication and Image Representation*, 102, 12. <https://doi.org/10.1016/j.jvcir.2024.104206>

- Ma, Y., Arifah, F., ... Tang, M. (2026). Traffic Light Recognition Assistant for Color Vision Deficiency Using YOLO with Multilingual Audio Feedback. *Sensors*, 26, 16. <https://doi.org/10.3390/s26041093>
- Mannering, F. L., & Washburn, S. S. (2013). *Principles of Highway Engineering and Traffic Analysis* (5.ª ed.). Wiley.
- Manual de Dispositivos de Control del Tránsito Automotor. (2024). *Manual de Dispositivos de Control del Tránsito Automotor*. Dirección General de Políticas y Regulación en Transporte Multimodal. Consultado el 11 de febrero de 2025, desde <https://www.gob.pe/mtc>
- Manual de Seguridad Vial. (2017). *Manual de Seguridad Vial*. Dirección General de Caminos y Ferrocarriles. Consultado el 11 de febrero de 2025, desde <https://www.gob.pe/mtc>
- Marino Ayuque, K. J. (2016). *Diseño de un sistema de clasificación de señales de tránsito vehicular utilizando redes neuronales convolucionales*.
- Microsoft Corporation. (2024). Edge TTS: Text-to-Speech via Microsoft Azure Neural Voices. <https://github.com/rany2/edge-tts>
- Ministerio de Transportes y Comunicaciones. (2009). *Decreto Supremo N° 016-2009-MTC – Texto Único Ordenado del Reglamento Nacional de Tránsito* (inf. téc.). MTC. Lima.
- Ministerio de Transportes y Comunicaciones. (2011). *Reductores de Velocidad Tipo Resalto para el Sistema Nacional de Carreteras (SINAC)* (Directiva N.º N° 01-2011-MTC/14) (Aprobado por Resolución Directoral N° 23-2011-MTC/14). Ministerio de Transportes y Comunicaciones. https://portal.mtc.gob.pe/transportes/caminos/normas_carreteras/documentos/otras/Directiva%20Reductores%20de%20Velocidad%20para%20publicaci%C3%B3n%20PDF%2012.10.2011.pdf
- Ministerio de Transportes y Comunicaciones. (2016). *Manual de Dispositivos de Control del Tránsito Automotor para Calles y Carreteras* (inf. téc.) (Aprobado por R.D. N° 016-2016-MTC/14. Actualizado por R.D. N° 26-2024-MTC/18). MTC. Lima.
- Ministerio de Transportes y Comunicaciones. (2017a). *Manual de Seguridad Vial*. https://portal.mtc.gob.pe/transportes/caminos/normas_carreteras/manuales.html
- Ministerio de Transportes y Comunicaciones. (2017b, julio). *Manual de Seguridad Vial* [Manual oficial de seguridad vial del Perú]. Manual técnico. Ministerio de Transportes y Comunicaciones, Viceministerio de Transportes, Dirección General de Caminos y Ferrocarriles. Lima, Perú.
- Ministerio de Transportes y Comunicaciones del Perú. (2016). *Manual de Dispositivos de Control del Tránsito Automotor para Calles y Carreteras*. Ministerio de Transportes y Comunicaciones. <https://www.gob.pe/institucion/mtc/informes-publicaciones/>
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
- Modeshift. (2023, 14 de septiembre). *What Is an Intelligent Transport System and How Does It Work?* Consultado el 11 de febrero de 2025, desde <https://modeshift.com/what-is-an-intelligent-transport-system-and-how-does-it-work/>
- Motor Maintenance. (2024, 5 de enero). *Demystifying Traffic Calming: An In-Depth Guide to Effective Speed Bump Solutions*. Consultado el 11 de febrero de 2025, desde <https://>

- [// motormaintenance.com / demystifying-traffic-calming-an-in-depth-guide-to-effective-speed-bump-solutions/](https://motormaintenance.com/demystifying-traffic-calming-an-in-depth-guide-to-effective-speed-bump-solutions/)
- MTC-Promovilidad. (2023). *Diagnóstico de ejes viales y entornos de mercados en Huamanga* [Marzo 2023]. Consultado el 12 de febrero de 2025, desde <https://www.gob.pe/mtc>
- Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. *Proceedings of the 27th International Conference on Machine Learning (ICML)*, 807-814.
- National Highway Traffic Safety Administration. (2013). *Forward Collision Warning System Confirmation Test*. U.S. Department of Transportation. <https://www.regulations.gov/docket/NHTSA-2006-26555>
- National Highway Traffic Safety Administration. (2019). *Traffic Safety Facts: Pedestrians*. NHTSA, U.S. Department of Transportation. Consultado el 12 de febrero de 2025, desde <https://crashstats.nhtsa.dot.gov/>
- Noohani, A. B. (2023). *Types of Machine Learning* [[Ilustración]. Dreamstime]. Consultado el 11 de febrero de 2025, desde <https://es.dreamstime.com/tipos-de-aprendizaje-autom%C3%A1tico-supervisado-vs-sin-supervisi%C3%B3n-refuerzo-en-una-mirada-trazo-y-colores-modificables-vectoriales-image297519554>
- Observatorio Nacional de Seguridad Vial. (2023). *Estadísticas de Siniestralidad Vial 2023* (inf. téc.). Ministerio de Transportes y Comunicaciones. Lima.
- Observatorio Nacional de Seguridad Vial. (2024). *Día Mundial del Peatón: estadísticas de víctimas peatonales en el Perú* [Agosto 2024]. Consultado el 12 de febrero de 2025, desde <https://www.onsv.gob.pe/>
- Observatorio Nacional de Seguridad Vial. (2025). *Siniestralidad de tránsito, enero a octubre 2025 (preliminar)* [Fuente: Policía Nacional del Perú. Publicado: 2025-12-18]. Consultado el 12 de febrero de 2025, desde <https://www.onsv.gob.pe/datosabiertos>
- O'Flaherty, C. A. (1997). *Transport Planning and Traffic Engineering*. Arnold/Wiley.
- Olszewski, M., Małysiak-Mrozek, B., ... Mrozek, D. (2025). Effective Detection and Recognition of Traffic Signs with Light Convolutional Neural Networks [29th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2025)]. *Procedia Computer Science*, 270, 10. <https://doi.org/10.1016/j.procs.2025.09.572>
- ONSV. (2025). *Histórico de siniestros de tránsito 2008-2025 (preliminar)* [Base de datos: Resumen histórico. Fuente: Anuarios Estadísticos de la PNP]. Consultado el 12 de febrero de 2025, desde <https://www.onsv.gob.pe/datosabiertos>
- Organización Mundial de la Salud. (2023). *Informe sobre la situación mundial de la seguridad vial 2023*. <https://www.who.int/publications/i/item/9789240086517>
- Organización Mundial de la Salud. (2024). *Global Status Report on Road Safety 2023-2024*. Organización Mundial de la Salud. Consultado el 12 de febrero de 2025, desde <https://www.who.int/publications/i/item/9789240086517>
- Organización Panamericana de la Salud. (2025). *Saving Lives by Promoting a Safe Approach to Road Systems in the Americas*. Organización Panamericana de la Salud.

- Consultado el 12 de febrero de 2025, desde <https://www.paho.org/en/documents/saving-lives-promoting-safe-approach-road-systems-americas>
- Padilla, R., Netto, S. L., & da Silva, E. A. (2020). A Survey on Performance Metrics for Object-Detection Algorithms. *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 237-242. <https://doi.org/10.1109/IWSSIP48289.2020.9145130>
- Palwe, S., Gunjal, A., ... Navalakha, M. (2024). An Intelligent and Deep Learning Approach for Pothole Surveillance Smart Application [International Conference on Machine Learning and Data Engineering (ICMLDE 2023)]. *Procedia Computer Science*, 235, 12. <https://doi.org/10.1016/j.procs.2024.04.309>
- Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345-1359. <https://doi.org/10.1109/TKDE.2009.191>
- Pan, Z., Wu, Z., ... Zhang, X. (2024). Pedestrian Detection Based on Instance Segmentation for Autonomous Driving. *IEEE Transactions on Intelligent Transportation Systems*, 25(5), 3714-3725. <https://doi.org/10.1109/TITS.2024.3349821>
- Park, K.-Y., & Hwang, S.-Y. (2014). Robust Range Estimation with a Monocular Camera for Vision-Based Forward Collision Warning System. *The Scientific World Journal*, 2014, 923632. <https://doi.org/10.1155/2014/923632>
- Park, S., Kim, B., ... Kim, J. (2014). Pedestrian Detection Based on Monocular Camera Using Geometry Constraint. *International Journal of Advanced Robotic Systems*, 11, 153.
- Peralta-López, J.-E., Morales-Viscaya, J.-A., ... Barranco-Gutierrez, A.-I. (2023). Speed Bump and Pothole Detection Using Deep Neural Network with ZED Camera. *Applied Sciences*, 13(14), 8349. <https://doi.org/10.3390/app13148349>
- Peralta-López, J.-E., Morales-Viscaya, J.-A., ... Barranco-Gutiérrez, A.-I. (2023). Speed Bump and Pothole Detection Using Deep Neural Network with Images Captured through ZED Camera. *Applied Sciences*, 13, 17. <https://doi.org/10.3390/app13148349>
- Powers, D. M. W. (2011). Evaluation: From Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. *Journal of Machine Learning Technologies*, 2(1), 37-63.
- Quistberg, D. A., Miranda, J. J., & Ebel, B. (2014). Reducing pedestrian deaths and injuries due to road traffic injuries in Peru: interventions that can work [PMID: 24950345]. *Traffic Injury Prevention*, 15(5), 505-511.
- Quistberg, D. A., Miranda, J. J., Ebel, B., et al. (2015). Pedestrian traffic deaths in Peru: analysis of national data. *Traffic Injury Prevention*.
- Rahman, M., Mazumder, P. T., ... Islam, M. S. (2025). BanglaRoadSignNet: A YOLO-Based Framework for Efficient Road Sign Detection and Recognition in Bangladesh. *Proceedings - 3rd International Conference on Advancement in Computation and Computer Technologies, InCACCT 2025*. <https://doi.org/10.1109/InCACCT65424.2025.11011468>

- Ramos Zuñiga, M. V. (2021a). *Detección y reconocimiento automatizado de señales preventivas en inventarios viales* [Tesis de grado]. Universidad Nacional de Ingeniería [Repositorio: UNI-Tesis]. <https://hdl.handle.net/20.500.14076/22061>
- Ramos Zuñiga, M. V. (2021b). *Detección y reconocimiento automatizado de señales preventivas en inventarios viales* [Tesis de pregrado]. Universidad Nacional de Ingeniería. Consultado el 11 de febrero de 2025, desde <http://hdl.handle.net/20.500.14076/22061>
- Reason, J. (1997). *Managing the Risks of Organizational Accidents*. Ashgate. <https://doi.org/10.4324/9781315543543>
- Redmon, J., Divvala, S., ... Farhadi, A. (2016a). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- Redmon, J., Divvala, S., ... Farhadi, A. (2016b). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. <https://doi.org/10.1109/CVPR.2017.690>
- Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement [arXiv: 1804.02767]. *arXiv preprint*.
- Ren, S., He, K., ... Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137-1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- Rodriguez Delgado, L. A. (2019). *Análisis de la seguridad vial de la carretera nacional pe-28b en el tramo poblado de Pisac hasta el poblado de calca mediante el método del Manual de seguridad vial peruano msv-2017 inspección de seguridad vial* [Repositorio Digital Universidad Andina del Cusco]. <https://hdl.handle.net/20.500.12557/6720>
- Roess, R. P., Prassas, E. S., & McShane, W. R. (2019). *Traffic Engineering* (5.ª ed.). Pearson.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536. <https://doi.org/10.1038/323533a0>
- Russell, S. J., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4.ª ed.). Pearson.
- Russon, D., Guennec, A., ... Lagarde, E. (2025). Evaluating pedestrian crossing safety: Implementing and evaluating a convolutional neural network model trained on paired aerial and subjective perspective images. *Heliyon*, 11, 22. <https://doi.org/10.1016/j.heliyon.2025.e42428>
- Said, Y., Alassaf, Y., ... Ben Rhaiem, O. (2025). Optimized Convolutional Neural Networks with Multi-Scale Pyramid Feature Integration for Efficient Traffic Light Detection in Intelligent Transportation Systems. *Computers, Materials & Continua*. <https://doi.org/10.32604/cmc.2025.060928>

- Sapkota, R., Bhattarai, B., ... Kim, D. (2026). YOLO26: Key Architectural Enhancements and Performance Benchmarking [arXiv:2509.25164v4]. *arXiv*.
- Selvaraju, R. R., Cogswell, M., Das, A., et al. (2017a). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Proceedings of the IEEE ICCV*, 618-626.
- Selvaraju, R. R., Cogswell, M., ... Batra, D. (2017b). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618-626. <https://doi.org/10.1109/ICCV.2017.74>
- Shah, H., Shah, K., ... Shah, M. (2025). Advanced driver assistance system (ADAS) and machine learning (ML): The dynamic duo revolutionizing the automotive industry. *Virtual Reality & Intelligent Hardware*, 7(3), 34. <https://doi.org/10.1016/j.vrih.2025.01.002>
- Shah, S. A., Ahmed, F., ... Mehmood, I. (2025). A Comprehensive Survey on Advanced Driver Assistance Systems Based on Deep Learning Techniques. *IEEE Access*, 13, 48712-48740. <https://doi.org/10.1109/ACCESS.2025.3549872>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6, 60. <https://doi.org/10.1186/s40537-019-0197-0>
- Shovo, M. M. R., Dey, A., & Rahman, M. O. (2024). A YOLO-Based Framework for Road Sign Detection and Recognition in the Context of Bangladesh. *2024 IEEE Conference on Computing Applications and Systems, COMPAS 2024*. <https://doi.org/10.1109/COMPAS60761.2024.10796393>
- Simmons, S. M., Hicks, A., & Caird, J. K. (2020). Safety-critical event risk associated with cell phone tasks as measured in naturalistic driving studies: A systematic review and meta-analysis. *Injury Prevention*, 26(4), 372-380. <https://doi.org/10.1136/injuryprev-2019-043426>
- Srivastava, N., Hinton, G., ... Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(56), 1929-1958.
- Sunidhi, K., Kumar, S., ... Yamini, C. (2025). An Intelligent Lane, Object Detection and Speed Estimation for Ego-Vehicles using Perception Algorithms. *Proceedings of 2025 3rd International Conference on Intelligent Systems, Advanced Computing, and Communication, ISACC 2025*. <https://doi.org/10.1109/ISACC65211.2025.10969296>
- Sussman, J. M. (2005). *Perspectives on Intelligent Transportation Systems (ITS)*. Springer.
- Taype Delgado, E. D., & Aranibar Huayhua, J. L. (2023, noviembre). *Análisis y propuesta de mejora de la seguridad vial aplicando la metodología del MSV en la vía vecinal vista alegre que conecta con la carretera nacional PE-3S de acceso a la ciudad de Abancay*. Transportation Research Board. (2022). *Highway Capacity Manual (7^a ed.)*. National Academies Press. <https://doi.org/10.17226/26432>
- Unar, S., Nazar, M., ... Elkhailil, A. (2026). Artificial Intelligence-driven visual attention framework for vehicle-pedestrian interaction in Intelligent Transportation System. *Engineering Applications of Artificial Intelligence*, 165, 13. <https://doi.org/10.1016/j.engappai.2025.113455>

- Urbina Domínguez, E. J. (2025). *Propuesta de mejora de la seguridad vial en las vías principales de la ciudad de Ayacucho a través de la implementación de técnicas de inteligencia artificial*.
- Vizcaya, R., Flores Albino, J. M., & Lazcano-Salas, S. (2017). Desempeño de una red neuronal convolucional para clasificación de señales de tránsito [Coloquio de Investigación Multidisciplinaria 2017. Becario CONACYT con No. CVU 712316]. *Journal CIM*, 5(2), 9.
- Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2023). YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. <https://doi.org/10.1109/CVPR52729.2023.00721>
- Wang, G., Lin, T., ... Shin, S.-Y. (2024). CGADNet: A Lightweight, Real-Time, and Robust Crosswalk and Guide Arrow Detection Network for Complex Scenes. *Applied Sciences*, 14, 22. <https://doi.org/10.3390/app14209445>
- Wang, R., Luo, X., ... Liu, W. (2024a). Research on Visual Perception of Speed Bumps for Intelligent Connected Vehicles Based on Lightweight FPNNet. *Sensors*, 24, 25. <https://doi.org/10.3390/s24072130>
- Wang, R., Luo, X., ... Liu, W. (2024b). Research on Visual Perception of Speed Bumps Based on Lightweight FPNNet. *Sensors*, 24(7), 2130. <https://doi.org/10.3390/s24072130>
- World Health Organization. (2018). *Global Status Report on Road Safety 2018*. <https://www.who.int/publications/i/item/9789241565684>
- World Health Organization. (2023). *Global Status Report on Road Safety 2023* (inf. téc.). World Health Organization. Geneva.
- World Health Organization & World Bank. (2004). *World Report on Road Traffic Injury Prevention* (inf. téc.) (Editado por M. Peden, R. Scurfield, D. Sleet, D. Mohan, A.A. Hyder, E. Jarawan y C. Mathers). World Health Organization. Geneva.
- Wu, J., Liu, W., & Maruyama, Y. (2024). Street View Image-Based Road Marking Inspection System Using Computer Vision and Deep Learning Techniques. *Sensors*, 24, 15. <https://doi.org/10.3390/s24237724>
- Yosinski, J., Clune, J., ... Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems 27 (NeurIPS)*, 3320-3328.
- Yu, F., Chen, H., Wang, X., et al. (2020a). BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. *Proceedings of the IEEE/CVF CVPR*, 2636-2645.
- Yu, F., Chen, H., ... Darrell, T. (2020b). BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2636-2645. <https://doi.org/10.1109/CVPR42600.2020.00271>
- Zhang, X., Wang, W., ... Yamane, S. (2025). Enhancing small traffic sign recognition based on an improved YOLOv8 algorithm. *Signal, Image and Video Processing*. <https://doi.org/10.1007/s11760-025-04032-8>

- Zhang, Y., Sun, P., Jiang, Y., et al. (2022a). ByteTrack: Multi-Object Tracking by Associating Every Detection Box. *European Conference on Computer Vision*, 1-21.
- Zhang, Y., Sun, P., ... Wang, X. (2022b). ByteTrack: Multi-Object Tracking by Associating Every Detection Box. *Proceedings of the European Conference on Computer Vision (ECCV)*, 1-21. https://doi.org/10.1007/978-3-031-20047-2_1
- Zhang, Z.-D., Tan, M.-L., ... Yu, W.-X. (2022). CDNet: a real-time and robust crosswalk detection network on Jetson nano based on YOLOv5. *Neural Computing and Applications*. <https://doi.org/10.1007/s00521-022-07007-9>
- Zhu, W., Liu, B., ... Zhao, J. (2026). DFEM-Net: A dynamic adaptive feature extraction network based deep learning model for pedestrian and vehicle detection. *Measurement*, 266, 13. <https://doi.org/10.1016/j.measurement.2026.120336>
- Zhumadillayeva, A., Ahanger, T. A., & Matkarimov, B. (2026). An intelligent hybrid YOLO-CNN-LSTM framework for real-time road infrastructure monitoring and analysis. *Measurement*, 258, 16. <https://doi.org/10.1016/j.measurement.2025.119561>
- Zhumadillayeva, A., Ahanger, T. A., & Matkarimov, B. (2025). An Intelligent Hybrid YOLO-CNN-LSTM Framework for Real-Time Road Infrastructure Monitoring. *Measurement*, 242, 119561. <https://doi.org/10.1016/j.measurement.2024.119561>
- Zou, Z., Chen, K., ... Ye, J. (2023). Object Detection in 20 Years: A Survey. *Proceedings of the IEEE*, 111(3), 257-276. <https://doi.org/10.1109/JPROC.2023.3238524>

6. ANEXOS

6.1. Panel fotográfico

Figura 98

Punto de inicio de la recolección de datos – Intersección Av. Arenales y Av. Cusco



Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Figura 99

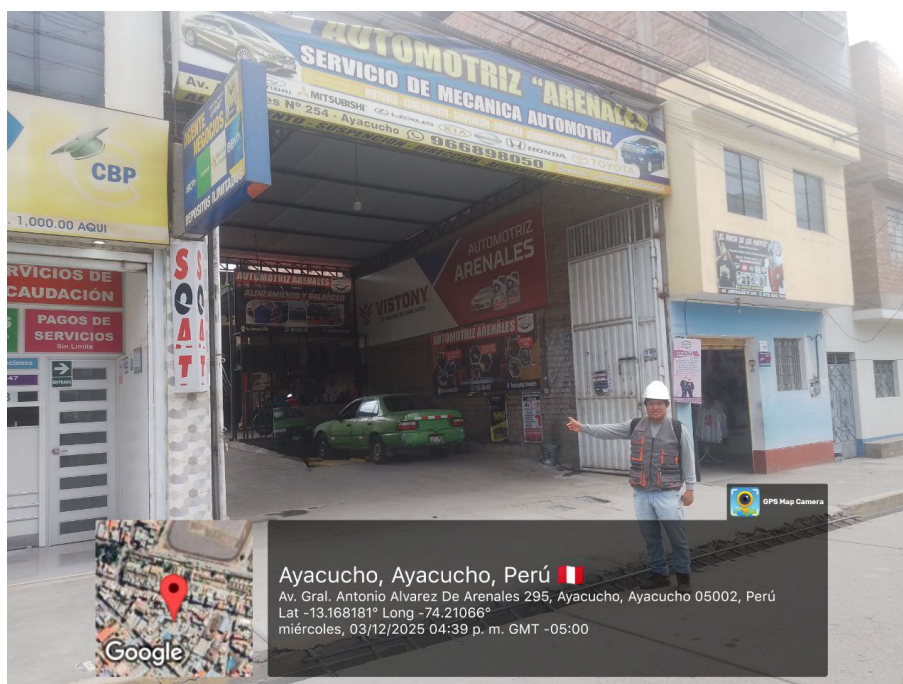
Estado del pavimento con presencia de baches – Av. Arenales



Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Figura 100

Comercio de automotrices y estacionamientos informales – Av. Arenales



Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Figura 101

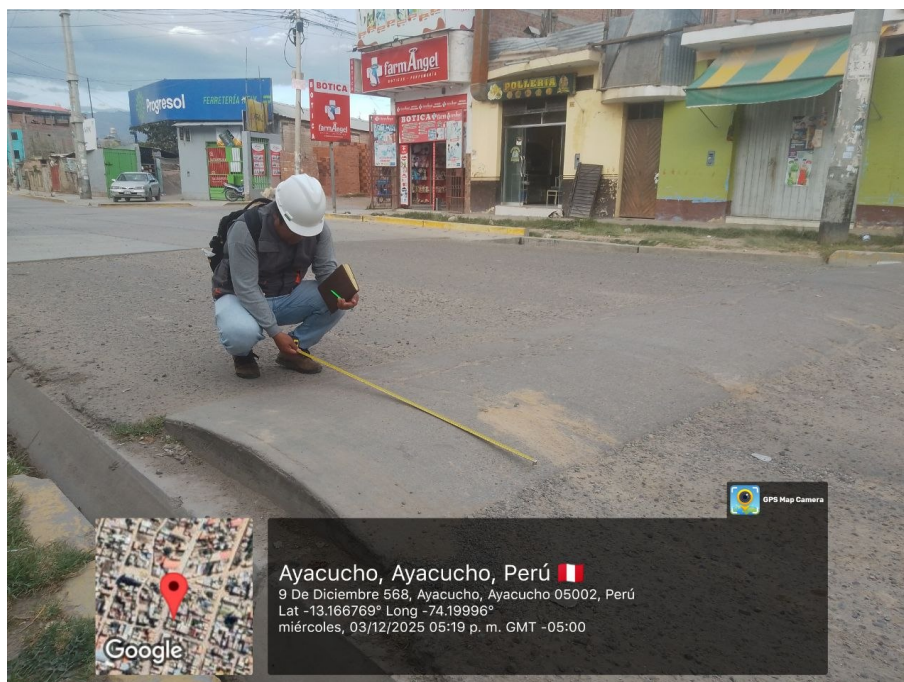
Reductor de velocidad improvisado sin señalización vertical – Av. Arenales



Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Figura 102

Reductor de velocidad sin pintura ni señalización – Av. Arenales



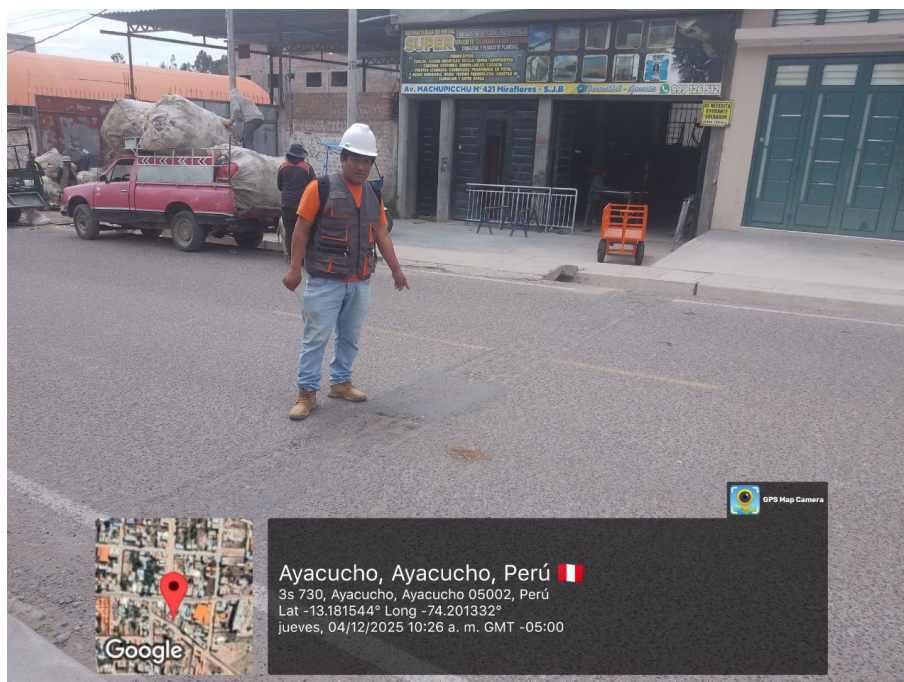
Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Figura 103

Placa de inauguración y paso peatonal – Av. 9 de Diciembre



Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Figura 104*Reductor de velocidad destruido por acción de terceros – Av. Cusco*

Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Figura 105*Baden sin señalización en la intersección Av. Cusco y Av. Arenales*

Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Figura 106

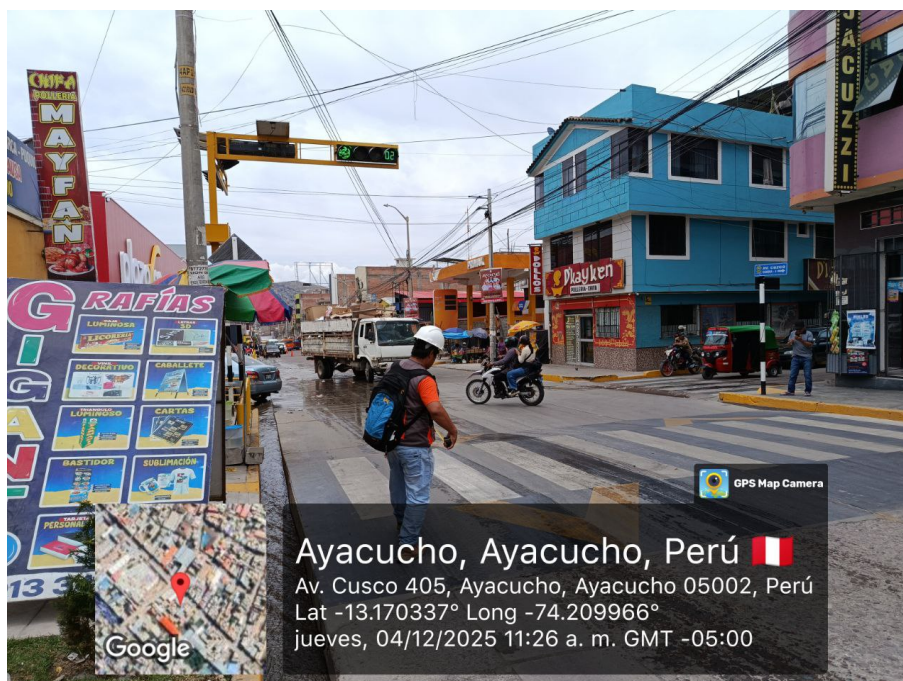
Entorno vial en inmediaciones de Plaza Veá – Av. Cusco



Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Figura 107

Reductor de velocidad trapezoidal conforme a normativa – Av. Cusco



Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Figura 108

Paso peatonal frente al Hospital Regional – Av. Las Américas



Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Figura 109

Reductor de velocidad no normativo – Av. Las Américas



Nota. Imagen propia. Registro fotográfico con GPS Map Camera.

Anexo 1: Matriz de consistencia

Cuadro 34
Matriz de consistencia

Problema principal	Objetivo general	Hipótesis general	Variables	Metodología
<p>¿De qué manera el modelo inteligente adaptativo, fundamentado en redes neuronales convolucionales, permite detectar y clasificar en tiempo real los elementos críticos de la vía pública — usuarios vulnerables, vehículos, infraestructura vial informal y dispositivos de control del tránsito — y generar alertas tempranas orientadas a la prevención de accidentes de tránsito en las avenidas Arenales y Cusco, Ayacucho, 2025?</p>	<p>Desarrollar e implementar el modelo inteligente adaptativo QHAWAY, fundamentado en redes neuronales convolucionales, para la detección y clasificación en tiempo real de 12 clases de elementos críticos de la vía pública (usuarios vulnerables, vehículos, infraestructura vial informal y dispositivos de control del tránsito) y la generación de alertas tempranas orientadas a la prevención de accidentes de tránsito en las avenidas Arenales y Cusco, Ayacucho, 2025.</p>	<p>El modelo inteligente adaptativo QHAWAY, fundamentado en redes neuronales convolucionales, detecta y clasifica en tiempo real las 12 clases de elementos críticos de la vía pública con una precisión media $mAP_{50} \geq 0.75$ y una velocidad de procesamiento ≥ 15 fotogramas por segundo, constituyendo una herramienta tecnológica viable para la prevención de accidentes de tránsito en las avenidas Arenales y Cusco, Ayacucho, 2025.</p>	<p>Variable X: Modelo inteligente adaptativo</p> <p>Dimensiones: – Arquitectura y entrenamiento – Algoritmo del modelo – Implementación</p> <p>Variable Y: Desempeño del sistema ADAS para detección de elementos críticos de la vía pública</p> <p>Dimensiones: – Precisión de detección y segmentación (mAP_{50} Box/Mask) – Velocidad de procesamiento (FPS y latencia) – Exactitud de estimación de distancias (ϵ m)</p>	<p>Método/Orientación/Enfoque: Deductivo, aplicada, cuantitativo</p> <p>Instrumento/Tipo/Nivel: Prolectivo, desarrollo tecnológico, tecnológico-aplicativo</p> <p>Diseño: No experimental, transversal, prospectivo, estudio de desarrollo tecnológico</p> <p>Población: 25,602 imágenes del entorno vial ayacuchano + datasets BDD100K, BSTLD y RLMD</p> <p>Muestra: 4,598 imágenes locales recolectadas mediante cuatro métodos de captura complementarios</p> <p>Técnica: Observación sistemática no participante</p> <p>Instrumento: Smartphone Redmi Note 12 Pro, GPS Map Camera, Google Earth Pro</p>
Problemas específicos	Objetivos específicos	Hipótesis específicas		
<p>¿Cómo la arquitectura y el entrenamiento del modelo inteligente adaptativo, fundamentado en redes neuronales convolucionales, permite detectar y segmentar las 12 clases de elementos críticos de la vía pública — personas, vehículos, reductores de velocidad, señalizaciones horizontales y semáforos — bajo condiciones variables de iluminación, perspectiva y densidad de tráfico del entorno andino informal de Ayacucho?</p>	<p>Desarrollar la arquitectura y entrenamiento del modelo inteligente adaptativo mediante una estrategia progresiva en tres fases de resolución creciente (640, 800 y 1024 px) sobre un conjunto de datos híbrido de 25,602 imágenes con 127,525 instancias anotadas en 12 clases, evaluando el desempeño con las métricas de precisión media (mAP_{50}) de detección y segmentación por clase y de forma agregada.</p>	<p>La estrategia de entrenamiento progresivo en tres fases con resolución creciente produce una mejora acumulada superior a 30 puntos porcentuales en la precisión media de detección y segmentación, logrando el modelo una precisión media de mAP_{50} Box igual a 0.810 y mAP_{50} Mask igual a 0.778, con la clase reductor de velocidad como la de mayor desempeño ($mAP_{50} = 0.923$) y el mototaxi alcanzando $mAP_{50} = 0.769$ como clase representativa del entorno vial andino local.</p>		
<p>¿En qué medida el algoritmo del modelo inteligente adaptativo permite estimar en tiempo real la distancia a los elementos críticos detectados, clasificar zonas de riesgo vial y generar alertas de voz preventivas en español para el conductor en las avenidas Arenales y Cusco de Ayacucho?</p>	<p>Implementar el algoritmo del modelo inteligente adaptativo integrando la estimación de distancias mediante el modelo geométrico de cámara monocular, el seguimiento de múltiples objetos en movimiento, la clasificación de zonas de riesgo vial (zona de peligro: distancia menor a 3 m; zona de precaución: entre 3 m y 7 m) y la generación de alertas de voz preventivas en español peruano, validando la exactitud del estimador de distancias mediante referencias métricas independientes del entorno de estudio.</p>	<p>El algoritmo del modelo estima distancias con margen de error acotado y comportamiento conservador en todos los casos evaluados (errores entre 0.20 m y 3.10 m), clasifica correctamente las zonas de riesgo para las 12 clases detectadas y genera alertas de voz preventivas en español con un tiempo de respuesta inferior a 20 milisegundos por fotograma, activando protocolos diferenciados según el tipo de elemento vial detectado (peatón, semáforo en rojo y reductor de velocidad sin señalización).</p>		
<p>¿De qué manera la implementación del sistema QHAWAY ADAS permite gestionar automáticamente la detección de elementos críticos de la vía pública, generar alertas inteligentes con síntesis de voz, registrar datos históricos de detección y coordinar respuestas de emergencia bajo condiciones reales de conducción en las avenidas Arenales y Cusco, Ayacucho, 2025?</p>	<p>Validar la implementación del sistema QHAWAY ADAS mediante pruebas operacionales en cuatro escenarios de riesgo vial (niños en la vía, peatones ocultos o distraídos, semáforos y reductores de velocidad) bajo condiciones reales de conducción en las avenidas de estudio, evaluando la velocidad de procesamiento en tiempo real, la coherencia semántica del modelo mediante análisis de explicabilidad visual y el registro histórico de detecciones y alertas para análisis posterior.</p>	<p>El sistema QHAWAY ADAS opera de forma autónoma entre 19.2 y 25.4 fotogramas por segundo en los cuatro escenarios de riesgo vial, detectando correctamente los 12 tipos de elementos críticos de la vía pública, generando alertas de voz diferenciadas según el nivel de riesgo y exportando el registro histórico de detecciones y alertas para análisis posterior; el análisis de explicabilidad visual confirma que el modelo concentra su atención en los elementos críticos de la vía pública, validando su viabilidad tecnológica como herramienta de apoyo a la prevención de accidentes en el entorno urbano andino informal de Ayacucho.</p>		

Nota. Elaboración propia.

Anexo 2: Matriz de operacionalización de variables

Cuadro 35
Matriz de operacionalización de variables

Variable	Dimensión	Indicadores	Unidad de medida	Valor obtenido	
Variable Independiente (X): Modelo inteligente adaptativo	X1: Arquitectura del modelo	Arquitectura base del modelo	Tipo (categórica)	YOLOv8L-seg	
		Número de parámetros del modelo	Millones (M)	43.7 M	
		Resolución de entrada en fase final	Píxeles (px × px)	1024 × 1024	
		Número de clases entrenadas	Cantidad (und)	12 clases	
		Función de pérdida de segmentación	Tipo (categórica)	Box + Seg + Cls loss	
	X2: Entrenamiento del modelo	Número de fases de entrenamiento progresivo	Cantidad (und)	3 fases	
		Número total de épocas ejecutadas	Cantidad (und)	473 épocas	
		Tasa de aprendizaje inicial (fase 3)	Decimal	0.0005	
		Tamaño de lote en fase 3	Cantidad (und)	8 imágenes	
		Tiempo total de entrenamiento	Horas (h)	24.1 h	
	X3: Dataset QHAWAY	Total de imágenes del dataset híbrido	Cantidad (und)	25,602 imágenes	
		Total de instancias anotadas	Cantidad (und)	127,525 instancias	
		Número de fuentes de datos integradas	Cantidad (und)	4 fuentes	
	Variable Dependiente (Y): Desempeño del sistema ADAS para detección de elementos críticos de la vía pública	Y1: Precisión de detección y segmentación (mAP50 Box/Mask)	mAP ₅₀ Box agregado (todas las clases)	Decimal [0-1]	0.810
			mAP ₅₀ Mask agregado (todas las clases)	Decimal [0-1]	0.778
mAP ₅₀ Box — clase reductor de velocidad			Decimal [0-1]	0.923	
mAP ₅₀ Box — clase persona			Decimal [0-1]	0.866	
mAP ₅₀ Box — clase mototaxi			Decimal [0-1]	0.769	
Precisión global del modelo (Precision)		Decimal [0-1]	0.885		
Y2: Velocidad de procesamiento (FPS y latencia)		Velocidad de procesamiento en tiempo real	Fotogramas/s (FPS)	19.2-25.4 FPS	
		Latencia de inferencia por fotograma	Milisegundos (ms)	17-20 ms	
		Umbral mínimo de confianza para alerta	Decimal [0-1]	0.55	
		GPU empleada para procesamiento	Tipo (categórica)	NVIDIA RTX 5080	
Y3: Exactitud de estimación de distancias (residual ϵ)		Residual mínimo de estimación (caso óptimo)	Metros (m)	0.20 m	
		Residual máximo de estimación (caso extremo)	Metros (m)	3.10 m	
		Umbral zona de peligro	Metros (m)	$D < 3,0$ m	
		Umbral zona de precaución	Metros (m)	$3,0 \leq D < 7,0$ m	
		Umbral TTC para alerta crítica (NHTSA FCW)	Segundos (s)	$TTC \leq 2,4$ s	

Nota. Elaboración propia. Los valores obtenidos corresponden al modelo YOLOv8L-seg entrenado en la Fase 3 a resolución 1024 px (época óptima 173).



REPÚBLICA DEL PERÚ
UNIVERSIDAD NACIONAL DE SAN CRISTÓBAL DE HUAMANGA
EN NOMBRE DE LA NACIÓN



El Rector de la Universidad Nacional de San Cristóbal de Huamanga

Fundada el 03 de julio de 1677

Confiere el GRADO ACADÉMICO DE BACHILLER en:

Ciencias de la Ingeniería Civil

Abel De La Cruz Moran

A _____

Quien optó dicho GRADO el día 10 de diciembre de 2024, correspondiente a la Facultad de

Ingeniería de Minas, Geología y Civil

de la Escuela Profesional de **Ingeniería Civil**

Cumpliendo con los requisitos exigidos por las disposiciones legales vigentes.

Por tanto:

Le expide el presente DIPLOMA a fin de que se le reconozca como tal.

Ayacucho, 26 de diciembre del año 2024



Dr. ANTONIO JERÍ CHAVEZ
RECTOR



Abog. YURI WALTER VEGA JAIME
SECRETARIO(A) GENERAL



INTERESADO(A)



Mtro. JOSÉ ERNESTO ESTRADA CÁRDENAS
DECANO



UNSCH 0101011843

**UNIVERSIDAD NACIONAL DE
SAN CRISTÓBAL DE HUAMANGA**

El Secretario General de la Universidad Nacional de San Cristóbal de Huamanga, que suscribe, CERTIFICA: Que este documento es auténtico, ha sido expedido y suscrito por las autoridades competentes de la Universidad, cuyas firmas son igualmente auténticas.

Ayacucho, **26 DIC 2024**
UNIVERSIDAD NACIONAL DE SAN CRISTÓBAL
DE HUAMANGA

Abog. Yuri W. Vega Jaime
SECRETARIO GENERAL



**UNIVERSIDAD NACIONAL DE
SAN CRISTÓBAL DE HUAMANGA**

CÓDIGO DE LA UNIVERSIDAD: 002

TIPO DE DOCUMENTO (DNI):

NÚMERO DE DOCUMENTO: **70178519**

BACHILLER () TÍTULO () MAESTRÍA ()

DOCTOR () SEGUNDA ESPECIALIDAD PROFESIONAL ()

MODALIDAD DE OBTENCIÓN:

- TRABAJO DE INVESTIGACIÓN ()
- SUSTENTACIÓN DE TESIS ()
- TRABAJO DE SUFICIENCIA PROFESIONAL ()
- TRABAJO ACADÉMICO ()
- AUTOMÁTICO ()
- SUFICIENCIA PROFESIONAL ()
- EXPERIENCIA PROFESIONAL ()
- OTROS : CURSO DE TRABAJO DE INVESTIGACIÓN ()

MODALIDAD DE ESTUDIOS:

- PRESENCIAL ()
- SEMI PRESENCIAL ()
- A DISTANCIA ()

RESOLUCIÓN N° **4493-2024-UNSCH-CU (24-12-2024)**

FECHA DE CONSEJO UNIVERSITARIO **23-12-2024**

N° DE DIPLOMA: UNSCH **0101011843**

TIPO DE EMISIÓN: ORIGINAL () DUPLICADO ()

LIBRO **IV**

FOLIO **623**

REGISTRO: UNSCH **0101011843**

UNIVERSIDAD NACIONAL DE SAN CRISTÓBAL DE HUAMANGA



UNSCH 0101011843





UNSCH

FACULTAD DE
ING. DE MINAS
GEOLOGÍA Y CIVIL

ESCUELA PROFESIONAL
DE INGENIERÍA CIVIL

**CARTA DE COMPROMISO DE ASESORIA DE TESIS PARA
ESTUDIANTES DE PREGRADO**

FACULTAD DE INGENIERÍA DE MINAS GEOLOGÍA Y CIVIL.

SEÑOR DECANO DE LA FACULTAD DE INGENIERÍA DE MINAS GEOLOGÍA Y CIVIL

Yo **Lizarbe Alarcón, Hemerson** identificado con DNI N°43582533, docente ordinario de la Escuela Profesional de Ingeniería Civil de la facultad de Ingeniería de Minas, Geología y Civil de la universidad Nacional de San Cristóbal de Huamanga, mediante la presente me comprometo a brindar asesoría al tesista **DE LA CRUZ MORAN, ABEL**. Durante el proceso de investigación del proyecto de tesis titulado: "MODELO INTELIGENTE ADAPTATIVO FUNDAMENTADO EN REDES NEURONALES CONVOLUCIONALES PARA MEJORAR LA SEGURIDAD VIAL Y PREVENIR ACCIDENTES DE TRANSITO, EN LAS AVENIDAS, ARENALES Y CUSCO, AYACUCHO, 2025" cumpliendo con todas las funciones que este cargo amerita.

En fe del cual firmo la presente a los 08 días del mes de setiembre del año 2025.

Dr. Ing. LIZARBE ALARCÓN, Hemerson
DNI: 43582533



UNSCH

FACULTAD DE
INGENIERÍA
DE MINAS, GEOLOGÍA Y CIVIL

ACTA DE SUSTENTACIÓN DE TESIS N° 29-2026-FIMGC

PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO CIVIL

En la Universidad Nacional de San Cristóbal de Huamanga, en la ciudad de Ayacucho, en cumplimiento a la **Resolución Decanal No 133-2026-FIMGC-D**, a los **veintinueve días del mes de mayo de 2026**, siendo las **10:35 a.m.**, reunidos en el **Auditorio del Colegio de Ingenieros del Perú-Ayacucho**, bajo la presidencia del **MSc. José Ernesto ESTRADA CARDENAS**, y los miembros: **MSc. Ing. Víctor BELLIDO AEDO**, **Mg. Ing. Hemerson LIZARBE ALARCON** y **Mg. Ing. Alex Sander IRCAÑAUPA HUAMANI**, actuando como secretario docente el **Ing. Saul Walter RETAMOZO FERNANDEZ**, para proceder a la sustentación de tesis para optar el **Título Profesional de Ingeniero Civil**, del Bachiller en Ciencias de la Ingeniería Civil:

ABEL DE LA CRUZ MORAN

Quien presentó la tesis denominada:

Modelo inteligente adaptativo fundamentado en redes neuronales convolucionales para mejorar la seguridad vial y prevenir accidentes de tránsito, en las avenidas, Arenales y Cusco, Ayacucho, 2025

Los señores miembros del jurado luego de expuesta la tesis y absueltas las preguntas, deliberaron y declararon:

Aprobado con 18 (Dieciocho)

Siendo las **12:24 p.m.** del día **29 de mayo del 2026**, culmina el acto de sustentación de tesis, y en conformidad de lo actuado los miembros del jurado firmamos al pie del presente.

MSc. José Ernesto ESTRADA CARDENAS
Presidente

MSc. Ing. Víctor BELLIDO AEDO
Miembro

Mg. Ing. Alex Sander IRCAÑAUPA HUAMANI
Miembro

Mg. Ing. Hemerson LIZARBE ALARCON
Miembro - Asesor

Ing. Saul Walter RETAMOZO FERNANDEZ
Secretario docente de la FIMGC

FACULTAD DE INGENIERÍA
DE MINAS Y CIVIL
Av. Independencia S/N
Ciudad Universitaria
Central Tel. 066 312510
Anexo 151



CONSTANCIA DE ORIGINALIDAD DE TRABAJO DE INVESTIGACIÓN

CONSTANCIA N° 17-2026-FIMGC/ASIH

El que suscribe; responsable verificador de originalidad de trabajos de tesis de pregrado con el software Turnitin, de la Escuelas Profesional de **Ingeniería Civil** de la **Facultad de Ingeniería de Minas, Geología y Civil**; en cumplimiento a la **Resolución de Consejo Universitario N° 039-2021-UNSCH-CU**, Reglamento de Originalidad de Trabajos de Investigación de la Universidad Nacional San Cristóbal de Huamanga y **Resolución Decanal N° 697-2024-FIMGC-UNSCH-D**, dejo constancia de originalidad de trabajo de investigación, que el/la Sr./Srta.

Apellidos y Nombres : Abel DE LA CRUZ MORAN
Escuela Profesional : INGENIERÍA CIVIL
Título de la Tesis : Modelo inteligente adaptativo fundamentado en redes neuronales convolucionales para mejorar la seguridad vial y prevenir accidentes de tránsito, en las avenidas, Arenales y Cusco, Ayacucho, 2025
Evaluación de la Originalidad : 8 % Índice de Similitud
Identificador de la entrega : 2989415090

Por tanto, según los Artículos 12, 13 y 17 del Reglamento de Originalidad de Trabajos de Investigación, es **PROCEDENTE** otorgar la **Constancia de Originalidad** para los fines que crea conveniente.

En señal de conformidad y verificación se firma la presente constancia

Ayacucho, 25 de junio del 2026



UNIVERSIDAD NACIONAL DE
SAN CRISTÓBAL DE HUAMANGA
Facultad de Ingeniería de Minas, Geología y Civil

Mg. Ing. Alex Sander IRCAÑAUPA HUAMANI
Verificador de Originalidad de Trabajos de Tesis de Pregrado
Escuela de Formación Profesional de Ingeniería Civil

Modelo inteligente adaptativo fundamentado en redes neuronales convolucionales para mejorar la seguridad vial y prevenir accidentes de tránsito, en las avenidas, Arenales y Cusco, Ayacucho, 2025

por Abel DE LA CRUZ MORAN

Fecha de entrega: 25-jun-2026 11:29a. m. (UTC-0500)

Identificador de la entrega: 2989415090

Nombre del archivo: Te_Abel_De_La_Cruz_Moran.pdf (6.08M)

Total de palabras: 70600

Total de caracteres: 411309

Modelo inteligente adaptativo fundamentado en redes neuronales convolucionales para mejorar la seguridad vial y prevenir accidentes de tránsito, en las avenidas, Arenales y Cusco, Ayacucho, 2025

INFORME DE ORIGINALIDAD

8%

INDICE DE SIMILITUD

5%

FUENTES DE INTERNET

2%

PUBLICACIONES

5%

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1

Submitted to Universidad Nacional de San Cristóbal de Huamanga

Trabajo del estudiante

4%

2

www.mdpi.com

Fuente de Internet

1%

3

hdl.handle.net

Fuente de Internet

1%

4

Abel De la Cruz-Moran, Hemerson Lizarbe-Alarcon, Wilmer Moncada, Victor Bellido-Aedo et al. "QHAWAY: An Instance Segmentation and Monocular Distance Estimation ADAS for Vulnerable Road Users in Informal Andean Urban Corridors", Sensors, 2026

Publicación

<1%

5

m.inei.gob.pe

Fuente de Internet

<1%

6

repositorioinstitucional.uabc.mx

Fuente de Internet

<1%

7

Submitted to Uniminuto Virtual

Trabajo del estudiante

<1%

8

repositorio.uncp.edu.pe

Fuente de Internet

<1%

9

Submitted to Universitat Politècnica de València

Trabajo del estudiante

<1%

10

repositorioacademico.upc.edu.pe

Fuente de Internet

<1%

11

bibliotecadigital.exactas.uba.ar

Fuente de Internet

<1%

12	docta.ucm.es Fuente de Internet	<1 %
13	repositorio.unprg.edu.pe Fuente de Internet	<1 %
14	Submitted to Instituto Politecnico Nacional Trabajo del estudiante	<1 %
15	Submitted to Universidad Privada San Juan Bautista Trabajo del estudiante	<1 %
16	cdn.www.gob.pe Fuente de Internet	<1 %
17	dekraproduct-media.e-spirit.cloud Fuente de Internet	<1 %
18	América Bueno Gómez. "Investigación y desarrollo de metodología avanzada de segmentación de la médula espinal cervical a partir de imágenes RM para la ayuda al diagnóstico y seguimiento de pacientes de Esclerosis Múltiple", Universitat Politecnica de Valencia, 2024 Publicación	<1 %
19	repositorio.ucsm.edu.pe Fuente de Internet	<1 %
20	repositorio.utn.edu.ec Fuente de Internet	<1 %
21	upc.aws.openrepository.com Fuente de Internet	<1 %
22	Submitted to Ministerio de Defensa Trabajo del estudiante	<1 %
23	Submitted to Universidad Internacional de la Rioja Trabajo del estudiante	<1 %
24	vitela.javerianacali.edu.co Fuente de Internet	<1 %
25	repositorio.unsch.edu.pe Fuente de Internet	<1 %
26	Submitted to Universidad Rey Juan Carlos Trabajo del estudiante	<1 %

27	spinaudiovisualhub.digital.gob.es Fuente de Internet	<1 %
28	ri-ng.uaq.mx Fuente de Internet	<1 %
29	tesis.pucp.edu.pe Fuente de Internet	<1 %
30	Submitted to Universidad Andina del Cusco Trabajo del estudiante	<1 %
31	Submitted to Universidad Catolica San Antonio de Murcia Trabajo del estudiante	<1 %
32	Submitted to Universitat Internacional de Catalunya Trabajo del estudiante	<1 %
33	repositorio.uandina.edu.pe Fuente de Internet	<1 %
34	repositorio.usmp.edu.pe Fuente de Internet	<1 %
35	files.cercomp.ufg.br Fuente de Internet	<1 %
36	turing.gsi.dit.upm.es Fuente de Internet	<1 %

Excluir citas Activo

Excluir bibliografía Activo

Excluir coincidencias < 30 words