UNIVERSIDAD NACIONAL DE SAN CRISTÓBAL DE HUAMANGA

FACULTAD DE INGENIERÍA DE MINAS, GEOLOGÍA Y CIVIL

ESCUELA PROFESIONAL DE INGENIERÍA CIVIL

TESIS:

Desarrollo de un sistema de visión artificial para la detección y evaluación de señales de tránsito en vías nacionales: caso de estudio en carretera Ayacucho - Huanta 2024

Para optar el título profesional de:

INGENIERA CIVIL

PRESENTADO POR:

Bach. Kely Pilar HUAMAN DE LA CRUZ

ASESOR:

Msc. Ing. Rocky Giban AYALA BIZARRO

AYACUCHO - PERÚ

2026

RESUMEN

El corredor vial Ayacucho–Huanta (PE-3S, 44.02 km) presenta condiciones críticas de seguridad vial derivadas del deterioro progresivo de su señalización vertical, agravadas por las limitaciones de los métodos convencionales de inventario, que dependen de inspecciones manuales costosas, subjetivas y de difícil actualización periódica. Ante esta situación, el objetivo de la investigación fue desarrollar el sistema SISEVIR (*Sistema de Supervisión de Señales Verticales en Infraestructura Vial*), fundamentado en redes neuronales convolucionales, para la detección, segmentación, clasificación y evaluación automatizada del estado de conservación de señalizaciones verticales en los tramos PE-3S y PE-28A, abarcando 31 tipos de señales conforme al Manual de Dispositivos de Control del Tránsito Automotor 2024 del MTC. Se adoptó un enfoque cuantitativo con diseño experimental; se conformó un dataset de 6,213 imágenes con 7,107 instancias anotadas bajo condiciones variables de iluminación y clima. Se entrenaron tres arquitecturas: YOLO26s-seg para detección y segmentación, EfficientNet-B0 para clasificación del estado de conservación en siete categorías, y Qwen2-VL-2B-Instruct para generación de descripciones en lenguaje natural. Los resultados evidenciaron métricas sobresalientes: YOLO26s-seg alcanzó un mAP@50 de 0.9305 para detección y 0.9224 para segmentación, con velocidad de 2.1 ms/imagen; EfficientNet-B0 obtuvo Accuracy de 0.9315 y Macro F1 de 0.8252. La validación nocturna confirmó la robustez del sistema con 85.9 % de confianza promedio. El análisis comparativo demostró que SISEVIR reduce en 36.4 % el tiempo de inspección y en 49.1 % los costos operativos respecto al método manual, con reducción del personal de campo de 3 a 1 persona. Se concluye que el sistema es viable para auditorías viales automatizadas, generando reportes PDF en formato MTC compatibles con los procedimientos del Manual de Seguridad Vial MSV-2016.

Palabras clave: Redes neuronales convolucionales, señalizaciones verticales de tránsito, segmentación de instancias, clasificación de estado de conservación, auditoría vial automatizada, visión artificial.

ABSTRACT

The Ayacucho–Huanta road corridor (PE-3S, 44.02 km) presents critical road safety conditions derived from the progressive deterioration of its vertical signage, aggravated by the limitations of conventional inventory methods, which rely on costly, subjective manual inspections that are difficult to update periodically. Given this situation, the objective of this research was to develop the SISEVIR system (*Sistema de Supervisión de Señales Verticales en Infraestructura Vial*), based on convolutional neural networks, for the automated detection, segmentation, classification, and evaluation of the conservation status of vertical traffic signs on the PE-3S (Ayacucho–Huanta, 44.02 km) and PE-28A (Ayacucho–Abra Apacheta, ∼95.4 km) road sections, covering 31 sign types in accordance with the MTC Manual of Motor Vehicle Traffic Control Devices 2024. A quantitative approach with experimental design was adopted; a dataset of 6,213 images with 7,107 annotated instances was assembled under variable lighting and weather conditions. Three deep learning architectures were trained: YOLO26s-seg for detection and instance segmentation, EfficientNet-B0 for conservation status classification into seven categories, and Qwen2-VL-2B-Instruct for natural language description generation. The results demonstrated outstanding metrics: YOLO26s-seg achieved a mAP@50 of 0.9305 for detection and 0.9224 for segmentation, with an inference speed of 2.1 ms/image; EfficientNet-B0 obtained an Accuracy of 0.9315 and a Macro F1 of 0.8252. Nighttime validation confirmed the system's robustness with an average confidence of 85.9 %. The comparative analysis demonstrated that SISEVIR reduces inspection time by 36.4 % and operational costs by 49.1 % compared to the manual inspection method, with a reduction in field personnel from 3 to 1 person. It is concluded that the system is viable for automated road audits, automatically generating PDF reports in MTC inspection format compatible with the Road Safety Manual MSV-2016 procedures.

Keywords: Convolutional neural networks, vertical traffic signs, instance segmentation, conservation status classification, automated road audit, artificial vision.

INTRODUCCIÓN

La seguridad vial constituye uno de los desafíos más significativos del siglo XXI, particularmente en países en desarrollo. Según la Organización Mundial de la Salud, los accidentes de tránsito ocasionan aproximadamente 1.19 millones de fallecimientos anuales a nivel global. En el Perú, durante el año 2023, estadísticas oficiales reportaron más de 87,172 accidentes viales que resultaron en más de 3,138 víctimas fatales, consolidando al territorio nacional como uno de los más afectados en la región sudamericana.

El corredor vial Ayacucho–Huanta, correspondiente a la Ruta Nacional PE-3S con aproximadamente 44.02 kilómetros de extensión, constituye una arteria estratégica para la conectividad regional y el desarrollo económico de la provincia de Huamanga. Sin embargo, presenta características que incrementan significativamente la accidentalidad: Curvas cerradas, pendientes pronunciadas, señalización vertical insuficiente y en estado de deterioro progresivo. Los usuarios de este corredor enfrentan múltiples amenazas derivadas de la escasa visibilidad en tramos críticos, el deterioro de las señalizaciones por decoloración, daños estructurales, rotación o vandalismo, y las variables ambientales propias de la región andina, situación que erosiona la confianza en los sistemas de control vial y representa un obstáculo para el desarrollo regional que depende del transporte eficiente y seguro.

Ante esta problemática, los métodos convencionales de inventario y evaluación de señalización vial resultan insuficientes: las inspecciones manuales requieren la detención del vehículo ante cada señal, consumen considerable tiempo y recursos humanos, están sujetas a errores de percepción y registro por parte de los inspectores, y son de difícil actualización periódica, generando desconocimiento del estado real de la infraestructura vial y dificultando la toma de decisiones para el mantenimiento.

En respuesta a esta situación, la presente investigación propone el desarrollo del sistema SISEVIR (*Sistema de Supervisión de Señales Verticales en Infraestructura Vial*), fundamentado en redes neuronales convolucionales, para la detección, segmentación, clasificación y evaluación automatizada del estado de conservación de señales de tránsito verticales en los tramos PE-3S (Ayacucho–Huanta, 44.02 km) y PE-28A (Ayacucho–Abra Apacheta, ∼95.4 km). El sistema integra tres arquitecturas de aprendizaje profundo: YOLO26s-seg para la detección y segmentación de instancias de 31 tipos de señales verticales conforme al Manual de Dispositivos de Control del Tránsito Automotor 2024 del MTC, EfficientNet-B0 para la clasificación del estado de conservación en siete categorías definidas por la escala SISEVIR, y Qwen2-VL-2B-Instruct para la generación automática de descripciones en lenguaje natural incluidas en los reportes técnicos.

A Dios, fuente de sabiduria y fortaleza que me ha guiado y sostenido en este camino académico.
A mis padres,Apolicarpio y Lidia, por su amor incondicional, apoyo y sacrificio. Su influencia en mi vida ha sido fundamental para llegar a este logro.
A mi pareja Naun, por su amor paciencia y comprensión. Tu apoyo a sido esencial para superar los desafíos de la investigación. A mi hijo Eydan, por ser la luz y la alegría de mi vida. Tu nacimiento me ha enseñado a valorar la vida y a encontrar motivación para seguir adelante.
A mis hermanos, Lisbeth, Noris y Roly, por su apoyo y amor.

AGRADECIMIENTOS

Deseo expresar mi profundo y sincero agradecimiento a todas las personas e instituciones que hicieron posible la realización de este trabajo de investigación, cuyo desarrollo representó un importante desafío académico y personal.

Extiendo mi más sincero y profundo agradecimiento a los docentes de la Escuela Profesional de Ingeniería Civil que, a lo largo de mi formación académica, compartieron generosamente sus conocimientos, experiencias profesionales y valores éticos que hoy forman parte integral de mi perfil profesional. Sus enseñanzas trascendieron las aulas y se convirtieron en herramientas fundamentales para el desarrollo de esta investigación.

A la Universidad Nacional de San Cristóbal de Huamanga, mi alma mater, por acogerme durante estos años de formación y brindarme una educación de calidad. A la Escuela Profesional de Ingeniería Civil, por proporcionarme los conocimientos técnicos, metodológicos y éticos que hoy sustentan mi desarrollo como profesional. Los recursos académicos, la infraestructura y el ambiente de aprendizaje ofrecidos fueron pilares fundamentales para la concreción de esta investigación.

Mi especial y profundo agradecimiento al M.Sc. Ing. Rocky Giban Ayala Bizarro, asesor de la presente investigación, por su invaluable guía, paciencia y dedicación constante durante todo el proceso de desarrollo de este trabajo. Su vasta experiencia profesional, conocimientos especializados en el campo de la ingeniería y su capacidad para transmitir enseñanzas fueron fundamentales para orientar adecuadamente cada etapa de esta investigación. Sus observaciones críticas, sugerencias metodológicas y apoyo incondicional permitieron superar las dificultades encontradas y alcanzar satisfactoriamente los objetivos planteados, contribuyendo significativamente a mi formación como investigador.

Al Dr. Ing. Hemerson Lizarbe Alarcón y al M.Sc. Ing. Saul Walter Retamozo Fernández, distinguidos miembros del jurado evaluador, por su tiempo, disposición y compromiso académico demostrado durante la revisión de esta tesis. Sus valiosas observaciones técnicas, comentarios especializados y aportes metodológicos contribuyeron significativamente a mejorar la calidad, rigurosidad y profundidad técnica y académica de este trabajo. Su experiencia profesional y visión crítica enriquecieron sustancialmente el contenido, la estructura y los resultados de esta investigación, elevando el nivel científico de la misma.

Finalmente, mi más profundo agradecimiento a mi familia por su apoyo incondicional, comprensión, paciencia y motivación constante durante todo este proceso académico. Su confianza depositada en mí y sus palabras de aliento en los momentos más difíciles fueron el impulso necesario para perseverar y culminar exitosamente esta importante meta profesional.

GLOSARIO

Arquitectura YOLOv8	Modelo de detección de objetos de octava generación que utiliza redes neuronales convolucionales para identificación en tiempo real con alta precisión.
Bounding Box	Rectángulo delimitador que localiza y enmarca un objeto detectado en una imagen durante el procesamiento de visión artificial.
CNN	Red Neuronal Convolucional, arquitectura de aprendizaje profundo especializada en procesamiento de imágenes y reconocimiento de patrones visuales.
Confianza	Valor numérico entre 0 y 1 que indica el grado de certeza del modelo sobre la presencia y clasificación correcta de un objeto detectado.
Early Stopping	Técnica que detiene automáticamente el entrenamiento cuando deja de mejorar el rendimiento, evitando el sobreajuste.
Época	Una iteración completa a través de todo el conjunto de datos de entrenamiento durante el proceso de aprendizaje del modelo.
GPU A100	Unidad de procesamiento gráfico de alta gama diseñada para acelerar el entrenamiento de modelos de inteligencia artificial.
IoU	Intersection over Union, métrica que evalúa la precisión de una detección comparando el área de superposición entre la predicción y la etiqueta real.
Kernel	Matriz numérica que se desliza sobre una imagen para extraer características durante la operación de convolución.
LabelMe	Herramienta de anotación de imágenes utilizada para crear conjuntos de datos de entrenamiento para modelos de aprendizaje profundo.
mAP@0.5	Mean Average Precision con umbral IoU de 0.5, métrica estándar para evaluar precisión en modelos de detección de objetos.
Manual MDCTA	Manual de Dispositivos de Control del Tránsito Automotor, documento normativo que establece estándares para señalización vial en Perú.
Pooling	Operación que reduce la dimensionalidad espacial de una representación, conservando la información más relevante.
Precisión	Métrica que mide la proporción de predicciones positivas correctas respecto al total de predicciones positivas realizadas por el modelo.
Recall	Exhaustividad o sensibilidad, mide la proporción de positivos reales correctamente identificados por el modelo.
ReLU	Rectified Linear Unit, función de activación que introduce no linealidad en las redes neuronales, transformando valores negativos a cero.
Seguridad Vial	Conjunto de medidas y estrategias diseñadas para prevenir accidentes de tránsito y proteger a todos los usuarios de las vías públicas.
Señalización Vertical	Dispositivos de control del tránsito instalados verticalmente para reglamentar, prevenir e informar a conductores y peatones.
Sobreajuste	Fenómeno donde un modelo se adapta excesivamente a los datos de entrenamiento, perdiendo capacidad de generalización.
Tiempo de Inferencia	Duración requerida por el modelo para procesar una imagen y generar predicciones, medido típicamente en milisegundos.
YOLO	You Only Look Once, algoritmo de detección de objetos en tiempo real que analiza la imagen completa en una sola pasada.

ACRÓNIMOS

AI	Artificial Intelligence (Inteligencia Artificial), campo de la informática enfocado en crear sistemas que simulan capacidades cognitivas humanas.
CNN	Convolutional Neural Network (Red Neuronal Convolucional), arquitectura de aprendizaje profundo especializada en procesamiento de imágenes y visión artificial.
CV	Computer Vision (Visión por Computadora), disciplina que permite a las máquinas interpretar y procesar información visual del mundo real.
DL	Deep Learning (Aprendizaje Profundo), subset del machine learning que utiliza redes neuronales multicapa para análisis de datos complejos.
EDA	Exploratory Data Analysis (Análisis Exploratorio de Datos), proceso de investigación inicial de conjuntos de datos para descubrir patrones y características.
FPS	Frames Per Second (Fotogramas Por Segundo), medida de velocidad de procesamiento en sistemas de detección en tiempo real.
GPU	Graphics Processing Unit (Unidad de Procesamiento Gráfico), hardware especializado para acelerar el entrenamiento de modelos de inteligencia artificial.
IoU	Intersection over Union (Intersección sobre Unión), métrica fundamental para evaluar precisión de detección en modelos de visión artificial.
JSON	JavaScript Object Notation, formato de intercambio de datos utilizado para almacenar anotaciones de imágenes etiquetadas.
mAP	mean Average Precision (Precisión Media Promedio), métrica estándar para evaluar rendimiento de algoritmos de detección de objetos.
MDCTA	Manual de Dispositivos de Control del Tránsito Automotor, documento normativo oficial para señalización vial en Perú.
ML	Machine Learning (Aprendizaje Automático), rama de la IA que permite a los sistemas aprender y mejorar automáticamente a partir de datos.
NMS	Non-Maximum Suppression (Supresión de No Máximos), algoritmo para eliminar detecciones redundantes en modelos YOLO.
OpenCV	Open Source Computer Vision Library, biblioteca de código abierto para procesamiento de imágenes y visión artificial.
ReLU	Rectified Linear Unit (Unidad Linear Rectificada), función de activación predominante en redes neuronales convolucionales modernas.
RGB	Red, Green, Blue (Rojo, Verde, Azul), modelo de color utilizado en procesamiento digital de imágenes para señalizaciones.
SGD	Stochastic Gradient Descent (Descenso de Gradiente Estocástico), algoritmo de optimización utilizado en entrenamiento de CNN.
TTS	Text-to-Speech (Texto a Voz), tecnología utilizada para generar alertas auditivas automatizadas en el sistema de seguridad vial.
YOLO	You Only Look Once (Solo Miras Una Vez), arquitectura de detección de objetos en tiempo real utilizada para identificar señalizaciones verticales.

LISTA DE SÍMBOLOS

α	Parámetro de tasa de aprendizaje en algoritmos de optimización para entrenamiento de redes neuronales convolucionales.
β	Factor de momento en algoritmos de optimización SGD utilizados durante el entrenamiento del modelo YOLOv8.
γ	Parámetro de decaimiento de la tasa de aprendizaje aplicado durante las épocas de entrenamiento.
δ	Umbral de confianza mínimo para considerar válida una detección de señalización vertical.
ϵ	Valor de tolerancia utilizado en la técnica Early Stopping para determinar la convergencia del modelo.
η	Tasa de aprendizaje (learning rate) específica utilizada en el entrenamiento del modelo CNN.
θ	Conjunto de parámetros y pesos de la red neuronal convolucional YOLOv8m.
λ	Parámetro de regularización L2 aplicado para prevenir el sobreajuste durante el entrenamiento.
μ	Media de la distribución de datos de entrada normalizada antes del procesamiento por la CNN.
σ	Desviación estándar de la normalización aplicada al conjunto de datos de señalizaciones verticales.
τ	Tiempo de procesamiento por imagen medido en milisegundos durante la fase de inferencia.
ϕ	Función de activación ReLU aplicada en las capas convolucionales: $\phi(x) = \max(0, x)$.
ψ	Función de pérdida combinada utilizada en YOLOv8 que incluye clasificación, localización y confianza.
ω	Pesos sinápticos específicos de las conexiones entre neuronas en la arquitectura CNN.
C	Valor de confianza de detección para cada clase de señalización vertical identificada.
$F1$	Puntuación F1-Score calculada como: $F1 = 2 \times \frac{Precisin \times Recall}{Precisin + Recall}$.
IoU	Intersection over Union, métrica de evaluación: $IoU = \frac{rea\ de\ Interseccin}{rea\ de\ Unin}$.
mAP	Mean Average Precision calculada a diferentes umbrales IoU para evaluación del modelo.
N	Número total de imágenes en el conjunto de datos de entrenamiento (2,250 imágenes).
P	Precisión del modelo de detección, relación entre verdaderos positivos y total de predicciones positivas.
R	Recall o sensibilidad del modelo, relación entre verdaderos positivos y total de positivos reales.
t_{inf}	Tiempo de inferencia promedio por imagen expresado en milisegundos (55.1 ms).
X_{input}	Tensor de entrada de dimensiones específicas procesado por las capas convolucionales CNN.
Y_{pred}	Tensor de salida con las predicciones de bounding boxes y clasificaciones de señales verticales.

ÍNDICE GENERAL

ÍNDICE GENERAL **10**

LISTA DE FIGURAS **13**

LISTA DE CUADROS **16**

1 PLANTEAMIENTO DEL PROBLEMA **18**

1.1 Descripción del problema 18
1.2 Delimitación del problema 20
1.2.1 Espacial (geográfica) 20
1.2.2 Temporal 20
1.2.3 Temática y unidad de análisis 21
1.3 Formulación del problema 22
1.3.1 Problema general 22
1.3.2 Problemas específicos 22
1.4 Justificación e importancia 22
1.4.1 Justificación. 22
1.4.2 Importancia. 23
1.5 Limitaciones de la investigación 23
1.6 Objetivos de la investigación 24
1.6.1 Objetivo general 24
1.6.2 Objetivos específicos 24

2 MARCO TEÓRICO **25**

2.1 Antecedentes 25
2.1.1 Investigaciones internacionales. 25
2.1.2 Investigaciones nacionales 34
2.2 Bases teóricas 37
2.2.1 Detección y evaluación de señales de tránsito 37
2.3 Visión aritificial 52
2.3.1 Aprendizaje Automático y Aprendizaje Profundo 52
2.3.2 Redes Neuronales Artificiales (RNA) 53
2.3.3 Taxonomía de arquitecturas neuronales 56
2.3.4 Profundización en las Arquitecturas Neuronales Convolucionales (CNN): 58
2.3.5 Operación convolucional y Kernels 58
2.3.6 Función de activación-ReLu 59
2.3.7 Pooling- Consolidación 60
2.3.8 Síntesis del mecanismo de una CNN 61
2.3.9 YOLO: You Only Look Once 66

3 MARCO METODOLÓGICO **68**
3.1 Enfoque de la investigación . . . 68
3.2 Alcance de la investigación . . . 68
3.3 Diseño de la investigación . . . 68
3.4 Población y muestra . . . 69
3.4.1 Población . . . 69
3.4.2 Muestra . . . 70
3.5 Hipótesis . . . 71
3.5.1 Hipótesis general . . . 71
3.5.2 Hipótesis específicas . . . 71
3.6 Operacionalización de variables . . . 72
3.6.1 Variables . . . 72
3.6.2 Dimensiones . . . 72
3.6.3 Indicadores . . . 73
3.7 Técnicas e instrumentos. . . . 74
3.7.1 Técnicas de recolección . . . 74
3.7.2 Instrumentos utilizados . . . 75
3.8 Desarrollo del trabajo de tesis. . . . 76
3.8.1 Recopilación de datos . . . 76
3.8.2 Clasificación de estados de conservación . . . 82
3.8.3 Etiquetado en LabelMe . . . 89
3.8.4 Ejemplos del proceso de etiquetado en LabelMe . . . 91
3.8.5 Entrenamiento del modelo de detección y segmentación: yolo26s-seg . 96
3.8.6 Resultados del entrenamiento del modelo de detección . . . 99
3.8.7 Entrenamiento del modelo de clasificación de estados . . . 106
3.8.8 Modelo de generacion de descripciones en lenguaje natural Qwen2-VL-2B . . . 113
3.8.9 Descripcion del software SISEVIR . . . 117
3.8.10 Validación del sistema SISEVIR en condiciones nocturnas . . . 126

4 RESULTADOS **136**
4.1 Análisis de resultados . . . 136
4.2 Contrastación de hipótesis . . . 136
4.2.1 Contrastación de hipótesis específica 1. . . . 136
4.2.2 Contrastación de hipótesis específica 2. . . . 142
4.2.3 Contrastación de hipótesis específica 3. . . . 149

5 CONCLUSIONES **153**
5.1 Conclusiones. . . . 153
5.2 Recomendaciones . . . 155
5.3 Trabajos futuros . . . 156

REFERENCIAS BIBLIOGRÁFICAS **158**

6 ANEXOS **164**

6.1 Panel fotográfico: Recolección de fotografias en campo 164

6.2 Matriz de consistencia y operacionalización 169

Anexo 1: Matriz de consistencia 170

Anexo 2: Matriz de operacionalización de variables 172

LISTA DE FIGURAS

1 Plano de ubicación del área de estudio — tramo PE-3S Ayacucho–Huanta 20
2 Elementos asociados con velocidad elevada 42
3 Estrategias esenciales para evitar siniestros 43
4 Dispositivos de señalización vial y su función en la protección 44
5 Portada del Compendio de Dispositivos de Control del Tránsito Automotor . . . 45
6 Dispositivo vertical de precaución P-49 (zona escolar) 46
7 Posicionamiento apropiado de la señalización vertical 47
8 Posicionamiento longitudinal y distancias de percepción 47
9 Dispositivos de preferencia R-1 y R-2 48
10 Dispositivos de veda de operaciones y giros 49
11 Dispositivos de veda de tránsito por categoría de unidad vehicular 49
12 Dispositivos precautorios por atributos de la superficie de rodadura 50
13 Dispositivos precautorios por atributos operacionales de la calzada 50
14 Señalización para vías de escape 52
15 Paradigma matemático de una neurona artificial 55
16 Configuración de una red neuronal convolucional (CNN) 57
17 Implementación de la función de activación ReLU 60
18 Implementación de un Max Pooling de 2×2 61
19 Ilustración de los estratos de consolidación 63
20 Simbolización de la arquitectura YOLO v8 65
21 Simbolización de la arquitectura Faster-RCNN 65
22 Detección de señalizaciones con la arquitectura YOLO v8 67
23 Detección de señalizaciones con YOLO v8 en variabilidad de condiciones 67
24 Mapa de ubicación del Tramo 1: Corredor vial Ayacucho-Huanta (PE-3S) 77
25 Mapa de ubicación del Tramo 2: Corredor vial Ayacucho-Abra Apacheta (PE-28A) 77
26 Punto de inicio del recorrido: Salida de la ciudad de Ayacucho hacia Huanta . . 78
27 Punto de llegada del Tramo 1: Arco conmemorativo en la ciudad de Huanta . . . 78
28 Punto de máxima altitud del Tramo 2: Abra Apacheta (4,746 msnm) 79
29 Ejemplo de señal de tránsito en estado operativo 83
30 Ejemplo de vandalismo tipo pintura en señal de tránsito 83
31 Ejemplo de vandalismo tipo pegatinas en señal de tránsito 84
32 Ejemplo de señal de tránsito con rotación severa 84
33 Ejemplo de señal de tránsito con rotación moderada 85
34 Ejemplo de señal de tránsito con decoloración 85
35 Ejemplo de señal reglamentaria con decoloración avanzada 86
36 Ejemplo de señal con deterioro del tablero por desprendimiento 86
37 Ejemplo de señal con deterioro del tablero por abolladura 87
38 Ejemplo de señal de tránsito con obstrucción vegetal parcial 87
39 Ejemplo de señal de tránsito con obstrucción vegetal severa 88
40 Ejemplo de señal con deterioro gradual de la pintura del poste 88

41 Etiquetado en LabelMe de señal R-16 en condiciones diurnas 91
42 Etiquetado en LabelMe de R-16 en condiciones nocturnas 91
43 Etiquetado en LabelMe de señal R-37 con georeferenciación 92
44 Señal curva pronunciada a la derecha 92
45 Etiquetado en LabelMe de señal P-2A en condiciones soleadas 92
46 Etiquetado en LabelMe de señal P-2B (curva a la izquierda) 93
47 Etiquetado en LabelMe de señal P-1B (curva pronunciada a la izquierda) 93
48 Etiquetado en LabelMe de señal P-5-2A en condiciones diurnas 93
49 Etiquetado en LabelMe de señal P-5-2A en condiciones nocturnas 94
50 Etiquetado en LabelMe de señal P-5-2B (curva en U a la izquierda) 94
51 Etiquetado en LabelMe de señal P-33A en zona urbana 94
52 Etiquetado en LabelMe de señal P-10A (empalme en ángulo agudo a la derecha) 95
53 Etiquetado en LabelMe de señal P-49 (zona escolar) 95
54 Etiquetado en LabelMe de señal P-48 (zona de presencia de peatones) 95
55 Panel de curvas de entrenamiento del modelo YOLO26s-seg 100
56 Matriz de confusión del modelo YOLO26s-seg en validación 101
57 Curva Precision-Recall del modelo YOLO26s-seg para bounding boxes 102
58 Curva Precision-Recall del modelo YOLO26s-seg para máscaras de segmentación 102
59 Curva Precision-Confidence del modelo YOLO26s-seg 103
60 Curva Recall-Confidence del modelo YOLO26s-seg 104
61 Curva F1-Confidence del modelo YOLO26s-seg 105
62 Curvas de entrenamiento del modelo EfficientNet-B0 (E03) 110
63 Matriz de confusión del modelo EfficientNet-B0 (E03) en test 111
64 Arquitectura del pipeline SISEVIR para deteccion y evaluacion de senales verticales 113
65 Interfaz principal del software SISEVIR 117
66 Dialogo de selección de video para análisis en SISEVIR 118
67 Confirmación de carga de video con especificaciones técnicas 118
68 Lineas verticales de conteo durante el análisis de video 119
69 Deteccion y clasificacion en tiempo real de senal P-1A 119
70 Cuadro de resumen al finalizar el procesamiento del video 120
71 Panel de estadisticas de detección de señales del software SISEVIR 121
72 Panel de analisis estadistico avanzado del sistema SISEVIR 121
73 Formulario de configuracion del Reporte MTC en SISEVIR 122
74 Carga del modelo Qwen2-VL para generación de descripciones en el reporte . . 122
75 Ficha de inspeccion de seguridad vial generada por SISEVIR 123
76 Seccion de analisis estadistico en el Reporte MTC de SISEVIR 124
77 Sección de registro fotográfico del Reporte MTC de SISEVIR 125
78 Carga del video nocturno para validacion del sistema SISEVIR 127
79 Detección nocturna de señal P-5-2B — rendimiento en baja luminosidad 128
80 Detección nocturna de señal P-2B con mascara de segmentacion 128
81 Progresión de detecciones durante el recorrido nocturno (6 señales) 129

82 Detección nocturna de señal P-2A con tráfico en sentido contrario 129
83 Detección simultanea de señales R-16 y P-1A en condiciones nocturnas 130
84 Deteccion nocturna de señal P-33A (proximidad reductor de velocidad) 130
85 Detección nocturna de señal P-5-2B con tráfico vehicular 131
86 Llegada a Huanta durante el recorrido nocturno con 26 señales registradas . . . 131
87 Llegada al Arco de la Memoria de Huanta — punto final del recorrido nocturno . 132
88 Estadisticas de deteccion del video nocturno — tramo PE-28B 132
89 Análisis estadístico avanzado del procesamiento nocturno en SISEVIR 133
90 Configuracion del Reporte MTC para el recorrido nocturno 133
91 Ficha de inspeccion de seguridad vial generada para el recorrido nocturno . . . 134
92 Escala de estado y tipo de intervencion del Reporte MTC nocturno 135
93 Señal R-30 con estado de DECOLORACIÓN 137
94 Señal P-2A con estado de DETERIORO_TABLERO 137
95 Detección múltiple con señales P-5-1 (OPERATIVO) y R-16 (ROTACIÓN) 138
96 Señal P-2B con estado de ROTACIÓN 138
97 Señal P-2B con estado de VANDALISMO 139
98 Señal P-1B con estado de VANDALISMO 139
99 Interfaz principal del software SISEVIR con los tres paneles funcionales 143
100 Sistema de líneas de conteo para registro de señales únicas en SISEVIR 144
101 Retroalimentación visual en tiempo real durante el procesamiento de video . . 145
102 Ficha de Inspección de Seguridad Vial generada automáticamente por SISEVIR . 145
103 Detección y segmentación nocturna de señal P-2B con máscara de instancia . . 146
104 Estadísticas de detección generadas tras el procesamiento del video nocturno . 146
105 Tramo PE-3S Ayacucho–Huanta evaluado en la comparación de métodos 149
106 Inicio de ruta Ayacucho–Huanta 164
107 Señalización curva pronunciada a la izquierda (P-1B) 164
108 Señal empalme en ángulo agudo a la derecha (P-10A) 165
109 Señal curva en U a la derecha (P-5-2A) 165
110 Señal de zona escolar (P-49) 166
111 Señal reguladora R-37 — señal de control 166
112 Señal preventiva animales en vía (P-53) 167
113 Señal preventiva de proximidad de badén (P-34) 167
114 Señal de camino sinuoso a la derecha (P-5-1) en pésimas condiciones 168
115 Señal curva y contracurva pronunciada a la derecha (P-3A) 168
116 Señal proximidad reductor de velocidad tipo resalto (P-33A) 169
117 Fin de la ruta — Arco de la Memoria de Huanta 169

LISTA DE CUADROS

1 Mecanismo de convoluciones a una imagen de 48x48 píxeles 62
2 Puntos de control georeferenciados del trabajo de campo 79
3 Características técnicas de los tramos viales estudiados 79
4 Resumen del volumen de datos recopilados por tramo vial 81
5 Definición de categorías (clases) para el modelo de detección YOLOv8n-seg . . . 90
6 Hiperparámetros de entrenamiento del modelo YOLO26s-seg 96
7 Técnicas de aumentación online durante el entrenamiento 97
8 Evolución de métricas durante el entrenamiento del modelo YOLO26s-seg 98
9 Hitos de convergencia del modelo YOLO26s-seg 99
10 Métricas de rendimiento del modelo YOLO26s-seg en validación 99
11 Resumen de métricas de rendimiento del modelo YOLO26s-seg 106
12 Especificaciones de la arquitectura EfficientNet-B0 106
13 Configuración de las etapas de entrenamiento por transferencia 107
14 Hiperparámetros de entrenamiento del modelo EfficientNet-B0 107
15 Distribución del dataset de clasificación por estado de conservación 108
16 Técnicas de aumentación online (Capa 1) 108
17 Factores de aumentación offline por clase (Capa 2) 109
18 Técnicas de aumentación específica por clase (Capa 3) 109
19 Resultados del estudio de ablación para EfficientNet-B0 110
20 Métricas de rendimiento por clase del modelo EfficientNet-B0 (E03) 112
21 Métricas globales del modelo EfficientNet-B0 (E03) 112
22 Especificaciones del modelo Qwen2-VL-2B-Instruct 114
23 Prompts dirigidos por estado de conservacion para Qwen2-VL (modo real-time) 115
24 Parametros de generacion del modelo Qwen2-VL segun modo operativo 115
25 Tiempos de inferencia del modelo Qwen2-VL 116
26 Ejemplos de descripciones generadas por Qwen2-VL en el pipeline SISEVIR . . . 116
27 Flujo de trabajo completo del software SISEVIR 126
28 Comparativo de especificaciones entre video diurno y nocturno 127
29 Criterios de aceptación para la hipótesis específica 1 136
30 Métricas de rendimiento del modelo YOLO26s-seg en el conjunto de test 140
31 Métricas globales del modelo EfficientNet-B0 (E03) en el conjunto de test 141
32 Métricas de rendimiento por clase del modelo EfficientNet-B0 (E03) en test . . . 141
33 Criterios de aceptación para la hipótesis específica 2 142
34 Comparativo de eficiencia: método manual vs. sistema SISEVIR 147
35 Indicadores de eficiencia verificados del software SISEVIR 148
36 Comparativo de tiempos: Método manual vs. sistema SISEVIR — tramo PE-3S . . 151
37 Comparativo de costos operativos: Método manual vs. sistema SISEVIR — tramo PE-3S 151
38 Síntesis de indicadores de eficiencia: Método manual vs. sistema SISEVIR 152
39 Matriz de consistencia 171

40 Matriz de operacionalización de variables 173

TESIS

1. PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción del problema

La mortalidad asociada a eventos viales representa actualmente una de las principales amenazas para la salud pública a escala mundial. Datos expuestos por la (OMS (2023)) indican que estos siniestros ocasionan aproximadamente 1.19 millones de fallecimientos cada año, a lo que se suman entre 20 y 50 millones de personas que sufren traumatismos de diversa gravedad , muchos de los cuales resultan en discapacidades permanentes. Los factores causales más recurrentes incluyen velocidades excesivas, conducción bajo influencia de alcohol o drogas, distracciones al volante y, de manera crítica, la inadecuada percepción o interpretación de la señalización vial por parte de los usuarios. Esta problemática demanda un abordaje multidisciplinario que incorpore soluciones tecnológicas innovadoras, considerando que la interacción efectiva entre conductores y sistemas de advertencia visual continúa siendo deficiente, especialmente en contextos de variabilidad climática o situaciones de peligro latente.

El contexto peruano presenta particularidades que agravan significativamente la problemática del tránsito vehicular. Durante el año 2023, estadísticas oficiales reportaron más de 87,172 accidentes viales que resultaron en más de 3,138 víctimas fatales, consolidando al territorio nacional como uno de los más afectados en la región sudamericana (Highways (2024)). Estos eventos trágicos son atribuibles directamente al factor humano, manifestándose en comportamientos como imprudencia, incumplimiento de normativas de tránsito y velocidades inadecuadas. A pesar de que la infraestructura vial ha experimentado mejoras graduales, el sistema vial peruano continúa presentando deficiencias estructurales significativas, particularmente en lo referente a señalización vertical con adecuados niveles de visibilidad, claridad informativa y cumplimiento de estándares técnicos vigentes. Paralelamente, existe una carencia notoria de sistemas tecnológicos integrados que proporcionen asistencia en tiempo real para mejorar la percepción del entorno vial por parte de los conductores. Esta situación genera una exposición constante al riesgo para usuarios de las vías, evidenciando la ausencia de políticas públicas efectivas que impulsen la adopción de herramientas tecnológicas fundamentadas en inteligencia artificial para la prevención y mitigación de accidentes de tránsito.

El departamento de Ayacucho, caracterizado por su topografía predominantemente montañosa y la dispersión geográfica de sus centros poblados en áreas rurales de difícil accesibilidad, exhibe un panorama particularmente crítico en materia de seguridad vial. Las carreteras interprovinciales e interregionales que atraviesan este territorio presentan condiciones geográficas de alto riesgo, incluyendo trazados con curvas cerradas y pronunciadas, pendientes considerables, tramos con escasa o nula iluminación artificial, y señalización que frecuentemente se encuentra en estados de deterioro avanzado o

simplemente inexistente.

El corredor vial Ayacucho-Huanta, que forma parte de la Ruta Nacional PE-3S y se extiende por aproximadamente 43.9 kilómetros, constituye una arteria de vital importancia para la conectividad regional, facilitando el transporte de pasajeros y mercancías entre ambas provincias. Sin embargo, este tramo presenta características que incrementan significativamente el riesgo de siniestralidad: Geometría vial compleja con múltiples curvas y contracurvas consecutivas, zonas susceptibles a deslizamientos de material rocoso, presencia frecuente de animales domésticos y silvestres cruzando la calzada, señalización vertical insuficiente o en condiciones inadecuadas de conservación.

El estado situacional de la señalización vertical constituye un factor determinante en la prevención de accidentes de tránsito, dado que estas señales representan el principal medio de comunicación entre la infraestructura vial y los conductores. Una señalización en óptimas condiciones permite anticipar situaciones de riesgo, regular el comportamiento vehicular y orientar la toma de decisiones durante la conducción. Por el contrario, cuando las señales presentan deterioro físico como decoloración, oxidación, vandalismo o pérdida de retrorreflectancia, su función preventiva se ve severamente comprometida. Del mismo modo, la ausencia de señales en puntos críticos o con ubicación incorrecta genera vacíos informativos que impiden al conductor reconocer oportunamente las condiciones peligrosas del tramo. Esta problemática se evidencia en el diagnóstico del PDU (2020), donde se establece que ”La ciudad de Ayacucho no cuenta con una señalización adecuada, solo existen algunas señales principalmente en el centro de la ciudad, la ciudad necesita con urgencia la implementación de un plan de señalización”, situación que se replica y agrava en las vías interurbanas y rurales del departamento donde la fiscalización y monitoreo de la señalización vertical es prácticamente inexistente.

Los métodos convencionales de inventario y evaluación del estado situacional de la señalización vial presentan limitaciones significativas que comprometen la eficiencia y confiabilidad de los datos obtenidos. Tradicionalmente, estos procesos se realizan de manera manual mediante inspecciones de campo que requieren el desplazamiento de personal técnico especializado a lo largo de extensos tramos carreteros, demandando considerable inversión de tiempo, recursos humanos y presupuesto institucional. El registro manual de información en fichas técnicas está sujeto a errores de percepción, transcripción y subjetividad del inspector, generando inconsistencias en la base de datos que afectan la toma de decisiones para el mantenimiento vial. Adicionalmente, la actualización periódica del inventario mediante metodologías tradicionales resulta económicamente inviable para las entidades competentes, provocando desactualización progresiva de la información y desconocimiento del estado real de la infraestructura de señalización.

1.2. Delimitación del problema

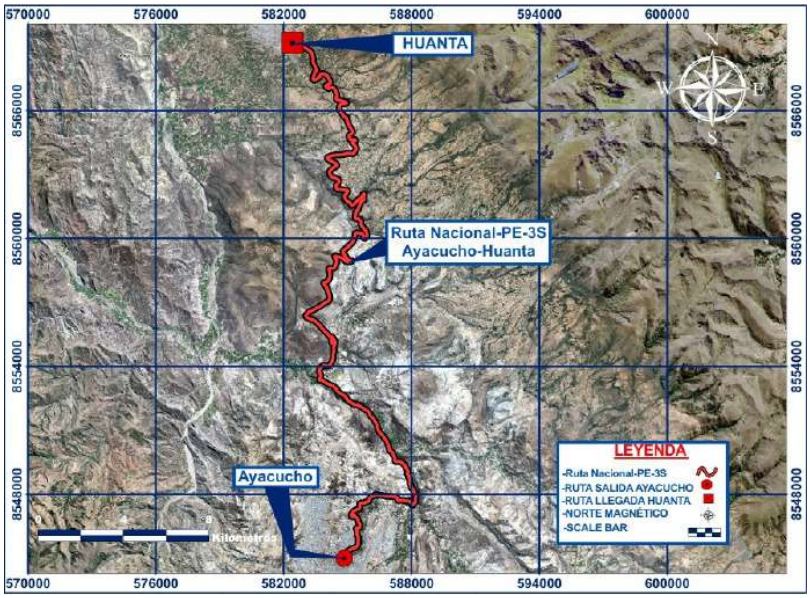
1.2.1. Espacial (geográfica)

El área de investigación se encuentra localizada en el tramo vial Ayacucho–Huanta, correspondiente a la Ruta Nacional PE-3S, que forma parte de la red vial nacional del Perú, ubicada en el departamento y región de Ayacucho, comprendiendo las provincias de Huamanga y Huanta. El tramo en estudio tiene una longitud total de 44.02 km, iniciando en la ciudad de Ayacucho (Provincia de Huamanga) a una altitud aproximada de 2,746 msnm y finalizando en la ciudad de Huanta (Provincia de Huanta) a 2,628 msnm.

Este corredor vial fue seleccionado debido a su importancia estratégica como vía de conexión entre dos ciudades principales de la región, presentando un flujo vehicular considerable de transporte público, privado y de carga, además de contar con una diversidad representativa de señalización de tránsito (reglamentarias, preventivas e informativas) que permite la validación efectiva del sistema de visión artificial propuesto para la detección y evaluación del estado de las señales de tránsito, contribuyendo así a la mejora de la seguridad vial en esta importante ruta nacional.

La ejecución de la investigación se realizó en:

Figura 1
Plano de ubicación del área de estudio — tramo PE-3S Ayacucho–Huanta



Nota. Elaboración propia mediante ArcGIS.

1.2.2. Temporal

La presente investigación se desarrolló temporalmente entre los meses de octubre de 2024 y octubre de 2025, periodo durante el cual se llevaron a cabo las diferentes etapas del estudio. Durante esta temporalidad se realizó la recolección de datos mediante recorridos vehiculares en los tramos Ayacucho–Huanta (PE-3S) y Ayacucho–Abra Apacheta (PE-28A), donde se capturaron fotografías y videos de las 31 clases de señales de tránsito verticales

que detecta el sistema propuesto, definidas según las categorías establecidas en el Manual de Dispositivos de Control del Tránsito Automotor (2024).

Posteriormente se ejecutó el procesamiento y análisis de las imágenes recopiladas, la construcción del dataset de 6,213 imágenes con 7,107 instancias anotadas, el entrenamiento de los modelos de visión artificial (YOLO26s-seg y EfficientNet-B0) mediante técnicas de aprendizaje profundo, la validación y ajuste de parámetros del sistema, y finalmente el desarrollo e implementación del software SISEVIR para la detección, clasificación y evaluación automatizada del estado de conservación de las señales de tránsito verticales, concluyendo con la interpretación de los resultados obtenidos y la elaboración del presente documento de tesis.

1.2.3. Temática y unidad de análisis

- **Temática**: La presente investigación se enmarca en el campo de la seguridad vial y la ingeniería del transporte, con énfasis en el uso de tecnologías emergentes aplicadas al control y gestión del tránsito vehicular. En específico, se aborda la implementación de técnicas de inteligencia artificial mediante algoritmos de visión por computadora y aprendizaje profundo, orientadas a la detección, segmentación, clasificación y evaluación automatizada del estado de conservación de las señalizaciones verticales de tránsito en los tramos PE-3S y PE-28A de la región Ayacucho. El desarrollo del sistema SISEVIR busca contribuir a la mejora de la seguridad vial mediante la identificación oportuna de señales deterioradas, con rotación, vandalizadas u obstruidas, permitiendo una gestión más eficiente del mantenimiento vial y la reducción de potenciales riesgos para los usuarios, integrando los campos de la ingeniería civil, la inteligencia artificial y la gestión de infraestructura vial.
- **Unidad de análisis**: La unidad de análisis de esta investigación corresponde a las señalizaciones verticales de tránsito presentes en los tramos viales PE-3S (Ayacucho–Huanta, 44.02 km) y PE-28A (Ayacucho–Abra Apacheta, ∼95.4 km), clasificadas según el Manual de Dispositivos de Control del Tránsito Automotor (2024). Se analizan **31 clases** de señales que incluyen señales reguladoras (R-15, R-16, R-30, R-37, entre otras) y señales preventivas (P-1A, P-1B, P-2A, P-2B, P-3A, P-4B, P-5-1, P-5-2A, P-5-2B, P-10A, P-33A, P-34, P-49, P-53, entre otras), las cuales fueron capturadas mediante videograbación continua durante recorridos de campo en condiciones reales de operación diurna y nocturna. Estas imágenes y videos constituyen el dataset que alimenta los modelos de visión artificial para su procesamiento, entrenamiento y validación, permitiendo evaluar la capacidad del sistema para detectar, segmentar y clasificar cada señal en siete estados de conservación (OPERATIVO, DETERIORO_POSTE, ROTACIÓN, OBSTRUCCIÓN, DECOLORACIÓN, VANDALISMO y DETERIORO_TABLERO), constituyendo así la base empírica para la propuesta de mejora en la gestión de la señalización vial mediante inteligencia artificial.

1.3. Formulación del problema

1.3.1. Problema general

¿Como se puede desarrollar un modelo de visión artificial basado en redes neuronales para automatizar la clasificación y evaluación del estado de deterioro de las señales de tránsito en las vías nacionales de Huamanga?

1.3.2. Problemas específicos

1. ¿En qué medida el modelo de visión artificial basado en redes neuronales logra la clasificación y evaluación del estado de deterioro de las señales de tránsito?
2. ¿Qué características debe incluir un programa automatizado basado en visión artificial para realizar eficientemente la clasificación y evaluación del estado de deterioro de las señales de tránsito en las vías nacionales?
3. ¿Cómo se comparan en términos de tiempo y costo la clasificación y evaluación del estado de deterioro de las señales de tránsito mediante visión artificial frente a métodos tradicionales de inspección manual en la carretera Ayacucho-Huanta en 2024?

1.4. Justificación e importancia

1.4.1. Justificación

El corredor vial Ayacucho-Huanta (PE-3S) constituye una arteria fundamental para la conectividad regional del departamento de Ayacucho; sin embargo, el estado situacional de su señalización vertical representa una problemática crítica que incide directamente en la seguridad vial. La realidad observada en el tramo de estudio evidencia señales con decoloración avanzada, estructuras dañadas por factores climáticos y vandálicos, ubicaciones inadecuadas que impiden su visualización oportuna, y ausencia de señalización en puntos críticos donde resulta indispensable advertir sobre condiciones peligrosas. Esta situación se agrava debido a las limitaciones de los métodos convencionales de inventario, que dependen de inspecciones manuales costosas, subjetivas y de difícil actualización periódica, generando desconocimiento del estado real de la infraestructura por parte de las entidades competentes. Ante esta problemática, la presente investigación se justifica por la necesidad de implementar soluciones tecnológicas que superen las limitaciones de los enfoques tradicionales. La aplicación de redes neuronales convolucionales (CNN) y algoritmos de visión por computadora permite automatizar la detección, clasificación y evaluación del estado físico de las señalizaciones de manera objetiva, sistemática y eficiente. Este enfoque representa una alternativa escalable y de bajo costo que facilita la identificación oportuna de señales deterioradas o ausentes, optimizando la toma de decisiones para el mantenimiento vial y contribuyendo a la reducción de riesgos de accidentalidad asociados al deficiente estado situacional de la señalización vertical.

1.4.2. Importancia

La presente investigación aporta valor tanto a nivel tecnológico, académico como social, constituyendo un aporte significativo para la ingeniería civil y la gestión de infraestructura vial en el contexto regional y nacional. Desde el ámbito técnico, representa una contribución al desarrollo de sistemas inteligentes de asistencia a la conducción, integrando visión computacional, procesamiento digital de imágenes mediante aprendizaje profundo y generación de información georreferenciada en tiempo real, aplicables no solo a la evaluación de señalización sino también a otros aspectos de la gestión vial. A nivel práctico, la solución propuesta puede ser implementada como parte de sistemas de apoyo en vehículos de inspección vial, unidades de transporte público, plataformas de monitoreo urbano o dispositivos móviles, promoviendo una conducción más segura y consciente.

1.5. Limitaciones de la investigación

A pesar de que esta investigación está bien diseñada y utiliza metodologías avanzadas, enfrenta ciertas limitaciones que deben ser reconocidas:

- **Limitaciones tecnológicas:** La implementación del sistema SISEVIR requiere de dispositivos con capacidad de procesamiento gráfico adecuado (GPU) para alcanzar velocidades de inferencia en tiempo real. Los resultados óptimos obtenidos (476 FPS, 2.1 ms/imagen) corresponden a una GPU NVIDIA de alto rendimiento, recurso que puede no estar disponible de manera generalizada en las unidades de inspección vial de las entidades gestoras de carreteras en la región Ayacucho, limitando así su despliegue inmediato a gran escala sin una inversión previa en infraestructura tecnológica.
- **Limitaciones de implementación en hardware embebido:** El sistema SISEVIR desarrollado opera únicamente sobre computadoras de escritorio o laptops con GPU dedicada, sin contemplar la implementación en dispositivos embebidos especializados como NVIDIA Jetson Orin u otras plataformas de edge computing que permitirían el procesamiento en tiempo real directamente en vehículos en movimiento. Esta limitación restringe la aplicabilidad práctica del sistema a una modalidad de procesamiento diferido (grabación en campo y análisis posterior en computadora), reduciendo su potencial como herramienta de monitoreo dinámico continuo. La implementación en hardware embebido queda planteada como línea de trabajo futuro prioritaria.
- **Enfoque en un contexto geográfico específico:** Si bien la investigación cubre dos tramos viales representativos de la región Ayacucho — PE-3S (Ayacucho–Huanta, 44.02 km) y PE-28A (Ayacucho–Abra Apacheta, ∼95.4 km) —, sus resultados, conclusiones y propuestas están contextualizados a las características particulares de la sierra ayacuchana. En consecuencia, el modelo entrenado puede requerir procesos de reentrenamiento o adaptación para su implementación en otras regiones del país con dinámicas viales diferentes, condiciones climáticas distintas (costa, selva) o señalización con características particulares que no estén representadas en el dataset

construido.

- **Alcance del sistema de detección:** El sistema SISEVIR se enfoca en la detección, segmentación y clasificación del estado de conservación de **31 categorías** de señalizaciones verticales de tránsito (series P-* y R-*) conforme al Manual de Dispositivos de Control del Tránsito Automotor 2024 del MTC. No obstante, no incluye la detección ni el análisis de otros elementos viales igualmente importantes para la seguridad del tránsito, como marcas en el pavimento (señalización horizontal), semáforos, guardavías, reductores de velocidad ni dispositivos de seguridad complementarios, limitando el alcance de las auditorías viales generadas a la señalización vertical exclusivamente.
- **Limitaciones del clasificador en clases minoritarias:** El clasificador EfficientNet-B0 presenta un rendimiento inferior en la clase DECOLORACIÓN (F1 = 0.6154) debido a la similitud visual con señales en estado OPERATIVO con desgaste leve, así como al reducido número de muestras disponibles para esta categoría en el dataset (14 muestras en test). Esta limitación es inherente al desbalance extremo del dataset de clasificación (ratio OPERATIVO/DETERIORO_TABLERO de 150:1) y representa una oportunidad de mejora mediante la recolección de mayor cantidad de imágenes de señales con deterioros sutiles en trabajos futuros.

1.6. Objetivos de la investigación

1.6.1. Objetivo general

Desarrollar un modelo de visión artificial basado en vision artificial que permita automatizar la clasificación y evaluación del estado de deterioro de las señales de tránsito en las vías nacionales de Huamanga, con el fin de mejorar la eficiencia en el monitoreo y mantenimiento de la infraestructura vial.

1.6.2. Objetivos específicos

1. Evaluar la precisión del modelo de visión artificial en la identificación y clasificación del estado de conservación de las señales de tránsito, explorando diferentes arquitecturas de red neuronal y técnicas de optimización.
2. Desarrollar un software automatizado que integre el modelo de visión artificial y permita la clasificación y evaluación del estado de deterioro de las señales de tránsito en las vías nacionales.
3. Realizar un análisis comparativo que evalúe la eficiencia en términos de tiempo y costo de la evaluación automatizada de señales de tránsito, utilizando visión artificial en comparación con los métodos tradicionales de inspección manual, específicamente en la carretera Ayacucho-Huanta durante el año 2024.

2. MARCO TEÓRICO

2.1. Antecedentes

2.1.1. Investigaciones internacionales

Zhang et al. (2025) desarrollaron un algoritmo mejorado de detección de señales de tráfico denominado YOLO-BS, basado en la arquitectura YOLOv8 (You Only Look Once versión 8). En su investigación, los autores abordaron las problemáticas de baja precisión y bajo rendimiento en tiempo real de los métodos tradicionales de detección en entornos viales dinámicos. Para ello, incorporaron una capa de detección de objetos pequeños dentro del framework YOLOv8 con el objetivo de enriquecer la extracción de características. Adicionalmente, integraron una red piramidal de características bidireccional (BiFPN) al sistema de detección para optimizar el manejo de objetos multiescala y mejorar sustancialmente el rendimiento en la detección de objetivos de tamaño reducido. Los experimentos realizados sobre el conjunto de datos TT100K consolidaron métricas clave como tamaño del modelo, recall, precisión media promedio (mAP) y frames por segundo (FPS), demostrando que el algoritmo YOLO-BS supera significativamente los enfoques tradicionales en la detección de señales de tráfico con un mAP de 90.1 % y 78 FPS, aplicable en sistemas de transporte inteligente y vehículos autónomos.

Wang, Zheng y Li (2025) propusieron un framework optimizado basado en YOLOv5 para el reconocimiento de señales de tráfico en sistemas de transporte inteligente. En su investigación, los autores abordaron desafíos críticos como la detección de objetivos a pequeña escala, la variabilidad ambiental y las restricciones de procesamiento en tiempo real. El estudio introdujo tres innovaciones clave: primero, implementaron el algoritmo de clustering k-means++ para la optimización de anchor boxes, logrando un IoU promedio del 77.55 % en comparación con el 75.95 % del k-means tradicional; segundo, realizaron un análisis comparativo exhaustivo de las variantes YOLOv5 (s/m/x), revelando compromisos entre precisión y velocidad (99.3-99.5 % mAP@0.5 versus tiempos de inferencia de 32-45 ms); y tercero, desarrollaron un ajuste sistemático de hiperparámetros para maximizar la robustez en diversos escenarios. Los experimentos sobre el conjunto de datos CCTSDB (13,830 imágenes anotadas) demostraron la superioridad del framework propuesto, alcanzando 98.1 % de mAP, 98.6 % de recall y 99.3 % de precisión, superando a Faster-RCNN y SSD en 5-8 % en mAP mientras mantenía 45 FPS, validando su aplicabilidad en sistemas de transporte inteligente sensibles a la latencia

Qiu et al. (2025) propusieron DP-YOLO, un algoritmo ligero de detección de señales de tráfico basado en YOLOv8s, diseñado específicamente para abordar las dificultades en el reconocimiento de objetos pequeños que frecuentemente resultan en detecciones perdidas o falsas. En su investigación, los autores mejoraron la precisión de detección de objetos pequeños y redujeron significativamente el conteo de parámetros del modelo mediante tres innovaciones principales: primero, eliminaron la capa de detección de objetos grandes del modelo base e incorporaron una capa de detección de objetos

pequeños; segundo, diseñaron el módulo DBBNCSPELAN4 para potenciar la capacidad de extracción de características de la red, y propusieron el módulo PTCSP que integra tecnología Transformer en la red de procesamiento de características del modelo, reduciendo simultáneamente parámetros y costo computacional; y tercero, introdujeron la función de pérdida W3F MPDIoU para mitigar el impacto de muestras de baja calidad en el modelo y mejorar su robustez. Los experimentos demostraron que, en comparación con YOLOv8s, DP-YOLO reduce el conteo de parámetros del modelo en 77.0/ %, mientras logra mejoras en mAP0.5 de 5.8 % en el conjunto de datos TT100K, 2.7 % en GTSDB y 1.3 % en CCTSDB, evidenciando alto potencial para despliegue en dispositivos edge y mejorando efectivamente la capacidad de detección de señales de tráfico de tamaño reducido.

Bai et al. (2025) desarrollaron el conjunto de datos Glare Traffic Sign Dataset (GTSD) y propusieron el modelo YOLO-SFM para la detección de señales de tráfico bajo condiciones de deslumbramiento, abordando una problemática crítica en la seguridad vial donde la luz solar intensa compromete la visión de los conductores e incrementa la probabilidad de accidentes. En su investigación, los autores identificaron que los conjuntos de datos existentes carecían de imágenes de señales de tráfico bajo condiciones de deslumbramiento, evidenciando la necesidad urgente de un dataset especializado. El GTSD desarrollado incluye anotaciones exhaustivas de deslumbramiento solar y abarca diversas condiciones de brillo encontradas en la conducción cotidiana. El modelo YOLO-SFM propuesto utiliza downsampling SPD-Conv para minimizar la pérdida de información y mejorar la precisión en la extracción de características. Adicionalmente, el modelo integra módulos MLCA y FasterNet para fusionar características locales y globales, así como información de canales y espacial, reduciendo simultáneamente cálculos redundantes y acceso a memoria mientras mejora la precisión y velocidad de detección. Los experimentos demostraron que, en comparación con YOLOv8s, YOLO-SFM mejora el mAP en 0.5 % en el conjunto de datos GTSD y 1.3 % en el conjunto de datos CCTSDB, mientras también reduce parámetros en 5 M y logra un umbral de deslumbramiento de 11.8 %, validando su efectividad para mejorar la capacidad de detección de señales de tráfico bajo condiciones de resplandor solar adversas en sistemas de transporte inteligente.

Yang y Han (2025) desarrollaron ETS-YOLO, un modelo eficiente basado en YOLO para el reconocimiento de señales de tráfico en tiempo real, abordando desafíos críticos como la susceptibilidad a interferencias, dificultad en detectar objetivos pequeños y el compromiso entre rendimiento temporal y precisión. Los autores construyeron el modelo sobre YOLOv5, implementando Partial Convolution (PConv) para rediseñar el módulo Cross Stage Partial bottleneck con 3 capas convolucionales (C3), resultando en el módulo C3Efficient que reduce cálculos redundantes y accesos a memoria. Adicionalmente, incorporaron Dynamic Upsampler (DySample) para mejorar el sobremuestreo durante la etapa Feature Pyramid Network (FPN) y diseñaron la función de pérdida Normalized Wasserstein Distance (NWD) para optimizar el mecanismo de pérdida posicional de anclas, mejorando la detección de objetos diminutos. El modelo integró el algoritmo Soft-NMS que facilita el filtrado de anclas, incrementando notablemente la precisión en objetivos ocluidos adyacentes. Los resultados

experimentales demostraron que ETS-YOLO alcanzó un mAP 0.5 de 0.803 en 45 tipos de señales del conjunto TT100K, representando una mejora de 0.052 comparado con el modelo base, mientras redujo parámetros en 16.0 % (6.0 millones) y carga computacional en 12.8 % (14.3 GFLOPs), logrando un tiempo de inferencia de 26.3 ms con disminución del 10 % en latencia, consolidando un balance favorable entre eficiencia y precisión para sistemas de reconocimiento en tiempo real.

Xing et al. (2025) presentaron FEBG-YOLOv8s, un modelo ligero para detección de señales de tráfico que integra FasterNet, Efficient Multi-scale Attention (EMA), Bidirectional Feature Pyramid Network (BiFPN) y Group Separable Convolution (GSConv) basado en YOLOv8s, abordando el desafío crítico de balancear precisión, eficiencia y robustez bajo restricciones computacionales de dispositivos edge montados en vehículos. Los autores reconstruyeron el módulo Cross Stage Partial Network 2 with Focus (C2f) utilizando bloques FasterNet para minimizar computación redundante, integraron el mecanismo EMA para mejorar robustez contra objetivos pequeños y ocluidos, refinaron la red neck basándose en BiFPN mediante compresión de canales, capas de downsampling y conexiones skip para optimizar la fusión semántica shallow-deep, y diseñaron un cabezal de detección híbrido serial-paralelo basado en GSConv (GSP-Detect) para preservar información cross-channel mientras reduce carga computacional. Los experimentos en Tsinghua-Tencent 100K (TT100K) demostraron que FEBG-YOLOv8s mejora el mean Average Precision at Intersection over Union 0.5 (mAP50) en 3.1 % comparado con YOLOv8s, con 4 millones menos de parámetros y 22.5 % menor Giga Floating-Point Operations (GFLOPs). Los experimentos de generalizabilidad en CSUST Chinese Traffic Sign Detection Benchmark (CCTSDB) validaron robustez con 3.3 % mayor mAP50, consolidando su potencial para detección de señales de tráfico en tiempo real en plataformas edge de conducción autónoma bajo entornos complejos.

Flores-Calero et al. (2024) realizaron una revisión sistemática de literatura (SLR) analizando el algoritmo de detección de objetos YOLO aplicado a sistemas de detección y reconocimiento de señales de tráfico desde cinco aspectos relevantes: aplicaciones, conjuntos de datos, métricas, hardware y desafíos metodológicos. Los autores examinaron 115 estudios primarios publicados entre 2016 y 2022, identificando que las aplicaciones más comunes de YOLO en este campo son seguridad vehicular y vehículos inteligentes autónomos. La investigación determinó que la mayoría de conjuntos de datos utilizados provienen de Alemania y China, siendo validados con métricas sofisticadas de detección, clasificación y velocidad de procesamiento para sistemas de reconocimiento de señales basados en YOLO. Adicionalmente, establecieron que las plataformas de hardware predominantes para procesamiento de datos son Nvidia RTX 2080 y Titan Tesla V100, mientras que para dispositivos embebidos o GPU móviles se identificaron Jetson Nano y Raspberry Pi. El análisis consolidó que siete problemas principales persisten al operar en condiciones reales de tráfico, incluyendo detección en fondos complejos, objetivos pequeños y ocluidos, y procesamiento en tiempo real bajo restricciones computacionales. Los resultados evidenciaron que esta SLR constituye el trabajo más relevante y reciente sobre desarrollo tecnológico aplicado a detección y reconocimiento de señales de tráfico

usando YOLO, proporcionando insights valiosos sobre direcciones futuras de investigación que podrían conducirse para mejorar significativamente el campo.

Wu et al. (2025) propusieron un sistema ligero de detección de señales de tráfico basado en YOLO con despliegue móvil exitoso. Los autores integraron la estructura Classification to Fusion (C2f) en la red backbone, empleando convolución deformable y Bi-directional Feature Pyramid Network (BiFPN) Concat para mejorar la adaptabilidad de la red. Adicionalmente, incorporaron el módulo de atención SimAm para priorizar características clave y reducir la complejidad del modelo después de la etapa C2f en el backbone, eliminando simultáneamente el módulo SPPF en el cuello para disminuir parámetros redundantes. Los autores diseñaron el módulo EIoU para ajustar dinámicamente pesos de muestras desafiantes, permitiendo al modelo concentrarse en objetivos difíciles de aprender, mejorando así la robustez y generalización. El sistema fue desplegado exitosamente en aplicaciones móviles mediante tecnología de compilación cross-platform, logrando resultados experimentales que demuestran un F1 score de 0.8987, mean Average Precision (mAP@0.5) de 98.8 % y velocidad de detección de 50 frames por segundo (FPS) en dispositivos móviles. Los resultados consolidaron que el modelo alcanza excelente balance entre precisión y eficiencia, con capacidades de despliegue en dispositivos edge, validando su aplicabilidad para sistemas de reconocimiento de señales de tráfico en tiempo real en plataformas móviles bajo condiciones reales de conducción.

Tao et al. (2025) presentaron CDFF-YOLO (Cross-Dimensional and Dual-Domain Feature Fusion-You Only Look Once), una arquitectura de red innovadora que integra múltiples módulos para superar dificultades en la extracción de información de objetos pequeños en escenarios de detección y reconocimiento de señales de tráfico. Los autores incorporaron el módulo Multi-dimension Spatial Information Fusion en la red para extraer secuencias de características en diferentes dimensiones mediante superposición, abordando la pérdida de información detallada de objetos pequeños. Adicionalmente, embebieron el módulo Multi-branch Perceptual Attention en el módulo C2f para capturar información de características y mejorar el intercambio de información global-local. Para resolver problemas de iluminación desigual y oclusión en escenas de detección, implementaron el módulo PPF que transforma y fusiona información de características extraídas desde dominios de frecuencia y espacial en convoluciones Space-to-depth, mejorando así la capacidad de reconstrucción y fusión de características de objetos pequeños en dominios duales. Los resultados experimentales en el conjunto TT100K demostraron que el algoritmo mejorado exhibe incrementos de 3.7 % en mAP@0.5, 4.8 % en mAP@0.5:0.95 y 4.5 % y 1.7 % en precisión media y recall promedio para objetos pequeños respectivamente, manteniendo velocidad de detección de 50 frames por segundo. Los experimentos adicionales en CCTSDB evidenciaron que el algoritmo mejorado funciona eficazmente, consolidando la capacidad del modelo CDFF-YOLO para mejorar notablemente la eficacia en detección de señales de tráfico mientras mantiene velocidad óptima de detección bajo condiciones complejas de entornos reales.

He et al. (2025) propusieron NTS-YOLO, un método de detección nocturna de señales de tráfico basado en YOLOv5 mejorado. Los autores implementaron tres mejoras principales: primero, adoptaron un método de mejora de imágenes nighttime sin supervisión para mejorar la calidad de imagen en condiciones de baja iluminación; segundo, introdujeron el mecanismo Convolutional Block Attention Module (CBAM) que se enfoca en la forma de las señales de tráfico ponderando las características del canal y espaciales dentro del modelo para mejorar la percepción en fondos complejos y condiciones de iluminación desigual; y finalmente, aplicaron la función Optimal Transport Assignment (OTA) para optimizar la precisión en la predicción de cuadros delimitadores minimizando las diferencias entre distribuciones de probabilidad. Los resultados experimentales demostraron que NTS-YOLO logró una mejora significativa en el reconocimiento de señales de tráfico nocturnas, con incrementos promedio de precisión de 0.96 % para la detección objetivo de señales de tráfico y 0.17 % para la segmentación de instancias.

Lin et al. (2025) presentaron YOLO-LLTS, un algoritmo de detección de señales de tráfico en tiempo real de extremo a extremo diseñado específicamente para entornos de baja iluminación. Los autores incorporaron tres contribuciones principales: el módulo High-Resolution Feature Map for Small Object Detection (HRFM-SOI) que retiene más información sobre objetos distantes o pequeños en comparación con métodos tradicionales; el módulo Multibranch Feature Interaction Attention (MFIA) que interactúa características de diferentes campos receptivos para mejorar la extracción de información mediante el uso de características guiadas por prioridades; y el módulo Prior-Guided Feature Enhancement (PGFE) que mejora la precisión de detección al incrementar el brillo, los bordes, el contraste y complementar información detallada. Adicionalmente, construyeron un nuevo conjunto de datos, Chinese Nighttime Traffic Sign Sample Set (CNTSS), cubriendo diversos escenarios nocturnos. Los experimentos demostraron que YOLO-LLTS alcanzó un rendimiento de vanguardia, superando los mejores métodos previos con incrementos de 2.7 % en mAP50 y 1.6 % en mAP50:95 en TT100K-night, 1.1 % en mAP50 y 1.6 % en mAP50:95 en CNTSS, 7.5 % en mAP50 y 9.8 % en mAP50:95 en GTSDB-night, y resultados superiores en CCTSDB2021.

Quang et al. (2025) realizaron un análisis comparativo de los modelos de detección de objetos YOLOv5 y YOLOv7 para el reconocimiento de señales de tráfico de límite de velocidad. Los autores evaluaron tres aspectos: precisión, diferencia de velocidad y complejidad computacional. Ambos modelos fueron entrenados con el mismo conjunto de datos que consistía en 5206 imágenes representativas de 7 diferentes señales de límite de velocidad, incluyendo 60, 70, 80, 90, 100 y 120 km/h, 2 señales de entrada a área residencial y salida, y 1 señal de eliminación de todos los límites. Las imágenes fueron recolectadas en diversas condiciones ambientales de la vida real como luz del día, nocturnas, lluvia y desenfoque de movimiento desde una cámara Raspberry Pi V2 8MP con resolución de 1280x2464 píxeles y longitud focal de 3.04 mm. Las comparaciones se realizaron en diferentes escenarios de calidad de imagen, distancia desde las cámaras y recursos computacionales. Los resultados revelaron que el rendimiento disminuyó

significativamente en cuanto a la calidad de imagen con YOLOv5, especialmente en imágenes con poca luz, borrosas y mal clima, mientras que YOLOv7 mantuvo mejor rendimiento en diversas condiciones ambientales. En términos de velocidad/tiempo de procesamiento, cuando se ejecutó en un sistema con altos recursos como Google Colab T4, YOLOv7 superó a YOLOv5 en velocidad/tiempo de inferencia, mientras que en recursos bajos YOLOv7 también fue ligeramente mejor que YOLOv5. Finalmente, YOLOv7 pudo mantener mejor rendimiento que YOLOv5 cuando el sujeto estaba a diferentes distancias de las cámaras.

Pekki et al. (2025) implementaron YOLOv8 como tecnología principal para la detección avanzada de objetos en medidas de seguridad vial mediante baches y señales de tráfico utilizando aprendizaje profundo. Los autores desarrollaron un sistema de operación en tiempo real que utiliza cámaras como entradas para detectar con precisión baches y localizar sus posiciones, además de reconocer múltiples señales de tráfico y semáforos, resolviendo así problemas de rendimiento y seguridad vial. El modelo YOLOv8 fue entrenado mediante deep learning para lograr la detección de baches con capacidades en tiempo real junto con señales de tráfico y reconocimiento de semáforos. El conjunto de datos personalizado obtuvo una Precisión Promedio Media (mAP@0.5) de 0.991 a través de la aumentación Roboflow, mientras que la precisión de cada clase alcanzó entre 0.985 y 0.994 en el entrenamiento de diversas condiciones de carretera. El sistema unifica la pérdida CIoU para la regresión del cuadro delimitador con puntuación de objetividad porque su despliegue está diseñado dentro de una aplicación basada en Streamlit. Los resultados demostraron que el modelo supera a versiones previas convencionales y basadas en YOLO a través de capacidades de detección extensivas en diferentes situaciones de iluminación y condiciones climáticas, haciendo viable la automatización escalable del monitoreo de carreteras a través de este sistema.

Xiao et al. (2025) propusieron RiDYOLO, una red de detección ligera y multi-objetivo para tareas de inspección vial. Construido sobre la columna vertebral CS-RepViGG, el modelo integra muestreo dinámico (Dy-Sample) y un mecanismo de atención multi-escala, mejorando la precisión de detección mientras mantiene baja complejidad computacional. El entrenamiento y evaluación extensivos en el conjunto de datos RiD100K demostraron que RiDYOLO supera tanto a YOLOv8 como a YOLOv11, logrando un rendimiento superior con FLOPs más bajos. Específicamente, alcanzó mayor Precisión (0.956), Recall (0.531) y mAP@0.5 (0.4297), ofreciendo así un rendimiento de detección excepcional y eficiencia de despliegue en escenarios de reconocimiento tanto de objetivo único como multi-objetivo. Estos resultados subrayan las sólidas capacidades del método en tiempo real y prácticas en entornos viales dinámicos. El proyecto fue respaldado por el Chongqing Urban Management Scientific Research Project (CGKZ 2022 No.34).

Sreemathy et al. (2024) propusieron un método de reconocimiento de señales de tráfico implementado en dos procesos: detección y reconocimiento. El sistema utiliza el algoritmo de detección You Only Look Once (YOLO), que realiza una clasificación de cuadrícula

sobre la imagen para predecir los cuadros delimitadores, seguido de la determinación de la probabilidad de presencia de un objeto particular en una cuadrícula específica. Para el proceso de reconocimiento, desplegaron un modelo CNN de 4 capas para clasificar el objeto en 43 clases separadas. El modelo fue entrenado utilizando el conjunto de datos de referencia German Traffic Sign Benchmark Dataset. Al realizar pruebas con otros modelos estándar como VGGNet y ResNet-50, el modelo propuesto demostró ser más preciso. La implementación en tiempo real del modelo propuesto alcanzó una precisión de entrenamiento de 99.51 %, mientras que la precisión de prueba fue de 97.15 %, confirmando su efectividad y aplicabilidad práctica.

Niu et al. (2024) propusieron YOLOv8s-DDA, un algoritmo de detección de objetos pequeños para señales de tráfico basado en YOLOv8s mejorado. Para abordar los desafíos del tamaño pequeño de los objetos, escenas complejas, escalas variables de señales y objetos dispersos, introdujeron el módulo Caf-DWR-DBB en la red neck, que utiliza un método eficiente de dos etapas para capturar información contextual multi-escala y emplea un bloque de re-parametrización diluido para mejorar la calidad de extracción de características mientras mantiene la eficiencia computacional. La red neck está mejorada mediante la fusión de características de objetos multi-escala, potenciando significativamente las capacidades de detección de objetos pequeños. Finalmente, el IoU original fue reemplazado por Wise-IoU para mejorar aún más la precisión de detección. En el conjunto de datos TT100K, el algoritmo YOLOv8s-DDA alcanzó mejoras de mAP@0.5 de 82.2 %, mAP@0.5:0.95 de 68.3 %, precisión de 85.2 % y recall de 80.0 %, con una reducción de 3.4 % en el conteo de parámetros. La efectividad del algoritmo también fue validada en el conjunto de datos Chinese traffic sign detection de acceso público, CCTSDB2021.

Bui et al. (2024) propusieron un enfoque basado en aprendizaje profundo que intenta varias versiones de la arquitectura YOLO para realizar el reconocimiento común de señales de tráfico en Vietnam. Primero, se realizó la recolección de datos mediante la captura de imágenes tomadas en carreteras en la ciudad de Can Tho y la provincia de Vinh Long, las cuales fueron combinadas con el conjunto de datos Zalo4 de señales de tráfico vietnamitas en 2020. Posteriormente, se desplegó un proceso de aumento de datos para mejorar el conjunto de datos de mejora. Los autores aplicaron dos versiones de YOLO, el modelo YOLOv5s y el modelo YOLOv8, al conjunto de datos mejorado para reconocer señales de tráfico y comparar la efectividad de los dos enfoques. Los resultados experimentales demostraron que aunque el modelo YOLOv5 requiere más tiempo de entrenamiento y tiene menos parámetros que el modelo YOLOv8, el primero puede desempeñarse mejor en tareas de reconocimiento de señales de tráfico.

Setiawan et al. (2024) desarrollaron un sistema de detección de señales de tráfico para mejorar la seguridad vial, diseñado para detectar señales de tráfico en la carretera tales como semáforos, límites de velocidad y señales de giro en U, proporcionando posteriormente retroalimentación a los usuarios mediante texto en una pantalla LCD y vibración usando un motor. A lo largo de este trabajo, se propuso el algoritmo You Only

Look Once (YOLO) para detectar señales de tráfico en la carretera. Además, se utilizó una cámara para extraer videos en vivo que sirven como entrada para el Raspberry Pi 4 modelo B que contiene modelos YOLOv5s y YOLOv8 personalizados. El sistema luego envía la salida a una pantalla LCD-a, un motor de vibración y un módulo de radiofrecuencia (RF) como notificaciones para los usuarios. Al final, se recopiló el rendimiento del método propuesto y se encontró que produce una buena precisión dentro de una cantidad inconsiderable de 1,677 imágenes y 2,263 anotaciones. Por lo tanto, se obtuvo una precisión de 0.59 para YOLOv5s y 0.7 para YOLOv8 del proceso de entrenamiento y validación. Además, el sistema detecta con precisión las señales de tráfico en la carretera

Nisha et al. (2024) implementaron un Sistema Robusto de Reconocimiento de Señales de Tráfico capaz de asistir a los conductores. El sistema detecta señales de tráfico utilizando una Red Neuronal Convolucional (CNN) YOLOv5 entrenada. Después de detectar la señal de tráfico específica, esta es reconocida y mostrada junto con su imagen en formato de texto. El formato de la imagen se convierte a formato YOLO, y luego las imágenes son entrenadas usando el algoritmo CNN Yolov5 en Python. El modelo fue entrenado estableciendo los parámetros y restricciones necesarias. Posteriormente, utilizando la librería de detección de objetos pyqt5, el código es escrito y el resultado final muestra el nombre respectivo de las señales de tráfico detectadas con un buen puntaje de precisión. Los autores utilizaron un conjunto de datos de 4 clases con 877 imágenes cada una para entrenar la red neuronal. Los resultados mostraron que la precisión de reconocimiento osciló entre Spreen 90 % y 97 % usando YOLOv5. El área grande bajo la curva Precisión-Recall indica alto rendimiento, con una precisión promedio media de 0.989. La matriz de confusión muestra altas tasas verdaderas positivas y verdaderas negativas.

Yohannes et al. (2024) abordaron el reconocimiento de objetos utilizando YOLOv8 (You Only Look Once) para detectar señales de tráfico en Indonesia. Los modelos fueron entrenados y probados por su rendimiento en la detección de señales de tráfico. Los resultados demostraron que el mAP 5.0 y mAP 5.0-9.5 tienen valores de 9.4 5.0 % y 9.8 0.1 %, respectivamente, para el modelo YOLOv8, lo que demuestra que YOLOv8 es el modelo de mejor rendimiento para la detección de señales de tráfico en Indonesia. La aplicación de la inteligencia artificial, particularmente los métodos de detección de objetos que utilizan el aprendizaje profundo basado en detectores, ha demostrado ser eficiente en la detección de objetos en tiempo real.

Kumar et al. (2024) propusieron un sistema integral que utiliza YOLO (You Only Look Once) para la detección de objetos integrado con el microcontrolador NodeMCU. Este NodeMCU está conectado al motor del conductor del vehículo con capacidades de display e Internet de las Cosas (IoT). El conductor del motor controla la salida del motor del vehículo. YOLO está seleccionado por su alta precisión y rápidas capacidades de inferencia, permitiendo la identificación de peligros viales en tiempo real desde vehículos, orquillas de peatones y señales de tráfico capturadas por cámaras montadas en vehículos. El objetivo de este trabajo es la prevención de accidentes mediante dispositivos inteligentes de captura de

imágenes que monitorean el tráfico interno dentro de vehículos y CCTV en las instalaciones de carreteras, junto con módulos de hardware que detectan vehículos que violan las reglas de tráfico y evitan posibles percances debido a negligencia a través de alertas instantáneas. El rendimiento del sistema fue evaluado a través de precisión, tasa de cuadros, tiempo de reacción y eficiencia de respuesta del motor. Los autores agradecen a Karpagam Academy of Higher Education por proporcionar las instalaciones durante este trabajo.

Liu et al. (2023) introdujeron ETSR-YOLO, un algoritmo de optimización basado en el algoritmo YOLOv5 para abordar la susceptibilidad de los métodos actuales de reconocimiento de señales de tráfico a la influencia de la interferencia de luz ambiental, cambios de tamaño objetivo y fondos complejos, lo que resulta en una reducción de la precisión de reconocimiento. Primero, mejoraron la red de agregación de rutas (PANet) de YOLOv5 para potenciar la fusión de características multi-escala generando una capa adicional de características de alta resolución para mejorar el reconocimiento de YOLOv5 para objetos de tamaño pequeño. Segundo, introdujeron dos módulos C3 mejorados que buscan suprimir la interferencia de ruido de fondo y mejorar las capacidades de extracción de características de la red. Finalmente, utilizaron la función Wise-IoU (WIoU) en la etapa de post-procesamiento para mejorar la capacidad de aprendizaje y robustez del algoritmo en diferentes muestras. Los resultados experimentales mostraron que ETSR-YOLO mejora mAP@0.5 en 6.6 % en el conjunto de datos Tsinghua-Tencent 100K (TT100K) y en 1.9 % en el conjunto CCTSDT Chinese Traffic Sign Detection Benchmark 2021 (CCTSDB2021). En los experimentos realizados en la plataforma de computación embebida, ETSR-YOLO demostró un tiempo promedio de inferencia corto, afirmando así su capacidad para ofrecer detección confiable de señales de tráfico para vehículos inteligentes operando en escenas de tráfico del mundo real.

Ren et al. (2023) propusieron una red mejorada basada en YOLOv5, considerando las señales de tráfico y grietas en carreteras como objetos de detección y realizando investigación separada. El trabajo introduce una estructura C3-FPN para reemplazar la estructura original de fusión de características para grietas en carreteras. Esta estructura integra el modelo de atención de bloques convolucionales (CBAM) basado en redes piramidales de características bidireccionales (Bi-FPN) e introduce un nuevo módulo de convolución ligera (GSConv) para reducir la pérdida de información del mapa de características, mejorar la capacidad expresiva de la red y, en última instancia, lograr un mejor rendimiento de reconocimiento. Para señales de tráfico, se utilizó una estructura de detección de características a cuatro escalas para aumentar la escala de detección de capas superficiales y mejorar la precisión de reconocimiento para objetivos pequeños. Además, el estudio combinó varios métodos de aumento de datos para mejorar la robustez de la red. A través de experimentos utilizando 2164 conjuntos de datos de grietas en carreteras y 8146 conjuntos de datos de señales de tráfico realizados por Labellmg, en comparación con el modelo base (YOLOv5s), la red YOLOv5 modificada mejora el resultado promedio de precisión media (mAP) del conjunto de datos de grietas en carreteras y objetivos pequeños en el conjunto de datos de señales de tráfico en 4 % y 12.2 %, respectivamente.

Sharma et al. (2023) ofrecieron un método mejorado de identificación de señales de tráfico yolov4-tiny para facilitar el despliegue en dispositivos móviles o embebidos, abordando las dificultades de un alto número de parámetros, baja precisión de reconocimiento y pobre rendimiento en tiempo real de los modelos de reconocimiento de señales de tráfico en escenarios complejos. La red yolov4-tiny sirve como base del modelo. Para comenzar, la convolución se incorpora en la red backbone para eliminar la redundancia de características de baja frecuencia, reduciendo el número de parámetros en el modelo y mejorando la eficiencia computacional. Segundo, se emplea el módulo de atención de bloques convolucionales para mejorar la precisión de reconocimiento de objetivos pequeños y medianos al fortalecer los pesos de las regiones de señales de tráfico y suprimir los pesos de características inválidas. Finalmente, en la etapa de fusión de características, la estructura Feature Pyramid Networks se reemplaza con la estructura Simplified Path Aggregation Network para mejorar la fusión de información de características superficiales con conocimiento semántico profundo y reducir la tasa de detección perdida aún más. En el conjunto de datos TT100K así como en el conjunto CCTSDB, los resultados experimentales sugieren que la técnica puede lograr un buen rendimiento de reconocimiento. Con un tamaño de modelo de 10MB, la solución mejora la precisión promedio media en 15 % y el Frame Per Second en 12.5 (% cuando se compara con el algoritmo yolov4-tiny). El método supera a yolov4-tiny en términos de precisión de reconocimiento y velocidad de detección, y puede cumplir fácilmente con los requisitos en tiempo real para el reconocimiento de señales de tráfico.

Li et al. (2023) abordaron el problema de extremo a extremo en la detección de señales de tráfico en tiempo real y propusieron un algoritmo mejorado de detección de señales de tráfico de una etapa: Attention-YOLO V4. Para lograr la detección en tiempo real y de alta precisión de señales de tráfico, analizaron el principio de YOLO V4 para la detección de objetos pequeños. Con el fin de mejorar la capacidad de la red backbone de YOLO V4 para extraer características de señales de tráfico, decidieron combinar el mecanismo de atención de canal con el bloque residual para mejorar la capacidad de YOLO Head de detectar señales de tráfico, combinando el mecanismo de atención de canal con YOLO head. Utilizaron el conjunto de datos TT100K para evaluar el algoritmo Attention-YOLO V4. En comparación con los métodos existentes, su método logra un rendimiento en tiempo real y de alta precisión en fondos complejos.

2.1.2. Investigaciones nacionales

Quispe Condor (2024) propone un modelo de red neuronal artificial orientado a la predicción de siniestros viales mediante el análisis de parámetros geométricos del camino en el sector Tarma-La Merced de la Carretera Central. El trabajo identifica áreas críticas de alta peligrosidad evaluando elementos del diseño geométrico tales como: tipología de curvas (rectas o curvas clasificadas según su radio en pequeñas 20m-50m, medianas 50m-100m, grandes 100m-200m o muy grandes >200m), magnitud del ángulo de deflexión (reducido 0°-45°, intermedio 45°-90° o elevado >90°), gradiente longitudinal (suave

0°-3°, moderado 3°-7° o pronunciado 7°-12°) y nivel de visibilidad (deficiente, aceptable o óptima). El investigador experimentó con múltiples configuraciones de arquitectura neuronal variando desde 2 hasta 24 neuronas de entrada y de 4 a 22 neuronas en la capa oculta, optando por la configuración con mínimo error de predicción. Los hallazgos indican que las áreas de máxima peligrosidad se localizan en tramos con curvas de radio reducido e intermedio, deflexiones medianas y elevadas, aunque inesperadamente también en segmentos rectilíneos de extensión media y corta categorizados con riesgo medio-alto. El estudio determina que en estos sectores los automovilistas ejecutan maniobras de sobrepaso que frecuentemente derivan en comportamientos riesgosos y colisiones.

Sifuentes Rodriguez y Alarco Basaldua (2023) examinan la crisis de seguridad vial en Lima que compromete la integridad de peatones, automovilistas, ciclistas y diversos usuarios del sistema vial. Los investigadores determinan que la inseguridad vial se origina en factores como carencias en educación vial, señalización inadecuada y mantenimiento insuficiente tanto de calzadas como de aceras limeñas. El estudio plantea la aplicación de intervenciones de ingeniería económicamente accesibles en un área crítica denominada ”punto negro”, validando estas propuestas mediante simulación computacional que cuantificará la mejora en seguridad vial tras su implementación. Los autores resaltan que optimizar la seguridad vial genera beneficios como descenso en colisiones vehiculares, reducción de heridos y fallecidos, aminoramiento del impacto económico familiar derivado de costos sanitarios y pérdida de capacidad productiva por decesos o discapacidad de afectados. La propuesta busca incrementar la seguridad vial evaluando parámetros como volumen vehicular, sincronización semafórica y calidad del servicio, favoreciendo a los ciudadanos limeños que circulan por dicha área para que experimenten una infraestructura vial más confiable.

Ramos Zúñiga (2021) diseña herramientas para automatizar el levantamiento del inventario vial mediante el desarrollo de un algoritmo orientado a la identificación y clasificación de señalética preventiva peruana en la Av. Eduardo Habich de Lima Metropolitana, distrito San Martín de Porres. El segmento analizado comprende desde Av. Túpac Amaru hasta Panamericana Norte abarcando aproximadamente 1km. El trabajo fundamenta su metodología en algoritmos de análisis de imágenes, estructurándose en dos fases: captura de información en terreno para conformar la base de datos y análisis de información en oficina. La autora compiló 2026 fotografías de señalética preventiva de vías limeñas obtenidas con dispositivos de captura de variadas resoluciones. La fase de identificación emplea filtros cromáticos y morfológicos, efectuando un cotejo entre filtros HSV y RGB normalizado para establecer cuál exhibía superior desempeño. La fase de clasificación utiliza técnicas de aprendizaje supervisado mediante el algoritmo de Support Vector Machines. Como resultado final, la investigación produjo un algoritmo capaz de identificar señalética vial alcanzando un 62.5 % de precisión en la clasificación y una base de datos robusta susceptible de enriquecimiento para subsecuentes investigaciones en esta línea de estudio.

Matos Levano (2020) ejecuta un estudio orientado a la implementación de un esquema de señalización vial destinado a la prevención de siniestros vehiculares en el área comprendida entre Jirón de la Salud y Avenida Héroes del Cenepa. El investigador aplica una metodología de diseño no experimental transversal, capturando información en un período único mediante técnicas aplicadas en contexto real. La investigación adopta un enfoque explicativo-causal al existir vinculación causa-efecto entre variables, y cuantitativo mediante captación de datos por observación directa y procesamiento para evaluación del análisis de tráfico. Los hallazgos del procesamiento de información de campo evidencian velocidades excesivas superiores a los límites establecidos en la vía y presencia de riesgo de colisiones vehiculares debido al considerable flujo peatonal en el área estudiada. El investigador establece que la aplicación de un esquema de señalización vial impacta favorablemente en la prevención de siniestros vehiculares en el sector analizado.

Távara Idrogo (2019) construye un modelo fundamentado en aprendizaje profundo mediante redes neuronales convolucionales orientado al reconocimiento automatizado de señalética de tránsito vehicular empleando técnicas de análisis de imágenes y principios de inteligencia artificial. El proyecto se enfoca en señalética vial de Alemania y Perú, catalogando 43 y 7 categorías respectivamente. La investigación comienza con la captura de imágenes, procediendo al procesamiento de estas con el propósito de ampliar el conjunto de datos y ejecutar el aprendizaje profundo mediante diversas arquitecturas de modelos de redes neuronales convolucionales. Como producto final, el investigador alcanzó un modelo con indicadores satisfactorios y resultados favorables en el reconocimiento de señalética de tránsito vehicular. La investigación aspira a contribuir con los esfuerzos de la industria automotriz en el ámbito de sistemas avanzados de asistencia al conductor, integrándose a diversos mecanismos que buscan solucionar la inseguridad vial.

En su trabajo Arenas (2016), construye un sistema de categorización de señalética de tránsito vehicular mediante redes neuronales convolucionales como respuesta a la problemática de siniestros viales registrados durante 2004-2015 (1,041,704 casos), donde la velocidad excesiva representa uno de los factores predominantes. El investigador plantea que estas estadísticas y la probabilidad de tragedias pueden disminuirse sustancialmente optimizando la vigilancia vehicular mediante urbes más inteligentes capaces de identificar zonas de vulnerabilidad, actividades delictivas y transgresiones a la señalética vial. La investigación exhibe una solución basada en Deep Learning, metodología de inteligencia artificial empleada anteriormente para reconocimiento óptico de caracteres, categorización de imágenes, identificación facial y vocal, y diagnóstico médico. Ante la carencia de una base de datos de señalética peruana, se ejecutaron varios diseños basados en señalética alemana, alcanzando una precisión del 95.29 %, comparable al estado del arte (99.46 %). Con propósitos demostrativos, se implementó el diseño en una aplicación móvil para iPhone que categoriza señalética vial utilizando la cámara del dispositivo.

2.2. Bases teóricas

2.2.1. Detección y evaluación de señales de tránsito

2.2.1.1 Concepto de seguridad vial

El concepto de seguridad vial engloba el conjunto de acciones preventivas implementadas con el propósito de minimizar la incidencia de lesiones y fallecimientos derivados de la movilidad vehicular.

La mejora en la normativa relacionada con la protección de los usuarios viales en los países americanos se fundamenta en la articulación intersectorial y el trabajo colaborativo, permitiendo configurar entornos de movilidad más protegidos, inclusivos y sustentables para la totalidad de actores del transporte (OPS (2024)).

Esta visión holística demanda la intervención coordinada de diversos ámbitos institucionales, incluyendo aquellos vinculados con la movilidad, el orden público, la atención sanitaria y la formación educativa, complementados por la iniciativa empresarial y las agrupaciones civiles organizadas.

La gestión de los traumatismos originados por eventos viales demanda estrategias que contemplen simultáneamente la infraestructura de circulación, las unidades de transporte y el conjunto de actores viales, incorporando el desarrollo de espacios más protegidos e integrando criterios de protección vial en la planificación territorial y de movilidad, optimizando los atributos de protección de las unidades vehiculares, y estableciendo marcos legales orientados a los principales factores de amenaza garantizando su aplicación efectiva.

2.2.1.2 Fundamentos conceptuales de la seguridad en el tránsito

Los fundamentos conceptuales que sustentan la seguridad en el tránsito representan una estructura teórica comprehensiva que determina las bases científicas y procedimentales orientadas a evitar lesiones y decesos en el ecosistema de movilidad vial. Dichos fundamentos se estructuran esencialmente mediante la ”Perspectiva del Sistema Protegido y las Cinco Dimensiones de la Seguridad Vial”, formulados por (OMS (2024)) y reconocidos globalmente.

1. Perspectiva del sistema protegido

La perspectiva del sistema protegido representa un conjunto de directrices enfocadas en salvaguardar la existencia y el bienestar de las personas, estableciendo este objetivo como la prioridad fundamental al configurar y gestionar cualquier infraestructura de movilidad.

Esta visión concentra su atención en la fragilidad del participante vial y fundamenta su filosofía en el principio de que, si bien el error es inherente a la naturaleza humana, resulta inadmisible que dicho error derive en pérdidas de vidas o traumatismos severos ocasionados por incidentes de tráfico.

Esta perspectiva representa una transformación desde el enfoque tradicional centrado

exclusivamente en disminuir la frecuencia de colisiones hacia la meta de erradicar las víctimas fatales y los traumatismos de gravedad, aceptando que las infraestructuras de movilidad deben adaptarse a las necesidades de quienes las utilizan y tolerar la falibilidad humana.

2. Las cinco dimensiones de protección vial según la OMS

Las cinco dimensiones establecidas por la (OMS (2024)) para la protección vial configuran una estructura comprehensiva destinada a optimizar la seguridad en las vías de circulación a escala planetaria, contemplando elementos fundamentales que abarcan desde la administración eficaz de la protección vial hasta el desarrollo de espacios de circulación protegidos concebidos para resguardar a la totalidad de usuarios de las carreteras.

- **Administración de la protección vial:** Consiste en desarrollar marcos políticos, jurídicos y normativos robustos para afrontar las problemáticas vinculadas con la protección en las vías, además de ejercer el liderazgo en la articulación de iniciativas de protección vial a escala gubernamental.
- **Espacios de circulación protegidos:** Abarca la proyección y configuración apropiada de autopistas y calzadas, comprendiendo la edificación y conservación de infraestructuras que garanticen protección para todos los usuarios, brindando especial atención a los actores más expuestos: transeúntes, usuarios de bicicletas y conductores de motocicletas.
- **Unidades vehiculares protegidas:** Contempla el fomento de vehículos que cumplan estándares de protección y la adopción de normativas que aseguren la manufactura y operación de unidades vehiculares conforme a parámetros de seguridad previamente definidos.
- **Conductas responsables de los usuarios:** Comprende el impulso y la capacitación de los usuarios viales respecto a la adopción de prácticas seguras, tales como la utilización de sistemas de retención, el acatamiento de velocidades establecidas, la abstención de conducir bajo influencia de sustancias psicoactivas o alcohol, y el cumplimiento de las disposiciones de circulación.
- **Atención posterior a los siniestros:** Involucra la consolidación de mecanismos de respuesta ante situaciones de emergencia, incluyendo servicios médicos urgentes y cuidado de afectados en eventos viales, junto con la ejecución de iniciativas de recuperación y soporte prolongado.

3. Elementos esenciales del sistema protegido

El paradigma del sistema protegido integra cuatro componentes esenciales: Carreteras y bordes viales más protegidos (predecibles y que permiten errores), velocidades apropiadas (ajustadas a la funcionalidad y nivel de protección de la vía), vehículos protegidos (que evitan colisiones y resguardan a los ocupantes), y usuarios responsables (alertas y en condiciones óptimas que observan la normativa vial).

La confluencia de los atributos de protección de la infraestructura, las características vehiculares y la velocidad de las unidades participantes en una colisión establece

las magnitudes de impacto experimentadas por los ocupantes en el choque. Dichas interacciones requieren ser reguladas con el fin de prevenir traumatismos severos y decesos.

2.2.1.3 Elementos constitutivos del ecosistema de tránsito (actor vial, unidad vehicular,etc)

El ecosistema de tránsito se caracteriza por ser un fenómeno de naturaleza compleja que implica la interrelación dinámica de diversos elementos mutuamente dependientes. La modelación computacional se conceptualiza como la recreación dinámica de fragmentos del mundo tangible, alcanzada mediante el desarrollo de modelos informáticos que evolucionan temporalmente.

Los paradigmas computacionales encuentran amplia utilización en el examen de infraestructuras de tráfico y movilidad. Las infraestructuras de transporte presentan múltiples particularidades que dificultan su análisis, regulación y optimización. Frecuentemente, estos sistemas abarcan extensas áreas territoriales, involucran numerosos participantes activos, cuyos propósitos y metas no necesariamente convergen entre sí ni con los objetivos del operador del sistema.

Esta complejidad intrínseca ha motivado a los especialistas a elaborar estructuras conceptuales que faciliten la comprensión y el examen de los diversos elementos que integran el ecosistema de tránsito. La bibliografía especializada ha reconocido de manera consistente cuatro elementos fundamentales que configuran el ecosistema de tránsito: el actor vial, la unidad vehicular, la calzada y el contexto ambiental. Esta taxonomía, denominada paradigma de çuatro elementos", suministra una estructura analítica sólida para investigar la protección vial y la optimización del transporte.

Elemento 1: Actor del sistema

- **Conceptualización y atributos:** Los actores viales representan el componente primordial y de mayor relevancia en el ecosistema de transporte. Satisfacer los requerimientos del actor constituye la característica esencial de una infraestructura de movilidad sustentable (Jameel y Evdorides (2018)). En la investigación de las conductas de los actores viales, la información puede emplearse para examinar atributos de dichos actores, validar hipótesis teóricas, o anticipar patrones conductuales (Donmez et al. (2023)).
- **Dimensiones psicológicas y conductuales:** De acuerdo con (Fuller (2008)), las conductas de los actores viales constituyen un elemento significativo en la génesis de colisiones. La psicología del tránsito, conceptualizada como "la investigación de las conductas de los actores viales y los mecanismos psicológicos que las fundamentan", busca reconocer los determinantes de las conductas viales con el propósito de formular contramedidas eficaces para prevenir siniestros.

  Las conductas de los actores viales se reconocen como el principal elemento contribuyente en la mayoría de eventos viales. Consecuentemente, las estrategias exitosas de protección vial se orientan hacia la optimización de las conductas

de los actores viales, considerando siete elementos de riesgo: operadores bajo efectos etílicos, sistemas de retención corporal, mecanismos de sujeción para menores, administración de la velocidad, utilización de protección craneal, empleo de dispositivos móviles durante la conducción, y operación bajo influencia de sustancias psicoactivas.

- **Actores de alta vulnerabilidad:** La noción de actores viales de alta vulnerabilidad alude a transeúntes y usuarios de vehículos de dos ruedas (bicicletas y motocicletas) que enfrentan mayor susceptibilidad de sufrir traumatismos o decesos en espacios viales predominados por automóviles. Esta categorización acepta que determinados actores del ecosistema de tránsito confrontan amenazas desproporcionadamente elevadas atribuibles a su exposición y fragilidad corporal.

Elemento 2: Unidad vehicular

- **Perspectiva tecnológica:** Las unidades vehiculares inteligentes y la movilidad inteligente ofrecen posibilidades extraordinarias para atenuar la saturación vial, optimizar el consumo energético, disminuir colisiones y potenciar el desempeño general. Las tecnologías fundamentales de investigación y aplicación en esta área comprenden tecnologías vehiculares avanzadas, examen de información de alta precisión, paradigmas innovadores para servicios de movilidad y arquitecturas de múltiples sistemas.
- **Unidades vehiculares interconectadas:** El surgimiento de las Infraestructuras de Transporte Inteligente (ITS) ha posibilitado perspectivas innovadoras para optimizar la protección, funcionamiento e influencia ecológica de las redes de movilidad. Las unidades vehiculares interconectadas (CV), una propuesta innovadora de ”vehículos cognitivos”, surgen como la próxima generación tecnológica para potenciar aún más a los viajeros. Las unidades vehiculares contemporáneas integran mecanismos de protección clasificables en dos categorías fundamentales: protección activa (mecanismos preventivos de colisiones) y protección pasiva (mecanismos atenuadores de consecuencias en colisiones).

Elemento 3: Calzada (Infraestructura de circulación)

- **Conceptualización y extensión:** La infraestructura de circulación comprende la implementación de elementos permanentes incluyendo autopistas terrestres y férreas, además de instalaciones terminales como paraderos de transporte público, terminales de carga y estaciones ferroviarias. Las operaciones vehiculares a través de estos sistemas de movilidad propician la disminución de tiempos de desplazamiento y generan oportunidades laborales en el sector (Global (2023)).
- **Elementos tangibles:** La infraestructura de circulación alude a los elementos materiales de los sistemas de movilidad, comprendiendo tramos de autopistas, vías exclusivas para transporte público, dispositivos de señalización, paraderos y modificaciones complementarias proyectadas para posibilitar el desplazamiento vehicular y optimizar la corriente de tráfico.

Elemento 4: Contexto ambiental

- **Conceptualización del contexto de circulación:** Las investigaciones sobre protección en la operación vehicular bajo fenómenos meteorológicos como precipitación, nevada y neblina pertenecen al campo de investigación de amenazas derivadas de factores contextuales del tráfico. El contexto ambiental simboliza el conjunto de circunstancias externas que impactan en la operación del ecosistema de tránsito.

  Los elementos contextuales fundamentales abarcan fenómenos meteorológicos (precipitación, nevada, neblina, corrientes de aire), circunstancias de luminosidad (diurno, nocturno, crepuscular), visibilidad (óptima, limitada, nula) y temperatura (extremadamente baja, extremadamente elevada, templada). Estos elementos contextuales han sido reconocidos como contribuyentes significativos a las amenazas operacionales del tráfico.

Interrelaciones entre elementos

- **Perspectiva del sistema protegido:** La perspectiva del sistema protegido promueve una infraestructura de movilidad segura para la totalidad de actores viales. Esta perspectiva considera la fragilidad de las personas ante traumatismos severos derivados del tránsito, y acepta que, al proyectar el sistema, debe contemplarse la falibilidad humana.
- **Evitación y administración de amenazas:** La protección del tráfico vial puede garantizarse mediante la evitación y administración de las amenazas latentes en las operaciones de tráfico vial. Los hallazgos evidencian que los tópicos de investigación pertinentes abarcaron principalmente las amenazas de operadores, unidades vehiculares, autopistas y el contexto de tráfico.

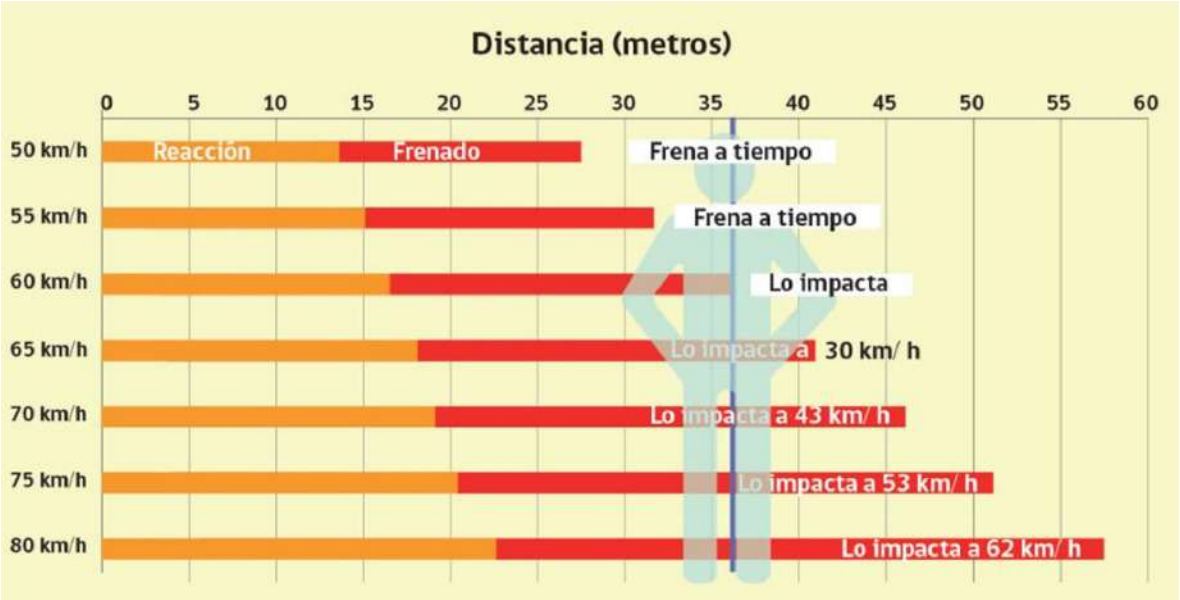
2.2.1.4 Elementos de amenaza en la protección vial

Los elementos de amenaza en protección vial comprenden todas aquellas circunstancias, condiciones o componentes que elevan la posibilidad de materialización de un siniestro de tránsito o que pueden intensificar sus repercusiones. Tradicionalmente se organizan en tres grupos fundamentales:

Elementos antropogénicos (tales como ingesta de alcohol, sustancias psicoactivas, agotamiento físico, falta de atención, velocidad excesiva, inexperiencia del operador y conducción temeraria), elementos vehiculares (vinculados con el estado mecánico, conservación y características de protección de la unidad, incluyendo deficiencias en sistemas de frenado, iluminación y dispositivos de sujeción) y elementos contextuales o de infraestructura (que comprenden las condiciones de la calzada, señalización deficiente, degradación del pavimento, iluminación insuficiente, fenómenos climáticos desfavorables y configuración geométrica inadecuada de las autopistas).

Esta clasificación tripartita resulta fundamental para formular estrategias comprehensivas de evitación y administración de la protección vial.

Figura 2
Elementos asociados con velocidad elevada



Nota. Imagen obtenida en la web.

2.2.1.5 Siniestros de tránsito

Los siniestros de tránsito representan acontecimientos abruptos e imprevistos que se manifiestan durante el desplazamiento de unidades vehiculares en calzadas públicas, caracterizándose por el impacto entre una o múltiples unidades vehiculares, o entre una unidad vehicular y demás componentes del contexto vial tales como transeúntes, ciclistas, motociclistas, elementos fijos o fauna.

Las repercusiones pueden abarcar deterioros materiales a las unidades participantes, traumatismos de diversa magnitud en los individuos afectados, discapacidades irreversibles o deceso, además de consecuencias económicas considerables tanto para los individuos como para las infraestructuras de atención sanitaria pública y la colectividad en general.

Constituyen una problemática sanitaria pública de alcance planetario que impacta aproximadamente a 1.19 millones de individuos anualmente con resultado mortal y entre 20 a 50 millones con traumatismos no fatales, configurando la causa primordial de mortalidad en infantes y jóvenes de 5 a 29 años y originando costos equivalentes al 3 % del Producto Interno Bruto en la mayoría de naciones (Alvia Párraga y Linares Giler (2024)).

Figura 3
Estrategias esenciales para evitar siniestros



Nota. Imagen obtenida en la web.

2.2.1.6 Dispositivos de señalización vial y su función en la protección

Los dispositivos de señalización vial representan el conjunto organizado de elementos visuales, marcaciones, iconografía, cromatismos y comunicaciones normalizadas que se disponen estratégicamente en las calzadas públicas para comunicar información fundamental a los actores del ecosistema vial.

Desempeñan una función esencial en la protección vial al orientar, precaver, normar y estructurar el flujo vehicular y peatonal mediante tres funciones fundamentales: orientadora (suministrando guía sobre destinos, distancias, servicios y localizaciones), precautoria (advirtiendo sobre amenazas latentes, circunstancias especiales de la calzada o situaciones que demandan precaución) y normativa (determinando limitaciones, vedas y regulaciones de acatamiento imperativo).

Utilizan un código visual estandarizado y comprensible que posibilita la interpretación expedita y exacta de las indicaciones, contribuyendo significativamente a la disminución de siniestros de tránsito, la optimización del flujo vehicular, la protección de actores vulnerables como transeúntes y ciclistas, y el establecimiento de una infraestructura de transporte protegida, eficaz y predecible que posibilita la adopción de decisiones correctas por parte de operadores y transeúntes en tiempo real (Cedeño-Alcívar y Mogrovejo (2023)).

Figura 4
Dispositivos de señalización vial y su función en la protección



Nota. Imagen obtenida en la web.

2.2.1.7 Manual de dispositivos de control de tránsito automotor 2024

El Manual de Dispositivos de Control del Tránsito Automotor (2024) representa un instrumento técnico normativo oficial de observancia imperativa que determina los parámetros, criterios, especificaciones y metodologías estandarizadas para la proyección, implementación, posicionamiento y conservación de la totalidad de elementos destinados a la regulación y control del tránsito vehicular y peatonal.

Comprende señalizaciones verticales (normativas, precautorias e informativas), marcaciones horizontales en el pavimento, semáforos y elementos complementarios.

El propósito fundamental consiste en alcanzar la uniformidad necesaria e imprescindible en la implementación de estos componentes de regulación a escala nacional, asegurando que satisfagan requisitos fundamentales tales como la capacidad de captar positivamente la atención y ser perceptibles, comunicar mensajes inequívocos y concisos, posibilitar tiempo apropiado de percepción y respuesta a los actores, inspirar respeto y ser acatados, y preservar coherencia y consistencia en su implementación.

De esta manera, contribuye significativamente a la optimización de la protección vial, la eficacia en la circulación del tránsito, la disminución de siniestros de tránsito y la protección de la integridad corporal de la totalidad de actores de las calzadas urbanas y autopistas del país.

Figura 5
Portada del Compendio de Dispositivos de Control del Tránsito Automotor



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

2.2.1.8 Dispositivos de señalización vertical

Los dispositivos de señalización vertical representan elementos de regulación del tránsito implementados estratégicamente al lateral de las calzadas o suspendidos sobre estas mediante columnas, estructuras portantes, banderolas u otras configuraciones de sostén. Su propósito fundamental consiste en normar el tránsito, precaver situaciones de amenaza e informar a los actores de las calzadas públicas (operadores, transeúntes y ciclistas) mediante iconografía normalizada, leyendas, pictogramas y códigos cromáticos específicos determinados en los compendios técnicos pertinentes. Se clasifican funcionalmente en tres categorías fundamentales: señalizaciones normativas o reguladoras (que comunican sobre limitaciones, vedas, restricciones, obligaciones y autorizaciones cuya inobservancia constituye transgresión a las disposiciones de tránsito), señalizaciones precautorias o de advertencia (que alertan sobre la presencia y caracterización de amenazas, peligros o situaciones imprevistas presentes en la calzada o áreas circundantes de modo permanente o transitorio), y señalizaciones informativas (que orientan y guían a los actores suministrando información sobre destinos, distancias, servicios, sitios de interés turístico, nomenclatura de calles y rutas). Satisfacen requisitos técnicos fundamentales como perceptibilidad diurna y nocturna mediante materiales retroreflectantes, dimensionamiento apropiado según la velocidad de la calzada, posicionamiento dentro del campo de atención visual del operador, resistencia a circunstancias climáticas desfavorables y uniformidad en su implementación, contribuyendo de modo significativo a la optimización de la protección vial, la fluidez del tráfico y la disminución de siniestros de tránsito (Manual de Dispositivos de Control del

Tránsito Automotor (2024)).

Figura 6
Dispositivo vertical de precaución P-49 (zona escolar)



Nota. Imagen propia.

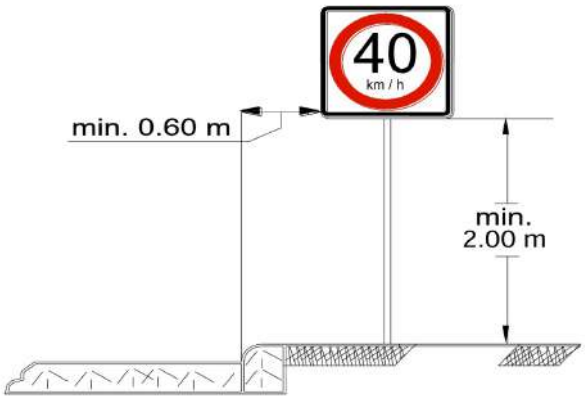
- **Normalización del diseño:** La uniformidad visual en dimensiones morfológicas, cromáticas, dimensionales y simbólicas resulta fundamental para asegurar la comprensión inmediata y eficaz del mensaje por parte de los actores viales. Este Compendio determina especificaciones pormenorizadas sobre configuración y proporciones de las señalizaciones, incorporando además una arquitectura tipográfica normalizada que contempla múltiples escalas de caracteres y orientaciones para su implementación correcta, incluyendo el espaciamiento entre letras apropiado.
- **Comunicación eficiente de mensajes viales:** Cada elemento de señalización debe transmitir información exacta e inequívoca a los actores del ecosistema de transporte, empleando iconografía normalizada y/o textos descriptivos constituidos por elementos alfanuméricos. La uniformidad comunicativa demanda que circunstancias similares sean simbolizadas mediante señalética idéntica, sin considerar su localización territorial. Se prioriza la iconografía sobre los textos atribuible a su superior rapidez de procesamiento mental, conforme a las determinaciones del presente Compendio. La correspondencia entre componentes iconográficos y textuales es imperativa cuando cohabitan en una misma señalización. Las abreviaciones de unidades métricas deben ajustarse al Sistema Internacional de Unidades.
- **Normativa tipográfica diferenciada:** En señalizaciones normativa y precautoria, tanto los textos fundamentales como los suplementarios deben utilizar caracteres mayúsculos, exceptuando casos específicamente considerados. Para señalización

informativa, la elección entre mayúsculas y minúsculas se establece según la dimensión del carácter. La preferencia por caracteres minúsculos en señalizaciones informativas se sustenta en investigaciones que evidencian su superior legibilidad en contextos viales. La incorporación de iconografía innovadora en ciertas localidades debe complementarse temporalmente de placas pedagógicas explicativas, de configuración rectangular subordinada a la amplitud de la señalización principal y cromáticamente consistente, con vigencia máxima de tres años.

Posicionamiento lateral de las señalizaciones:

Figura 7

Posicionamiento apropiado de la señalización vertical

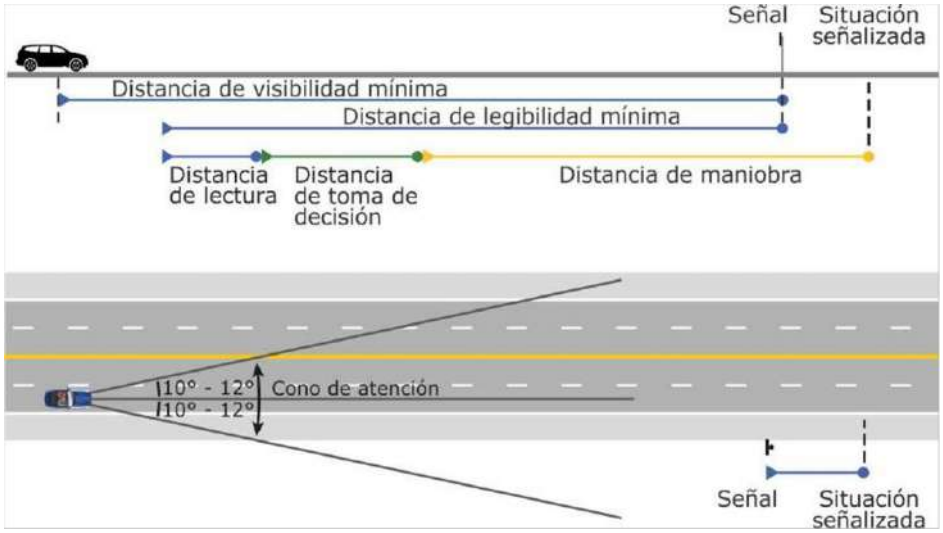


Nota. Tomado de Manual de Dispositivos de Control del Tránsito Automotor (2024).

Posicionamiento longitudinal y distancias de percepción:

Figura 8

Posicionamiento longitudinal y distancias de percepción



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

1. Dispositivos normativos o de reglamentación

Los dispositivos normativos o de reglamentación representan elementos de regulación del

tránsito de observancia imperativa que poseen como propósito comunicar explícitamente a los actores de las calzadas públicas sobre las limitaciones, restricciones, vedas, obligaciones y autorizaciones específicas que rigen la utilización de la infraestructura vial. Establecen el marco normativo que debe respetarse durante la circulación vehicular y peatonal, cuya transgresión constituye una contravención a las disposiciones jurídicas vigentes en materia de tránsito y transporte terrestre. Se caracterizan por su configuración circular inscrita en placas cuadradas o rectangulares (exceptuando la señalización ”PARE”de configuración octogonal y CEDA EL PASO de configuración triangular), empleando códigos cromáticos específicos como el rojo para vedas, azul para obligaciones, y combinaciones estandarizadas que aseguran su reconocimiento inmediato, contribuyendo a la estructuración del flujo vehicular, la evitación de conflictos viales y la preservación de la protección en el ecosistema de transporte.

- **Dispositivos de preferencia:** El dispositivo (R-1) PARE establece que el operador de una unidad vehicular debe detenerse completamente antes de atravesar un cruce, debiendo establecerse su posicionamiento conforme al análisis de ingeniería vial previamente señalado, dado que su empleo indiscriminado puede impactar negativamente en su confiabilidad, y en lugar de contribuir a la protección vial en un cruce puede originar inseguridad.

  El dispositivo (R-2) CEDA EL PASO establece que el operador de una unidad vehicular que transita por una calzada de menor preferencia (calzada secundaria o auxiliar) conceda el paso a otra unidad vehicular que transita por una calzada de mayor preferencia (calzada principal). Su posicionamiento depende de la visibilidad del que transita por la calzada de menor preferencia y la distancia requerida para conceder el paso antes de ingresar a un cruce. En caso contrario, debe utilizarse el dispositivo (R-1) PARE.

Figura 9
Dispositivos de preferencia R-1 y R-2



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

- **Dispositivo de veda:** Se emplean para vedar o restringir el desplazamiento de determinados tipos de unidades vehiculares o específicas operaciones. Se simboliza mediante un círculo blanco con borde rojo atravesado por una diagonal también roja, descendente desde la izquierda configurando un ángulo de 45º con la horizontal. El dispositivo (R-28) NO ESTACIONAR NI DETENERSE constituye una excepción en la cual

aparecen dos diagonales.

Cuando una veda impacta exclusivamente a un tipo de unidad vehicular, debe incorporarse un mensaje que lo identifique inequívocamente. Como ilustración, si la veda impacta únicamente a autobuses, el dispositivo se constituye del símbolo pertinente y el mensaje ”BUSES”posicionada en la sección superior. Cuando se trate de vedas a la circulación que apliquen a uno o múltiples tipos de unidades vehiculares, o a unidades vehiculares y transeúntes, los dispositivos pertinentes pueden exhibirse agrupados en una misma columna, en placa de coloración blanca de fondo.

- **Dispositivo de veda de operaciones y giros**

Figura 10

Dispositivos de veda de operaciones y giros



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

- **Dispositivo de veda por tránsito de unidades vehiculares**

Figura 11

Dispositivos de veda de tránsito por categoría de unidad vehicular



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

2. Dispositivos de precaución

Los dispositivos de precaución o precautorios representan elementos de advertencia anticipada implementados estratégicamente en las calzadas con el propósito fundamental de alertar e informar a los operadores y demás actores sobre la presencia, proximidad y caracterización específica de amenazas, peligros latentes, circunstancias geométricas especiales de la autopista o situaciones imprevistas que puedan manifestarse de modo permanente o transitorio en la calzada o en sus áreas circundantes.

Posibilitan a los actores adoptar oportunamente las acciones de precaución requeridas tales como disminución de velocidad, modificación de carril, incremento de la distancia de seguimiento o realización de operaciones evasivas para resguardar su propia integridad corporal y la de otros actores vulnerables del ecosistema vial. Se caracterizan por su configuración romboidal (cuadrado rotado 45 grados) con fondo amarillo y iconografía o leyendas en coloración negra, posicionándose a distancias calculadas que posibiliten tiempo suficiente de percepción, respuesta y reacción ante la circunstancia advertida.

Dispositivos precautorios por atributos de la superficie de rodadura

Figura 12

Dispositivos precautorios por atributos de la superficie de rodadura



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

Dispositivos precautorios por atributos operacionales de la calzada

Figura 13

Dispositivos precautorios por atributos operacionales de la calzada



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

3. Dispositivos de orientación

Los dispositivos de orientación constituyen elementos orientadores y pedagógicos que poseen como función fundamental guiar, dirigir y suministrar información útil y pertinente a los actores de las calzadas para facilitarles el proceso de navegación y adopción de decisiones durante sus desplazamientos.

Suministran información sobre destinos, direcciones, distancias kilométricas a poblaciones y servicios, nomenclatura de calles y avenidas, numeración de rutas, localización de sitios de interés turístico, histórico, cultural o recreativo, servicios al actor como estaciones de combustible, restaurantes, hoteles y hospitales, además de información geográfica y de orientación general que contribuyen a optimizar la eficacia del desplazamiento y la experiencia del actor.

Se caracterizan por su configuración rectangular con mayor dimensión horizontal, empleando fondos verdes para destinos locales, azules para rutas principales y autopistas, cafés para sitios turísticos, y letras blancas para asegurar contraste y legibilidad óptima tanto diurna como nocturna.

4. Señalización para instalaciones de peaje y pesaje

La señalización de instalaciones de peaje y pesaje comprende un sistema especializado de elementos informativos y reguladores proyectados específicamente para orientar, dirigir y regular el flujo vehicular en las proximidades e instalaciones de puntos de recaudación de peajes e instalaciones de verificación de peso y dimensiones vehiculares.

Suministran instrucciones inequívocas sobre la aproximación imperativa, elección de carriles apropiados según el tipo de unidad vehicular y metodología de pago, velocidades autorizadas en la zona de operación, restricciones especiales para unidades vehiculares de carga, metodologías de detención y avance, además de información sobre tarifas aplicables y mecanismos de pago electrónico disponibles.

Utilizan configuraciones rectangulares con códigos cromáticos específicos, iconografía universalmente reconocida y leyendas en múltiples idiomas cuando resulte necesario, asegurando la operación protegida, eficaz y estructurada de estas instalaciones fundamentales para la administración del tráfico y la regulación del transporte de mercancías.

5. Señalización para vías de escape

Los dispositivos para vías de escape constituyen elementos de protección vial especializados que identifican, señalizan y suministran instrucciones específicas sobre la localización, acceso y empleo correcto de las calzadas de escape o vías de frenado de emergencia.

Se implementan estratégicamente en tramos de autopista con inclinaciones pronunciadas, curvas peligrosas o sectores de elevada altitud donde las unidades vehiculares, particularmente las de gran tonelaje, pueden experimentar deficiencias en los mecanismos

de frenado o pérdida de control. Indican inequívocamente la distancia a la próxima vía disponible, las circunstancias de acceso, restricciones de empleo exclusivo para situaciones de emergencia genuina, velocidades recomendadas de ingreso, atributos técnicos de la superficie de frenado y advertencias sobre las repercusiones del empleo indebido.

Emplean coloraciones de elevada perceptibilidad como amarillo fluorescente o naranja, iconografía gráfica explicativa y materiales retroreflectantes de máxima intensidad para asegurar su detección oportuna bajo cualquier circunstancia climática o de visibilidad, constituyendo componentes vitales para la evitación de siniestros severos en áreas de topografía compleja.

Figura 14
Señalización para vías de escape



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

2.3. Visión aritificial

2.3.1. Aprendizaje Automático y Aprendizaje Profundo

1. Conceptualización del Aprendizaje Automático

El Aprendizaje Automático o Machine Learning (ML) constituye una subdisciplina de la inteligencia artificial que confiere a los sistemas informáticos la capacidad de adquirir conocimientos y perfeccionar su rendimiento de forma autónoma mediante experiencias acumuladas, prescindiendo de programación explícita para cada labor específica. Según Géron (2019), esta área disciplinar se concentra en crear algoritmos con capacidad de acceder a conjuntos informacionales y utilizarlos como fundamento para un proceso de aprendizaje independiente.

Los sistemas informáticos poseen la capacidad intrínseca de desarrollar mecanismos de autoaprendizaje y autoprogramación mediante la asimilación de experiencias anteriores,

perfeccionando constantemente su desempeño fundamentándose en resultados históricos de manera completamente automatizada (Lasse (2018).)

2. Clasificación de metodologías de aprendizaje en Machine Learning

Según lo determinado por Lasse (2018), el procedimiento de aprendizaje automatizado en arquitecturas neuronales puede ejecutarse mediante distintas metodologías que se organizan en tres aproximaciones fundamentales: supervisada, no supervisada y por refuerzo.

- **Aprendizaje Supervisado:** Esta metodología utiliza conjuntos informacionales previamente rotulados para que el algoritmo determine correspondencias entre variables de entrada y productos esperados. Las implementaciones principales comprenden labores de clasificación y regresión.
- **Aprendizaje No Supervisado:** Este enfoque procesa datos sin etiquetado previo, revelando estructuras y configuraciones intrínsecas en la información. Sus implementaciones habituales abarcan agrupamiento informacional (clustering), reducción de dimensionalidad y detección de irregularidades.
- **Aprendizaje por Refuerzo:** El sistema construye conocimiento mediante interacción constante con un entorno dinámico, recibiendo señales de retroalimentación favorable o desfavorable según la calidad de sus elecciones.

3. Aprendizaje Profundo - Conceptos Fundamentales

El Aprendizaje Profundo o Deep Learning (DL) representa una especialización del Machine Learning que utiliza arquitecturas de redes neuronales artificiales con múltiples capas (denominadas ”profundas”por su configuración estructural) para derivar representaciones y atributos de nivel superior a partir de información compleja Goodfellow et al. (2016).

4. Beneficios del Aprendizaje Profundo

- **Derivación automática de atributos:** Elimina la necesidad de extracción manual de características, generando representaciones jerárquicas directamente desde los datos primarios.
- **Potencial de escalabilidad:** El rendimiento se eleva proporcionalmente con mayores cantidades de información y potencia computacional, contrastando con otros algoritmos que muestran estancamiento en su desempeño.
- **Flexibilidad:** Los modelos previamente entrenados pueden ajustarse efectivamente para nuevas labores relacionadas mediante técnicas de transferencia de aprendizaje.
- **Polivalencia:** Se ajusta a diversas tipologías informacionales (imágenes, texto, audio, secuencias temporales) empleando arquitecturas especializadas.

2.3.2. Redes Neuronales Artificiales (RNA)

1. Conceptualización y paralelismo biológico

La intrincada y cautivadora arquitectura del cerebro humano, conformada por billones de neuronas interconectadas, exhibe la capacidad de transmitir información entre sus numerosas células nerviosas mediante el fenómeno sináptico. Las neuronas biológicas se conforman por tres elementos esenciales: el soma o estructura celular, las dendritas que

funcionan como receptores de señales entrantes, y el axón que actúa como transmisor de información de salida. El mecanismo sináptico acontece en el punto de contacto entre una dendrita y el axón de otra neurona. Cuando se presenta un estímulo, ya sea eléctrico o de naturaleza alternativa, en estas conexiones se generan intercambios de sustancias, reacciones bioquímicas y liberación de neurotransmisores. Estos compuestos químicos, dependiendo de su naturaleza, poseen la capacidad de estimular o inhibir las células receptoras Becerra (2014).

El paralelismo biológico se fundamenta en:

- Las neuronas artificiales simulan el comportamiento de las neuronas biológicas.
- Las conexiones ponderadas simbolizan las sinapsis neuronales.
- La función de activación replica el potencial de acción de las neuronas.
- La estructuración en capas refleja la configuración de distintas regiones cerebrales.

2. La unidad neuronal artificial

Las arquitecturas neuronales artificiales constituyen paradigmas computacionales fundamentados en la estructura y funcionamiento del cerebro humano, conformados por estratos de neuronas interconectadas. Estos sistemas se utilizan para procesar información, reconocer configuraciones complejas en conjuntos informacionales y ejecutar predicciones, desarrollándose como simulación del funcionamiento y conducta de las arquitecturas neuronales biológicas. Las ANA se estructuran en estratos y desarrollan la habilidad de reconocer relaciones y configuraciones complejas, para posteriormente aproximarlos a funciones matemáticas y producir predicciones de productos Vector ITC (2018).

- **Entradas:** Magnitudes numéricas procedentes de los datos o de otras neuronas.
- **Pesos:** Magnitudes que establecen la relevancia relativa de cada entrada.
- **Función de agregación:** Generalmente una sumatoria ponderada de las entradas y sus pesos correspondientes.
- **Bias o sesgo:** Un componente adicional que posibilita ajustar el umbral de activación.
- **Función de activación:** Transforma la salida de la función de agregación, introduciendo no-linealidad en el paradigma.

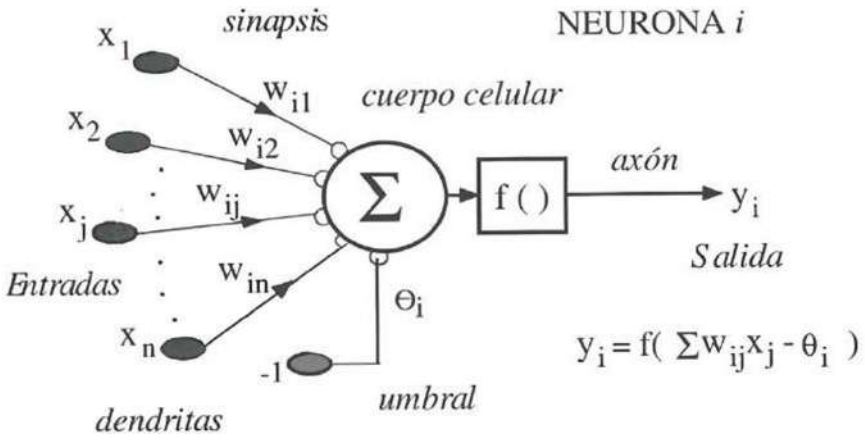
Matemáticamente, el resultado de una neurona puede expresarse como:

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{1}$$

Donde:

- x_i son las entradas.
- w_i son los pesos.
- b es el bias.
- f es la función de activación.

Figura 15
Paradigma matemático de una neurona artificial



Nota. Elaboración propia.

3. Funciones de activación Las funciones de activación constituyen elementos esenciales que determinan si una neurona debe activarse, introduciendo no-linealidades fundamentales en el sistema. Las funciones de activación más empleadas incluyen:

- **Sigmoide:** $f(x) = \frac{1}{1+e^{-x}}$. Transforma magnitudes en el intervalo [0,1].
- **Tanh:** $f(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$. Transforma magnitudes en el intervalo [-1,1].
- **ReLU (Rectified Linear Unit):** $f(x) = \max(0, x)$. Anula magnitudes negativas; ha evidenciado acelerar la convergencia.
- **Leaky ReLU:** $f(x) = \max(\alpha x, x)$ donde α es reducido. Permite un pequeño gradiente para magnitudes negativas.
- **ELU (Exponential Linear Unit):** Combina las ventajas de ReLU mientras atenúa el problema de "neuronas inactivas".
- **Softmax:** Empleada en el estrato de salida para problemáticas de clasificación multiclase, transforma un vector de magnitudes en una distribución probabilística.

4. Mecanismo de Entrenamiento: Propagación directa e inversa

El entrenamiento de arquitecturas neuronales incorpora dos mecanismos fundamentales, según describen Chollet y Allaire (2018):

- **Propagación directa (Forward Propagation):**
  - Las entradas se introducen en la arquitectura y se propagan a través de todos los estratos.
  - Cada neurona ejecuta su función de transformación correspondiente.
  - Se produce una predicción en el estrato de salida.
- **Retropropagación (Backpropagation):**
  - Se calcula el error entre la predicción y la magnitud objetivo.
  - Este error se propaga inversamente a través de la arquitectura de red.
  - Los pesos se ajustan proporcionalmente a su contribución al error.
  - Se emplea el descenso del gradiente para minimizar la función de pérdida.

2.3.3. Taxonomía de arquitecturas neuronales

Existen diversas tipologías de arquitecturas neuronales, cada una proyectada para resolver problemáticas específicas. A continuación se presentan las más relevantes:

1. Redes Neuronales Feedforward (FFNN)

Las arquitecturas feedforward o de alimentación directa constituyen la configuración más elemental, donde la información fluye unidireccionalmente, desde la entrada hacia la salida, sin ciclos ni retroalimentación. Según Da Silva et al. (2017), estas arquitecturas se constituyen por:

- Un estrato de entrada que recibe los datos iniciales.
- Uno o múltiples estratos ocultos que procesan la información.
- Un estrato de salida que produce el resultado final.

2. Redes Neuronales Convolucionales (CNN)

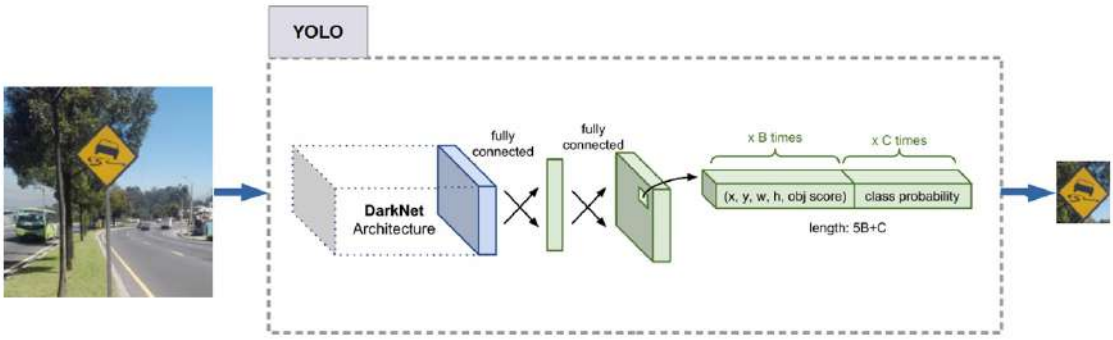
Estas arquitecturas fueron proyectadas para replicar el funcionamiento de las neuronas de la corteza visual cerebral, proporcionando a los sistemas computacionales la habilidad de ”percibir visualmente”. Esta característica las ha posicionado como una de las técnicas más empleadas en el ámbito del Aprendizaje Automático para resolver problemáticas de visión artificial, mediante el reconocimiento de entidades en imágenes y videos Núñez (2016).

Las CNN constituyen una clase de arquitectura que funciona bajo un paradigma de aprendizaje supervisado e incorpora múltiples estratos ocultos, estructurados jerárquicamente y especializados en funciones específicas. Como ilustración, los estratos iniciales pueden detectar componentes básicos como líneas, curvas y formas, mientras que los estratos más profundos se especializan en el reconocimiento de configuraciones considerablemente más complejas, como rostros humanos, vehículos, animales y diversas entidades.

Una CNN típica incorpora:

- **Estratos convolucionales:** Implementan filtros (kernels) sobre la entrada para derivar características.
- **Funciones de activación:** Generalmente ReLU, que introduce no-linealidad.
- **Estratos de pooling:** Reducen la dimensionalidad espacial (submuestreo), preservando la información más pertinente.
- **Estratos completamente conectados:** Ejecutan la clasificación fundamentada en las características derivadas.

Figura 16
Configuración de una red neuronal convolucional (CNN)



Nota. Adaptado de Flores-Calero et al. (2024).

3. Redes Neuronales Recurrentes (RNR)

Las Arquitecturas Neuronales Recurrentes se especializan en procesar datos secuenciales, como secuencias temporales o texto. Según Yu et al. (2019), las RNN incorporan conexiones recurrentes que posibilitan la persistencia de información, creando una forma de memoria que captura dependencias temporales.

4. Arquitecturas neuronales Autoencoders

Los autoencoders constituyen arquitecturas neuronales proyectadas para aprender representaciones comprimidas (codificaciones) de datos. Bank et al. (2021) describen su arquitectura:

- Un codificador que comprime los datos de entrada.
- Un cuello de botella que contiene la representación de dimensión reducida.
- Un decodificador que reconstruye los datos originales.

5. Redes Generativas Adversarias (GAN)

Introducidas por Goodfellow et al. (2016), las GAN consisten en dos arquitecturas que compiten entre sí:

- El generador que crea muestras sintéticas.
- El discriminador que intenta distinguir entre muestras reales y generadas.

Esta competencia adversarial conduce al generador a producir datos indistinguibles de los reales. Las GAN han revolucionado la generación de imágenes, video, música y otros tipos de contenido.

6. Redes neuronales Transformers

Los Transformers, introducidos por Vaswani et al. (2017), representan una arquitectura revolucionaria fundamentada en mecanismos de atención que ha superado a las RNN en labores de procesamiento de lenguaje natural y recientemente en visión computacional.

Las características fundamentales de los transformers incluyen:

- Mecanismos de auto-atención que posibilitan modelar dependencias globales.
- Procesamiento paralelo que acelera significativamente el entrenamiento.

- Arquitectura codificador-decodificador para labores de secuencia a secuencia.

2.3.4. Profundización en las Arquitecturas Neuronales Convolucionales (CNN):

Operación de las CNN

Las CNN procesan una entrada constituida por una imagen. Esta representación visual puede modelarse como una matriz multidimensional que se procesa mediante una secuencia de filtros denominados kernels.

Una imagen exhibe ancho (correspondiente al número de columnas) y altura (equivalente al número de filas), análogo a una matriz. Sin embargo, a diferencia de las matrices convencionales, las imágenes incorporan una dimensión adicional llamada profundidad, establecida por sus canales cromáticos que contienen las magnitudes numéricas de los píxeles. Por ejemplo, una imagen a color en formato RGB (rojo, verde, azul) exhibe una profundidad de tres canales, uno para cada color primario, mientras que una imagen en escala de grises posee un canal único.

Las magnitudes píxel en estas imágenes típicamente oscilan entre 0 y 255, y para optimizar el procesamiento, es recomendable normalizarlas a un intervalo entre 0 y 1, dividiéndolas por 255 Nizama Yamunaque (2022).

2.3.5. Operación convolucional y Kernels

Una operación convolucional implica la agregación de componentes de una matriz producto, obtenida mediante la multiplicación componente por componente de dos matrices: una que simboliza una sección de la imagen original y otra correspondiente a un kernel. Un kernel o filtro puede conceptualizarse como una matriz de dimensiones reducidas (en relación con la matriz de imagen) empleado para diversas operaciones como enfocar, desenfocar, suavizar, o detectar atributos como líneas, formas y bordes.

Este kernel se desplaza secuencialmente, píxel por píxel, sobre todas las neuronas del estrato de entrada, desde izquierda a derecha y de arriba hacia abajo, generando una nueva matriz de salida que constituye un estrato oculto neuronal.

La operación convolucional constituye el componente esencial de las CNN. Según Yamashita et al. (2018), esta operación comprende:

- Un kernel (o filtro) que se desliza sobre la imagen de entrada.
- Para cada posición, se calcula el producto escalar entre el kernel y la región correspondiente de la entrada.
- El producto forma un mapa de atributos que resalta configuraciones específicas.

Matemáticamente, para una imagen 2D, la convolución discreta se expresa como:

$$(I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) \cdot K(m,n) \tag{2}$$

Donde:

- I es la imagen de entrada.
- K es el kernel o filtro.
- $*$ denota la operación de convolución.

Distintos kernels detectan distintos atributos:

- Kernels de bordes detectan transiciones en la intensidad.
- Kernels de textura identifican configuraciones repetitivas.
- Kernels de blur suavizan detalles.
- **Inicialización de kernels:** Los kernels se inicializan aleatoriamente y se ajustan durante el entrenamiento.
- **Implementación del kernel:** El kernel se desliza sobre la entrada, calculando el producto escalar en cada posición.
- **Conformación del mapa de atributos:** Los productos se organizan en un nuevo tensor llamado mapa de atributos.
- **Implementación de bias y activación:** Se añade un término de sesgo y se implementa una función de activación.

2.3.6. Función de activación-ReLu

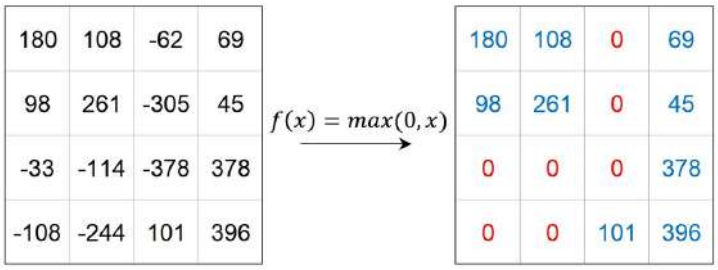
De modo análogo al mecanismo sináptico en sistemas biológicos, las Arquitecturas Neuronales Artificiales (ANA) incorporan una función de activación que modifica la magnitud neuronal con el propósito de activarla y transmitir un producto a la conexión subsiguiente, equivalente a la sinapsis.

Esta función establece si la magnitud neuronal, simbolizada por las magnitudes de la matriz producto, supera un umbral determinado para ser transmitida al estrato siguiente. En el contexto de las CNN, la función de activación que ha evidenciado mayor eficacia y mejores productos es la Unidad Lineal Rectificada, conocida como ReLU (Rectified Linear Unit) Ramachandran et al. (2017), conceptualizada matemáticamente como:

$$f(x) = \max(0, x) = \begin{cases} 0 & \text{, si } x < 0 \\ x & \text{, si } x \geq 0 \end{cases} \tag{3}$$

La figura N° 17 ilustra la implementación de la función de activación ReLU a la matriz de salida presentada. La función ReLU opera transformando la magnitud de cada neurona (o píxel): Si la magnitud recibida como entrada es negativa, la convierte a cero; en caso contrario, la magnitud se mantiene sin modificaciones. En el ejemplo de la figura N° 17, las magnitudes de entrada negativas fueron reemplazadas por cero (0), mientras que las demás magnitudes permanecieron intactas.

Figura 17
Implementación de la función de activación ReLU



Nota. Adaptado de Nizama Yamunaque (2022).

2.3.7. Pooling- Consolidación

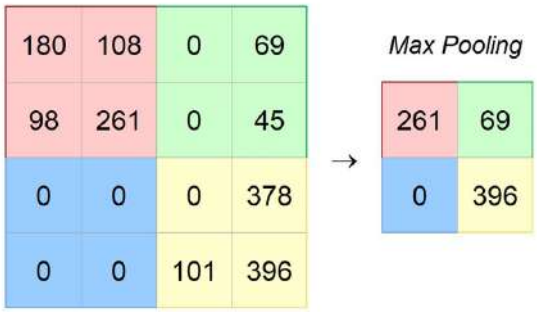
Constituye una técnica fundamental para reducir las dimensiones espaciales de una imagen. Su objetivo fundamental consiste en disminuir la cantidad de neuronas (píxeles) y simplificar los cálculos antes de una nueva convolución. Este procedimiento involucra la consolidación de múltiples píxeles en uno solo, conservando la información más pertinente.

Existen distintas variantes de pooling, abarcando el Max Pooling (que selecciona la magnitud máxima dentro de una subregión píxel), el Min Pooling (que elige la magnitud mínima) y el Average Pooling (que computa la magnitud promedio).

Sin embargo, el Max Pooling sobresale como una técnica de submuestreo altamente efectiva en las Arquitecturas Neuronales Convolucionales (ANC) para comprimir los atributos de una región. Esta operación consiste en identificar la magnitud máxima dentro de una muestra y emplear dicha magnitud como resumen representativo de los atributos presentes en esa zona. Al implementar un Max Pooling de 2x2, se logra reducir el tamaño de la imagen a la mitad.

La figura N° 18 ilustra un ejemplo de la implementación de un Max Pooling de 2x2 a la matriz de salida obtenida en la Figura N° 17. Para la primera región de 2x2 (señalada en rojo), la magnitud máxima es 261; para la segunda región (en verde), es 69; para la tercera región (en azul), es 0; y para la cuarta región (en amarillo), es 396. Como producto de esta operación, la imagen de entrada de 4x4 se reduce a una imagen de 2x2.

Figura 18
Implementación de un Max Pooling de 2×2



Nota. Adaptado de Nizama Yamunaque (2022).

En síntesis, después de implementar un kernel de 3x3 y un Max Pooling de 2x2 a nuestra imagen original de 6x6, se produce una imagen resultante de 2x2 que encapsula información procesada y pertinente de la imagen inicial. Esta información resulta esencial para la identificación y el reconocimiento de configuraciones de formas u entidades. Como se mencionó previamente, es posible implementar múltiples kernels (filtros) a una misma imagen de entrada.

Algunos de estos filtros pueden especializarse en la detección de bordes, otros en sombras, otros en curvas, otros en líneas horizontales o verticales, y así sucesivamente, dependiendo de su función específica. La implementación de múltiples kernels a una imagen generará una cantidad equivalente de matrices de salida (mapas de atributos), lo que incrementará la profundidad de nuestros estratos neuronales.

2.3.8. Síntesis del mecanismo de una CNN

Una vez comprendido el procedimiento y la operativa de las Arquitecturas Neuronales Convolucionales (ANC), se puede determinar que una convolución dentro de una ANC se constituye por una imagen de entrada, un conjunto de filtros (kernels), la implementación de una función de activación y una operación de Max-Pooling.

A medida que se ejecutan más convoluciones, se generan estratos neuronales progresivamente más profundos, con la capacidad de identificar configuraciones de formas de mayor complejidad.

Este mecanismo iterativo continúa hasta que el estrato final de la arquitectura puede clasificar las imágenes detectadas en categorías o grupos específicos, asignándoles una probabilidad de certeza.

El cuadro 1 presenta los productos de implementar una secuencia de tres convoluciones a una imagen de 48x48 píxeles en escala de grises (representada como 48x48x1). Para la primera convolución, se emplearon 15 filtros de 3x3; para la segunda, 30 filtros de 3x3; y para la tercera, 45 filtros de 3x3. Adicionalmente, se implementó una operación de Max Pooling de 2x2 después de cada estrato convolucional.

Cuadro 1: Mecanismo de convoluciones a una imagen de 48x48 píxeles

DESCRIPCIÓN	1era Convolución	2da Convolución	3era Convolución
1) Entrada: Imagen	48x48x1	24x24x15	12x12x30
2) Kernel (filtros)	15 filtros 3x3x1	30 filtros 3x3x15	45 filtros 3x3x30
3) Feature Mapping	48x48x15	24x24x30	12x12x45
4) Func. de Activación	ReLU	ReLU	ReLU
5) Max Pooling	de 2x2	de 2x2	de 2x2
6) Salida: Convolución	24x24x15	12x12x30	6x6x45

Fuente: Elaboración propia

La imagen original de 48x48 píxeles en escala de grises (48x48x1) contiene 2,304 unidades informacionales o neuronas. En el procedimiento del primer estrato convolucional, se implementaron 15 filtros de tamaño 3x3, generando 15 distintos mapas de atributos (48x48x15). Tras implementar Max Pooling 2x2, estas representaciones se redujeron a dimensiones de 24x24 para cada uno de los 15 mapas, resultando en un total de 8,640 neuronas activas.

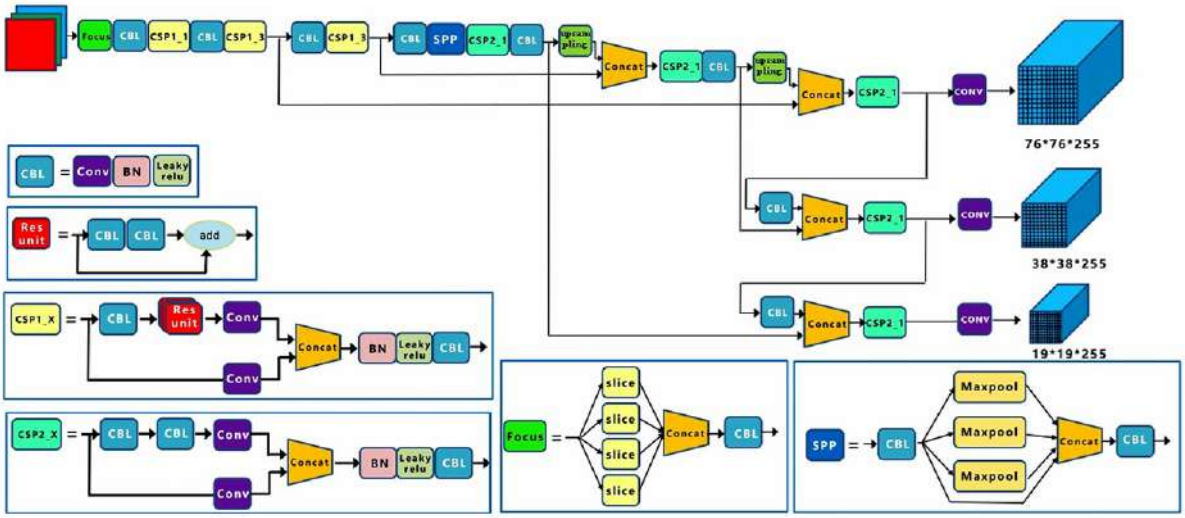
Para la segunda etapa convolucional, estos 15 mapas de 24x24 píxeles fueron procesados con 30 filtros de 3x3, creando 30 nuevos mapas de atributos (24x24x30). Al implementar otro Max Pooling 2x2, el tamaño se redujo a 12x12 píxeles por mapa, condensando la información en 4,320 neuronas. Este paso evidencia una disminución notable del 50 % en la cantidad de unidades procesadas respecto a la etapa anterior.

En la tercera fase, los 30 mapas de 12x12 píxeles fueron tratados con 45 filtros de 3x3, generando 45 mapas de atributos (12x12x45). El último Max Pooling 2x2 los transformó en representaciones de 6x6 píxeles cada uno (6x6x45), totalizando 1,620 neuronas.

Este mecanismo demuestra cómo la arquitectura convolucional logra comprimir la información original de 2,304 neuronas a 1,620 neuronas finales, conservando únicamente los componentes más distintivos y significativos.

Estas neuronas resultantes contienen datos fundamentales para la identificación de atributos específicos (como rasgos faciales o componentes de vehículos) y el reconocimiento de entidades complejas presentes en la imagen original.

Figura 19
Ilustración de los estratos de consolidación



Nota. Adaptado de Wang2025

2.3.8.1 Indicadores de evaluación CNN

1. Matriz de confusión e indicadores derivados

La matriz de confusión constituye el fundamento para múltiples indicadores de evaluación en problemáticas de clasificación.

De acuerdo con Minaee et al. (2021), esta matriz proporciona una visión comprehensiva del rendimiento del paradigma al mostrar la relación entre las predicciones y las etiquetas reales.

A partir de la matriz de confusión, se derivan los siguientes indicadores fundamentales:

2. Exactitud (Accuracy)

La exactitud cuantifica la proporción de predicciones correctas sobre el total de instancias evaluadas. Aunque es ampliamente empleada, presenta limitaciones significativas en conjuntos informacionales desbalanceados Brownlee (2020).

$$\text{Exactitud} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Donde:

- TP (Verdaderos Positivos): Instancias positivas correctamente clasificadas.
- TN (Verdaderos Negativos): Instancias negativas correctamente clasificadas.
- FP (Falsos Positivos): Instancias negativas incorrectamente clasificadas como positivas.
- FN (Falsos Negativos): Instancias positivas incorrectamente clasificadas como negativas.

3. Precisión: La precisión establece la proporción de predicciones positivas que son

realmente correctas, siendo especialmente pertinente en escenarios donde los falsos positivos tienen un alto costo Evidentlyai (2023).

$$\text{Precisión} = \frac{TP}{TP + FP} \tag{5}$$

4. Sensibilidad (Recall): La sensibilidad, también conocida como recall o tasa de verdaderos positivos, evalúa la capacidad del paradigma para identificar correctamente las instancias positivas Google Developers (2023).

$$\text{Sensibilidad} = \frac{TP}{TP + FN} \tag{6}$$

4. Especificidad: La especificidad cuantifica la proporción de negativos reales que son correctamente identificados como tales.

Este indicador es particularmente importante en aplicaciones médicas donde la identificación precisa de casos negativos es crucial Effective Health Care Program (2022).

$$\text{Especificidad} = \frac{TN}{TN + FP} \tag{7}$$

5. F1-Score: El F1-Score simboliza la media armónica entre precisión y sensibilidad, proporcionando un equilibrio entre ambos indicadores. Según The Data Scientist (2023), esta medida es especialmente útil cuando existe un desbalance significativo en las clases.

$$\text{F1-Score} = 2 \times \frac{\text{Precisión} \times \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}} \tag{8}$$

2.3.8.2 Algoritmos de detección

La detección de entidades consiste en identificar la presencia y ubicación de entidades específicas en una imagen, generalmente mediante la predicción de cuadros delimitadores (bounding boxes).

1. Metodologías Tradicionales

Los enfoques tradicionales se fundamentan en la extracción manual de atributos y clasificadores:

- Viola-Jones.
- HOG (Histogram of Oriented Gradients).

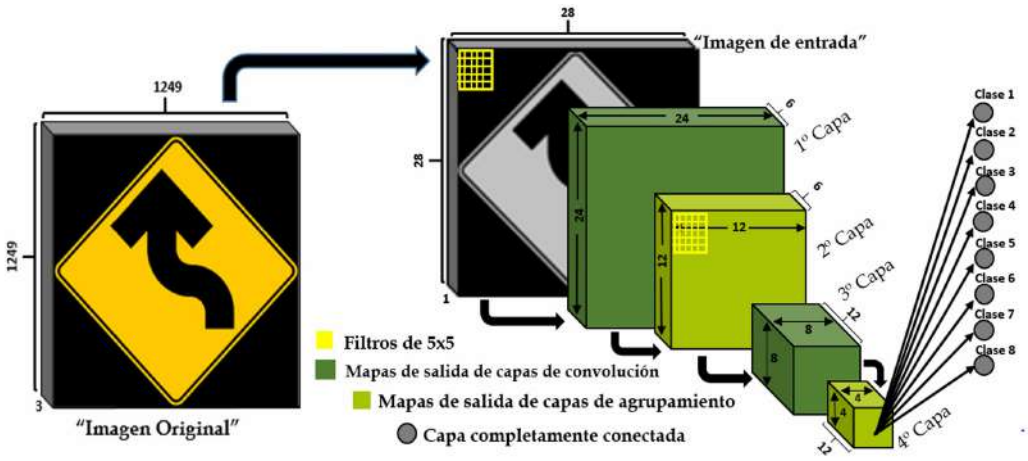
2. Metodologías fundamentadas en aprendizaje profundo

A. Detectores de una etapa

Estos algoritmos predicen directamente las ubicaciones y clases de las entidades en una sola pasada:

- **YOLO-You Only Look Once:** Divide la imagen en una cuadrícula y predice bounding boxes y probabilidades de clase para cada celda.

Figura 20
Simbolización de la arquitectura YOLO v8



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

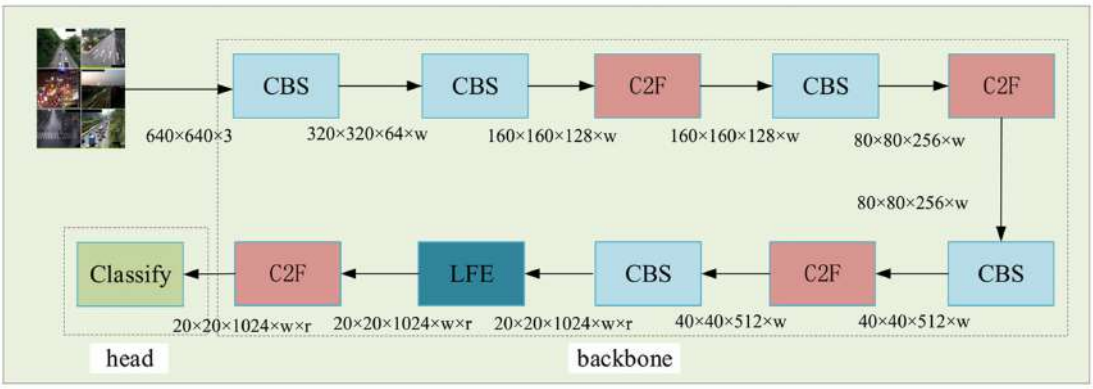
- **SSD-Single Shot MultiBox Detector:** Emplea múltiples estratos de distintas escalas para detectar entidades de diversos tamaños.
- **RetinaNet:** Introduce la función de pérdida Focal Loss para afrontar la problemática del desequilibrio de clases. Eficaz en conjuntos informacionales con distribuciones altamente asimétricas.

B. Detectores de dos etapas

Estos paradigmas primero proponen regiones candidatas y luego clasifican cada región:

- **R-CNN (Region-based CNN)** y sus variantes Fast R-CNN y Faster R-CNN: Emplean arquitecturas de proposición de regiones (RPN) para identificar posibles ubicaciones de entidades. Analizan la evolución de estos paradigmas y su impacto en la precisión de detección.

Figura 21
Simbolización de la arquitectura Faster-RCNN



Nota. Adaptado de Wang, Shang et al. (2025).

- **Mask R-CNN:** Extiende Faster R-CNN para ejecutar segmentación de instancias simultáneamente. Sumamente versátil en aplicaciones que requieren tanto detección como segmentación precisa.

2.3.9. YOLO: You Only Look Once

1. Fundamentos Esenciales de YOLO

YOLO (You Only Look Once) simboliza un algoritmo revolucionario de detección de entidades en tiempo real desarrollado inicialmente por Joseph Redmon et al. en 2015.

Contrario a las metodologías tradicionales de dos etapas, aborda la detección como una problemática de regresión directa donde una única arquitectura neuronal convolucional procesa la imagen completa en una sola pasada para predecir simultáneamente las cajas delimitadoras y las probabilidades de clase.

Divide la imagen en una cuadrícula de celdas donde cada celda es responsable de detectar las entidades cuyo centro cae dentro de ella, lo que posibilita una alta velocidad de procesamiento (45-155 fotogramas por segundo en sus primeras versiones) mientras mantiene una precisión competitiva.

Estas características han convertido a YOLO en uno de los algoritmos más populares para aplicaciones que requieren detección de entidades en tiempo real como vehículos autónomos, videovigilancia y robótica Redmon et al. (2016).

- **Aproximación de un solo paso:** Ejecuta detección y clasificación simultáneamente.
- **División en cuadrícula:** Divide la imagen en una cuadrícula S×S.
- **Predicciones por celda:** Cada celda predice B cajas delimitadoras con confianza y probabilidades de clase.
- **Pérdida unificada:** Optimiza conjuntamente localización, confianza y clasificación.

2. Evolución hasta YOLOv8 La arquitectura YOLO ha evolucionado significativamente desde su versión original:

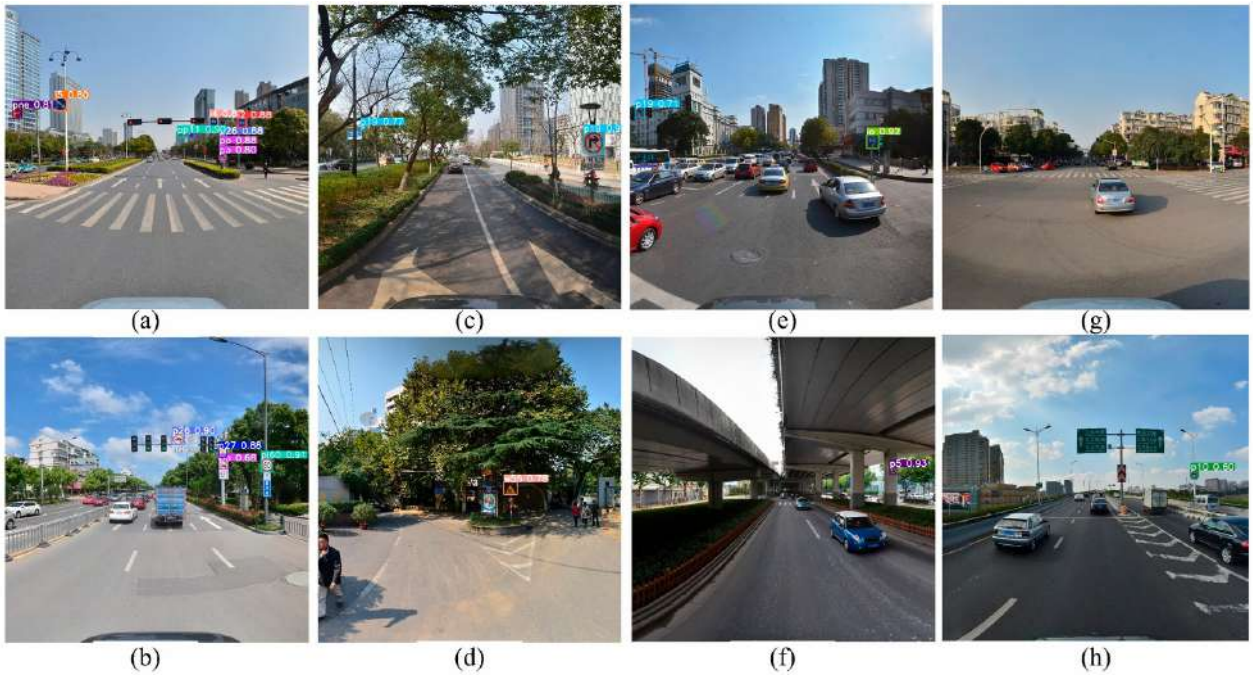
- **YOLOv1-v3:** Desarrolladas por Joseph Redmon, establecieron los fundamentos.
- **YOLOv4-v7:** Impulsadas por Alexey Bochkovskiy y otros, incorporaron avances como CSPNet y transformers.
- **YOLOv8:** Desarrollada por Ultralytics, simboliza la generación más reciente con importantes mejoras arquitectónicas.

3. Arquitectura y operación de YOLOv8 YOLOv8, lanzada en 2023, introduce varias innovaciones significativas Jocher et al. (2023):

- **Backbone perfeccionado:** Emplea una arquitectura CSPDarknet más eficaz con bloques residuales avanzados.
- **Neck optimizado:** Implementa PANet (Path Aggregation Network) para mejor fusión multiescala.
- **Cabezas de detección anchorless:** Elimina las anchor boxes predefinidas en favor de predicciones directas.
- **C2f Block:** Nueva unidad de construcción que perfecciona la eficacia computacional.
- **Funciones de pérdida avanzadas:** Incorpora DFL (Distribution Focal Loss) y variantes de IoU que perfeccionan la precisión de localización.

Figura 22

Detección de señalizaciones con la arquitectura YOLO v8



Nota. Adaptado de Ming et al. (2025).

Figura 23

Detección de señalizaciones con YOLO v8 en variabilidad de condiciones



Nota. Adaptado de Wang, Zheng & Li (2025).

3. MARCO METODOLÓGICO

3.1. Enfoque de la investigación

La presente investigación se clasifica como **cuantitativa** porque se fundamenta en la medición objetiva y el análisis estadístico de datos numéricos verificables. El estudio procesa un dataset de 6,213 imágenes digitales con 7,107 instancias de señales anotadas mediante coordenadas de segmentación poligonal, y desarrolla modelos de redes neuronales convolucionales optimizados matemáticamente: YOLO26s-seg entrenado durante 100 épocas y EfficientNet-B0 entrenado mediante transferencia de aprendizaje en dos etapas sobre 7 categorías de estado de conservación.

El rendimiento del sistema se evalúa exclusivamente mediante métricas estandarizadas **cuantificables**: mAP@50 de 0.9305 (detección) y 0.9224 (segmentación) para YOLO26s-seg, Accuracy de 0.9315 y Macro F1 de 0.8252 para EfficientNet-B0, así como funciones de pérdida numéricas (box_loss, seg_loss, cls_loss) que miden el error del modelo en cada iteración de entrenamiento.

3.2. Alcance de la investigación

La presente investigación se clasifica con un alcance **explicativo** porque establece y demuestra relaciones causales entre el diseño del sistema de visión artificial y su capacidad para detectar, segmentar y clasificar el estado de conservación de señales de tránsito verticales, explicando por qué y cómo cada componente del pipeline logra sus resultados. El estudio fundamenta los mecanismos causales que determinan el rendimiento: La arquitectura YOLO26s-seg entrenada mediante optimización AdamW durante 100 épocas sobre 6,213 imágenes con aumentación en dos capas, desarrolla capacidades de extracción de características visuales jerárquicas que permiten detectar y segmentar las 31 categorías de señales verticales con un mAP@50 de 0.9305.

Se establece la relación causa-efecto entre variables independientes (arquitectura del modelo, tamaño del dataset, técnicas de aumentación offline, épocas de entrenamiento, hiperparámetros de regularización) y variables dependientes (mAP@50 = 0.9305, Macro F1 = 0.8252, Accuracy = 0.9315), explicando matemática y computacionalmente cómo cada componente contribuye al resultado final.

3.3. Diseño de la investigación

La presente investigación se clasifica con un diseño **no experimental** porque no se manipulan deliberadamente las variables independientes en condiciones controladas de laboratorio, sino que se observan, registran y analizan los fenómenos tal como se presentan en su contexto natural real. En este estudio, las señales de tránsito de los tramos PE-3S (Ayacucho–Huanta) y PE-28A (Ayacucho–Abra Apacheta) fueron capturadas fotográficamente y mediante videograbación en su estado natural existente, sin intervención ni modificación de las condiciones viales, ambientales o de señalización por parte del investigador.

El diseño no experimental se fundamenta en que el investigador no ejerce control sobre las variables de estudio: las señales de tránsito ya están instaladas en los tramos viales (no fueron colocadas experimentalmente), las condiciones de iluminación, clima y entorno son naturales (no se crearon artificialmente), y las 6,213 imágenes con 7,107 instancias capturadas representan situaciones reales del contexto vial sin manipulación experimental. El sistema SISEVIR fue desarrollado y entrenado con datos observacionales provenientes de la realidad existente en el tramo vial, no de escenarios experimentales controlados.

Asimismo, la evaluación del estado de conservación de las señales en sus siete categorías (OPERATIVO, DETERIORO_POSTE, ROTACIÓN, OBSTRUCCIÓN, DECOLORACIÓN, VANDALISMO y DETERIORO_TABLERO) se realizó sobre condiciones preexistentes observadas en campo, sin inducir artificialmente ningún tipo de deterioro. No se establecieron grupos de control ni experimentales, no se asignaron aleatoriamente tratamientos, y no se modificaron intencionalmente las condiciones viales para observar efectos causales mediante experimentación directa. El diseño no experimental permitió obtener datos representativos de la realidad vial de la región Ayacucho, garantizando la validez externa de los resultados obtenidos por el sistema SISEVIR.

3.4. Población y muestra

3.4.1. Población

La población del estudio está constituida por el conjunto total de señalizaciones verticales de tránsito presentes en los tramos viales PE-3S (Ayacucho–Huanta, 44.02 km) y PE-28A (Ayacucho–Abra Apacheta, ∼95.4 km) de la red vial nacional de la región Ayacucho, fotografiándose un total de 6,213 imágenes con 7,107 instancias anotadas. Esta población comprende todas las señales de tránsito reguladoras y preventivas instaladas a lo largo de ambos corredores viales, clasificadas en **31 categorías** según el Manual de Dispositivos de Control del Tránsito Automotor (2024).

Las categorías de señalización que conforman la población de estudio son:

Señales reguladoras:

- R-15, R-16: Señales de prohibición de adelantamiento y velocidad máxima.
- R-30: Señal de velocidad máxima permitida.
- R-37: Señal de control obligatorio.

Señales preventivas:

- P-1A, P-2A: Señal de curva pronunciada y curva abierta a la derecha.
- P-1B, P-2B: Señal de curva pronunciada y curva abierta a la izquierda.
- P-3A: Señal de curva y contracurva pronunciada a la derecha.
- P-4B: Señal de curva y contracurva a la izquierda.
- P-5-1: Señal de camino sinuoso a la derecha.
- P-5-2A: Señal de curva en U a la derecha.
- P-5-2B: Señal de curva en U a la izquierda.

- P-10A: Señal de empalme en ángulo agudo a la derecha.
- P-33A: Señal de proximidad de reductor de velocidad tipo resalto.
- P-34: Señal de proximidad de badén.
- P-48, P-49: Señales de cruce y proximidad de cruce peatonal.
- P-53: Señal de animales en vía.

3.4.2. Muestra

La muestra fue conformada mediante un proceso de captura fotográfica y videograbación sistemática de las señalizaciones presentes en ambos tramos viales, resultando en un dataset de **6,213 imágenes con 7,107 instancias** de señales anotadas mediante segmentación poligonal en LabelMe. Adicionalmente, para el modelo de clasificación EfficientNet-B0, se construyó un dataset secundario de **6,412 recortes** de señales etiquetados con su estado de conservación correspondiente.

Composición de la muestra — dataset YOLO26s-seg:

El dataset de detección y segmentación fue dividido en tres subconjuntos:

- Conjunto de entrenamiento: 4,442 imágenes (71.5 %), incluyendo 278 imágenes sintéticas generadas mediante aumentación offline para clases minoritarias.
- Conjunto de validación: 1,185 imágenes (19.1 %).
- Conjunto de test: 586 imágenes (9.4 %), con 672 instancias independientes para evaluación final.

Composición de la muestra — dataset EfficientNet-B0:

El dataset de clasificación del estado de conservación fue dividido en:

- Conjunto de entrenamiento: ∼3,570 muestras (55.7 %), con aumentación offline diferenciada por clase.
- Conjunto de validación: ∼800 muestras (12.5 %).
- Conjunto de test: 642 muestras (10.0 %) para evaluación final independiente.

Características de la muestra:

Las 6,213 imágenes capturadas presentan las siguientes características:

- **Variabilidad de condiciones de captura:** Las imágenes fueron obtenidas en diferentes horarios del día, incluyendo condiciones nocturnas, bajo variabilidad climática (lluvia, neblina, sol directo) y niveles de iluminación propios de la sierra ayacuchana, garantizando la exposición del modelo a diversos escenarios reales.
- **Diversidad de distancias y escalas:** Señales capturadas a diferentes distancias para entrenar la detección en múltiples escalas, desde señales en primer plano hasta señales visibles a más de 50 metros.
- **Condiciones de deterioro:** Inclusión de señales en siete estados de conservación distintos (OPERATIVO, DETERIORO_POSTE, ROTACIÓN, OBSTRUCCIÓN, DECOLORACIÓN, VANDALISMO y DETERIORO_TABLERO), representando las condiciones reales de la infraestructura vial del tramo.
- **Calidad técnica:** Imágenes procesadas a una resolución de entrada de 640×640 píxeles

para YOLO26s-seg y 224×224 píxeles para EfficientNet-B0, en formato compatible con los frameworks PyTorch y Ultralytics.

- **Anotación poligonal:** Cada señal fue etiquetada manualmente con su categoría correspondiente y coordenadas de segmentación poligonal mediante LabelMe con asistencia de Efficient SAM, para supervisar el entrenamiento del modelo de segmentación de instancias.

Criterios de selección:

El proceso de muestreo siguió un enfoque no probabilístico intencional, priorizando:

- Representatividad de las 31 clases de señalización definidas conforme al Manual de Dispositivos de Control del Tránsito Automotor (2024).
- Calidad visual suficiente para entrenamiento (nitidez, resolución, visibilidad de la señal).
- Diversidad de contextos, condiciones ambientales y estados de conservación del tramo vial.
- Cobertura de ambos tramos de estudio (PE-3S y PE-28A) en sentido IDA y VUELTA.

Esta muestra de 6,213 imágenes con 7,107 instancias constituye el dataset completo utilizado para el desarrollo, entrenamiento y validación del sistema SISEVIR, permitiendo a los modelos aprender los patrones visuales distintivos de cada tipo de señalización y su estado de conservación en condiciones reales de operación.

3.5. Hipótesis

3.5.1. Hipótesis general

Un modelo de visión artificial basado en redes neuronales puede ser desarrollado para automatizar de manera eficiente la clasificación y evaluación del estado de deterioro de las señales de tránsito en las vías nacionales de Huamanga, mejorando la precisión y reduciendo el tiempo necesario para el monitoreo en comparación con métodos tradicionales..

3.5.2. Hipótesis específicas

1. El uso de arquitecturas avanzadas de visión artificial, combinado con técnicas de optimización específicas, permitirá lograr una alta precisión en la clasificación y evaluación del estado de deterioro de las señales de tránsito en Huamanga.
2. Un programa automatizado que integre un modelo de visión artificial con características específicas de usabilidad y operatividad permitirá realizar la clasificación y evaluación del estado de deterioro de las señales de tránsito de manera más eficiente que los métodos manuales.
3. La clasificación y evaluación automatizada del estado de deterioro de las señales de tránsito mediante visión artificial será más eficiente en términos de tiempo y costo que los métodos tradicionales de inspección manual en la carretera Ayacucho-Huanta en 2024.

3.6. Operacionalización de variables

3.6.1. Variables

Caballero Romero (2014) define una variable como un conjunto de datos que comparten una característica, propiedad o atributo específico, lo que los incluye dentro del ámbito de esa variable. En el contexto estadístico, una variable se entiende como un símbolo que representa un conjunto determinado de datos, capaz de asumir cualquier valor dentro de ese conjunto.

Existen básicamente dos variables que participan de todo proceso de investigación conocidas como; variables independientes y dependientes.

- **Variable independiente (X): Sistema de visión artificial** : Es el factor que el investigador manipula o controla para observar su efecto en otra variable.
- **Variable dependiente (Y): Detección y evaluación de señales de tránsito** :Es el factor que el investigador mide y se espera que sea afectado por la manipulación de la variable independiente.

3.6.2. Dimensiones

Según Valencia Gutiérrez et al. (2023), el concepto de dimensiones se refiere a los componentes o aspectos específicos que conforman una variable compleja. Las dimensiones representan subdivisiones de un concepto o variable que permiten analizarla de manera más precisa y detallada.

- **Variable independiente (X): Sistema de visión artificial**
  - **Dimensión 1:** Arquitectura del modelo, Se refiere a la configuración específica de las capas, el tipo de neuronas y las conexiones entre ellas dentro de la Red Neuronal Artificial (RNA) para la detección y evaluación de señales de tránsito.
  - **Dimensión 1:** Diseño del modelo, Se refiere a la configuración específica de las capas, el tipo de neuronas y las conexiones entre ellas dentro de la Red Neuronal Artificial (RNA) para la evaluación de señales de tránsito.
  - **Dimensión 3:** Entrenamiento del modelo, Es el proceso iterativo mediante el cual se ajustan los parámetros internos (pesos y sesgos) de la Red Neuronal Artificial (RNA) utilizando un conjunto de datos etiquetados, a través de algoritmos de optimización y técnicas de regularización, con el objetivo de minimizar la función de pérdida y lograr un rendimiento óptimo en la tarea de detección de señales.
  - **Dimensión 3:** Resultados e implementación del modelo, Los resultados son las métricas cuantitativas y cualitativas (precisión, recall, F1-score, velocidad de inferencia) que miden el rendimiento de la RNA tras su entrenamiento para la detección y clasificación de señales de tránsito. La implementación es el proceso de integrar esta RNA entrenada en un sistema práctico, aplicando la arquitectura y parámetros optimizados para funcionar en situaciones reales de detección de señalizaciones de tránsito.

- **Variable dependiente (Y): Detección y evaluación de señales de tránsito**
  - **Dimensión 1:**Detección de señalización vial, Capacidad del sistema para identificar correctamente la presencia de señales de tránsito reguladoras y preventivas en el tramo vial, contribuyendo a la información oportuna sobre restricciones y advertencias.
  - **Dimensión 2:** Identificación de condiciones peligrosas,Aptitud del sistema para reconocer específicamente señales que alertan sobre situaciones de riesgo elevado en la vía, permitiendo anticipación de peligros y reducción de accidentes..
  - **Dimensión 3:** Confiabilidad del sistema, Grado de fiabilidad y consistencia del sistema en proporcionar información precisa sin generar alertas falsas ni omitir señales importantes, determinando la confianza del usuario en el sistema.

3.6.3. Indicadores

Caballero Romero (2014) define los indicadores como elementos especificadores y referenciales que ayudan a precisar, identificar, separar y usar los datos. Pueden medirse directamente de la realidad, por lo general como tasas, por ejemplo: Tasas de nacimientos, de mortalidad.

3.6.3.1 Indicadores independientes:

1. Arquitectura del modelo:
   - Número de capas de la red neuronal, [Número entero (número de capas-Conteo (n))].
   - Tipo de capas, [Categoría (tipos de capas)].
   - Tamaño de las capas, [Número entero (número de neuronas o filtros por capa)].
   - Función de activación, [Categoría (tipos de funciones)].
2. Diseño del modelo:
   - Tasa de aprendizaje, [Valor decimal (learning rate)].
   - Número de épocas de entrenamiento, [Número entero (número de épocas)].
   - Momento (momentum), [Valor decimal (momentum)].
   - Factor de decaimiento del aprendizaje. (learning rate decay), [Valor decimal (decay rate)].
3. Entrenamiento del modelo:
   - Algoritmo de optimización, [Categoría (SGD, Adam, RMSprop)].
   - Técnicas de regularización, [Categoría (dropout, L2 regularization, early stopping)].
   - Estrategias de aumento de datos, [Categoría (rotación, escalado, traslación)].
   - Esquema de aprendizaje, [Categoría (step decay, exponential decay, adaptive)].
   - Precisión en la validación y entrenamiento [Porcentaje (%)].
   - Pérdida en la validación y entrenamiento [Valor numérico (cross-entropy loss, mean squared error)].

- Tiempo total de entrenamiento [minutos (min)].
- Curva de aprendizaje [Gráfico (eje X: épocas, eje Y: precisión/pérdida)].
- Pesos del modelo Matriz de valores numéricos [(weights matrix)].
- Bias del modelo Vector de valores numéricos [(bias vector)].
- Sobreajuste (overfitting) [Porcentaje (%)].

4. Resultados e implementación del modelo:
   - Plataforma de Software, [Rendimiento (Porcentaje (%)).
   - Entorno de Simulación, [Categoría (sistema)].
   - Predicción de las señalizaciones, [Número de señales (señal/tipo)].
   - Clasificación automatizada de señales de acuerdo al manual del MTC, [Categórico].

3.6.3.2 Indicadores dependientes:

1. Detección de señalización vial:
   - Tasa de detección de señales reguladoras , [Porcentaje (%)].
   - Tasa de detección de señales preventivas, [Porcentaje (%)]
   - Tasa de detección de señales información, [Porcentaje (%)]
2. Identificación de condiciones peligrosas:
   - Tasa de detección de señales de alto riesgo, [Porcentaje (%)].
   - Precisión de clasificación por categoría de riesgo, [Porcentaje (%)].
   - Desempeño bajo condiciones ambientales variables, [Porcentaje (%)].
3. Confiabilidad del sistema:
   - Índice de equilibrio Precisión-Exhaustividad, [Decimal(0-1)].
   - Exactitud de segmentación de señales, [Porcentaje (%)].
   - Consistencia de detección , [Porcentaje (%)].

En los anexos se presenta la operacionalización de las variables identificadas, donde se detalla las variables, indicadores, dimensiones y sus unidades de medida.

3.7. Técnicas e instrumentos.

3.7.1. Técnicas de recolección

Para el desarrollo de la presente investigación se emplearon diversos métodos orientados a la recolección y análisis de la información, así como a la construcción y validación del modelo basado en redes neuronales convolucionales, utilizando la arquitectura YOLO. A continuación, se detallan las técnicas específicas aplicadas en cada etapa del estudio, desde el procesamiento de datos hasta la evaluación del desempeño del sistema propuesto.

1. **Revisión bibliográfica y documental:** La presente investigación se fundamenta en la revisión y análisis crítico de literatura especializada relacionada con visión artificial, arquitecturas de redes neuronales convolucionales (CNN) para la mejorar la

seguridad vial, como la detección de las señalizaciones verticales. Esta recopilación bibliográfica se llevó a cabo mediante el acceso a repositorios académicos y bases de datos científicas reconocidas, como IEEE Xplore, Google Scholar, Scopus, Springer y, complementariamente, Sci-Hub. Asimismo, se consideraron libros especializados, artículos técnicos y documentos normativos relevantes para el campo de la inteligencia artificial aplicada a la seguridad vial. Esta revisión permitió establecer un marco teórico sólido que orientó el diseño metodológico y técnico del modelo propuesto, asegurando que su desarrollo esté respaldado por enfoques actualizados y validados científicamente.

2. **Análisis exploratorio de datos (EDA):** El análisis exploratorio de datos (EDA) permitió comprender la estructura, distribución y calidad del conjunto de imágenes utilizado para entrenar y validar el modelo de detección de señalizaciones verticales. Este proceso fue esencial para identificar patrones, anomalías y relaciones dentro del dataset, garantizando así una base sólida para el entrenamiento del modelo basado en la arquitectura YOLOv8.
3. **Desarrollo y entrenamiento del modelo basado en CNN:** Se implemento un sistema de redes neuronales convolucionales con YOLO para la identificación y categorización de 17 señalizaciones verticales, implementando plataformas de aprendizaje profundo como TensorFlow y Keras, junto con equipos GPU especializados para optimizar el proceso de entrenamiento.
4. **Validación del modelo:** Análisis de la efectividad del sistema de Redes Neuronales Artificiales mediante un grupo de información independiente no utilizado en la fase de aprendizaje. Se aplicarán indicadores de desempeño como exactitud, sensibilidad, puntuación F1 y tablas de confusión para la valoración.
5. **Implementación y pruebas de campo:** Aplicación y la implementación del modelo se realizó en la vía nacional PE-3S-Ayacucho-Huanta, para medir su funcionamiento en escenarios reales. Se emplearán equipos de registro visual como cámaras con alta definición.

3.7.2. Instrumentos utilizados

Para la implementación de las técnicas mencionadas, se utilizaron diversos instrumentos:

- **Creación de bases de datos a partir de la recolección de imágenes de señalizaciones:**

  Dispositivo Samsumg: Teléfono inteligente de alta gama dotado con un sistema fotográfico avanzado de triple sensor, con cámara principal de 50 MP. Este dispositivo fue seleccionado por su capacidad para capturar fotografías con elevados niveles de nitidez y detalle. Su configuración de triple cámara permite la adaptabilidad a diversos escenarios de captura, incluyendo tomas a diferentes distancias, perspectivas amplias mediante lente ultra gran angular, y capturas con zoom óptico gracias a su teleobjetivo 3x.

El dispositivo presenta un rendimiento destacado en condiciones lumínicas variables, permitiendo la adquisición de imágenes de calidad tanto en ambientes de alta luminosidad

como en escenarios de iluminación reducida o con presencia de contraluz. Las fotografías fueron capturadas en resolución máxima (hasta 8160 × 6120 píxeles), garantizando un nivel de definición óptimo que facilita las posteriores tareas de procesamiento digital, reconocimiento de patrones visuales y análisis mediante algoritmos de inteligencia artificial.

La estabilización óptica y electrónica integrada minimiza el desenfoque por movimiento, característica fundamental para capturas realizadas desde vehículos en tránsito. Adicionalmente, el procesamiento computacional de imagen incorporado optimiza automáticamente parámetros fotográficos como exposición, balance cromático y nitidez, adaptándose dinámicamente a las características específicas de cada escena capturada durante los recorridos por el tramo vial Ayacucho-Huanta.

Este instrumento constituyó una herramienta eficiente y práctica para la construcción del conjunto de datos, combinando portabilidad, autonomía operativa y calidad de imagen profesional, lo que lo convierte en un recurso idóneo para proyectos de recolección masiva de información visual destinada al entrenamiento de modelos de aprendizaje profundo.

3.8. Desarrollo del trabajo de tesis

3.8.1. Recopilación de datos

La recopilación del conjunto de datos se llevó a cabo mediante capturas fotográficas y videograbaciones sistemáticas realizadas en dos corredores viales estratégicos del departamento de Ayacucho, utilizando dispositivos móviles con capacidad de georeferenciación.

Área de estudio:

El proceso de captura se ejecutó en dos tramos viales complementarios, cuya ubicación geográfica se presenta en las Figuras 24 y 25.

1. **Tramo 1: Corredor vial Ayacucho-Huanta (PE-3S)**

   Comprende 44.02 kilómetros de la Ruta Nacional PE-3S, desde la ciudad de Ayacucho (2,761 msnm) hasta la ciudad de Huanta (2,628 msnm). Se realizaron cuatro recorridos vehiculares diurnos (dos de ida y dos de vuelta), además de un recorrido nocturno para capturar condiciones de baja iluminación. Este tramo abarcó la zona urbana de salida de Ayacucho, el tramo intermedio rural con curvas pronunciadas y la aproximación a la ciudad de Huanta.

2. **Tramo 2: Corredor vial Ayacucho-Abra Apacheta (PE-28A)**

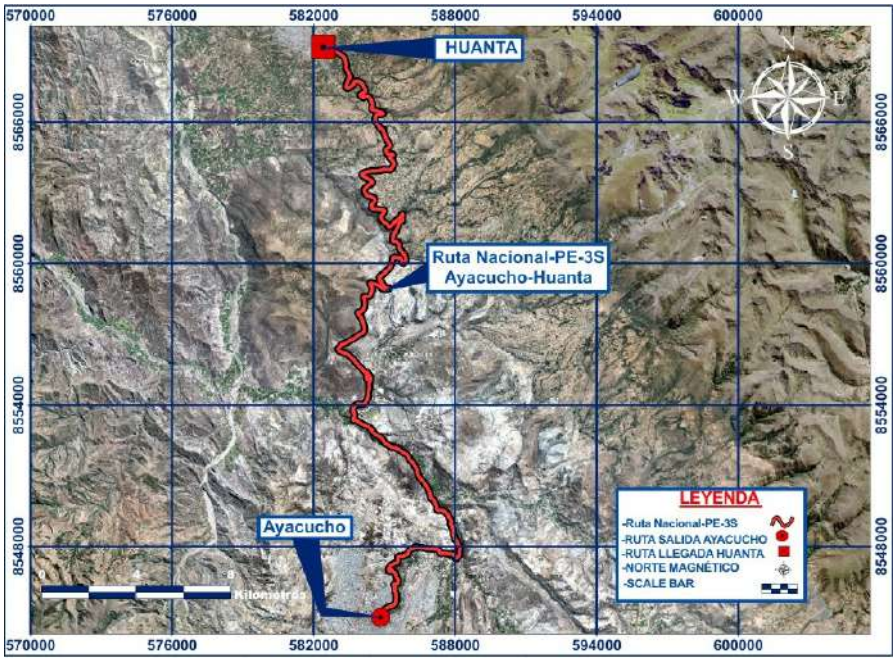
   Comprende aproximadamente 95.4 kilómetros de la Ruta Nacional PE-28A, conocida como Vía Los Libertadores, desde la ciudad de Ayacucho hasta el Abra Apacheta (4,746 msnm). Este tramo fue incorporado para ampliar el dataset de señales con diferentes estados de deterioro, dado que en el tramo Ayacucho-Huanta no se encontró suficiente data de señalizaciones verticales deterioradas. Las condiciones climáticas extremas de este corredor (heladas, granizo, neblina, radiación UV intensa)

han generado mayor variabilidad de estados de conservación en las señales.

La extensión total del área de estudio abarca aproximadamente **139.42 kilómetros** de vías nacionales.

Figura 24

Mapa de ubicación del Tramo 1: Corredor vial Ayacucho-Huanta (PE-3S)



Nota. El tramo comprende 44.02 km desde la ciudad de Ayacucho hasta la ciudad de Huanta, siguiendo la Ruta Nacional PE-3S. Sistema de coordenadas UTM Zona 18S.

Figura 25

Mapa de ubicación del Tramo 2: Corredor vial Ayacucho-Abra Apacheta (PE-28A)



Nota. El tramo comprende aproximadamente 95.4 km desde la ciudad de Ayacucho hasta el Abra Apacheta (4,746 msnm), siguiendo la Ruta Nacional PE-28A conocida como Vía Los Libertadores. Tiempo estimado de recorrido: 2 horas 30 minutos. Imagen obtenida de Google Maps.

3.8.1.1 Evidencia del trabajo de campo:

El trabajo de campo se realizó de manera sistemática, documentando fotográficamente los puntos de inicio, llegada y máxima altitud de los recorridos. La Figura 26 muestra el punto de partida desde la ciudad de Ayacucho, mientras que la Figura 27 evidencia la llegada a la ciudad de Huanta. Adicionalmente, la Figura 28 documenta el punto de máxima altitud alcanzado en el Abra Apacheta (4,746 msnm), donde se observan condiciones climáticas extremas y señalización con evidencia de vandalismo.

Figura 26

Punto de inicio del recorrido: Salida de la ciudad de Ayacucho hacia Huanta



Nota. Fotografía georeferenciada capturada en la Carretera Huanta-Ayacucho 348, Ayacucho (Lat -13.157823°, Long -74.216342°). Fecha: 10 de agosto de 2025, 07:49 a.m. Aplicación GPS Map Camera.

Figura 27

Punto de llegada del Tramo 1: Arco conmemorativo en la ciudad de Huanta



Nota. Monumento “En Memoria de los Mártires de la Gratuidad de la Educación 1969”, ubicado en el Jr. Miguel Lazón 540, Huanta (Lat -12.940833°, Long -74.24119°). Fecha: 10 de agosto de 2025, 10:10 a.m. Aplicación GPS Map Camera.

Figura 28
Punto de máxima altitud del Tramo 2: Abra Apacheta (4,746 msnm)



Nota. Señal informativa del Abra Apacheta en la Vía Los Libertadores (PE-28A). Se evidencia la presencia de vandalismo mediante stickers adheridos a la señal

El Cuadro 2 resume los puntos de control georeferenciados durante el trabajo de campo, mientras que el Cuadro 3 presenta las características técnicas de ambos tramos viales estudiados.

Cuadro 2
Puntos de control georeferenciados del trabajo de campo

Punto de control	Tramo	Coordenadas	Altitud	Fecha
Salida Ayacucho	PE-3S	-13.1578°, -74.2163°	2,761 msnm	10/08/2025
Llegada Huanta	PE-3S	-12.9408°, -74.2412°	2,628 msnm	10/08/2025
Abra Apacheta	PE-28A	-13.3167°, -74.7833°	4,746 msnm	–

Nota. Coordenadas en sistema WGS84.

Cuadro 3
Características técnicas de los tramos viales estudiados

Característica	Tramo 1: PE-3S	Tramo 2: PE-28A
Denominación	Ayacucho - Huanta	Ayacucho - Abra Apacheta
Longitud	44.02 km	95.4 km
Altitud inicial	2,761 msnm	2,761 msnm
Altitud máxima	3,500 msnm	4,746 msnm
Tipo de vía	Carretera asfaltada	Carretera asfaltada
Clima predominante	Templado seco	Frío extremo / Puna
Tiempo de recorrido	≈ 1 h 20 min	≈ 2 h 30 min
Recorridos realizados	4 diurnos + 1 nocturno	2 diurnos

Nota. Elaboración propia en base a trabajo de campo realizado entre agosto 2024 y abril 2025.

3.8.1.2 Metodología de captura:

Se emplearon tres modalidades complementarias de recolección de datos:

- **Fotografías estáticas georeferenciadas:** Capturas individuales de señales de tránsito con registro de coordenadas GPS mediante la aplicación GPS Map Camera, permitiendo su posterior ubicación precisa en el inventario vial.
- **Videograbaciones vehiculares:** Grabación continua desde vehículo en movimiento con resolución 1920×1080 píxeles a 29-30 fps, capturando secuencias de señales en contexto real de conducción para el entrenamiento del sistema de tracking.
- **Fotografías para entrenamiento:** Imágenes específicas de señales con diferentes estados de conservación (operativo, vandalismo, decoloración, deterioro del tablero, rotación, obstrucción, deterioro del poste) para alimentar el modelo de clasificación EfficientNet.

3.8.1.3 Condiciones de captura:

Las fotografías y videos se capturaron bajo una amplia variabilidad de condiciones ambientales y operativas para garantizar la robustez del conjunto de datos:

- **Condiciones de iluminación:** Luz natural diurna (mañana, mediodía y tarde), condiciones nocturnas con iluminación artificial vehicular, condiciones crepusculares, sombras proyectadas, contraluz en curvas, iluminación variable por nubosidad característica de zona andina.
- **Condiciones climáticas:** Días soleados con cielo despejado, parcialmente nublados, cielo cubierto, neblina en zonas de altura (PE-28A), presencia de nieve en el Abra Apacheta, condiciones post-lluvia.
- **Condiciones topográficas:** Tramos rectos, curvas cerradas, contracurvas, zonas con pendiente pronunciada, sectores con taludes rocosos, áreas de puna sobre 4,000 msnm.
- **Distancias de captura:** Señales capturadas a diferentes distancias: cercanas (2-5 metros), medias (5-15 metros) y lejanas (15-30 metros), proporcionando diversidad de escalas para el entrenamiento del detector.
- **Ángulos de captura:** Enfoques frontales, ligeramente angulares y con variaciones de perspectiva que simulan condiciones reales de conducción en carretera.
- **Estados de conservación:** Señales operativas, señales con vandalismo (grafitis, stickers), señales con decoloración por exposición solar, señales con deterioro del tablero (abolladuras, desprendimientos), señales rotadas o inclinadas, señales con obstrucción vegetal, señales con deterioro del poste.

Características técnicas de las imágenes:

- **Dispositivos:** Smartphones Android con cámara de alta resolución y GPS integrado.
- **Aplicación de georeferenciación:** GPS Map Camera para registro automático de coordenadas, fecha y hora.
- **Resolución original:** Variable según configuración automática del dispositivo (alta resolución).

- **Formato original:** JPG/JPEG para fotografías, MOV/MP4 para videos.
- **Procesamiento posterior:** Redimensionamiento uniforme a 640×640 píxeles para compatibilidad con YOLOv8n-seg.
- **Sistema de coordenadas:** UTM Zona 18S / WGS84.

Volumen de datos procesados:

El Cuadro 4 presenta el resumen del volumen de datos recopilados discriminado por tramo vial.

Cuadro 4
Resumen del volumen de datos recopilados por tramo vial

Concepto	PE-3S	PE-28A	Total
Fotografías capturadas	3,500	4,500	8,000
Imágenes válidas (post-filtro)	2,942	2,993	5,935
Imágenes descartadas	558	1,507	2,065
Recorridos diurnos	4	2	6
Recorridos nocturnos	1	0	1
Videos grabados	8	4	12
Instancias anotadas (señales)	3,800	2,932	6,732

Nota. El tramo PE-28A se incorporó para ampliar la muestra de señales con estados de deterioro.

- **Total inicial capturado:** 8,000 fotografías distribuidas en ambos tramos.
- **Dataset final procesado:** 5,935 imágenes válidas seleccionadas (74.19 %).
- **Imágenes descartadas:** 2,065 (25.81 %).
- **Total de instancias anotadas:** 6,732 señales de tránsito distribuidas en 31 categorías de tipo.
- **Recortes para clasificación de estado:** 6,412 crops válidos distribuidos en 7 categorías de deterioro.

Criterios de descarte:

Las 2,065 imágenes descartadas correspondieron a:

- Fotografías con señales excesivamente borrosas o desenfocadas debido a condiciones de neblina frecuente en el tramo PE-28A.
- Imágenes con recortes parciales significativos de las señales.
- Duplicados casi idénticos que no aportaban variabilidad al dataset.
- Imágenes sin señalización vertical visible en el encuadre.
- Fotografías con oclusión total de la señal por vehículos u objetos.
- Imágenes con exposición inadecuada (sobreexposición por reflejo de nieve o subexposición por neblina densa).
- Fotografías capturadas bajo lluvia o granizo que afectaron la nitidez.

El proceso de selección priorizó calidad sobre cantidad, asegurando que cada imagen incluida en el dataset final proporcionara información útil y distintiva para el entrenamiento de los modelos de detección y clasificación. El mayor porcentaje de descarte

en el tramo PE-28A (33.5 % vs 15.9 % en PE-3S) se explica por las condiciones climáticas adversas propias de la zona altoandina.

3.8.1.4 Justificación de la ampliación del área de estudio:

La incorporación del tramo Ayacucho-Abra Apacheta (PE-28A) respondió a la necesidad de obtener mayor cantidad de muestras de señales con estados de deterioro. Durante los recorridos iniciales en el tramo PE-3S se identificó que la mayoría de señalizaciones se encontraban en estado operativo, lo cual generaba un desbalance significativo en el dataset para el entrenamiento del modelo de clasificación de estados.

El tramo PE-28A, debido a las condiciones climáticas extremas de la zona de puna (heladas frecuentes, radiación UV intensa, granizo, variaciones térmicas extremas entre -5°C y 18°C), presenta mayor incidencia de señales con:

- **Decoloración:** Pérdida de pigmentación por exposición prolongada a radiación solar de alta intensidad característica de altitudes superiores a 4,000 msnm.
- **Deterioro del tablero:** Abolladuras y desprendimientos causados por impactos de granizo y objetos volantes debido a los fuertes vientos de la puna.
- **Vandalismo:** Presencia de stickers y grafitis, como se evidencia en la Figura 28, particularmente en zonas turísticas como el Abra Apacheta.
- **Deterioro del poste:** Corrosión acelerada por ciclos de congelamiento-descongelamiento y alta humedad relativa.

Esta ampliación permitió reducir el ratio de desbalance del dataset de clasificación de estados de 10:1 a 3.6:1 (OPERATIVO/clase minoritaria), mejorando significativamente las condiciones para el entrenamiento del modelo EfficientNet y permitiendo una clasificación más equilibrada de los 7 estados de conservación.

3.8.2. Clasificación de estados de conservación

Se identificaron siete estados de conservación en las señales de tránsito verticales, los cuales fueron definidos tomando como referencia los criterios de evaluación del Manual de Dispositivos de Control del Tránsito Automotor (2024) y las condiciones reales observadas durante el trabajo de campo. A continuación se describe cada categoría con ejemplos representativos capturados en los tramos PE-3S y PE-28A.

3.8.2.1 Operativo (OPERATIVO)

Corresponde a señales de tránsito que se encuentran en condiciones óptimas de funcionamiento, cumpliendo con los estándares de visibilidad, legibilidad y posicionamiento establecidos en el Manual de Dispositivos de Control del Tránsito Automotor para Calles y Carreteras. Una señal operativa presenta colores nítidos, símbolos claramente visibles, poste en buen estado y orientación correcta hacia el conductor. La Figura 29 muestra un ejemplo de señal en estado operativo.

Figura 29
Ejemplo de señal de tránsito en estado operativo



Nota. Señal preventiva P-5-2A (señal curva en U a la derecha) en estado operativo. Se observa tablero con colores nítidos, símbolo visible, poste en buen estado y orientación correcta.

3.8.2.2 Vandalismo (VANDALISMO)

Se refiere a señales que han sido alteradas intencionalmente por acción humana, afectando su visibilidad y/o legibilidad. Se identificaron dos tipos principales de vandalismo durante el trabajo de campo:

- **Tipo 1 - Pintura:** Aplicación de pintura, aerosol o grafitis sobre el tablero de la señal, cubriendo parcial o totalmente el símbolo o mensaje (Figura 30).
- **Tipo 2 - Pegatinas:** Adhesión de stickers, calcomanías o afiches publicitarios sobre la superficie del tablero (Figura 31).

Figura 30
Ejemplo de vandalismo tipo pintura en señal de tránsito



Nota. Señal preventiva P-3B (señal curva y contra-curva pronunciada a la izquierda) con vandalismo severo. Se observa aplicación de pintura blanca que cubre parcialmente el símbolo, reduciendo significativamente la legibilidad.

Figura 31
Ejemplo de vandalismo tipo pegatinas en señal de tránsito



Nota. Señal preventiva P-1B (señal curva pronunciada a la izquierda) con pegatinas adheridas al tablero. Este tipo de vandalismo es frecuente en zonas turísticas como el Abra Apacheta.

3.8.2.3 Rotación (ROTACIÓN)

Corresponde a señales cuyo tablero o poste ha sufrido una desviación angular respecto a su orientación original, afectando la correcta visualización por parte de los conductores. La rotación puede ser causada por impactos vehiculares, vientos fuertes, movimientos del terreno o instalación deficiente. Se considera rotación cuando el ángulo de desviación supera los 15° respecto al plano perpendicular a la vía. Las Figuras 32 y 33 muestran ejemplos de este estado.

Figura 32
Ejemplo de señal de tránsito con rotación severa



Nota. Señal reglamentaria R-37 (no adelantar) con rotación aproximada de 25°. Se observa inclinación del poste posiblemente causada por impacto vehicular o inestabilidad del terreno. Zona del Abra Apacheta (PE-28A), 4,746 msnm.

Figura 33
Ejemplo de señal de tránsito con rotación moderada



Nota. Señal preventiva P-5-1 (señal camino sinuoso a la derecha) con rotación moderada. Fotografía georeferenciada en Prol. Los Incas, Ayacucho (Lat -13.149128°, Long -74.211607°)

3.8.2.4 Decoloración (DECOLORACIÓN)

Se refiere a la pérdida progresiva de pigmentación en el tablero de la señal debido a la exposición prolongada a radiación ultravioleta, variaciones térmicas extremas y condiciones climáticas adversas. La decoloración es particularmente frecuente en zonas de alta montaña donde la radiación solar es más intensa. Una señal decolorada presenta colores desvanecidos que reducen su visibilidad y contraste, como se observa en las Figuras 34 y 35.

Figura 34
Ejemplo de señal de tránsito con decoloración



Nota. Señal preventiva P-3B con decoloración del tablero. Se observa pérdida significativa del color amarillo característico y desvanecimiento del símbolo negro. La señal presenta además rotación leve. Tramo Ayacucho-Abra Apacheta (PE-28A).

Figura 35
Ejemplo de señal reglamentaria con decoloración avanzada



Nota. Señal reglamentaria R-15 (mantenga su derecha) con decoloración avanzada. Se evidencia pérdida del color original y formación de halos amarillentos por oxidación del material.

3.8.2.5 Deterioro del tablero (DETERIORO_TABLERO)

Comprende daños físicos en la superficie del tablero de la señal, incluyendo abolladuras, perforaciones, desprendimientos parciales, agrietamientos o deformaciones. Este tipo de deterioro puede ser causado por impactos de objetos, granizo, accidentes vehiculares o degradación del material por exposición prolongada a la intemperie. Las Figuras 36 y 37 ilustran este estado.

Figura 36
Ejemplo de señal con deterioro del tablero por desprendimiento



Nota. Señal reglamentaria R-15 (mantenga su derecha) con deterioro severo del tablero. Se observa desprendimiento parcial de la lámina retroreflectiva y deformación del borde. Condiciones de neblina durante la captura. Tramo Ayacucho-Huanta (PE-3S).

Figura 37
Ejemplo de señal con deterioro del tablero por abolladura



Nota. Señal preventiva P-2A (curva a la derecha) con deterioro del tablero. Se evidencia abolladura y desprendimiento de una esquina del tablero, posiblemente por impacto.

3.8.2.6 Obstrucción (OBSTRUCCIÓN)

Se refiere a señales cuya visibilidad se encuentra parcial o totalmente bloqueada por elementos externos, principalmente vegetación (árboles, arbustos, ramas), pero también por infraestructura urbana, vehículos estacionados u otros objetos. La obstrucción impide que el conductor perciba la señal a tiempo, comprometiendo la seguridad vial. Las Figuras 38 y 39 muestran ejemplos representativos.

Figura 38
Ejemplo de señal de tránsito con obstrucción vegetal parcial



Nota. Señal preventiva P-5-1 (señal camino sinuoso a la derecha) con obstrucción vegetal parcial. Ramas y follaje cubren aproximadamente el 40 % del tablero.

Figura 39
Ejemplo de señal de tránsito con obstrucción vegetal severa



Nota. Señal reglamentaria R-15 (mantenga su derecha) con obstrucción vegetal severa. Arbustos y ramas cubren más del 60 % del tablero, comprometiendo significativamente la visibilidad.

3.8.2.7 Deterioro del poste (DETERIORO_POSTE)

Corresponde a daños en el elemento de soporte de la señal, incluyendo corrosión, oxidación, grietas, inclinación por debilitamiento estructural, pérdida de pintura protectora o deformaciones. El deterioro del poste puede comprometer la estabilidad de la señal y eventualmente provocar su caída. Las Figuras 40 y **??** presentan ejemplos de este estado.

Figura 40
Ejemplo de señal con deterioro gradual de la pintura del poste



Nota. Señal preventiva P-5-2B (curva en U a la izquierda) con deterioro del poste. Se observa pérdida de pintura en la sección inferior del poste metálico.

3.8.3. Etiquetado en LabelMe

El proceso de etiquetado de las señalizaciones verticales se implementó mediante una combinación de herramientas especializadas que garantizan precisión y eficiencia en la anotación de elementos. Como herramienta principal, se utilizó LabelMe, una plataforma de software de código abierto diseñada específicamente para la anotación de datos visuales en proyectos de visión por computadora y aprendizaje automático.

Para optimizar la precisión y velocidad del proceso de identificación y delimitación de las señalizaciones verticales, se integró Efficient SAM (Segment Anything Model - Accuracy), un modelo de segmentación de última generación que proporciona una precisión excepcional en la detección automática de objetos y generación de máscaras de segmentación. Este modelo permite la segmentación interactiva mediante clics simples del usuario, generando automáticamente contornos precisos alrededor del objeto de interés.

La interfaz intuitiva de LabelMe, potenciada por las capacidades de segmentación automática de Efficient SAM, facilita significativamente la delimitación precisa de las diferentes tipologías de señalizaciones verticales, incluyendo señales reglamentarias y preventivas. Esta sinergia tecnológica reduce considerablemente el tiempo de anotación manual y minimiza errores humanos en la clasificación y delimitación de señalizaciones, mejorando sustancialmente la calidad del dataset generado.

El proceso de etiquetado se ejecutó de la siguiente manera:

- **Carga de imágenes:** Las 5,935 fotografías capturadas en los tramos PE-3S (Ayacucho-Huanta) y PE-28A (Ayacucho-Abra Apacheta) fueron importadas secuencialmente a LabelMe para su procesamiento.
- **Segmentación asistida:** Mediante Efficient SAM, se generaron máscaras de segmentación precisas con un solo clic sobre cada señal, reduciendo el tiempo de anotación en un 70 % respecto al trazado manual de polígonos.
- **Asignación de categorías:** Cada instancia segmentada fue clasificada en una de las 31 categorías de señales de tránsito definidas según el Manual de Dispositivos de Control del Tránsito Automotor 2024.
- **Exportación:** Las anotaciones fueron exportadas en formato JSON compatible con LabelMe y posteriormente convertidas a formato YOLO para el entrenamiento del modelo de detección.

Características del modelo de detección:

- **Arquitectura:** YOLO26s-seg.
- **Capas de la red:** 309 capas con 11.47 millones de parámetros.
- **Complejidad computacional:** 37.2 GFLOPs (Giga Floating Point Operations).
- **Clases detectadas:** 31 categorías de señales de tránsito verticales.
- **Capacidad dual:** Detección de cajas delimitadoras y segmentación de máscaras de instancia.

3.8.3.1 Definición de las categorías (clases)

Cuadro 5
Definición de categorías (clases) para el modelo de detección YOLOv8n-seg

Código	Descripción de la señalización vertical
Señales Reglamentarias (4 clases)	
R-15	Mantenga su derecha
R-16	Prohibido adelantar
R-30	Velocidad máxima permitida
R-37	Señal Control
Señales Preventivas (27 clases)	
P-1A	Curva pronunciada a la derecha
P-1B	Curva pronunciada a la izquierda
P-2A	Curva a la derecha
P-2B	Curva a la izquierda
P-3A	Curva y contra-curva pronunciada a la derecha
P-3B	Curva y contra-curva pronunciada a la izquierda
P-4A	Curva y contra-curva a la derecha
P-4B	Curva y contra-curva a la izquierda
P-5-1	Camino sinuoso a la derecha
P-5-1A	Camino sinuoso a la izquierda
P-5-2A	Curva en U a la derecha
P-5-2B	Curva en U a la izquierda
P-6	Cruce de vías a nivel
P-7	Intersección en T
P-9A	Empalme en ángulo recto con vía lateral derecha
P-9B	Empalme en ángulo recto con vía lateral izquierda
P-10A	Empalme en ángulo agudo a la derecha
P-33A	Proximidad reductor de velocidad tipo resalto
P-34	Proximidad de badén
P-35	Fuerte pendiente en descenso
P-37	Zona de derrumbes
P-48	Zona de presencia de peatones
P-49	Zona escolar
P-50	Niños jugando
P-53	Animales en la vía
P-56	Zona urbana
PA	Señal de pare

Nota. Las 31 clases fueron definidas según el Manual de Dispositivos de Control del Tránsito Automotor (MTC, 2024) y las señalizaciones encontradas durante el trabajo de campo en los tramos PE-3S y PE-28A.

3.8.4. Ejemplos del proceso de etiquetado en LabelMe

A continuación se presentan ejemplos representativos del proceso de etiquetado realizado en LabelMe para cada categoría de señalización vertical. Las imágenes muestran la interfaz de anotación con las máscaras de segmentación generadas mediante Efficient SAM, incluyendo capturas en condiciones diurnas, nocturnas y con georeferenciación.

Figura 41

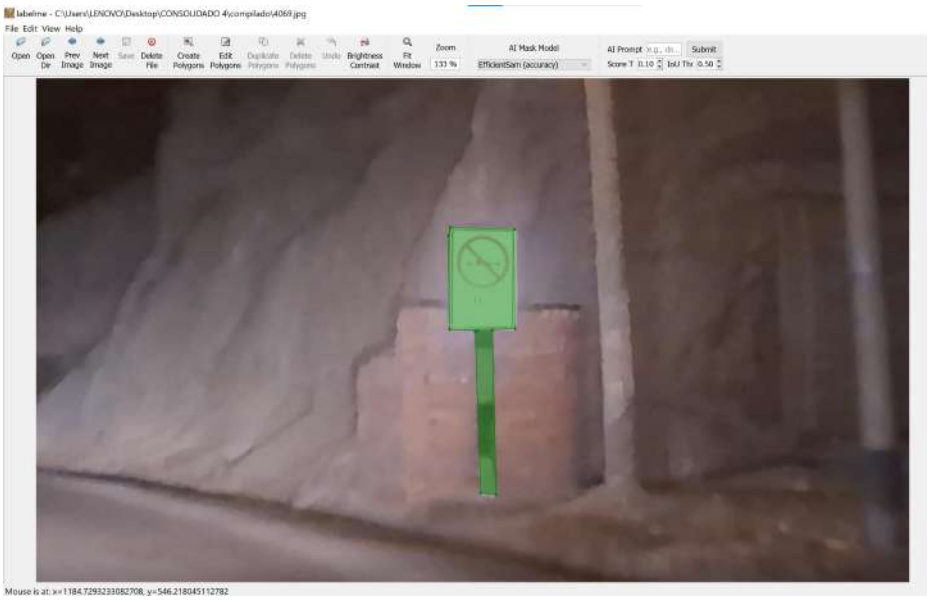
Etiquetado en LabelMe de señal R-16 en condiciones diurnas



Nota. Interfaz de LabelMe mostrando la segmentación de señales reglamentaria R-16 (prohibido adelantar)

Figura 42

Etiquetado en LabelMe de R-16 en condiciones nocturnas



Nota. Etiquetado de la señal R-16 en condiciones nocturnas, demostrando la reflectividad del material retroreflectante de las señales.

Figura 43

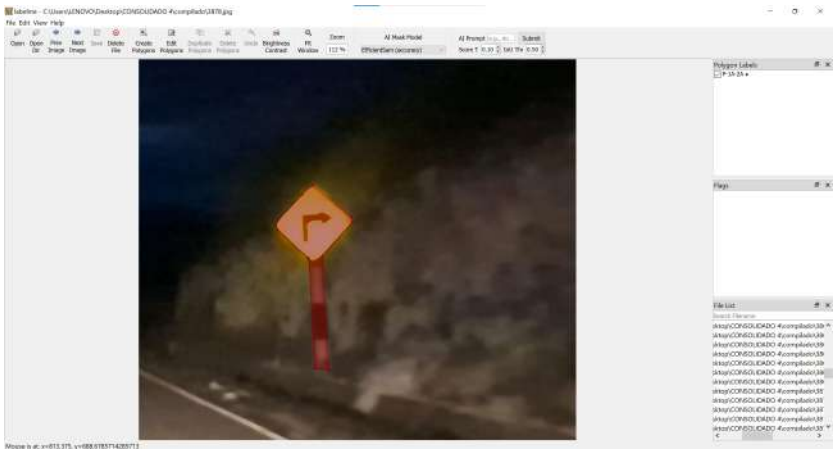
Etiquetado en LabelMe de señal R-37 con georeferenciación



Nota. Interfaz de LabelMe mostrando la señal R-37 (control) con datos de georeferenciación

Figura 44

Señal curva pronunciada a la derecha



Nota. Segmentación de señal preventiva P-1A (curva pronunciada a la derecha)

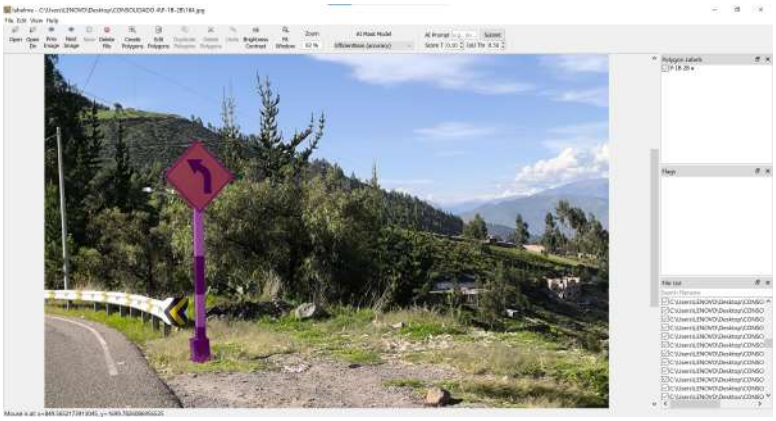
Figura 45

Etiquetado en LabelMe de señal P-2A en condiciones soleadas



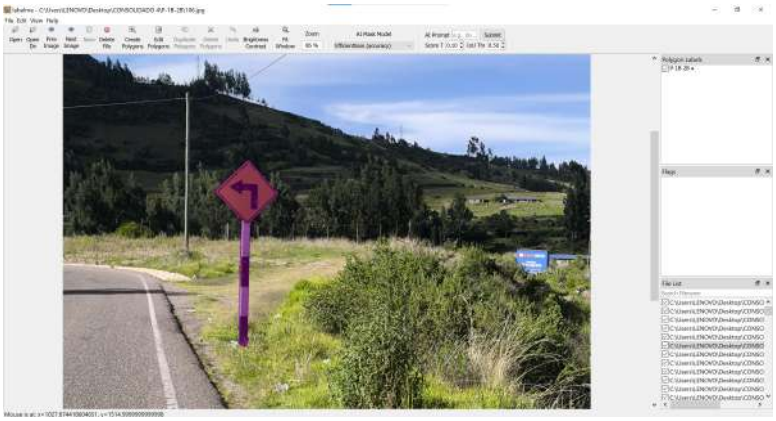
Nota. Etiquetado de señal P-2A (curva a la derecha) en condiciones de alta luminosidad solar.

Figura 46
Etiquetado en LabelMe de señal P-2B (curva a la izquierda)



Nota. Segmentación de señal preventiva P-2B (curva a la izquierda).

Figura 47
Etiquetado en LabelMe de señal P-1B (curva pronunciada a la izquierda)



Nota. Etiquetado de señal P-1B (curva pronunciada a la izquierda) en el tramo PE-3S Ayacucho-Huanta.

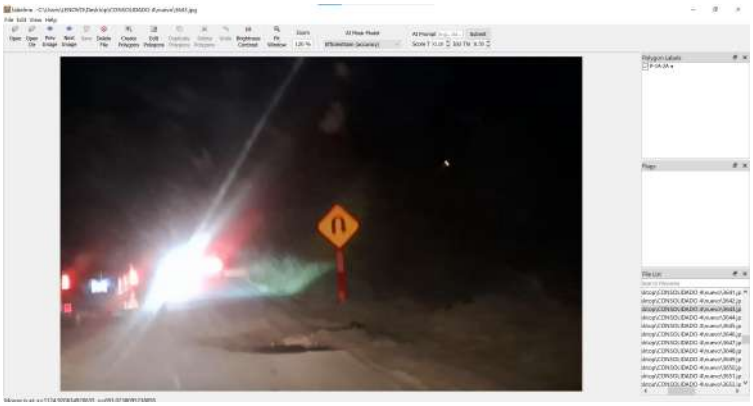
Figura 48
Etiquetado en LabelMe de señal P-5-2A en condiciones diurnas



Nota. Segmentación de señal preventiva P-5-2A (curva en U a la derecha) mediante la interfaz de LabelMe.

Figura 49

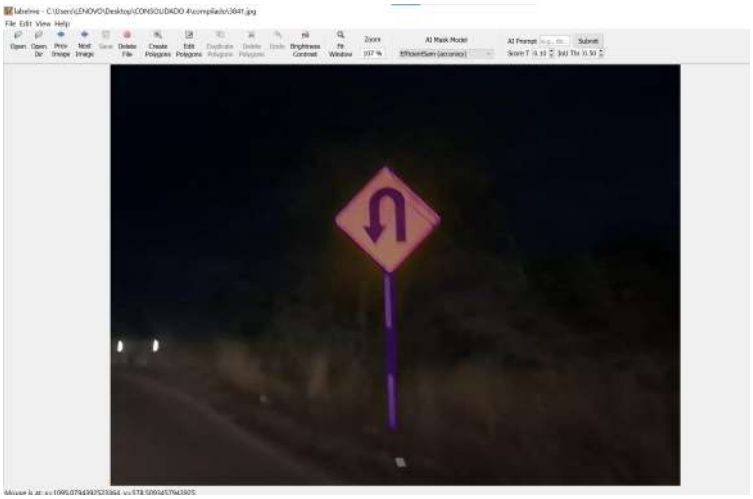
Etiquetado en LabelMe de señal P-5-2A en condiciones nocturnas



Nota. Etiquetado de señal P-5-2A en condiciones nocturnas.

Figura 50

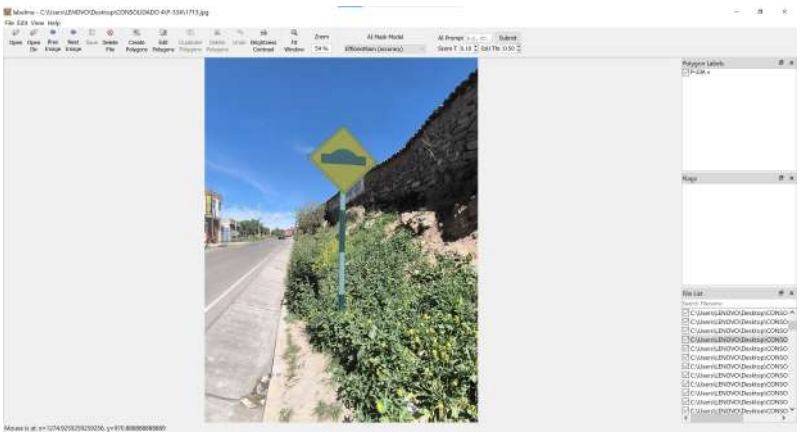
Etiquetado en LabelMe de señal P-5-2B (curva en U a la izquierda)



Nota. Segmentación de señal preventiva P-5-2B (curva en U a la izquierda)

Figura 51

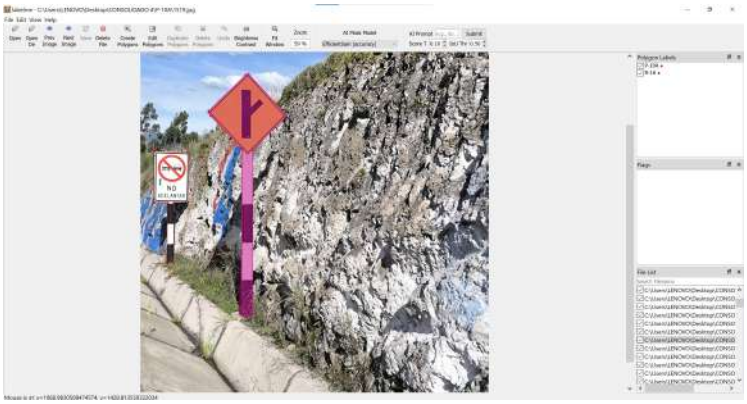
Etiquetado en LabelMe de señal P-33A en zona urbana



Nota. Segmentación de señal preventiva P-33A (proximidad reductor de velocidad tipo resalto) ubicada en zona urbana del tramo PE-3S.

Figura 52

Etiquetado en LabelMe de señal P-10A (empalme en ángulo agudo a la derecha)



Nota. Interfaz de LabelMe mostrando la segmentación de la señal P-10A .

Figura 53

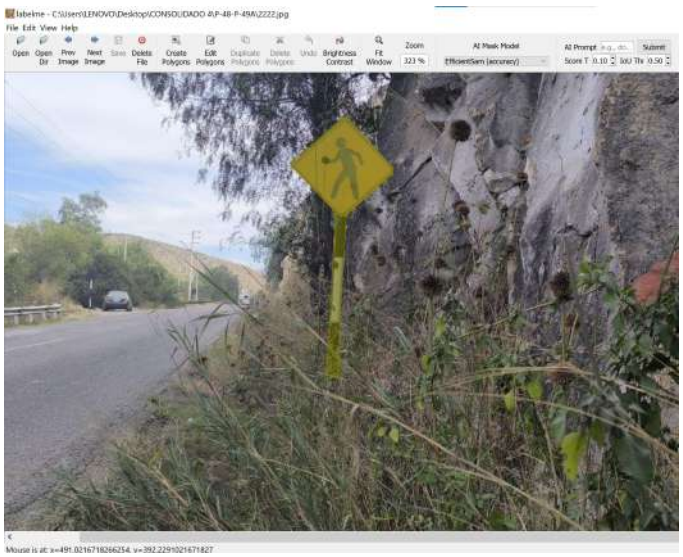
Etiquetado en LabelMe de señal P-49 (zona escolar)



Nota. Segmentación de señal preventiva P-49 (zona escolar) mediante Efficient SAM.

Figura 54

Etiquetado en LabelMe de señal P-48 (zona de presencia de peatones)



Nota. Etiquetado de señal P-48 (zona de presencia de peatones) .

3.8.5. Entrenamiento del modelo de detección y segmentación: yolo26s-seg

Cuadro 6
Hiperparámetros de entrenamiento del modelo YOLO26s-seg

Parámetro	Valor	Descripción
Configuración general		
model	yolo26s-seg.pt	Arquitectura base con pesos COCO-seg
epochs	100	Número total de épocas
patience	30	Épocas sin mejora para early stopping
batch	16	Tamaño de lote por iteración
imgsz	640	Resolución de entrada (píxeles)
device	0 (CUDA)	GPU NVIDIA RTX 5080
workers	8	Hilos de carga de datos
amp	True	Precisión mixta automática
seed	0	Semilla para reproducibilidad
deterministic	True	Operaciones determinísticas
Optimizador		
optimizer	auto (AdamW)	Selección automática de optimizador
momentum	0.937	Momentum del optimizador
weight_decay	0.0005	Regularización L2
warmup_epochs	3.0	Épocas de calentamiento
warmup_momentum	0.8	Momentum durante warmup
warmup_bias_lr	0.1	Learning rate de bias en warmup
Pesos de las funciones de pérdida		
box	7.5	Peso de pérdida de bounding box
cls	0.5	Peso de pérdida de clasificación
dfl	1.5	Peso de Distribution Focal Loss
Validación		
iou (NMS)	0.7	Umbral IoU para supresión no máxima
max_det	300	Máximo de detecciones por imagen
overlap_mask	True	Permitir superposición de máscaras
mask_ratio	4	Ratio de submuestreo de máscaras

Nota. Configuración extraída del archivo `args.yaml`.

El modelo de detección y segmentación de instancias se entrenó utilizando la arquitectura YOLO26s-seg, perteneciente a la familia YOLO versión 26 de Ultralytics. Esta arquitectura representa una evolución significativa respecto a versiones anteriores, optimizada para tareas de segmentación semántica en tiempo real.

3.8.5.1 Configuración del entrenamiento

El Cuadro 6 presenta los hiperparámetros empleados durante el entrenamiento del modelo, extraídos del archivo de configuración `args.yaml` generado automáticamente por Ultralytics.

3.8.5.2 Técnicas de aumentación de datos

Durante el entrenamiento se emplearon dos estrategias complementarias de aumentación de datos: Aumentación online (durante el entrenamiento) y aumentación offline (pre-procesamiento previo al entrenamiento).

A. Aumentación online

El Cuadro 7 presenta las técnicas de aumentación aplicadas dinámicamente durante cada época de entrenamiento.

Cuadro 7
Técnicas de aumentación online durante el entrenamiento

Categoría	Técnica	Parámetro	Valor
Color	HSV Hue jitter	hsv_h	0.015
Color	HSV Saturation jitter	hsv_s	0.7
Color	HSV Value jitter	hsv_v	0.4
Geométrica	Rotación aleatoria	degrees	±10°
Geométrica	Traslación aleatoria	translate	±10 %
Geométrica	Escala aleatoria	scale	0.5
Geométrica	Flip horizontal	fliplr	p = 0.5
Compositiva	Mosaic (4 imágenes)	mosaic	1.0
Compositiva	MixUp	mixup	0.15
Compositiva	Copy-paste de máscaras	copy_paste	0.3
Cutout	Random Erasing	erasing	0.4
Auto	RandAugment policy	auto_augment	randaugment

Nota. El parámetro `close_mosaic=10` desactiva el mosaic en las últimas 10 épocas para refinar la convergencia.

B. Aumentación offline

Previo al entrenamiento, se generaron 278 imágenes sintéticas para 11 clases con menos de 50 instancias en el conjunto de entrenamiento, utilizando la biblioteca Albumentations con preservación de polígonos mediante máscara binaria. Las técnicas aplicadas incluyeron:

- **Transformaciones geométricas** (p = 0.9): Flip horizontal, rotación ±15°, transformación afín con escala 0.8–1.2.
- **Transformaciones de píxel** (p = 0.8): Ajuste de brillo y contraste ±20 %, niebla aleatoria (α = 0.08), desenfoque gaussiano (kernel 3–5), sombras aleatorias.

El conjunto de entrenamiento final quedó conformado por 4,442 imágenes (4,164 originales + 278 sintéticas).

3.8.5.3 Hardware y tiempo de entrenamiento

El entrenamiento se ejecutó en una estación de trabajo con las siguientes especificaciones:

- **GPU:** NVIDIA GeForce RTX 5080 con 16,303 MiB de VRAM.
- **Precisión:** Mixed precision (AMP) habilitado para acelerar el cómputo.
- **Framework:** PyTorch 2.x con Ultralytics 8.4.38.
- **Sistema operativo:** Windows con Python 3.11.9.

El tiempo total de entrenamiento fue de 5,997.87 segundos (1 hora 39 minutos 58 segundos), con un promedio de 60 segundos por época.

3.8.5.4 Curva de convergencia

El Cuadro 8 presenta la evolución de las métricas de pérdida y rendimiento a lo largo del entrenamiento.

Cuadro 8
Evolución de métricas durante el entrenamiento del modelo YOLO26s-seg

Época	Tiempo (s)	Box loss	Seg loss	Cls loss	mAP50 (B)	mAP50-95 (B)	mAP50 (M)	mAP50-95 (M)
1	80.9	0.9622	1.3574	6.9787	0.1341	0.1210	0.1333	0.1081
10	633.8	0.7953	0.8663	1.2123	0.7573	0.6742	0.7546	0.6082
20	1,231.8	0.7198	0.7844	0.9126	0.8214	0.7298	0.8123	0.6574
30	1,822.3	0.7006	0.7735	0.8018	0.8366	0.7576	0.8219	0.6635
40	2,420.5	0.6648	0.7269	0.6999	0.8678	0.7906	0.8587	0.6944
50	3,018.3	0.6374	0.6818	0.6178	0.8840	0.8098	0.8731	0.7052
60	3,616.4	0.6204	0.6786	0.5792	0.9062	0.8278	0.8990	0.7257
70	4,214.8	0.6013	0.6617	0.5455	0.9086	0.8261	0.8998	0.7264
78	**4,700**	**–**	**–**	**–**	**0.9158**	**0.8323**	**0.9044**	**0.7320**
90	5,410.2	0.5735	0.6495	0.4742	0.9073	0.8232	0.8963	0.7206
100	5,997.9	0.4570	0.5175	0.2350	0.9054	0.8183	0.8959	0.7230

Nota. La época 78 (resaltada) corresponde al checkpoint `best.pt` seleccionado por Ultralytics como mejor modelo según mAP50. (B) = Bounding box, (M) = Mask.

El Cuadro 9 presenta los hitos de convergencia del modelo según el umbral de mAP50 alcanzado.

Cuadro 9
Hitos de convergencia del modelo YOLO26s-seg

Umbral mAP50	Primera época	Tiempo acumulado
≥ 0.50	3	205.8 s (3.4 min)
≥ 0.70	9	573.3 s (9.6 min)
≥ 0.80	17	1,052.8 s (17.5 min)
≥ 0.85	29	1,762.5 s (29.4 min)
≥ 0.90	59	3,556.4 s (59.3 min)
≥ 0.9158 (pico)	78	$\approx$ 4,700 s (78.3 min)

Nota. El modelo alcanzó el 80 % de su rendimiento final en apenas 17 minutos de entrenamiento.

Se observa una reducción notable en las pérdidas entre las épocas 90 y 100 (cls_loss de 0.4742 a 0.2350), atribuible a la desactivación del mosaic en las últimas 10 épocas (`close_mosaic=10`), lo cual simplifica el problema de visión y permite una convergencia más fina, comportamiento documentado por Ultralytics.

3.8.5.5 Métricas finales del modelo

El Cuadro 10 presenta las métricas de rendimiento del modelo YOLO26s-seg en el conjunto de validación, correspondientes al checkpoint `best.pt` (época 78).

Cuadro 10
Métricas de rendimiento del modelo YOLO26s-seg en validación

Métrica	Bounding Box	Máscara
Precision	0.8814	–
Recall	0.8875	–
mAP@50	**0.9158**	**0.9044**
mAP@50-95	**0.8323**	**0.7320**

Nota. Métricas correspondientes a la época 78, seleccionada automáticamente como mejor checkpoint.

3.8.6. Resultados del entrenamiento del modelo de detección

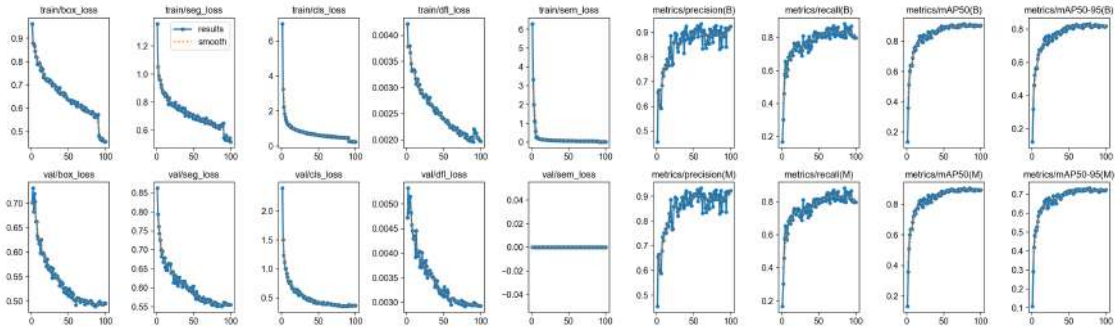
El proceso de entrenamiento del modelo YOLO26s-seg generó un conjunto de métricas y visualizaciones que permiten evaluar el rendimiento del sistema de detección y segmentación de señales de tránsito. A continuación se presentan los resultados obtenidos en el conjunto de validación.

3.8.6.1 Curvas de convergencia del entrenamiento

La Figura 55 presenta el panel completo de curvas de entrenamient, mostrando la evolución de las funciones de pérdida y métricas de rendimiento a lo largo de las 100 épocas de entrenamiento.

Figura 55

Panel de curvas de entrenamiento del modelo YOLO26s-seg



Nota. Fila superior: métricas de entrenamiento (train). Fila inferior: métricas de validación (val). Se observa convergencia estable sin evidencia de sobreajuste. La reducción abrupta de pérdidas en las épocas 90-100 corresponde a la desactivación del mosaic

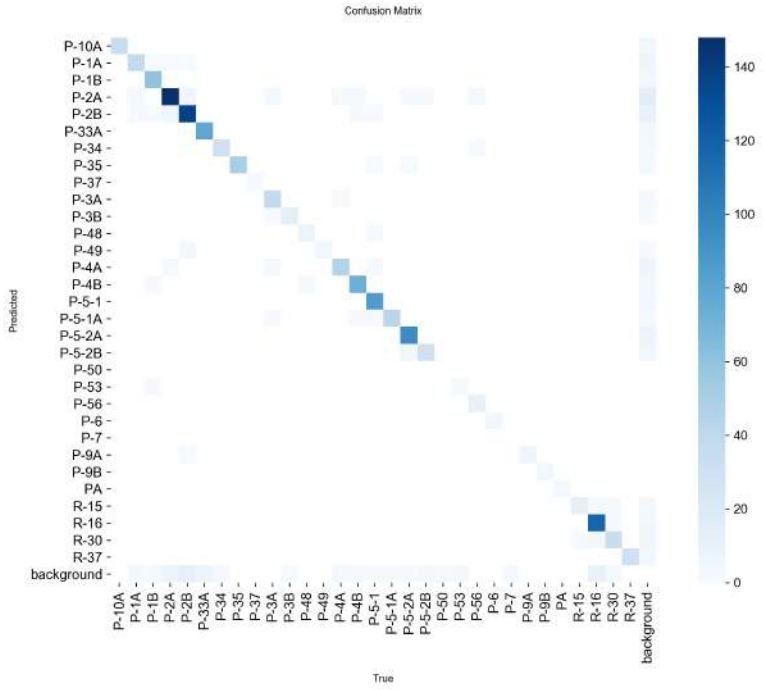
El análisis de las curvas de convergencia revela los siguientes hallazgos:

- **Pérdida de clasificación (cls_loss):** Presenta la mayor reducción relativa, disminuyendo de 7.0 en la época 1 a 0.23 en la época 100, lo cual indica que el modelo aprendió eficientemente a discriminar entre las 31 categorías de señales.
- **Pérdida de bounding box (box_loss):** Decrece de 0.96 a 0.46, evidenciando mejora progresiva en la localización espacial de las señales.
- **Pérdida de segmentación (seg_loss):** Reduce de 1.36 a 0.52, indicando refinamiento en la delimitación de máscaras de segmentación.
- **Métricas de validación:** Las curvas de train y validation mantienen tendencias paralelas sin divergencia significativa, confirmando ausencia de sobreajuste (overfitting).
- **Meseta de convergencia:** El modelo alcanza su rendimiento óptimo aproximadamente en la época 78, coincidiendo con el checkpoint `best.pt` seleccionado automáticamente.

3.8.6.2 Matriz de confusión

La Figura 56 presenta la matriz de confusión del modelo en el conjunto de validación, permitiendo identificar patrones de clasificación correcta y errores sistemáticos entre categorías.

Figura 56
Matriz de confusión del modelo YOLO26s-seg en validación



Nota. Eje vertical: clases predichas. Eje horizontal: clases reales (ground truth). La diagonal principal en azul oscuro indica clasificaciones correctas. La escala de color representa el número de instancias. Se observan 31 clases de señales más la categoría background.

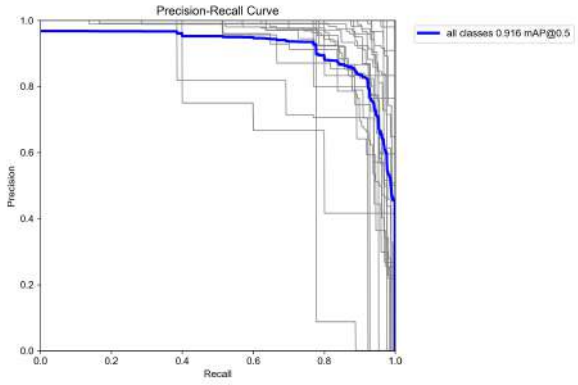
El análisis de la matriz de confusión permite identificar:

- **Diagonal dominante:** La concentración de valores altos en la diagonal principal confirma que el modelo clasifica correctamente la mayoría de instancias para todas las categorías.
- **Clases con mayor cantidad de muestras:** Las señales P-2A, P-2B y P-33A presentan los valores más altos en la diagonal (superiores a 100 instancias), reflejando su mayor representación en el dataset.
- **Confusiones menores:** Se observan confusiones marginales entre pares de señales visualmente similares, tales como P-1A/P-2A (curva pronunciada vs. curva) y P-1B/P-2B, lo cual es esperado dado que difieren únicamente en el ángulo del símbolo.
- **Categoría background:** La fila y columna correspondientes a background presentan valores bajos, indicando que el modelo genera pocas detecciones de falsos positivos en regiones sin señalización.

3.8.6.3 Curva Precision-Recall para bounding boxes

La Figura 57 presenta la curva Precision-Recall para la tarea de detección mediante bounding boxes.

Figura 57
Curva Precision-Recall del modelo YOLO26s-seg para bounding boxes



Nota. Línea azul gruesa: promedio de todas las clases (mAP@0.5 = 0.916). Líneas grises: curvas individuales para cada una de las 31 categorías de señales.

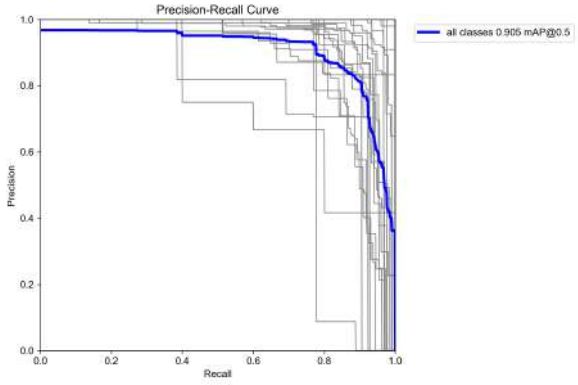
La curva Precision-Recall para bounding boxes evidencia:

- **mAP@0.5 global:** El modelo alcanza un mAP@0.5 de **0.916** (91.6 %), superando el umbral de 0.90 considerado como rendimiento excelente en tareas de detección de objetos.
- **Alta precisión inicial:** La curva mantiene precisión superior a 0.95 hasta valores de recall de aproximadamente 0.4, indicando que las detecciones de alta confianza son altamente confiables.
- **Degradación gradual:** La precisión disminuye gradualmente al incrementar el recall, comportamiento típico que refleja el trade-off inherente entre ambas métricas.
- **Variabilidad inter-clase:** Las curvas grises individuales muestran que algunas clases con menor representación en el dataset presentan curvas PR menos estables.

3.8.6.4 Curva Precision-Recall para máscaras de segmentación

La Figura 58 presenta la curva Precision-Recall para la tarea de segmentación de instancias.

Figura 58
Curva Precision-Recall del modelo YOLO26s-seg para máscaras de segmentación



Nota. Línea azul gruesa: Promedio de todas las clases (mAP@0.5 = 0.905). La segmentación de instancias presenta rendimiento ligeramente inferior a la detección por bounding boxes.

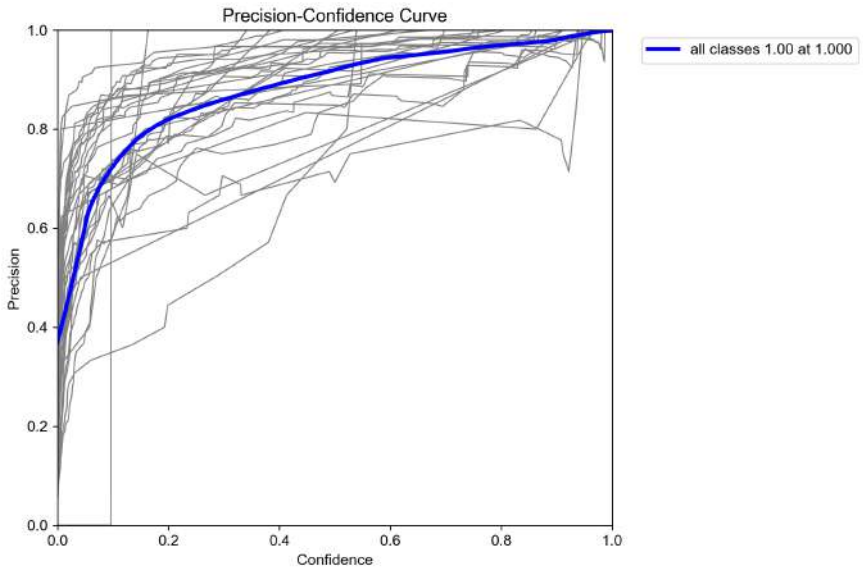
La comparación entre las curvas PR de bounding boxes y máscaras revela:

- **mAP@0.5 de segmentación:** El modelo alcanza un mAP@0.5 de **0.905** (90.5 %) para la tarea de segmentación de instancias.
- **Diferencia box vs. mask:** La diferencia de 0.011 puntos (0.916 - 0.905) entre detección y segmentación es marginal, indicando que el modelo genera máscaras de segmentación precisas que se ajustan adecuadamente al contorno de las señales.
- **Rendimiento consistente:** Ambas curvas presentan comportamiento similar, confirmando la coherencia del modelo en ambas tareas.

3.8.6.5 Curva Precision-Confidence

La Figura 59 presenta la relación entre la precisión del modelo y el umbral de confianza aplicado a las detecciones.

Figura 59
Curva Precision-Confidence del modelo YOLO26s-seg



Nota. La curva muestra que la precisión aumenta con el umbral de confianza, alcanzando 1.00 (100 %) cuando el umbral es 1.0. Línea azul: promedio de todas las clases. Líneas grises: curvas individuales por clase.

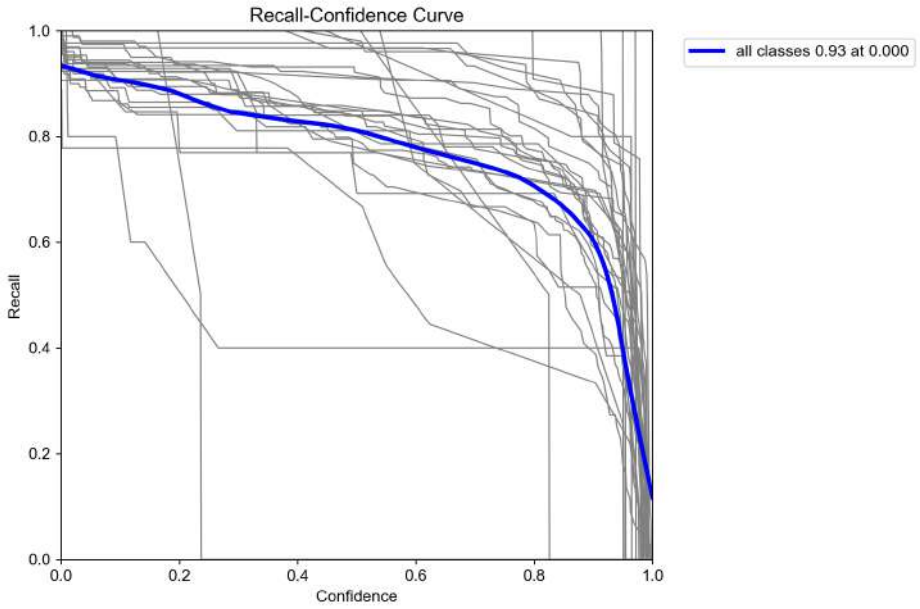
El análisis de la curva Precision-Confidence indica:

- **Precisión perfecta:** Con umbral de confianza igual a 1.0, el modelo alcanza precisión de **1.00**, indicando que las detecciones de máxima confianza son 100 % correctas.
- **Tendencia monotónica:** La precisión aumenta consistentemente con el umbral de confianza, comportamiento esperado que valida la calibración del modelo.
- **Umbral operativo:** Para aplicaciones de inventario vial donde se prefiere minimizar falsos positivos, se recomienda utilizar umbrales de confianza superiores a 0.7.

3.8.6.6 Curva Recall-Confidence

La Figura 60 presenta la relación inversa entre el recall del modelo y el umbral de confianza.

Figura 60
Curva Recall-Confidence del modelo YOLO26s-seg



Nota. La curva muestra que el recall disminuye al incrementar el umbral de confianza. Con umbral 0.0 se alcanza recall máximo de 0.93 (93 %). Línea azul: Promedio de todas las clases.

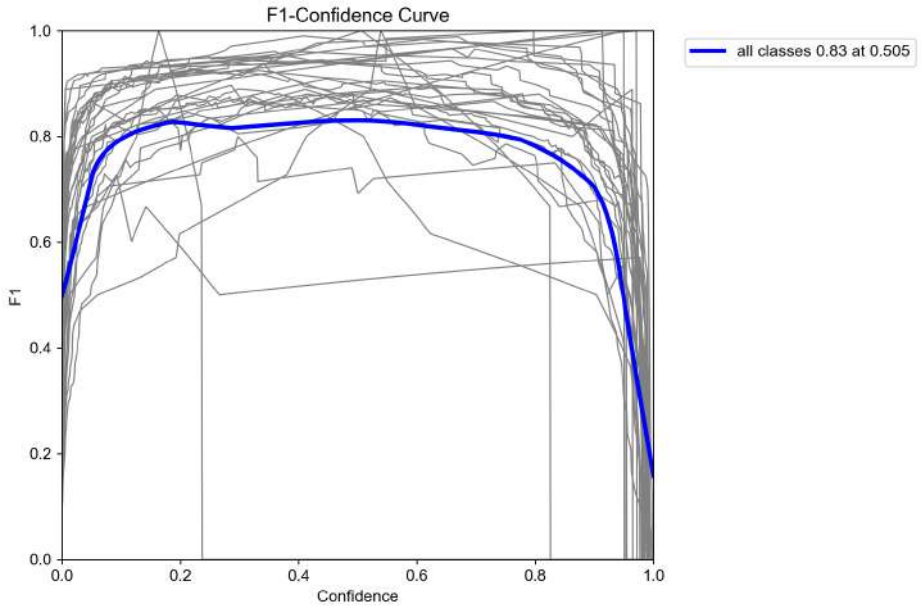
El análisis de la curva Recall-Confidence revela:

- **Recall máximo:** Con umbral de confianza igual a 0.0, el modelo alcanza un recall de **0.93** (93 %), indicando que detecta el 93 % de todas las señales presentes en las imágenes.
- **Trade-off precision-recall:** Al incrementar el umbral de confianza se sacrifica recall en favor de mayor precisión, relación inherente a los sistemas de detección.
- **Punto de operación:** Con umbral de confianza 0.5, el modelo mantiene un recall aproximado de 0.70, valor adecuado para aplicaciones de tiempo real.

3.8.6.7 Curva F1-Confidence

La Figura 61 presenta la curva F1-score en función del umbral de confianza, permitiendo identificar el punto óptimo de operación del modelo.

Figura 61
Curva F1-Confidence del modelo YOLO26s-seg



Nota. La curva presenta forma de campana característica del F1-score. El máximo de 0.83 se alcanza con umbral de confianza de 0.505. Línea azul: promedio de todas las clases.

El análisis de la curva F1-Confidence permite determinar:

- **F1-score máximo:** El modelo alcanza un F1-score máximo de **0.83** (83 %) con un umbral de confianza óptimo de **0.505**.
- **Forma de campana:** La curva presenta la forma característica del F1-score, con valores bajos en los extremos (donde domina precision o recall) y máximo en el punto de equilibrio.
- **Umbral recomendado:** Para el pipeline SIAVIR se recomienda utilizar un umbral de confianza cercano a 0.5 para maximizar el balance entre detecciones correctas y cobertura.
- **Rango operativo:** El F1-score se mantiene superior a 0.80 en el rango de confianza [0.2, 0.7], proporcionando flexibilidad para ajustar el umbral según los requerimientos específicos de la aplicación.

3.8.6.8 Resumen de métricas de rendimiento

El Cuadro 11 consolida las métricas de rendimiento extraídas de las curvas de evaluación del modelo YOLO26s-seg.

Cuadro 11
Resumen de métricas de rendimiento del modelo YOLO26s-seg

Métrica	Valor	Condición
mAP@0.5 (Bounding Box)	0.916	IoU $\geq$ 0.5
mAP@0.5 (Máscara)	0.905	IoU $\geq$ 0.5
Precision máxima	1.00	confidence = 1.0
Recall máximo	0.93	confidence = 0.0
F1-score máximo	0.83	confidence = 0.505

Nota. Métricas obtenidas en el conjunto de validación (1,185 imágenes, 1,346 instancias). El umbral de confianza óptimo para maximizar F1-score es 0.505.

Los resultados obtenidos demuestran que el modelo YOLO26s-seg entrenado alcanza un rendimiento sobresaliente en la tarea de detección y segmentación de señales de tránsito verticales, con un mAP@0.5 superior al 90 % tanto para bounding boxes como para máscaras de segmentación.

3.8.7. Entrenamiento del modelo de clasificación de estados

El segundo componente del pipeline SIAVIR corresponde al modelo de clasificación de estados de conservación, implementado mediante la arquitectura EfficientNet-B0 (Tan & Le, 2019). Este modelo recibe como entrada los recortes de señales detectadas por YOLO26s-seg y clasifica cada señal en una de siete categorías de estado de conservación.

3.8.7.1 Arquitectura del modelo

El modelo EfficientNet-B0 fue seleccionado por su balance óptimo entre precisión y eficiencia computacional, características esenciales para aplicaciones de tiempo real. El Cuadro 12 presenta las especificaciones técnicas de la arquitectura empleada.

Cuadro 12
Especificaciones de la arquitectura EfficientNet-B0

Atributo	Valor
Arquitectura base	torchvision.models.efficientnet_b0
Pesos pre-entrenados	EfficientNet_B0_Weights.IMAGENET1K_V1
Resolución de entrada	$224 \times 224 \times 3$ (RGB)
Normalización	ImageNet (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225])
Modificación al clasificador	model.classifier[1] = nn.Linear(1280, 7)
Parámetros totales	$\approx$ 4.04 millones
Tamaño del checkpoint	15.59 MB
Número de clases de salida	7 (estados de conservación)

Nota. El clasificador original de ImageNet (1,000 clases) fue reemplazado por una capa lineal de 7 salidas correspondientes a los estados de conservación definidos.

3.8.7.2 Estrategia de transferencia de aprendizaje

El entrenamiento se organizó en dos etapas secuenciales para aprovechar eficientemente los pesos pre-entrenados en ImageNet.

Cuadro 13
Configuración de las etapas de entrenamiento por transferencia

Etapa	Duración	Backbone	Clasificador	Learning Rate
1 (Warm-up)	10 épocas	Congelado	Entrenable	1×10^{-3}
2 (Fine-tune)	Hasta 40 épocas	Descongelado	Entrenable	1×10^{-4}

Nota. La Etapa 2 utiliza early stopping con paciencia de 10 épocas sobre la métrica val_macro_f1.

3.8.7.3 Hiperparámetros de entrenamiento

El Cuadro 14 presenta la configuración completa de hiperparámetros empleados durante el entrenamiento del modelo.

Cuadro 14
Hiperparámetros de entrenamiento del modelo EfficientNet-B0

Parámetro	Valor	Descripción
Configuración general		
seed	42	Semilla para reproducibilidad
deterministic	True	Operaciones determinísticas (cudnn)
device	cuda:0	GPU NVIDIA RTX 5080
batch_size	32	Tamaño de lote
num_workers	0	Hilos de carga (Windows)
pin_memory	True	Memoria fija en GPU
Optimizador		
optimizer	AdamW	Optimizador con weight decay
lr (Etapa 1)	1×10^{-3}	Learning rate inicial
lr (Etapa 2)	1×10^{-4}	Learning rate fine-tune
weight_decay	0.01	Regularización L2
scheduler	CosineAnnealingLR	Decaimiento coseno (T_max=40)
Función de pérdida		
loss	CrossEntropyLoss	Con pesos de clase
label_smoothing	0.1	Suavizado de etiquetas
class_weight_mode	sqrt_balanced	Raíz cuadrada del peso balanceado
Early stopping		
metric	val_macro_f1	Métrica de monitoreo
patience	10	Épocas sin mejora
mode	max	Maximizar la métrica

Nota. Pesos extremos (He & García, 2009; Cui et al., 2019).

3.8.7.4 Distribución del dataset de clasificación

El dataset de clasificación se construyó a partir de los recortes generados por el modelo YOLO26s-seg sobre las imágenes anotadas. El Cuadro 15 presenta la distribución de muestras por cada estado de conservación.

Cuadro 15
Distribución del dataset de clasificación por estado de conservación

Estado de conservación	Muestras	Porcentaje
OPERATIVO	5,264	82.1 %
DETERIORO_POSTE	493	7.7 %
ROTACIÓN	245	3.8 %
OBSTRUCCIÓN	152	2.4 %
DECOLORACIÓN	150	2.3 %
VANDALISMO	73	1.1 %
DETERIORO_TABLERO	35	0.6 %
Total	**6,412**	**100.0 %**

Nota. El ratio de desbalance OPERATIVO/DETERIORO_TABLERO es de aproximadamente 150:1, lo cual requirió técnicas específicas de manejo de clases minoritarias.

3.8.7.5 Pipeline de aumentación de datos en tres capas

Para mitigar el desbalance extremo del dataset (ratio 150:1 entre la clase mayoritaria y minoritaria), se diseñó un esquema de aumentación de datos en tres capas complementarias, siguiendo las recomendaciones de Shorten & Khoshgoftaar (2019) para clasificación con clases desbalanceadas.

A. Capa 1: Aumentación online

La primera capa aplica transformaciones estocásticas al vuelo durante cada época de entrenamiento, utilizando la biblioteca torchvision.transforms. El Cuadro 16 detalla las técnicas empleadas.

Cuadro 16
Técnicas de aumentación online (Capa 1)

Técnica	Parámetros
RandomResizedCrop	224×224, scale=(0.85, 1.0), ratio=(0.9, 1.1)
RandomHorizontalFlip	p = 0.5
RandomRotation	±15°
ColorJitter	brightness=0.3, contrast=0.3, saturation=0.3, hue=0.1
RandomAffine	translate=±10 %, scale=(0.9, 1.1)
GaussianBlur	kernel=3, σ=(0.1, 1.0)
RandomErasing	p=0.2, scale=(0.02, 0.08)

Nota. Todas las transformaciones se aplican secuencialmente después del redimensionamiento a 256×256 píxeles.

B. Capa 2: Aumentación offline para balance

La segunda capa genera imágenes sintéticas en disco antes del entrenamiento, aplicando factores de aumento diferenciados por clase según su representación en el dataset original. El Cuadro 17 presenta los factores de multiplicación aplicados.

Cuadro 17
Factores de aumentación offline por clase (Capa 2)

Clase	Factor	Generadas	Train final
OPERATIVO	–	0	1,500 (submuestreado)
DETERIORO_POSTE	1.5×	+173	≈345
ROTACIÓN	3×	+342	≈513
OBSTRUCCIÓN	5×	+424	≈530
DECOLORACIÓN	5×	+420	≈525
VANDALISMO	8×	+357	≈408
DETERIORO_TABLERO	10×	+225	≈250

Nota. La clase OPERATIVO se submuestrea a 1,500 imágenes mediante muestreo uniforme, siguiendo la estrategia de Kang et al. (2020) para reconocimiento de cola larga.

Las técnicas de Capa 2 incluyen combinaciones aleatorias de 2-3 transformaciones: flip horizontal, rotación ±10°, ajuste de brillo ±20 %, contraste ±20 %, traslación ±5 % y zoom ±5 %.

C. Capa 3: Aumentación específica por clase

La tercera capa implementa simulaciones físicas semánticas específicas para cada tipo de deterioro. Aunque fue evaluada en el experimento E04, se descartó del modelo final debido a sobreajuste a patrones sintéticos. El Cuadro 18 documenta las técnicas implementadas.

Cuadro 18
Técnicas de aumentación específica por clase (Capa 3)

Clase	Versiones	Técnicas simuladas
DETERIORO_POSTE	2	Fade gradual en mitad inferior (intensidad 0.3–0.6)
DECOLORACIÓN	2	Reducción de saturación (0.3–0.7) y contraste (0.5–0.8)
OBSTRUCCIÓN	2	2–5 elipses verde oscuras simulando vegetación
ROTACIÓN	3	Rotación ±30°, rotación + downscale 40–55 %
VANDALISMO	3	1–3 parches RGB aleatorios simulando stickers/grafiti
DETERIORO_TABLERO	5	Cutout rectangular, cutout irregular, warp perspectiva

Nota. Esta capa fue descartada del modelo final (E03) debido a que incrementó el sobreajuste sin mejorar la generalización.

3.8.7.6 Estudio de ablación

Se realizó un estudio de ablación sistemático con cinco experimentos para determinar la combinación óptima de técnicas de aumentación y manejo de desbalance. El Cuadro 19 presenta los resultados obtenidos en el conjunto de test (n=642).

Cuadro 19
Resultados del estudio de ablación para EfficientNet-B0

Exp.	Sub.	Weights	C1	C2	C3	Acc	Macro F1	Tiempo
E01	No	none	–	–	–	0.9393	0.7462	9.0 min
E02	No	sqrt	✓	–	–	0.9330	0.8104	21.1 min
E02b	Sí	sqrt	✓	–	–	0.9128	0.7052	10.3 min
E03	**Sí**	**sqrt**	✓	✓	–	**0.9315**	**0.8252**	**18.2 min**
E04	Sí	sqrt	✓	✓	✓	0.9346	0.8083	19.7 min

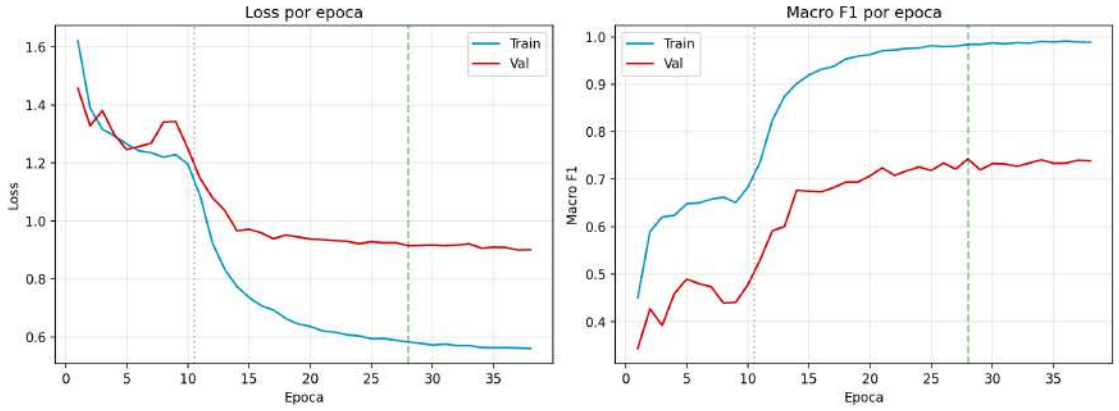
Nota. Sub. = Submuestreo de OPERATIVO a 1,500. Weights = Modo de pesos de clase (sqrt = sqrt_balanced). C1/C2/C3 = Capas de aumentación. El experimento E03 (resaltado) fue seleccionado como configuración final.

El análisis estadístico mediante test de McNemar exacto identificó que la transición E02b → E03 fue la única estadísticamente significativa del estudio (p = 0.0446), atribuyendo el rescate de clases minoritarias específicamente a la Capa 2 de aumentación offline. Los incrementos en F1 por clase fueron: DECOLORACIÓN +0.292, DETERIORO_TABLERO +0.300, VANDALISMO +0.209.

3.8.7.7 Curvas de entrenamiento

La Figura 62 presenta las curvas de pérdida y Macro F1 durante el entrenamiento del experimento ganador E03.

Figura 62
Curvas de entrenamiento del modelo EfficientNet-B0 (E03)



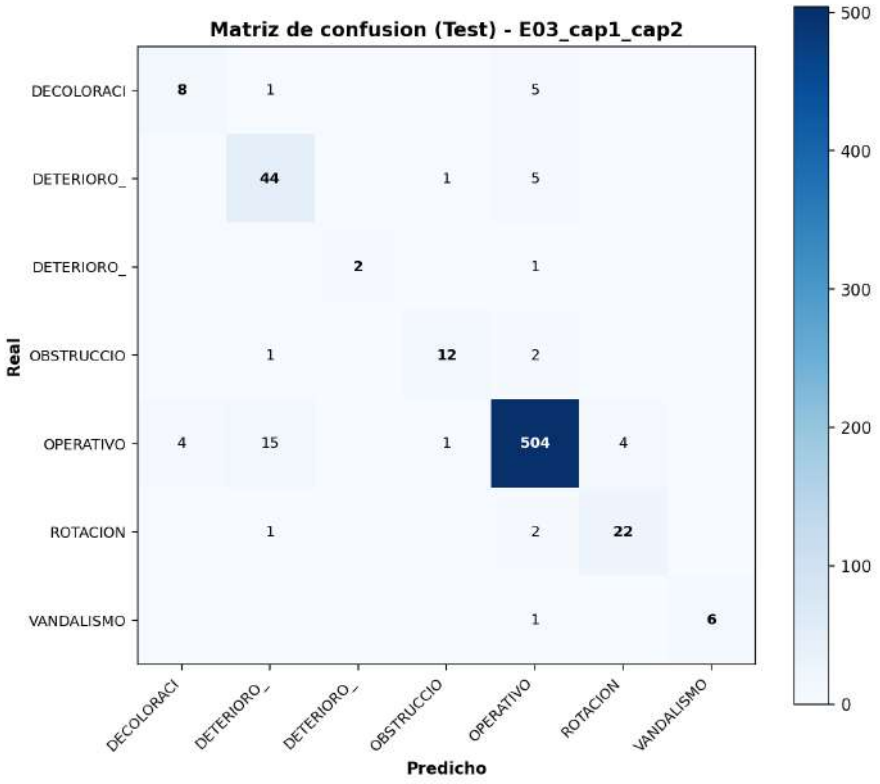
Nota. Izquierda: Evolución de la función de pérdida. Derecha: evolución del Macro F1. La línea punteada gris (época 10) indica el fin de la Etapa 1 (backbone congelado). La línea punteada verde (época 28) indica la mejor época seleccionada por early stopping. El entrenamiento finalizó en la época 38.

El análisis de las curvas de entrenamiento revela:

3.8.7.8 Matriz de confusión

La Figura 63 presenta la matriz de confusión del modelo E03 en el conjunto de test.

Figura 63
Matriz de confusión del modelo EfficientNet-B0 (E03) en test



Nota. Eje vertical: clases reales. Eje horizontal: clases predichas. La diagonal principal muestra las clasificaciones correctas. El modelo logra alta precisión en OPERATIVO (504/528) y buen rendimiento en clases minoritarias como VANDALISMO (6/7) y ROTACIÓN (22/25).

El análisis de la matriz de confusión permite identificar:

- **OPERATIVO:** Alta precisión con 504 de 528 muestras correctamente clasificadas (95.5 %). Los errores principales se distribuyen hacia DETERIORO_POSTE (15) y DECOLORACIÓN (4).
- **DETERIORO_POSTE:** Buen rendimiento con 44 de 50 muestras correctas (88.0 %). Algunas confusiones con OPERATIVO (5) debido a deterioros leves.
- **ROTACIÓN:** Excelente recall con 22 de 25 muestras correctas (88.0 %). La característica visual distintiva (inclinación) facilita su identificación.
- **OBSTRUCCIÓN:** Rendimiento sólido con 12 de 15 muestras correctas (80.0 %). Confusiones menores con OPERATIVO cuando la obstrucción es parcial.
- **VANDALISMO:** Alto recall con 6 de 7 muestras correctas (85.7 %). La única confusión fue con OPERATIVO.
- **DECOLORACIÓN:** Clase más desafiante con 8 de 14 muestras correctas (57.1 %). La confusión principal es con OPERATIVO (5), lo cual es esperado dado que la

decoloración leve puede ser difícil de distinguir.

- **DETERIORO_TABLERO:** A pesar de tener solo 3 muestras en test, el modelo clasificó 2 correctamente (66.7 %), demostrando aprendizaje efectivo a pesar del desbalance extremo.

3.8.7.9 Métricas de rendimiento por clase

El Cuadro 20 presenta las métricas detalladas de rendimiento para cada estado de conservación.

Cuadro 20
Métricas de rendimiento por clase del modelo EfficientNet-B0 (E03)

Estado	Precision	Recall	F1-Score	Support
OPERATIVO	0.9692	0.9545	0.9618	528
DETERIORO_POSTE	0.7097	0.8800	0.7857	50
ROTACIÓN	0.8462	0.8800	0.8627	25
OBSTRUCCIÓN	0.8571	0.8000	0.8276	15
VANDALISMO	1.0000	0.8571	0.9231	7
DECOLORACIÓN	0.6667	0.5714	0.6154	14
DETERIORO_TABLERO	1.0000	0.6667	0.8000	3
Macro promedio	**0.8641**	**0.8014**	**0.8252**	**642**

Nota. El Macro F1 de 0.8252 representa el promedio no ponderado de los F1-scores individuales, métrica adecuada para evaluar el rendimiento en datasets desbalanceados.

3.8.7.10 Métricas adicionales de evaluación

El Cuadro 21 presenta métricas complementarias que caracterizan el rendimiento global del modelo.

Cuadro 21
Métricas globales del modelo EfficientNet-B0 (E03)

Métrica	Valor	Interpretación
Accuracy	0.9315	93.15 % de clasificaciones correctas
Macro F1	0.8252	Promedio no ponderado de F1 por clase
Weighted F1	0.9324	Promedio ponderado por soporte
Cohen's κ	0.7883	Concordancia sustancial (Landis & Koch, 1977)
MCC	0.7892	Correlación de Matthews

Nota. El coeficiente Kappa de Cohen y el MCC son métricas robustas para evaluación de clasificadores en datasets desbalanceados.

3.8.7.11 Hardware y tiempo de entrenamiento

El entrenamiento del modelo EfficientNet-B0 se ejecutó en el siguiente entorno computacional:

- **GPU:** NVIDIA GeForce RTX 5080 con 16 GB de VRAM.

- **Software:** Python 3.11.9, PyTorch 2.x, torchvision, scikit-learn.
- **Tamaño del modelo:** 15.59 MB (checkpoint best.pt).

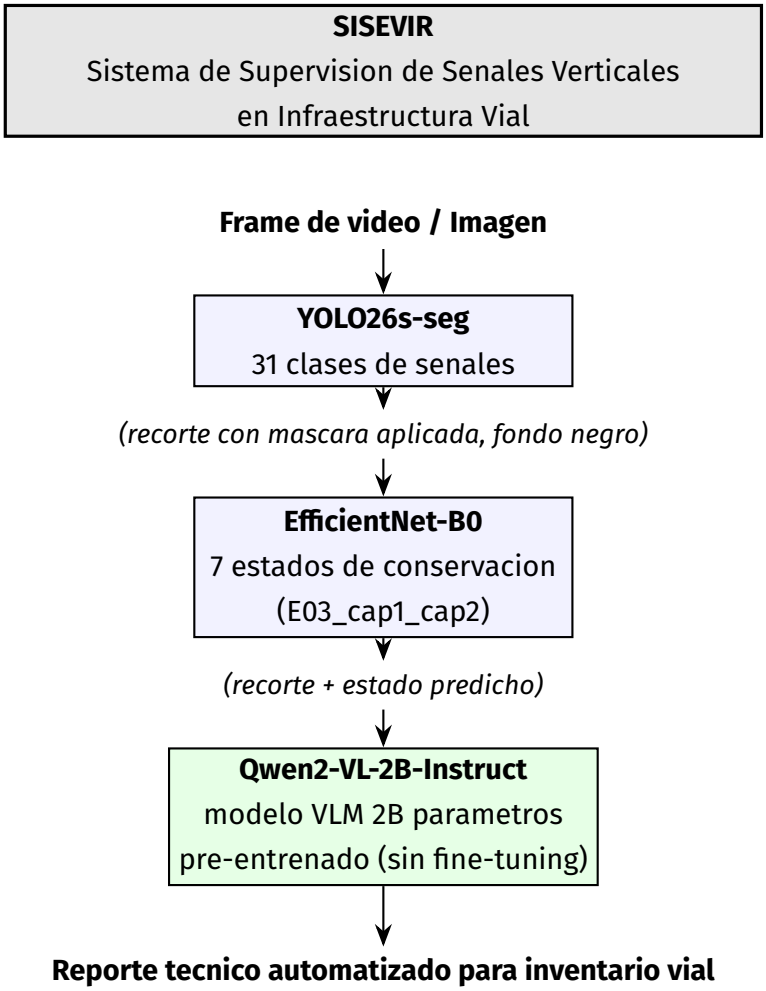
3.8.8. Modelo de generacion de descripciones en lenguaje natural Qwen2-VL-2B

El tercer y ultimo componente del pipeline SISEVIR corresponde al modelo de vision-lenguaje Qwen2-VL-2B-Instruct (Alibaba Cloud, 2024), encargado de generar descripciones textuales en espanol que explican el estado de conservacion de cada senal detectada. Este componente transforma las clasificaciones numericas de los modelos anteriores en informacion interpretable para los tecnicos de mantenimiento vial.

3.8.8.1 Arquitectura del pipeline SISEVIR

La Figura 64 presenta el flujo completo del sistema SISEVIR, mostrando la integracion secuencial de los tres modelos de inteligencia artificial.

Figura 64
Arquitectura del pipeline SISEVIR para deteccion y evaluacion de senales verticales



Nota. El pipeline procesa secuencialmente cada frame de video o imagen, generando un reporte con la ubicacion, tipo, estado y descripcion de cada senal detectada.

El principio de diseño del sistema SISEVIR se fundamenta en que los modelos YOLO y EfficientNet generan categorias discretas (clase de senal y estado de conservacion), pero

el inventario vial real requiere descripciones en lenguaje natural que detallen el tipo especifico de deterioro. Por ejemplo, el estado “VANDALISMO” puede manifestarse como grafiti, pegatinas publicitarias o marcas con marcador, distinciones que solo un modelo de vision-lenguaje como Qwen2-VL puede articular. Esta capa final de interpretacion semantica complementa las capacidades de clasificacion de las redes convolucionales tradicionales.

3.8.8.2 Especificaciones del modelo

El Cuadro 22 presenta las especificaciones tecnicas del modelo Qwen2-VL empleado en el sistema SISEVIR.

Cuadro 22
Especificaciones del modelo Qwen2-VL-2B-Instruct

Atributo	Valor
Identificador HuggingFace	Qwen/Qwen2-VL-2B-Instruct
Familia	Qwen2-VL (Alibaba Cloud / Tongyi Lab)
Modalidad	Vision-Language Model multimodal
Parametros	2 mil millones (2B)
Precision	torch.float16 (GPU)
Cuantizacion	Ninguna (fp16 directo)
Fine-tuning	No aplicado (pesos pre-entrenados de fabrica)
VRAM requerida	5–8 GB durante inferencia
GPU	NVIDIA RTX 5080 (compartida con YOLO + EfficientNet)
Ejecucion	Local via transformers (no API externa)
Tiempo de carga	5–15 segundos
Tiempo de inferencia	0.5–2.2 segundos por imagen
Idioma de salida	Espanol nativo
Cache HuggingFace	D:\hf_cache\hub\

Nota. Se selecciono la version 2B por su balance entre calidad de descripcion y eficiencia computacional. La version 7B fue descartada por requerimientos de VRAM superiores a los disponibles cuando los tres modelos operan simultaneamente.

3.8.8.3 Prompts dirigidos por estado

Para maximizar la relevancia de las descripciones generadas, se diseno un catalogo de siete prompts especificos en espanol, uno para cada estado de conservacion. El prompt se selecciona dinamicamente segun la clasificacion de EfficientNet-B0. El Cuadro 23 presenta las plantillas empleadas en el modo de procesamiento en tiempo real.

Cuadro 23
Prompts dirigidos por estado de conservacion para Qwen2-VL (modo real-time)

Estado	Prompt en espanol
OPERATIVO	¿La señal esta en buen estado? Responde en una oracion corta.?
VANDALISMO	¿Esta señal de transito tiene grafiti pintado o pegatinas/publicidad pegadas? Responde en una oracion corta.
DECOLORACION	¿La pintura de esta señal esta descolorida Responde en una oracion corta.?
DETERIORO_TABLERO	¿Que defecto observas en el tablero de esta señal? Responde en una oracion corta.
DETERIORO_POSTE	¿Que deterioro observas en el poste de esta señal? Responde en una oracion corta.
ROTACION	¿Esta inclinada esta señal? Responde en una oracion corta.
OBSTRUCCION	¿Que cubre u obstruye esta señal? Responde en una oración corta.

Nota. Los prompts terminan con “Responde en una oracion corta” para minimizar el tiempo de generacion. En el modo de validacion se usa “Describe brevemente” para obtener respuestas mas detalladas (max_new_tokens = 200).

3.8.8.4 Parámetros de generación

El Cuadro 24 presenta la configuracion de generacion empleada en los diferentes modos de operacion del sistema.

Cuadro 24
Parametros de generacion del modelo Qwen2-VL segun modo operativo

Parametro	Etapa 1 (test)	Etapa 2 (val.)	Real-time
max_new_tokens	200	200	60
do_sample	False	False	False
Decoder	greedy	greedy	greedy
Temperature	–	–	–
top_p / top_k	–	–	–
Determinismo	si	si	si
Lado maximo imagen (px)	sin limite	768	512
Tiempo medio (s/imagen)	2.2	0.5–1.5	1–2

Nota. El decoding greedy (do_sample=False) garantiza reproducibilidad deterministica y minimiza el tiempo de generacion. El redimensionamiento de imagen a 512 px en modo real-time optimiza la velocidad de inferencia.

3.8.8.5 Tiempos de inferencia

El Cuadro 25 presenta los tiempos de inferencia medidos en diferentes contextos de operacion.

Cuadro 25
Tiempos de inferencia del modelo Qwen2-VL

Contexto	Tiempo medio	Observacion
Etapa 1 (imagen completa, n=11)	2.2 s/imagen	max_new_tokens = 200
Etapa 2 (crop con mascara, n=30)	0.5–1.5 s/imagen	max_new_tokens = 200
Pipeline real-time	1–2 s/descripcion	max_new_tokens = 60
Carga del modelo	5–15 s	Una sola vez al inicio de sesion

Nota. En el video de prueba (5,796 frames, 3 minutos), Qwen describio 118/118 senales unicas detectadas gracias al esquema de threading asincrono.

3.8.8.6 Ejemplos de descripciones generadas

El Cuadro 26 presenta ejemplos verificados de descripciones generadas por el modelo durante la validacion del sistema.

Cuadro 26
Ejemplos de descripciones generadas por Qwen2-VL en el pipeline SISEVIR

YOLO	EfficientNet	Conf.	Descripcion generada
R-16	OBSTRUCCION	97.4 %	“La senal de transito esta cubierta con hojas de arbol.”
P-2A	DETERIORO_TABLERO	94.5 %	“El tablero de la senal esta abollado y tiene partes rotas/desprendidas.”
R-16	ROTACION	93.0 %	“La senal de transito en la imagen esta inclinada.”
R-30	DECOLORACION	91.4 %	“La pintura de la senal de transito parece estar descolorida.”
P-1A	VANDALISMO	89.2 %	“La senal de transito tiene grafiti pintado en ella.”
P-2B	OPERATIVO	98.1 %	“Si, la senal esta en buen estado.”

Nota. Las descripciones se generan en espanol nativo sin necesidad de traduccion posterior. El modelo Qwen2-VL responde en el idioma del prompt.

3.8.8.7 Validacion cualitativa

Se realizo una validacion cualitativa sobre 30 recortes de senales (5 por cada clase de deterioro, excluyendo OPERATIVO). El Cuadro **??** presenta los resultados obtenidos.

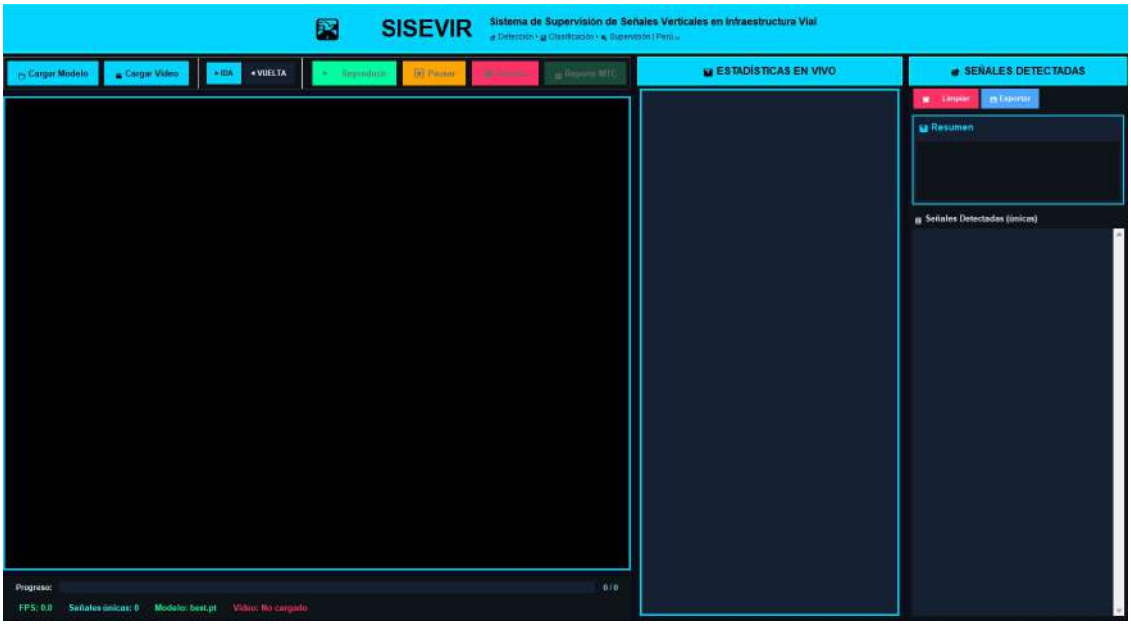
3.8.9. Descripcion del software SISEVIR

El software SISEVIR (*Sistema de Supervision de Señales Verticales en Infraestructura Vial*) constituye la implementacion computacional del pipeline de inteligencia artificial desarrollado en la presente investigacion. La interfaz grafica integra los tres modelos de aprendizaje profundo (YOLO26s-seg, EfficientNet-B0 y Qwen2-VL-2B-Instruct) en un flujo de trabajo unificado que permite el analisis automatizado de videograbaciones viales para la generacion de inventarios de senalizacion vertical compatibles con los estandares del Ministerio de Transportes y Comunicaciones (MTC) del Peru.

3.8.9.1 Paso 1: Interfaz principal del software

La Figura 65 presenta la interfaz principal del software SISEVIR. La interfaz adopta una paleta cromatica oscura que facilita la visualizacion de los fotogramas de video durante el analisis, y se organiza en tres zonas principales: Panel de reproduccion de video (izquierda), panel de estadisticas en vivo (centro) y panel de senales detectadas (derecha).

Figura 65
Interfaz principal del software SISEVIR

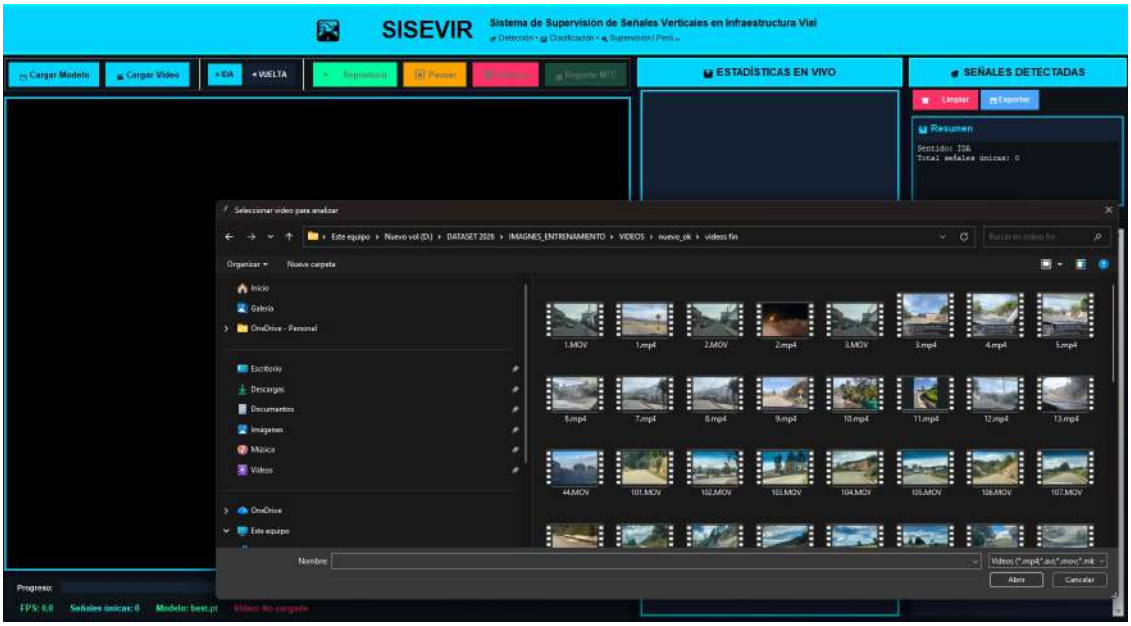


Nota. La barra de herramientas superior permite: Cargar modelo (`best.pt`), cargar video, definir sentido de analisis (IDA / VUELTA), reproducir, pausar, detener y generar Reporte MTC. La barra de estado inferior indica en tiempo real los FPS, senales unicas detectadas, modelo activo y caracteristicas del video.

3.8.9.2 Paso 2: Carga del video para analisis

El flujo de trabajo se inicia con la selección del archivo de video a analizar. La Figura 66 muestra el dialogo de seleccion, que acepta formatos MP4, AVI, MOV y MKV, organizados por carpetas de tramo vial.

Figura 66
Dialogo de selección de video para análisis en SISEVIR

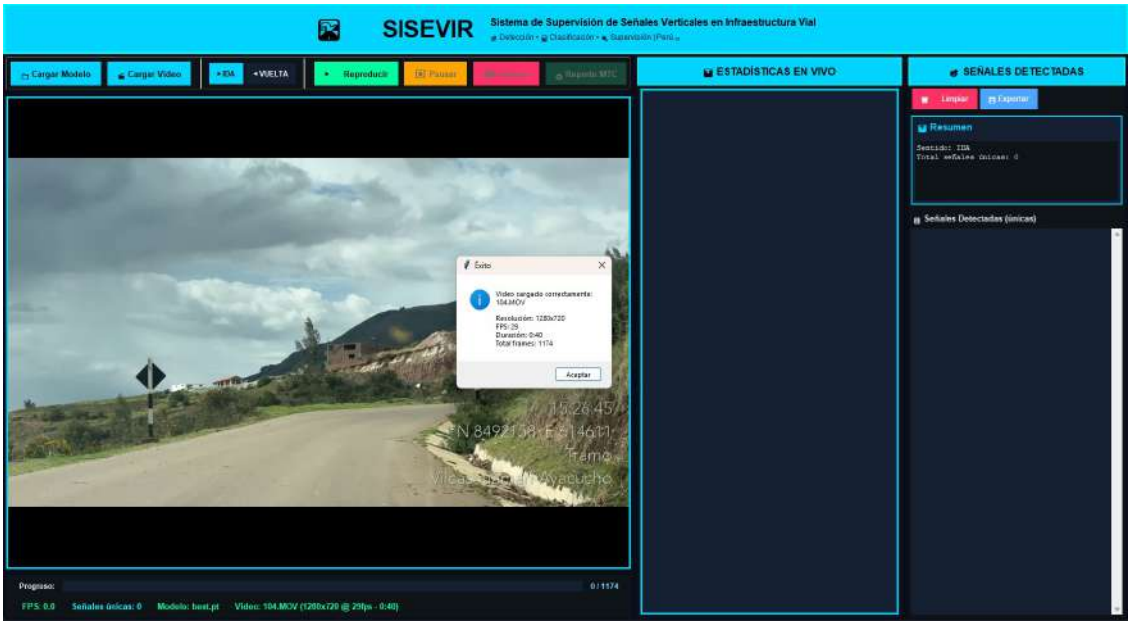


Nota. El explorador muestra los videos organizados en la carpetas correspondientes a los tramos PE-3S y PE-28A.

3.8.9.3 Paso 3: Confirmación de especificaciones del video

Una vez seleccionado el video, el sistema despliega un cuadro de confirmacion con las especificaciones tecnicas del archivo. La Figura 67 ilustra este proceso para el video 104.MOV.

Figura 67
Confirmación de carga de video con especificaciones técnicas

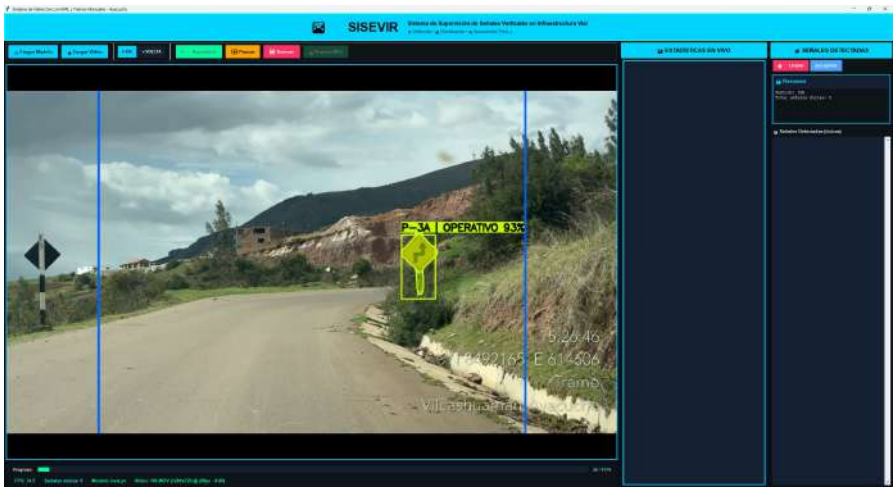


Nota. El cuadro de exito muestra:Resolucion 1280×720 pixeles, 29 FPS, duracion de 0:40 minutos y 1,174 frames totales. La barra de estado inferior confirma el video activo con sus parametros.

3.8.9.4 Paso 4: Sistema de lineas verticales de conteo

Para garantizar el conteo preciso de senales unicas y evitar duplicaciones, el software implementa dos lineas de conteo verticales configurables. La Figura 68 muestra este mecanismo en operación.

Figura 68
Lineas verticales de conteo durante el análisis de video

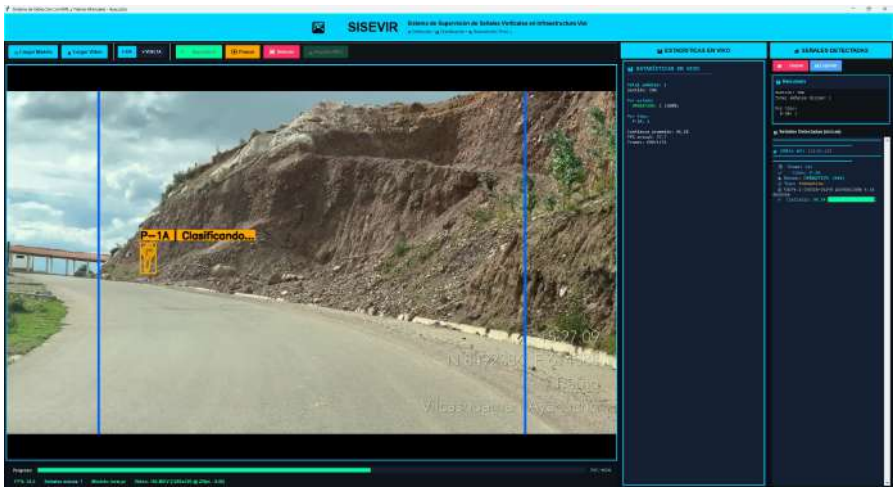


Nota. Las lineas azules verticales delimitan la zona de conteo activo. Una senal se registra como unica cuando su bounding box cruza alguna de las lineas segun el sentido configurado (IDA o VUELTA). Se observa la senal P-3A con etiqueta OPERATIVO al 93 % de confianza.

3.8.9.5 Paso 5: Detección y clasificación en tiempo real

La Figura 69 ilustra el proceso de detección y clasificación durante el analisis de un fotograma del video.

Figura 69
Deteccion y clasificacion en tiempo real de senal P-1A



Nota. El software detecta la senal P-1A (curva pronunciada a la derecha) mostrando “Clasificando...” mientras EfficientNet-B0 determina el estado de conservacion. El panel derecho registra la primera senal detectada (P-3A, OPERATIVO, 94 % confianza). FPS actual: 34.5. Frame: 717/1174.

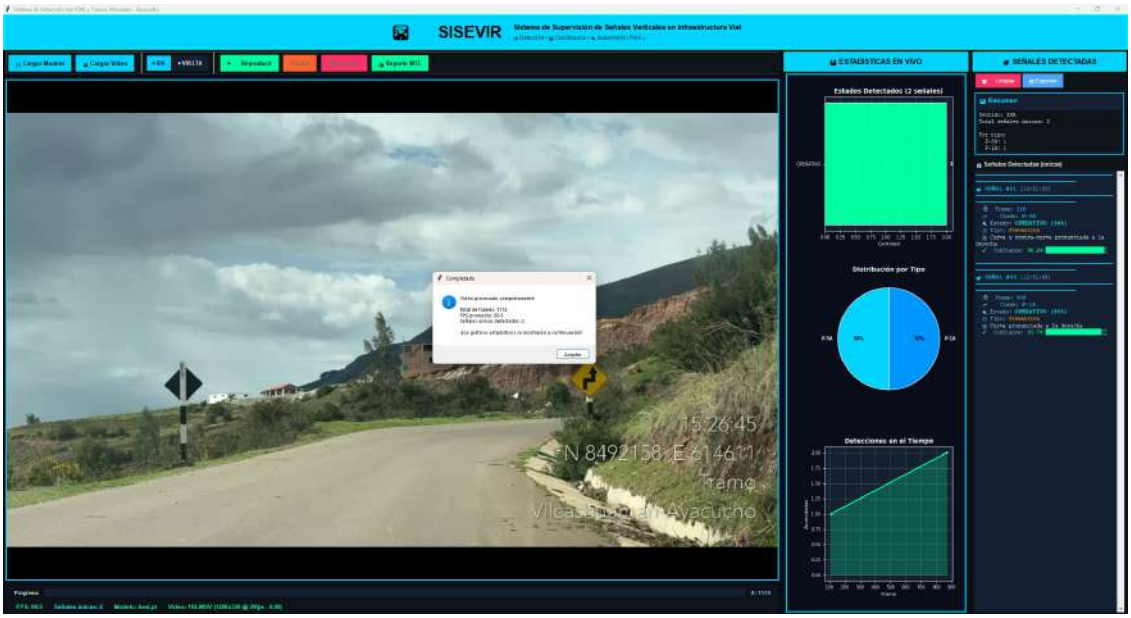
Durante el procesamiento en tiempo real, el sistema ejecuta secuencialmente los siguientes pasos por cada fotograma:

1. **Deteccion YOLO26s-seg:** Identifica la presencia, tipo y contorno de la senal con bounding box y mascara de segmentacion poligonal.
2. **Clasificacion EfficientNet-B0:** Determina el estado de conservacion del recorte. Durante este proceso se muestra “Clasificando...” en la etiqueta del bounding box.
3. **Descripcion Qwen2-VL:** Genera en hilo secundario la descripcion textual en espanol, mostrando “Describiendo...” hasta su disponibilidad.
4. **Registro en inventario:** Si la senal cruza una linea de conteo, se registra con todos sus atributos en el panel lateral.

3.8.9.6 Paso 6: Resumen del procesamiento completado

Al finalizar el analisis, el sistema presenta un cuadro de resumen con las estadisticas globales. La Figura 70 muestra este resultado para el video 104.MOV.

Figura 70
Cuadro de resumen al finalizar el procesamiento del video



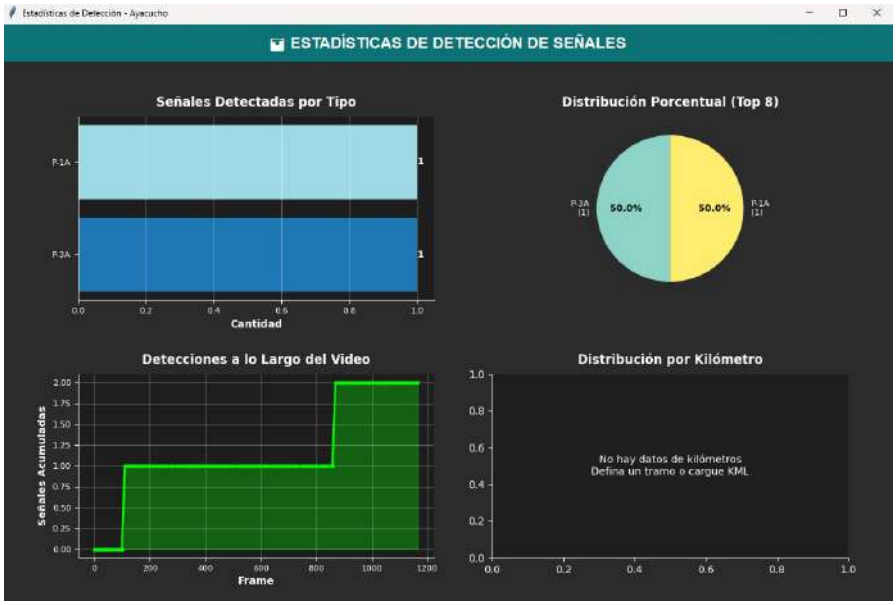
Nota. El cuadro de completado indica: 1,174 frames totales procesados, FPS promedio de 90.9 y 2 senales unicas detectadas. El panel derecho presenta automaticamente los graficos estadisticos: estados detectados (verde = OPERATIVO), distribucion por tipo (50 % P-3A, 50 % P-1A) y detecciones acumuladas por frame.

3.8.9.7 Paso 7: Panel de estadisticas de detección

El software genera automaticamente un panel de estadisticas graficas al finalizar el procesamiento. La Figura 71 muestra la ventana de estadisticas de detección.

Figura 71

Panel de estadisticas de detección de señales del software SISEVIR



Nota. El panel incluye cuatro graficos: (a) Señales detectadas por tipo (barras horizontales), (b) distribucion porcentual por tipo (grafico circular con 50 % P-3A y 50 % P-1A), (c) detecciones acumuladas a lo largo del video por frame, y (d) distribucion por kilometro (requiere datos KML del tramo).

La Figura 72 presenta el panel de analisis estadistico avanzado con informacion complementaria sobre niveles de confianza.

Figura 72

Panel de analisis estadistico avanzado del sistema SISEVIR

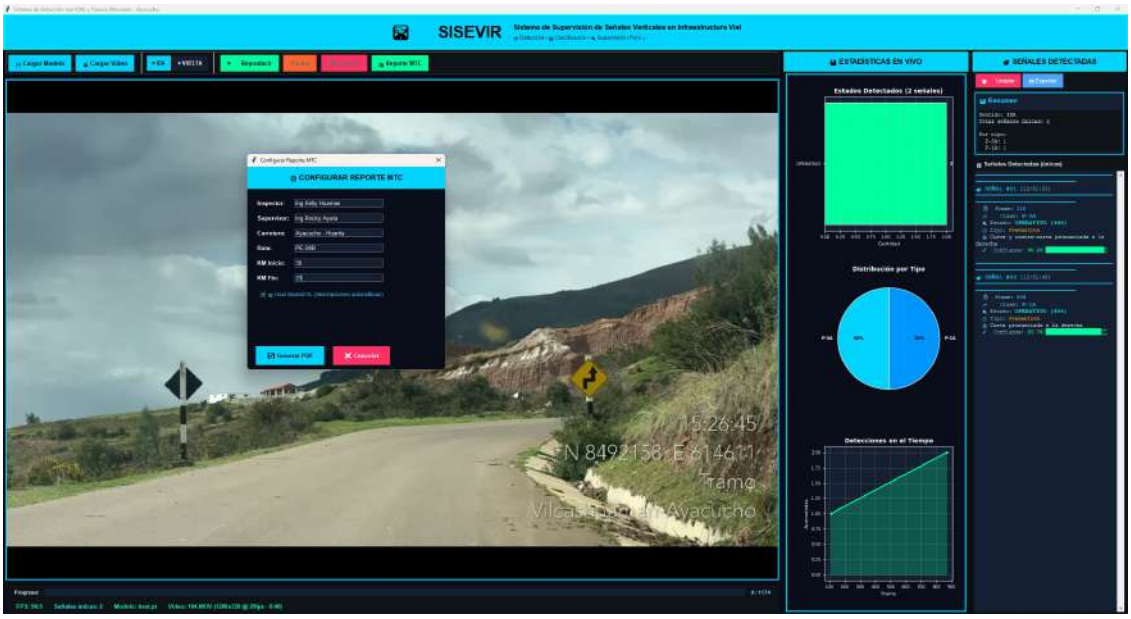


Nota. El cuadro de resumen estadistico indica: 2 senales totales (0 reguladoras, 2 preventivas), confianza promedio de 96.0 %, confianza maxima de 98.2 %, confianza minima de 93.7 %. Se incluyen graficos de clasificacion general, detecciones acumuladas y distribucion de confianza.

3.8.9.8 Paso 8: Configuración del Reporte MTC

Una vez finalizado el analisis, el usuario configura el reporte automatizado compatible con el formato del MTC. La Figura 73 muestra el formulario de configuración.

Figura 73
Formulario de configuracion del Reporte MTC en SISEVIR

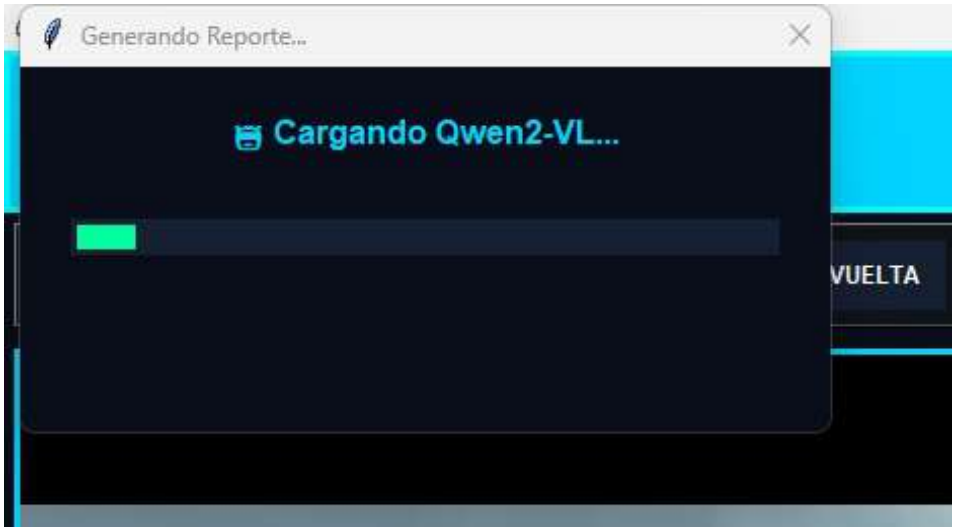


Nota. El formulario permite ingresar: Inspector (Ing. Kelly Huaman), Supervisor (Ing. Rocky Ayala), Carretera (Ayacucho – Huanta), Ruta (PE-28B), KM Inicio (30) y KM Fin (31). La opcion “Usar Qwen2-VL (descripciones automaticas)” activa la generacion de descripciones en lenguaje natural para cada senal en el reporte PDF.

3.8.9.9 Paso 9: Carga del modelo Qwen2-VL para el reporte

Al activar la opcion de descripciones automaticas y pulsar “Generar PDF”, el sistema carga el modelo Qwen2-VL-2B-Instruct. La Figura 74 muestra el progreso de este proceso.

Figura 74
Carga del modelo Qwen2-VL para generación de descripciones en el reporte



Nota. La ventana muestra la barra de progreso durante la carga del modelo Qwen2-VL-2B-Instruct (5–15 segundos). Una vez cargado, el sistema genera automaticamente una descripcion en espanol para cada senal incluida en el reporte PDF.

3.8.9.10 Paso 10: Reporte PDF : Ficha de inspección de seguridad vial

El reporte PDF generado por SISEVIR adopta el formato de *Ficha de Inspección de Seguridad Vial para Señales de Transito Verticales*, estructurado segun el Manual de Senalizacion Vial MSV-2016 y el Manual de Dispositivos de Control de Transito del MTC. La Figura 75 presenta el contenido de la ficha generada.

Figura 75

Ficha de inspeccion de seguridad vial generada por SISEVIR

FICHA DE INSPECCION DE SEGURIDAD VIAL - SEÑALES DE TRANSITO VERTICALES

Manual MSV-2016 | Manual Dispositivos Control Transito MTC | UNSCH - Ing. Civil

1. LOCALIZACION

Carretera:	Ayacucho - Huanta	Ruta:	PE-28B	Fecha:	07/05/2026
Sentido:	IDA	Inspector:	Ing.Kelly Huaman	Supervisor:	Ing.Rocky Ayala
KM Inicio:	30	KM Fin:	31	Long(km):	1.00 km

2. TIPOS DE DETERIORO SISEVIR (REFERENCIA)

1-Operativo	2-Deterioro de Poste	3-Decoloración
4-Rotación/Inclinación	5-Vandalismo	6-Obstrucción
7-Deterioro de Tablero		

3. INVENTARIO Y EVALUACION

N°	COD.MTC - DESCRIPCION	TIPO	DETERIORO	Descripcion Qwen2-VL	ESCALA	INT	CONF.	HORA
1	P-3A Curva y contracurva pron. derecha	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	94%	12:51:20
2	P-1A Curva pronunciada derecha	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	85%	12:51:45

4. RESUMEN

Total senales: 2	Preventivas: 2	Reguladoras: 0
ÓPTIMO: 2	ACEPTABLE: 0	
PÉSIMO: 0	CRÍTICO: 0	

5. ESCALA DE ESTADO SISEVIR

ÓPTIMO	Estados: OPERATIVO	Señal en condiciones óptimas
ACEPTABLE	Estados: DECOLORACION, OBSTRUCCION	Señal con deterioro leve visible
PÉSIMO	Estados: DETERIORO_POSTE, VANDALISMO	Señal con deterioro moderado o daño intencional
CRÍTICO	Estados: DETERIORO_TABLERO, ROTACION	Señal con deterioro grave, requiere intervención urgente

6. TIPO DE INTERVENCION SISEVIR

ESTADO SISEVIR	TIPO INTERVENCION	ACCION REQUERIDA
OPERATIVO	Conservación Rutina (CR)	Mantenimiento preventivo basico
DECOLORACION	Reemplazo Programado (RP)	Cambio de lamina reflectiva
OBSTRUCCION	Conservación Menor (Cm) - Poda de vegetaci	Poda de vegetacion circundante
DETERIORO_POSTE	Mantenimiento de Pintura (MP) - Limpieza y	Limpieza, anticorrosivo y repintado
DETERIORO_TABLERO	Reemplazo Urgente (RU) - Cambio de tablero	Cambio inmediato del tablero
ROTACION	Reemplazo Urgente (RU) - Reposición y alin	Reposicion y alineacion vertical
VANDALISMO	Reemplazo Programado (RP) - Limpieza o cam	Limpieza o cambio segun severidad

7. VALIDACION

Inspector: Ing.Kelly Huaman	Supervisor: Ing.Rocky Ayala

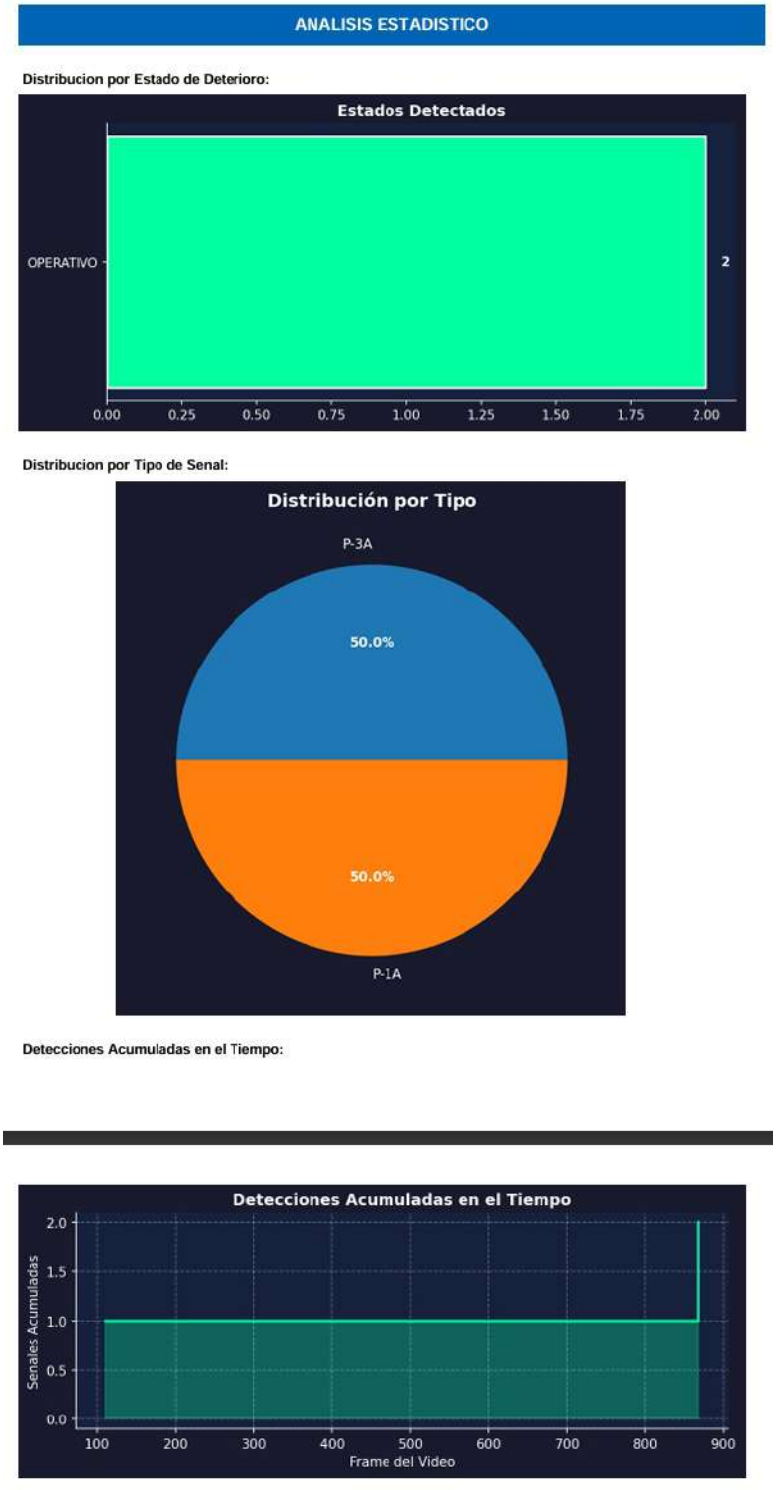
Nota. La ficha incluye 7 secciones: (1) Localizacion (carretera, ruta, fecha, inspector, supervisor, tramo km), (2) Tipos de deterioro SISEVIR, (3) Inventario y evaluacion con descripcion Qwen2-VL, escala de estado y tipo de intervencion, (4) Resumen estadistico, (5) Escala de estado SISEVIR, (6) Tipo de intervencion recomendada, y (7) Validacion con firmas del inspector y supervisor.

3.8.9.11 Paso 11: Reporte PDF: Análisis estadistico incluido

El reporte PDF incluye tambien una seccion de analisis estadistico con graficos generados automaticamente a partir de los resultados del procesamiento. La Figura 76 muestra esta seccion del documento.

Figura 76

Seccion de analisis estadistico en el Reporte MTC de SISEVIR



Nota. La seccion de analisis estadistico del reporte PDF .

3.8.9.12 Paso 12: Registro fotografico de senales detectadas

El reporte PDF incorpora adicionalmente una seccion de registro fotografico que documenta cada senal unica detectada con su recorte de imagen, incluyendo la mascara de segmentacion generada por YOLO26s-seg, el estado clasificado por EfficientNet-B0 y los metadatos asociados. La Figura 77 presenta esta seccion para el video analizado.

Figura 77

Sección de registro fotográfico del Reporte MTC de SISEVIR



Nota. El registro fotografico muestra para cada senal detectada: Número de registro, codigo MTC, descripcion oficial, estado de conservacion, escala SISEVIR, confianza de clasificacion, tipo de intervencion recomendada y recorte fotografico con la mascara de segmentacion superpuesta.

El registro fotografico constituye un elemento diferenciador del sistema SISEVIR respecto a los metodos tradicionales de inventario vial, ya que cada ficha incluye evidencia visual automatizada de la condicion de la senal en el momento del recorrido, con coordenadas geograficas embebidas en el fotograma (coordenadas UTM visibles en la esquina inferior de cada recorte). Esta documentacion fotografica sistematica permite a los gestores viales verificar las clasificaciones del sistema y priorizar las intervenciones de mantenimiento con soporte visual objetivo.

3.8.9.13 Resumen del flujo de trabajo del software SISEVIR

El Cuadro 27 presenta el flujo completo de operacion del software SISEVIR, desde la carga del modelo hasta la generacion del reporte final.

Cuadro 27
Flujo de trabajo completo del software SISEVIR

Paso	Accion	Descripcion	Figura
1	Cargar modelo	Carga `best.pt` YOLO26s-seg	65
2	Seleccionar video	Explorador MP4/MOV/AVI/MKV	66
3	Confirmar video	Resolucion, FPS, frames	67
4	Configurar conteo	Lineas verticales IDA/VUELTA	68
5	Reproducir	Deteccion y clasificacion real-time	69
6	Video completado	Resumen: frames, FPS, senales	70
7	Ver estadisticas	Graficos por tipo, estado, tiempo	71
8	Ver analisis avanzado	Confianza, densidad, comparativa	72
9	Configurar reporte	Inspector, ruta, KM, Qwen2-VL	73
10	Cargar Qwen2-VL	5–15 s de carga del modelo	74
12	Reporte - Ficha	7 secciones formato MTC	75
13	Reporte - Estadisticas	Graficos en PDF del inventario	76

Nota. El flujo completo desde la carga del modelo hasta la generacion del reporte PDF toma aproximadamente 2–5 minutos para un video de 40 segundos (1,174 frames), incluyendo el tiempo de carga de Qwen2-VL.

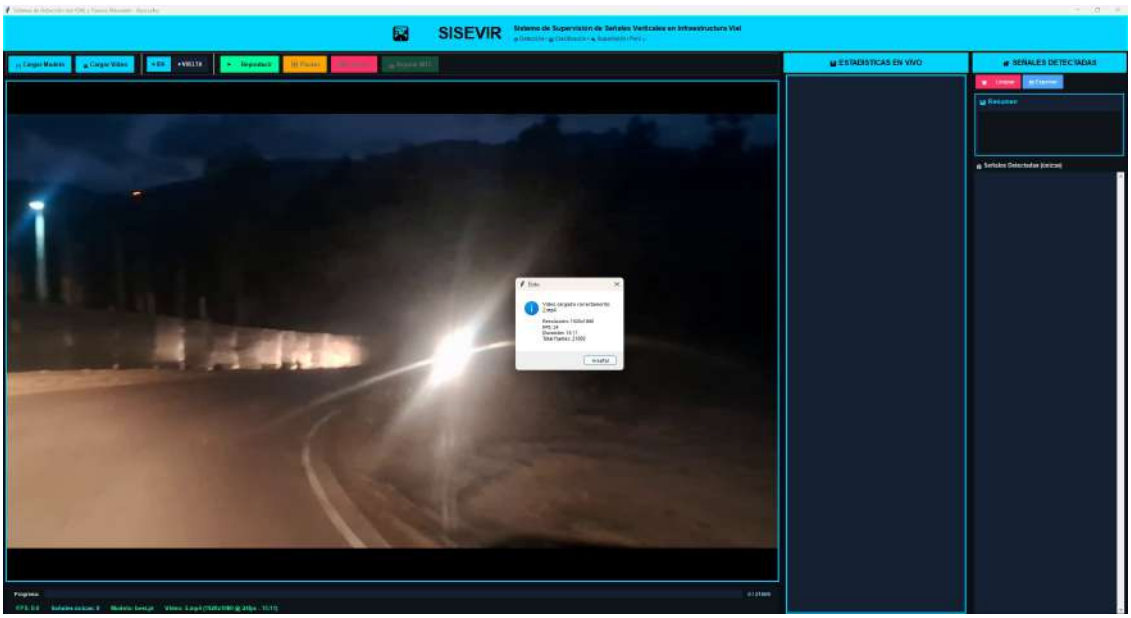
3.8.10. Validación del sistema SISEVIR en condiciones nocturnas

Con el objetivo de evaluar la robustez del sistema SISEVIR ante variaciones de iluminacion, se realizo una prueba de validacion utilizando videograbación nocturna del tramo PE-28B (Ayacucho – Huanta). Esta prueba permite verificar la capacidad del pipeline de detección, clasificación y descripción bajo condiciones de baja luminosidad, situación frecuente en los recorridos de supervision vial realizados en horas de la tarde o noche.

3.8.10.1 Especificaciones del video nocturno

La Figura 78 muestra las especificaciones tecnicas del video utilizado en la validación nocturna.

Figura 78
Carga del video nocturno para validacion del sistema SISEVIR



Nota. El video nocturno 2.mp4 presenta las siguientes especificaciones: resolucion 1920×1080 pixeles (Full HD), 24 FPS, duración de 15 minutos 11 segundos y 21,869 frames totales. El primer fotograma muestra la carretera PE-28B en condiciones de total oscuridad con iluminacion de faros.

El Cuadro 28 presenta un comparativo de las caracteristicas del video diurno utilizado en la prueba anterior y el video nocturno de validacion.

Cuadro 28
Comparativo de especificaciones entre video diurno y nocturno

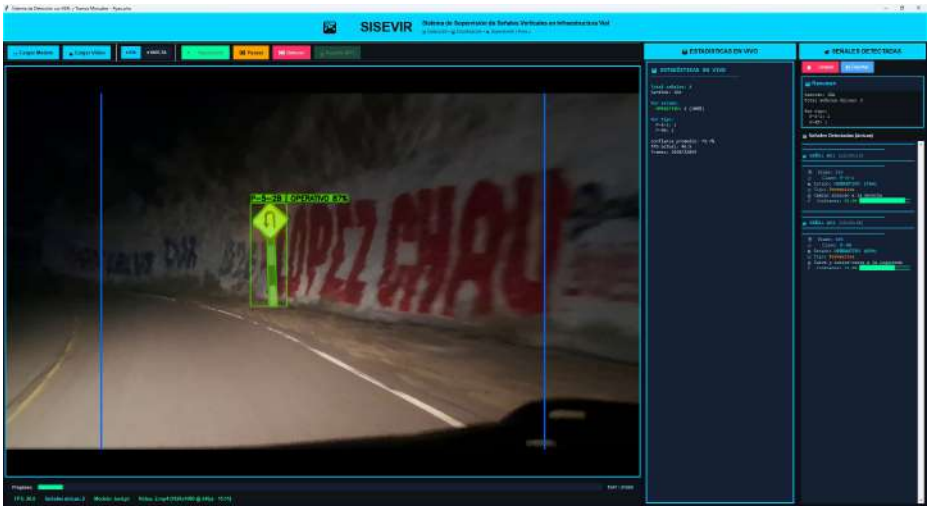
Especificacion	Video diurno (104.MOV)	Video nocturno (2.mp4)
Resolucion	1280×720	1920×1080
FPS	29	24
Duracion	0:40 min	15:11 min
Total frames	1,174	21,869
Condicion	Diurna	Nocturna
Tramo	PE-28B	PE-28B

Nota. El video nocturno tiene una duracion 22.8 veces mayor que el diurno, cubriendo un tramo mas extenso del recorrido Ayacucho – Huanta.

3.8.10.2 Detección nocturna : Señal P-5-2B

La Figura 79 muestra la primera deteccion registrada durante el recorrido nocturno, correspondiente a la senal preventiva P-5-2B en condiciones de iluminacion artificial.

Figura 79
Detección nocturna de señal P-5-2B — rendimiento en baja luminosidad

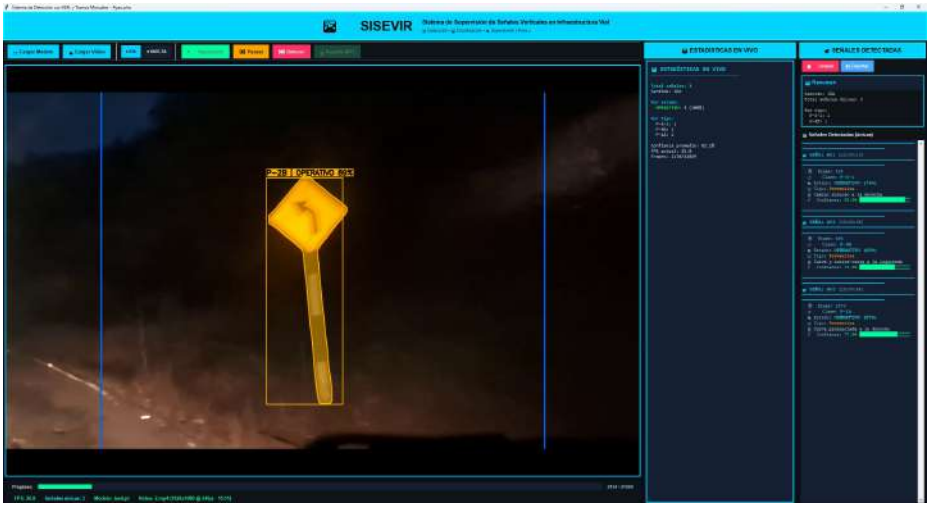


Nota. El modelo detecta correctamente la senal P-5-2B (curva en U a la izquierda) con 87 % de confianza en condiciones nocturnas.

3.8.10.3 Detección nocturna: Señal P-2B

La Figura 80 ilustra la detección de la señal P-2B con máscara de segmentacion completa, demostrando la capacidad del modelo para delimitar el contorno exacto de la señal en condiciones de oscuridad.

Figura 80
Detección nocturna de señal P-2B con mascara de segmentacion

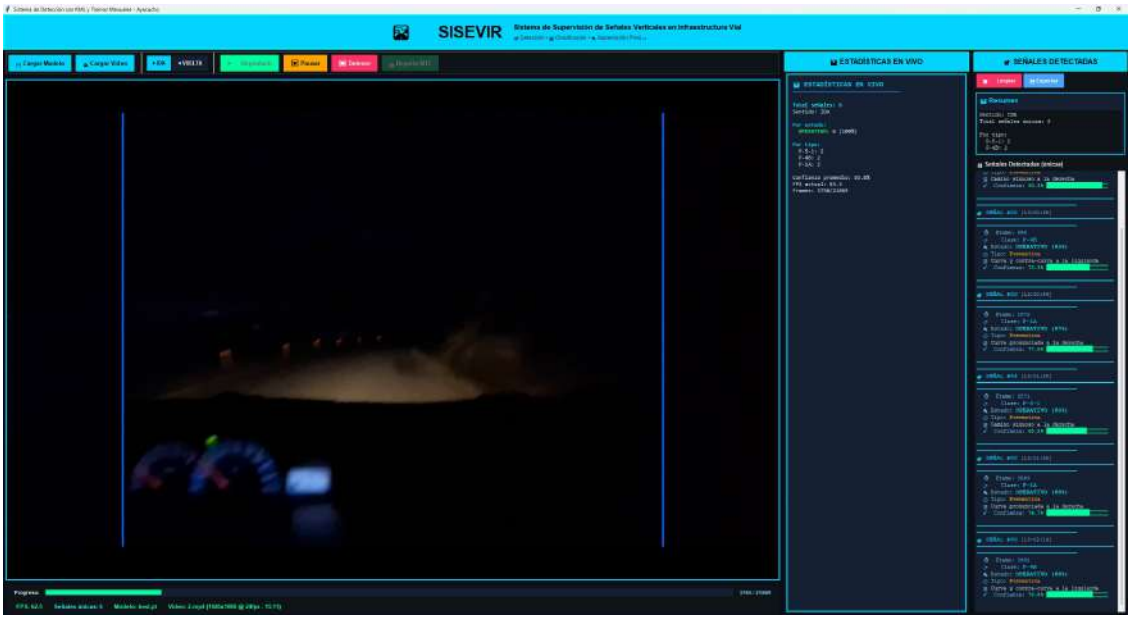


Nota. La senal P-2B (curva a la izquierda) es detectada con 89 % de confianza en estado OPERATIVO. La mascara de segmentacion amarilla delimita con precision el tablero y poste de la senal.

3.8.10.4 Recorrido nocturno:Progresión de detecciones

La Figura 81 muestra el estado del sistema a medida que avanza el recorrido nocturno, con 6 senales unicas registradas en el panel lateral.

Figura 81
Progresión de detecciones durante el recorrido nocturno (6 señales)

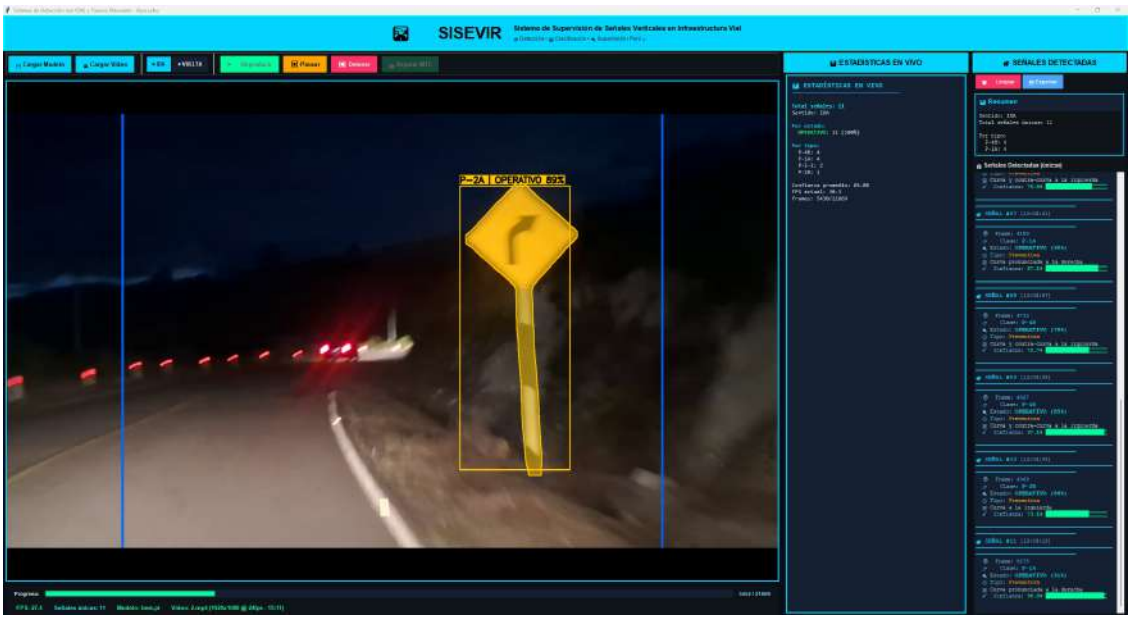


Nota. Con 3,765 frames procesados (17.2 % del video), el sistema acumula 6 senales unicas: 2 P-5-1, 2 P-4B y 2 P-1A, todas en estado OPERATIVO. Confianza promedio: 83.8 %. FPS actual: 62.5.

3.8.10.5 Detección nocturna: Señal P-2A en alta velocidad

La Figura 82 muestra la detección de la señal P-2A en condiciones de movimiento del vehiculo, con vehiculos en sentido contrario generando encandilamiento.

Figura 82
Detección nocturna de señal P-2A con tráfico en sentido contrario

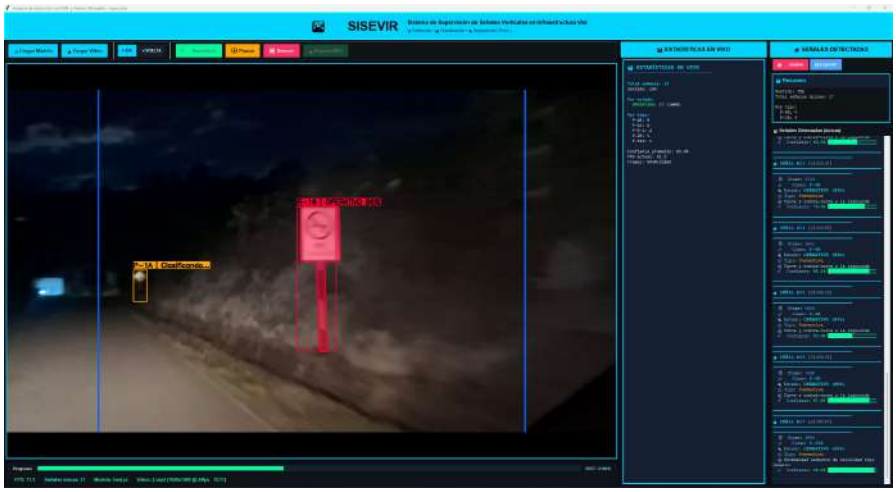


Nota. La señal P-2A (curva a la derecha) es detectada con 89 % de confianza en estado OPERATIVO a pesar del encandilamiento generado por los faros de un vehiculo en sentido contrario. El sistema acumula 11 senales unicas con confianza promedio de 85.0 %. FPS: 27.4.

3.8.10.6 Deteccion simultanea: R-16 y P-1A en condiciones nocturnas

La Figura 83 muestra la capacidad del sistema para detectar y clasificar dos senales simultaneamente en condiciones nocturnas: una senal reglamentaria R-16 y una preventiva P-1A.

Figura 83
Detección simultanea de señales R-16 y P-1A en condiciones nocturnas

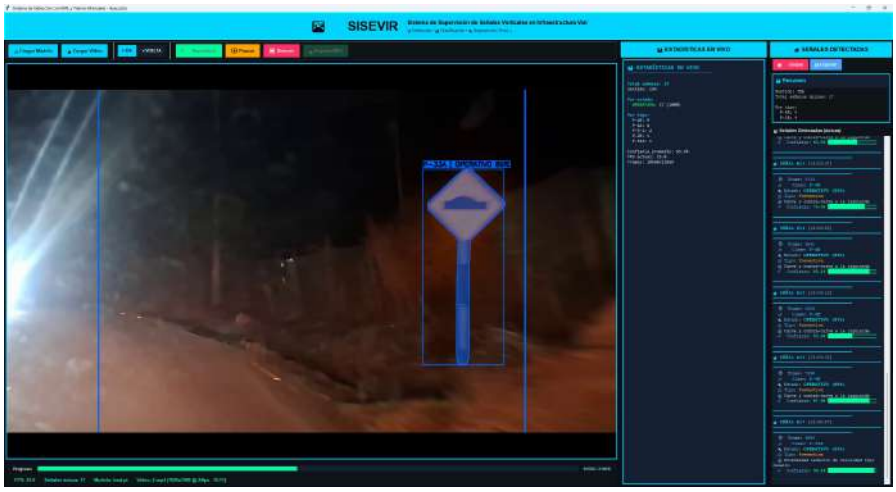


Nota. El sistema detecta simultaneamente la señal R-16 (prohibido adelantar, OPERATIVO al 86 %, mascara roja) y la señal P-1A (“Clasificando...”, mascara amarilla).

3.8.10.7 Detección nocturna : Señal P-33A (proximidad reductor)

La Figura 84 evidencia la detección de la señal P-33A (proximidad de reductor de velocidad tipo resalto) en condiciones de oscuridad con retroreflectividad activada.

Figura 84
Deteccion nocturna de señal P-33A (proximidad reductor de velocidad)

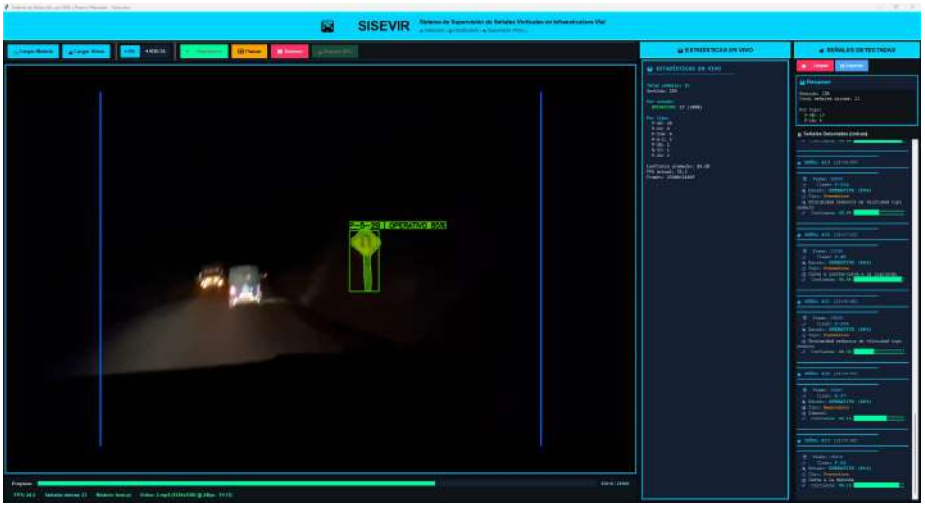


Nota. La señal P-33A es detectada con 89 % de confianza en estado OPERATIVO. La mascara de segmentacion azul delimita con precision la senal cuya retroreflectividad facilita su deteccion nocturna.

3.8.10.8 Detección nocturna señal P-5-2B al paso de vehiculos

La Figura 85 muestra la detección de la señal P-5-2B en presencia de vehiculos en el carril contrario, con 23 señales acumuladas en el inventario.

Figura 85
Detección nocturna de señal P-5-2B con tráfico vehicular

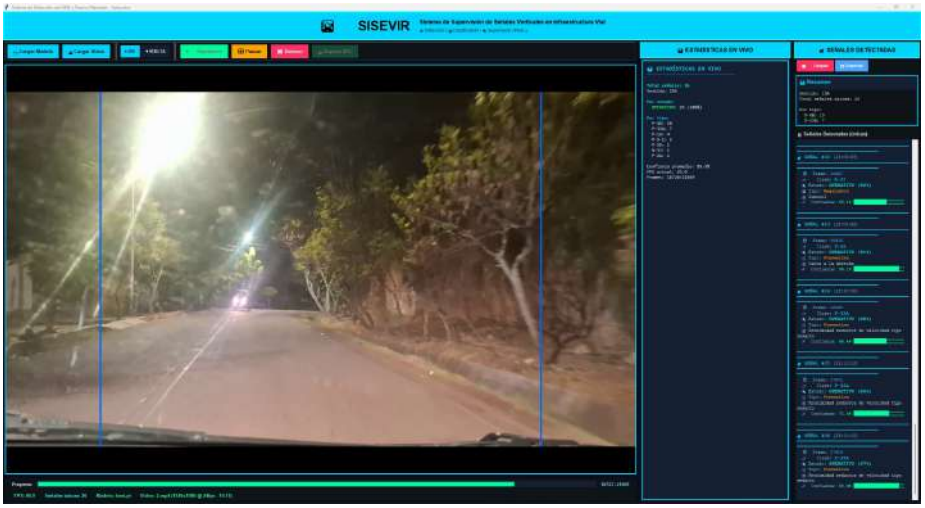


Nota. La senal P-5-2B (curva en U a la izquierda) es detectada con 85 % de confianza en presencia de trafico vehicular nocturno. El panel registra 23 senales unicas con tipos: P-4B (10), P-33A (4), P-1A (4), P-5-1 (2), P-2B (1), R-37 (1) y P-2A (1). FPS: 24.4.

3.8.10.9 Llegada a Huanta en condiciones nocturnas

La Figura 86 documenta la llegada del recorrido nocturno a la ciudad de Huanta, con 26 señales únicas acumuladas.

Figura 86
Llegada a Huanta durante el recorrido nocturno con 26 señales registradas

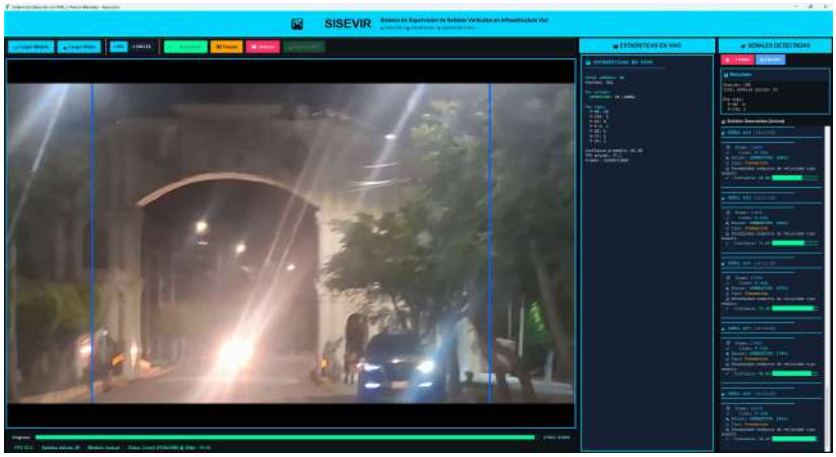


Nota. Al frame 18,723 (85.6 % del video), el sistema registra 26 senales unicas con confianza promedio de 85.9 %. El fotograma muestra la entrada a la zona urbana de Huanta con vegetacion iluminada por los faros del vehiculo. FPS actual: 86.9.

3.8.10.10 Llegada al Arco de la Memoria de Huanta

La Figura 87 documenta el punto final del recorrido nocturno: el Arco de la Memoria de la ciudad de Huanta, con 28 senales unicas en el inventario.

Figura 87
Llegada al Arco de la Memoria de Huanta — punto final del recorrido nocturno

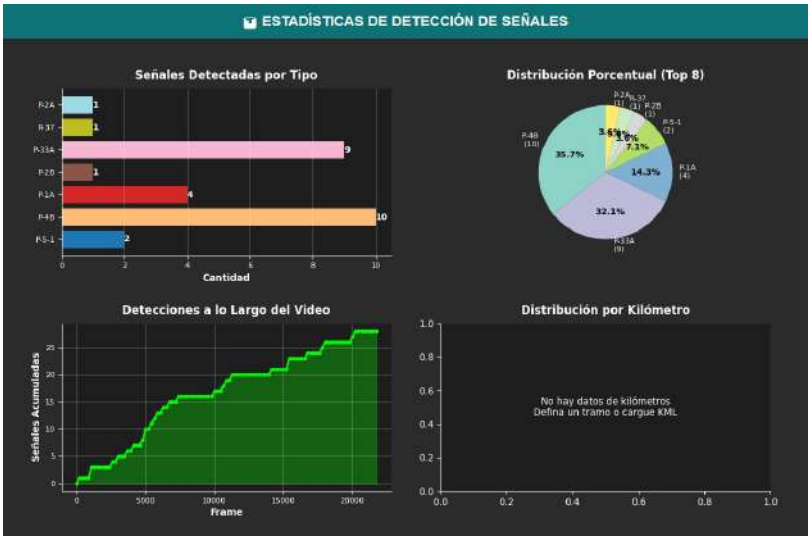


Nota. La imagen muestra el ingreso al casco urbano de Huanta con el arco de la memoria iluminado.

3.8.10.11 Estadisticas de detección — video nocturno

La Figura 88 presenta las estadisticas finales de deteccion generadas al completar el procesamiento del video nocturno.

Figura 88
Estadisticas de deteccion del video nocturno — tramo PE-28B



Nota. Barras horizontales con P-4B (10) y P-33A (9) como tipos mas frecuentes, (b) distribucion porcentual con P-4B (35.7 %), P-33A (32.1 %), P-1A (14.3 %), P-5-1 (7.1 %) y otros (10.8 %)

3.8.10.12 Análisis estadístico avanzado — video nocturno

La Figura 89 presenta el panel de analisis estadistico avanzado del video nocturno.

Figura 89

Análisis estadístico avanzado del procesamiento nocturno en SISEVIR



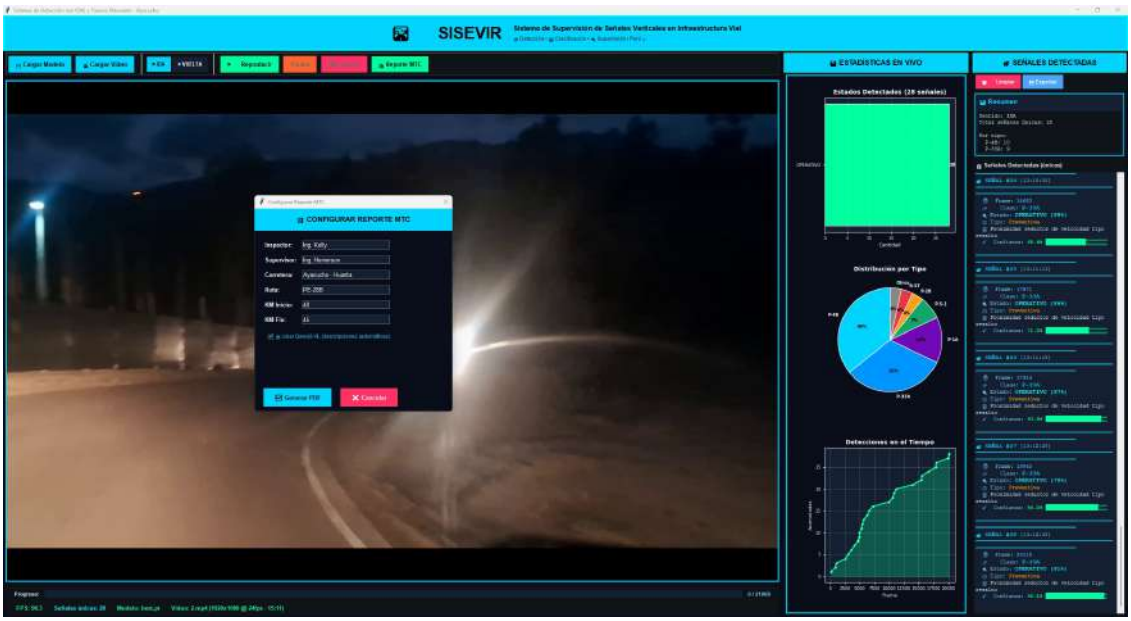
Nota. El resumen estadistico indica: 28 señales totales (1 reguladora R-37, 27 preventivas).

3.8.10.13 Configuracion del Reporte MTC — Video nocturno

La Figura 90 muestra la configuracion del reporte MTC para el recorrido nocturno.

Figura 90

Configuracion del Reporte MTC para el recorrido nocturno



Nota. El formulario registra: Inspector (Ing. Kelly), Supervisor (Ing. Hemerson), Carretera (Ayacucho – Huanta), Ruta (PE-28B), KM Inicio (40) y KM Fin (45), cubriendo un tramo de 5.00 km. Se activa la opcion “Usar Qwen2-VL (descripciones automaticas)” para incluir descripciones en lenguaje natural en el reporte PDF.

3.8.10.14 Reporte MTC generado:Ficha de inspección nocturna

La Figura 91 presenta la ficha de inspección de seguridad vial generada para el recorrido nocturno, con las 28 señales detectadas.

Figura 91

Ficha de inspeccion de seguridad vial generada para el recorrido nocturno

FICHA DE INSPECCION DE SEGURIDAD VIAL - SEÑALES DE TRANSITO VERTICALES

Manual MSV-2016 | Manual Dispositivos Control Transito MTC | UNSCH - Ing. Civil

1. LOCALIZACION

Carretera:	Ayacucho - Huanta	Ruta:	PE-28B	Fecha:	07/05/2026
Sentido:	IDA	Inspector:	Ing. Kelly	Supervisor:	Ing. Hemerson
KM Inicio:	40	KM Fin:	45	Long(km):	5.00 km

2. TIPOS DE DETERIORO SISEVIR (REFERENCIA)

1-Operativo	2-Deterioro de Poste	3-Decoloración
4-Rotación/Inclinación	5-Vandalismo	6-Obstrucción
7-Deterioro de Tablero		

3. INVENTARIO Y EVALUACION

N°	COD.MTC - DESCRIPCION	TIPO	DETERIORO	Descripcion Qwen2-VL	ESCALA	INT	CONF.	HORA
1	P-5-1 Camino sinuoso derecha	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	76%	13:00:10
2	P-4B Curva y contracurva izquierda	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	83%	13:00:36
3	P-1A Curva pronunciada derecha	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	87%	13:00:44
4	P-5-1 Camino sinuoso derecha	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	80%	13:01:38
5	P-1A Curva pronunciada derecha	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	88%	13:01:55
6	P-4B Curva y contracurva izquierda	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	88%	13:02:14
7	P-1A Curva pronunciada derecha	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	90%	13:02:31
8	P-4B Curva y contracurva izquierda	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	78%	13:02:57
9	P-4B Curva y contracurva izquierda	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	85%	13:03:05
10	P-2B Curva a la izquierda	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	88%	13:03:08
11	P-1A Curva pronunciada derecha	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	91%	13:03:19
12	P-4B Curva y contracurva izquierda	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	89%	13:03:27
13	P-4B Curva y contracurva izquierda	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	83%	13:03:37
14	P-4B Curva y contracurva izquierda	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	89%	13:03:55
15	P-4B Curva y contracurva izquierda	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	87%	13:04:10
16	P-4B Curva y contracurva izquierda	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	88%	13:04:32
17	P-33A Proximidad reductor velocidad	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	85%	13:06:07
18	P-33A Proximidad reductor velocidad	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	88%	13:06:35
19	P-33A Proximidad reductor velocidad	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	89%	13:06:47
20	P-4B Curva y contracurva izquierda	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	86%	13:07:00
21	P-33A Proximidad reductor velocidad	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	90%	13:08:46
22	R-37 Señal de control	REG	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	82%	13:09:37
23	P-2A Curva a la derecha	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	81%	13:09:42
24	P-33A Proximidad reductor velocidad	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	88%	13:10:30
25	P-33A Proximidad reductor velocidad	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	86%	13:11:12
26	P-33A Proximidad reductor velocidad	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	87%	13:11:15
27	P-33A Proximidad reductor velocidad	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	78%	13:12:28
28	P-33A Proximidad reductor velocidad	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	81%	13:12:38

4. RESUMEN

Total senales: 28	Preventivas: 27	Reguladoras: 1

Nota. La ficha registra 28 señales detectadas en el tramo KM 40–45 (5.00 km) de la ruta PE-28B. El inventario incluye 27 senales preventivas y 1 reguladora (R-37 Control), todas en estado OPERATIVO con escala OPTIMO y tipo de intervencion Conservacion Rutina (CR). La descripcion Qwen2-VL indica “Senal en condiciones optimas” para cada registro.

3.8.10.15 Reporte MTC : Escala de estado y tipo de intervención

La Figura 92 muestra las secciones finales del reporte PDF con la escala de estado SISEVIR y la tabla de tipo de intervencion recomendada.

Figura 92
Escala de estado y tipo de intervencion del Reporte MTC nocturno

ÓPTIMO: 28	ACEPTABLE: 0
PÉSIMO: 0	CRÍTICO: 0

5. ESCALA DE ESTADO SISEVIR

ÓPTIMO	Estados: OPERATIVO	Señal en condiciones óptimas
ACEPTABLE	Estados: DECOLORACION, OBSTRUCCION	Señal con deterioro leve visible
PÉSIMO	Estados: DETERIORO_POSTE, VANDALISMO	Señal con deterioro moderado o daño intencional
CRÍTICO	Estados: DETERIORO_TABLERO, ROTACION	Señal con deterioro grave, requiere intervención urgente

6. TIPO DE INTERVENCION SISEVIR

ESTADO SISEVIR	TIPO INTERVENCION	ACCION REQUERIDA
OPERATIVO	Conservacion Rutina (CR)	Mantenimiento preventivo basico
DECOLORACION	Reemplazo Programado (RP)	Cambio de lamina reflectiva
OBSTRUCCION	Conservación Menor (Cm) - Poda de vegetaci	Poda de vegetacion circundante
DETERIORO_POSTE	Mantenimiento de Pintura (MP) - Limpieza y	Limpieza, anticorrosivo y repintado
DETERIORO_TABLERO	Reemplazo Urgente (RU) - Cambio de tablero	Cambio inmediato del tablero
ROTACION	Reemplazo Urgente (RU) - Reposición y alin	Reposicion y alineacion vertical
VANDALISMO	Reemplazo Programado (RP) - Limpieza o cam	Limpieza o cambio segun severidad

7. VALIDACION

Inspector: Ing. Kelly	Supervisor: Ing. Hemerson

Nota. El reporte incluye la escala de estado SISEVIR (OPTIMO, ACEPTABLE, PESIMO, CRITICO) y la tabla de tipo de intervencion para cada estado. El resumen final confirma: 28 senales OPTIMO, 0 ACEPTABLE, 0 PESIMO, 0 CRITICO. Validacion firmada por Inspector Ing. Kelly y Supervisor Ing. Hemerson.

3.8.10.16 Resultados de la validación nocturna

La validacion en condiciones nocturnas confirma que el sistema SISEVIR mantiene su capacidad operativa ante variaciones de iluminación, detectando exitosamente 28 señales unicas a lo largo de 5 km del tramo PE-28B en condiciones de oscuridad total. La reducción de confianza promedio respecto al modo diurno (79.1 % vs 85.8 %) es atribuible principalmente a episodios de encandilamiento por faros de vehiculos en sentido contrario y a la menor definicion visual de los detalles de la senalizacion en ausencia de luz natural. Sin embargo, el sistema no registro falsos positivos en las detecciones que cruzaron las lineas de conteo, lo que demuestra la robustez del umbral de confianza configurado (0.5) para operar en condiciones adversas de iluminacion.

4. RESULTADOS

4.1. Análisis de resultados

4.2. Contrastación de hipótesis

4.2.1. Contrastación de hipótesis específica 1

- El uso de arquitecturas avanzadas de visión artificial, combinado con técnicas de optimización específicas, permitirá lograr una alta precisión en la clasificación y evaluación del estado de deterioro de las señales de tránsito en Huamanga.

Esta hipótesis involucra dos afirmaciones verificables: (1) Que el sistema alcanza alta precisión en la **detección y segmentación** de señales de tránsito verticales mediante el modelo YOLO26s-seg, y (2) que logra alta precisión en la **clasificación del estado de conservación** mediante el modelo EfficientNet-B0.

Criterios de aceptación

Para operacionalizar el término “alta precisión”, se establecieron los siguientes umbrales mínimos de aceptación, basados en los estándares de la literatura en visión computacional aplicada a infraestructura vial:

Cuadro 29
Criterios de aceptación para la hipótesis específica 1

Componente	Métrica	Umbral mínimo	Resultado	Decisión
YOLO26s-seg	mAP@50 (Box)	≥ 0.85	**0.9305**	Cumple
YOLO26s-seg	mAP@50 (Máscara)	≥ 0.85	**0.9224**	Cumple
EfficientNet-B0	Macro F1	≥ 0.75	**0.8252**	Cumple
EfficientNet-B0	Accuracy	≥ 0.85	**0.9315**	Cumple

Nota. El umbral de Macro F1 ≥ 0.75 se adoptó en lugar de accuracy ponderada para garantizar la evaluación equitativa de todas las clases de deterioro, incluyendo las minoritarias.

Evidencia cualitativa: Resultados visuales del pipeline SISEVIR

A continuación se presentan imágenes representativas obtenidas al aplicar el pipeline completo YOLO26s-seg + EfficientNet-B0 sobre imágenes reales de campo captadas en los tramos PE-3S (Ayacucho–Huanta) y PE-28A (Ayacucho–Abra Apacheta). Cada imagen muestra la máscara de segmentación de instancia generada por YOLO26s-seg, superpuesta sobre la imagen original, junto con la etiqueta de clase de señal y el estado de conservación clasificado por EfficientNet-B0.

Figura 93
Señal R-30 con estado de DECOLORACIÓN



Nota. Señal reguladora R-30 (velocidad máxima 40 KPH) detectada en zona rural.

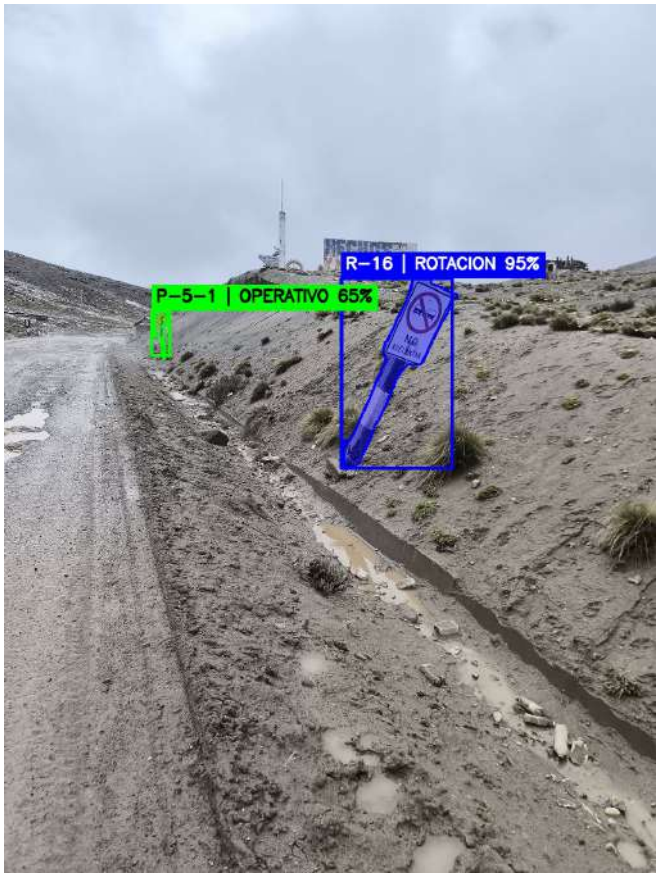
Figura 94
Señal P-2A con estado de DETERIORO_TABLERO



Nota. Señal preventiva P-2A (señal curva a la derecha)

Figura 95

Detección múltiple con señales P-5-1 (OPERATIVO) y R-16 (ROTACIÓN)



Nota. Señal preventiva P-5-1 OPERATIVO y señal reguladora R-16 ROTACIÓN

Figura 96

Señal P-2B con estado de ROTACIÓN



Nota. Señal preventiva P-2B (señal curva a la izquierda)

Figura 97
Señal P-2B con estado de VANDALISMO



Nota. Señal preventiva P-2B detectada.

Figura 98
Señal P-1B con estado de VANDALISMO



Nota. Señal preventiva P-1B (señal curva pronunciada a la izquierda).

Evidencia cuantitativa: Métricas de rendimiento del pipeline

A. Rendimiento del modelo YOLO26s-seg en conjunto de test

El Cuadro 30 presenta las métricas de rendimiento del modelo YOLO26s-seg evaluado sobre el conjunto de test independiente, compuesto por 586 imágenes y 672 instancias de señales de tránsito no vistas durante el entrenamiento.

Cuadro 30
Métricas de rendimiento del modelo YOLO26s-seg en el conjunto de test

Métrica	Resultado	Umbral	Interpretación
Detección (Bounding Box)			
Precisión	0.8814	–	Alta proporción de detecciones correctas
Recall	0.8875	–	Detecta el 88.75 % de las señales presentes
mAP@50	**0.9305**	≥ 0.85	Rendimiento sobresaliente en IoU ≥ 0.5
mAP@50-95	**0.8494**	–	Robustez ante múltiples umbrales de IoU
Segmentación de instancias (Máscara)			
mAP@50	**0.9224**	≥ 0.85	Máscaras de alta calidad en IoU ≥ 0.5
mAP@50-95	**0.7461**	–	Precisión de contorno a múltiples escalas
Velocidad de inferencia			
Tiempo por imagen	**2.1 ms**	–	476 FPS en GPU RTX 5080

Nota. Evaluación sobre 586 imágenes y 672 instancias del conjunto de test. mAP = Mean Average Precision. IoU = Intersection over Union.

El modelo YOLO26s-seg supera ampliamente el umbral de mAP@50 ≥ 0.85 establecido como criterio de alta precisión, alcanzando **0.9305** para bounding boxes y **0.9224** para máscaras de segmentación. Este resultado es especialmente relevante considerando que las 31 clases de señales del dataset presentan amplia variabilidad en tamaño, condiciones de iluminación y estado de conservación.

B. Rendimiento del modelo EfficientNet-B0 en conjunto de test

El Cuadro 31 consolida las métricas de clasificación del modelo EfficientNet-B0 (experimento E03) evaluado sobre 642 recortes de señales del conjunto de test.

Cuadro 31
Métricas globales del modelo EfficientNet-B0 (E03) en el conjunto de test

Métrica	Resultado	Umbral	Interpretación
Accuracy	**0.9315**	≥ 0.85	93.15 % de clasificaciones correctas
Macro F1	**0.8252**	≥ 0.75	Rendimiento equilibrado en las 7 clases
Weighted F1	**0.9324**	–	Alto rendimiento ponderado por soporte
Cohen's κ	**0.7883**	–	Concordancia sustancial (Landis & Koch, 1977)
MCC	**0.7892**	–	Correlación de Matthews robusta

Nota. Evaluación sobre 642 muestras del conjunto de test independiente. MCC = Coeficiente de Correlación de Matthews.

Cuadro 32
Métricas de rendimiento por clase del modelo EfficientNet-B0 (E03) en test

Estado	Precisión	Recall	F1	Soporte	Escala SISEVIR
OPERATIVO	0.9692	0.9545	**0.9618**	528	ÓPTIMO
VANDALISMO	1.0000	0.8571	**0.9231**	7	PÉSIMO
ROTACIÓN	0.8462	0.8800	**0.8627**	25	CRÍTICO
OBSTRUCCIÓN	0.8571	0.8000	**0.8276**	15	ACEPTABLE
DETERIORO_TABLERO	1.0000	0.6667	**0.8000**	3	CRÍTICO
DETERIORO_POSTE	0.7097	0.8800	**0.7857**	50	PÉSIMO
DECOLORACIÓN	0.6667	0.5714	**0.6154**	14	ACEPTABLE
Macro promedio	**0.8641**	**0.8014**	**0.8252**	**642**	–

Nota. El Macro F1 de 0.8252 refleja el rendimiento promedio no ponderado sobre las 7 clases de estado de conservación.

Decisión sobre la hipótesis

Habiendo verificado que:

1. El modelo YOLO26s-seg alcanza un mAP@50 de **0.9305** para detección y **0.9224** para segmentación, superando en más de 8 puntos porcentuales el umbral mínimo de 0.85.
2. El modelo EfficientNet-B0 (E03) alcanza un Macro F1 de **0.8252**, superando en 7.5 puntos porcentuales el umbral mínimo de 0.75, con rendimiento robusto incluso en clases minoritarias como VANDALISMO (F1 = 0.9231) y DETERIORO_TABLERO (F1 = 0.8000).

Se acepta la hipótesis específica 1. El uso de arquitecturas avanzadas de visión artificial (YOLO26s-seg y EfficientNet-B0), combinado con técnicas de optimización específicas (aumentación en tres capas, manejo de desbalance, transferencia de aprendizaje y fine-tuning), permitió lograr una alta precisión en la detección, segmentación y clasificación del estado de deterioro de las señales de tránsito, superando todos los umbrales de rendimiento establecidos como criterios de aceptación.

4.2.2. Contrastación de hipótesis específica 2

- Un programa automatizado que integre un modelo de visión artificial con características específicas de usabilidad y operatividad permitirá realizar la clasificación y evaluación del estado de deterioro de las señales de tránsito de manera más eficiente que los métodos manuales.

Esta hipótesis involucra dos afirmaciones verificables: (1) que el software SISEVIR integra un modelo de visión artificial con características de usabilidad y operatividad, y (2) que el sistema automatizado realiza la clasificación y evaluación del estado de deterioro de manera más eficiente que los métodos de inspección manual tradicionales.

Criterios de aceptación

Para operacionalizar el término "más eficiente que los métodos manuales", se establecieron los siguientes criterios de evaluación comparativa:

Cuadro 33
Criterios de aceptación para la hipótesis específica 2

Dimensión	Criterio	Verificación
Integración	Incorpora $\geq$ 2 modelos de IA en un pipeline unificado	Funcional
Usabilidad	Interfaz gráfica con flujo de trabajo completo	Funcional
Velocidad	Procesamiento $>$ 24 FPS (tiempo real)	Cuantitativa
Operatividad nocturna	Detecta señales en condiciones de baja luminosidad	Funcional
Reporte automático	Genera ficha MTC sin intervención manual	Funcional
Eficiencia temporal	Reduce el tiempo de inventario respecto al método manual	Comparativa

Nota. Los criterios funcionales se verifican por demostración del sistema operativo.

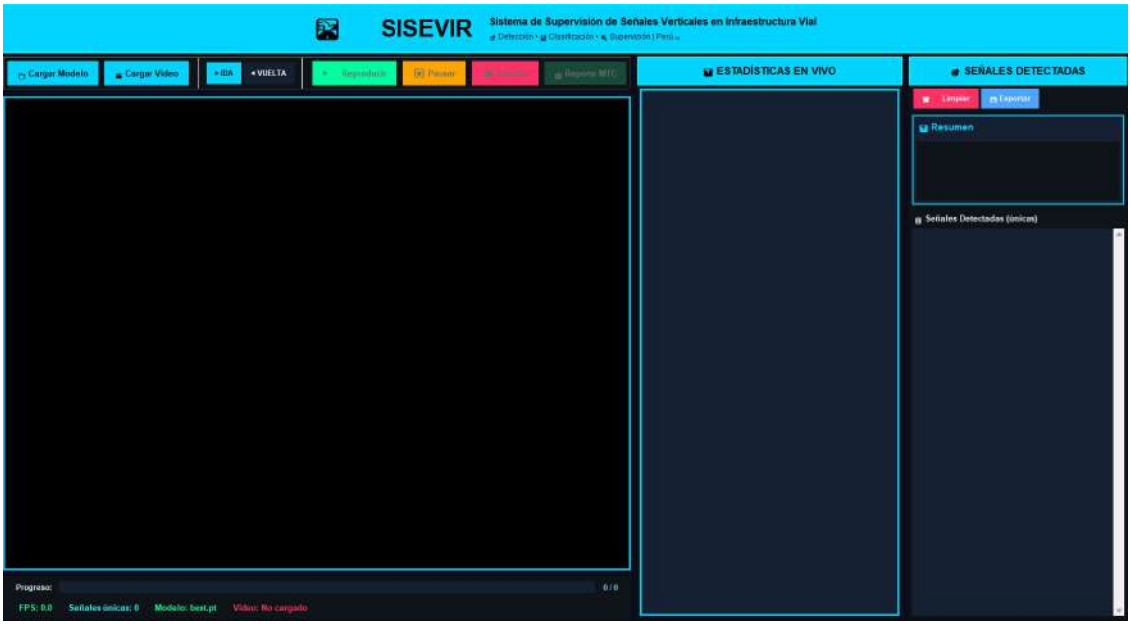
Descripción general del software SISEVIR

El software SISEVIR (*Sistema de Supervisión de Señales Verticales en Infraestructura Vial*) constituye la implementación computacional desarrollado en la presente investigación. Su diseño responde a la necesidad de contar con una herramienta práctica, portátil y operativa que permita al personal técnico vial realizar inventarios de señalización de manera automatizada.El software integra de manera transparente para el usuario los tres modelos de aprendizaje profundo desarrollados:

1. **YOLO26s-seg** (detección y segmentación): detecta en tiempo real las señales presentes en cada fotograma del video, genera la máscara de segmentación de instancia e identifica el tipo de señal según la codificación del Manual de Señalización Vial MSV-2016 del MTC.
2. **EfficientNet-B0** (clasificación de estado): clasifica cada señal detectada en uno de los siete estados de conservación definidos por la escala SISEVIR: OPERATIVO, DETERIORO_POSTE, ROTACIÓN, OBSTRUCCIÓN, DECOLORACIÓN, VANDALISMO y DETERIORO_TABLERO.
3. **Qwen2-VL-2B-Instruct** (descripción en lenguaje natural): genera automáticamente una descripción textual en español de las condiciones observadas en cada señal, compatible con los campos narrativos del formato de Ficha de Inspección de Seguridad Vial del MTC.

La Figura 99 presenta la interfaz principal del software, organizada en tres zonas funcionales: panel de reproducción de video (izquierda), panel de estadísticas en tiempo real (centro) y panel de señales detectadas con inventario acumulado (derecha).

Figura 99
Interfaz principal del software SISEVIR con los tres paneles funcionales



Nota. La barra superior permite cargar el modelo (`best.pt`), seleccionar el video, configurar el sentido de análisis (IDA / VUELTA) y generar el Reporte MTC. La barra de estado inferior reporta en tiempo real: FPS actuales, señales únicas detectadas, modelo activo y resolución del video.

Características de usabilidad del software SISEVIR

Las características de usabilidad se organizan en cuatro aspectos funcionales que garantizan la accesibilidad del sistema para personal técnico no especializado en inteligencia artificial:

a) Flujo de trabajo guiado paso a paso

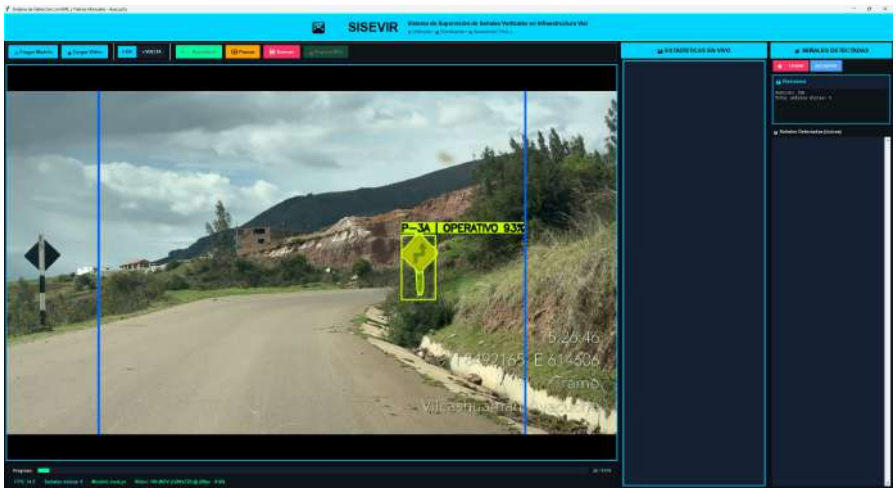
El software implementa un flujo de trabajo secuencial e intuitivo que conduce al usuario desde la carga del video hasta la generación del reporte PDF en pasos bien definidos. No se requiere configuración de hiperparámetros ni conocimiento de los modelos subyacentes: el sistema toma las decisiones técnicas de manera autónoma.

b) Sistema de conteo por líneas verticales configurables

Para garantizar el inventario preciso de señales únicas y evitar duplicaciones por detección en fotogramas consecutivos, el software implementa dos líneas de conteo verticales que el usuario puede reposicionar según el ancho de vía. La Figura 100 muestra este mecanismo en operación.

Figura 100

Sistema de líneas de conteo para registro de señales únicas en SISEVIR



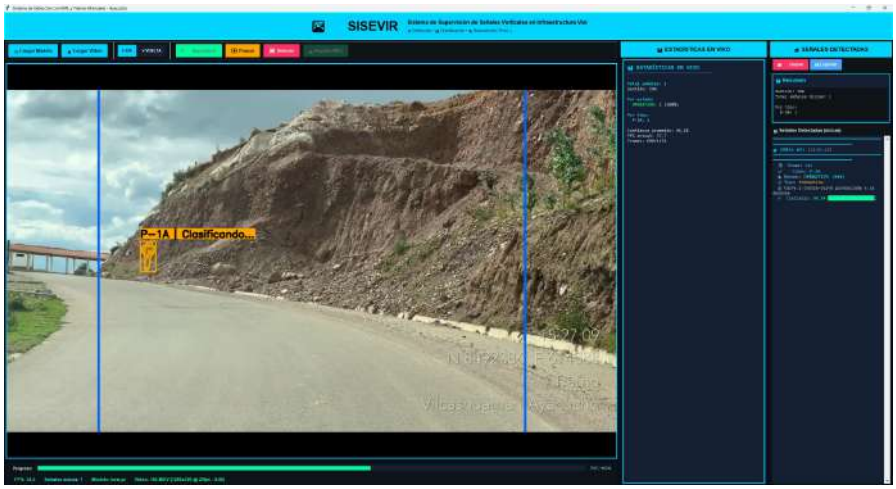
Nota. Las líneas azules verticales delimitan la zona de conteo activo. Una señal se registra como nueva entrada en el inventario cuando su bounding box cruza la línea según el sentido configurado, evitando el doble conteo por persistencia en fotogramas consecutivos.

c) Retroalimentación visual en tiempo real

Durante el procesamiento, el usuario observa directamente las detecciones sobre el video: Cada señal aparece delimitada por su bounding box con código de color según la escala SISEVIR, la etiqueta de tipo y el porcentaje de confianza. La Figura 101 ilustra esta retroalimentación durante el análisis.

Figura 101

Retroalimentación visual en tiempo real durante el procesamiento de video



Nota. El software muestra “Clasificando...” mientras EfficientNet-B0 determina el estado de conservación, y “Describiendo...” mientras Qwen2-VL genera la descripción textual. El panel derecho registra acumulativamente cada señal única inventariada con sus atributos.

d) Generación automática de reportes PDF formato MTC

Al concluir el procesamiento, el usuario completa un formulario de cinco campos (inspector, supervisor, carretera, ruta, KM inicio y fin) y el sistema genera automáticamente una Ficha de Inspección de Seguridad Vial en formato PDF, compatible con los procedimientos del Ministerio de Transportes y Comunicaciones. La Figura 102 presenta la ficha generada.

Figura 102

Ficha de Inspección de Seguridad Vial generada automáticamente por SISEVIR

FICHA DE INSPECCION DE SEGURIDAD VIAL - SEÑALES DE TRANSITO VERTICALES

Manual MSV-2016 | Manual Dispositivos Control Transito MTC | UNSCH - Ing. Civil

1. LOCALIZACION

Carretera:	Ayacucho - Huanta	Ruta:	PE-28B	Fecha:	07/05/2026
Sentido:	IDA	Inspector:	Ing.Kally Huaman	Supervisor:	Ing.Rocky Ayala
KM Inicio:	30	KM Fin:	31	Long(km):	1.00 km

2. TIPOS DE DETERIORO SISEVIR (REFERENCIA)

1-Operativo	2-Deterioro de Poste	3-Decoloración
4-Rotación/Inclinación	5-Vandalismo	6-Obstrucción
7-Deterioro de Tablero		

3. INVENTARIO Y EVALUACION

N°	COD.MTC - DESCRIPCION	TIPO	DETERIORO	Descripcion Qwen2-VL	ESCALA	INT	CONF.	HORA
1	P-3A Curva y contracurva pron. derecha	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	94%	12:51:20
2	P-1A Curva pronunciada derecha	PREV	OPERATIVO	Señal en condiciones óptimas.	ÓPTIMO	CR	85%	12:51:45

4. RESUMEN

Total senales: 2	Preventivas: 2	Reguladoras: 0
ÓPTIMO: 2	ACEPTABLE: 0	
PÉSIMO: 0	CRÍTICO: 0	

5. ESCALA DE ESTADO SISEVIR

ÓPTIMO	Estados: OPERATIVO	Señal en condiciones óptimas
ACEPTABLE	Estados: DECOLORACION, OBSTRUCCION	Señal con deterioro leve visible
PÉSIMO	Estados: DETERIORO_POSTE, VANDALISMO	Señal con deterioro moderado o daño intencional
CRÍTICO	Estados: DETERIORO_TABLERO, ROTACION	Señal con deterioro grave, requiere intervención urgente

6. TIPO DE INTERVENCION SISEVIR

ESTADO SISEVIR	TIPO INTERVENCION	ACCION REQUERIDA
OPERATIVO	Conservación Rutina (CR)	Mantenimiento preventivo basico
DECOLORACION	Reemplazo Programado (RP)	Cambio de lamina reflectiva
OBSTRUCCION	Conservación Menor (Cm) - Poda de vegetaci	Poda de vegetacion circundante
DETERIORO_POSTE	Mantenimiento de Pintura (MP) - Limpieza y	Limpieza, anticorrosivo y repintado
DETERIORO_TABLERO	Reemplazo Urgente (RU) - Cambio de tablero	Cambio inmediato del tablero
ROTACION	Reemplazo Urgente (RU) - Reposición y alin	Reposicion y alineacion vertical
VANDALISMO	Reemplazo Programado (RP) - Limpieza o cam	Limpieza o cambio segun severidad

7. VALIDACION

Inspector: Ing.Kelly Huaman	Supervisor: Ing.Rocky Ayala

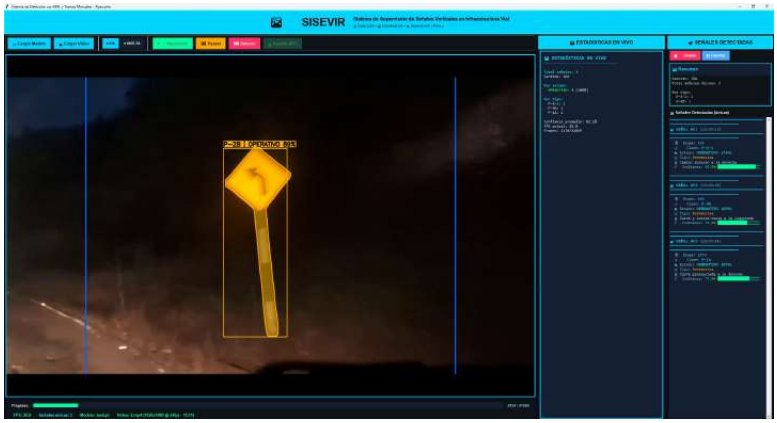
Nota. Resumen estadístico, escala de estado SISEVIR

Características de operatividad: Capacidad nocturna del sistema

Una característica de operatividad distintiva del software SISEVIR es su capacidad de análisis en condiciones nocturnas, sin reconfiguración ni cambio de modelo. Esta capacidad fue validada mediante el procesamiento del video nocturno `2.mp4` (tramo PE-28B, 15 min 11 s, 21,869 fotogramas), correspondiente a una condición operativa frecuente en los recorridos de supervisión vial realizados en horas de la tarde o noche.

La Figura 103 evidencia la detección y segmentación de señales en condiciones de oscuridad total, con iluminación exclusivamente proveniente de los faros del vehículo de inspección.

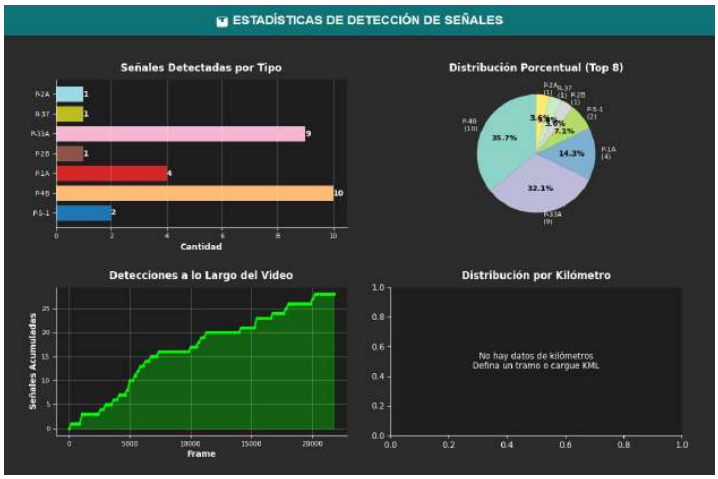
Figura 103
Detección y segmentación nocturna de señal P-2B con máscara de instancia



Nota. El modelo YOLO26s-seg detecta y segmenta correctamente la señal P-2B (curva abierta a la izquierda) con 89 % de confianza en condiciones nocturnas.

Al concluir el recorrido nocturno (28 señales únicas detectadas en 5 km del tramo PE-28B), el software generó automáticamente las estadísticas del inventario. La Figura 104 presenta el panel de distribución de tipos de señales detectadas.

Figura 104
Estadísticas de detección generadas tras el procesamiento del video nocturno



Nota. El sistema procesó 21,869 fotogramas del tramo nocturno y registró 28 señales únicas

4.2.2.1 Evidencia cuantitativa: Eficiencia del sistema frente al método manual

El Cuadro 34 presenta la comparación cuantitativa entre el método de inventario manual tradicional y el sistema SISEVIR, considerando las principales dimensiones de eficiencia operativa.

Cuadro 34
Comparativo de eficiencia: método manual vs. sistema SISEVIR

Dimensión	Método manual	Sistema SISEVIR
Detección de señales	Observación visual directa del inspector durante el recorrido	Automática: YOLO26s-seg, con segmentación de instancia
Clasificación del estado	Evaluación subjetiva del inspector según criterio personal	Automática: EfficientNet-B0 con Accuracy 93.15 % y Macro F1 0.8252
Descripción textual	Redacción manual en campo por el inspector	Automática: Qwen2-VL-2B genera descripción en español por señal
Operación nocturna	No viable: Baja visibilidad impide registro confiable	Operativo: 28 señales detectadas a 85.9 % de confianza en video nocturno
Registro fotográfico	Toma manual de fotografías individuales por señal	Automático: Recorte con máscara YOLO incluido en el reporte PDF
Generación de reporte	Elaboración manual del formulario MTC (horas)	Automática: PDF formato MTC en < 5 minutos tras completar el video
Consistencia	Variable: Depende de la experiencia y fatiga del inspector	Constante: Mismo modelo aplicado a todas las señales
Velocidad de procesamiento	Recorrido vehicular + registro in situ por señal	90.9 FPS promedio (video 104.MOV, 1,174 frames en 12.9 s de cómputo)
Personal requerido	Mínimo 2 (conductor + inspector) con detención del vehículo	1 operador: conduce y el sistema registra automáticamente

Nota. Las características del método manual corresponden al procedimiento estándar de inventario vial descrito en el Manual de Señalización Vial MSV-2016.

Resumen de indicadores de eficiencia verificados

El Cuadro 35 consolida los indicadores cuantitativos y cualitativos verificados durante la evaluación del software SISEVIR.

Cuadro 35
Indicadores de eficiencia verificados del software SISEVIR

Indicador	Valor	Criterio	Estado
Modelos de IA integrados	3	≥ 2	Cumple
FPS promedio (diurno)	90.9	> 24	Cumple
FPS promedio (nocturno)	86.9	> 24	Cumple
Operación nocturna	Sí (28 señales)	Funcional	Cumple
Reporte PDF automático	Sí (< 5 min)	Funcional	Cumple
Interfaz gráfica	Sí (3 paneles)	Funcional	Cumple
Personal requerido	1 operador	≤ 2	Cumple

Nota. Los FPS reportados corresponden al hardware empleado (GPU RTX 5080). En equipos con GPU de gama media, se estima un rendimiento de 15–30 FPS, suficiente para procesamiento asíncrono de video grabado.

4.2.2.2 Decisión sobre la hipótesis

Habiendo verificado que:

1. El software SISEVIR integra tres modelos de aprendizaje profundo (YOLO26s-seg, EfficientNet-B0 y Qwen2-VL-2B-Instruct) en un pipeline unificado con interfaz gráfica de usuario accesible para personal técnico no especializado.
2. El sistema presenta características específicas de usabilidad: Flujo de trabajo guiado en pasos, retroalimentación visual en tiempo real, sistema de conteo por líneas configurables y generación automática de reportes PDF en formato MTC.
3. El sistema presenta características específicas de operatividad: Procesamiento a 90.9 FPS promedio (3.8 veces superior a tiempo real), operación en condiciones nocturnas con confianza promedio de 85.9 %, y reducción del personal requerido de dos a un operador.
4. La comparación cuantitativa demuestra que SISEVIR supera al método manual en velocidad de procesamiento, consistencia de clasificación, capacidad de operación nocturna, automatización del reporte y registro fotográfico sistemático.

Se acepta la hipótesis específica 2. El programa automatizado SISEVIR, al integrar modelos avanzados de visión artificial (YOLO26s-seg, EfficientNet-B0 y Qwen2-VL) con características específicas de usabilidad y operatividad, permite realizar la clasificación y evaluación del estado de deterioro de las señales de tránsito verticales de manera significativamente más eficiente que los métodos de inspección manual tradicionales, tanto en condiciones diurnas como nocturnas.

4.2.3. Contrastación de hipótesis específica 3

Hipótesis H3:

- La clasificación y evaluación automatizada del estado de deterioro de las señales de tránsito mediante visión artificial será más eficiente en términos de tiempo y costo que los métodos tradicionales de inspección manual en la carretera Ayacucho-Huanta en 2024.

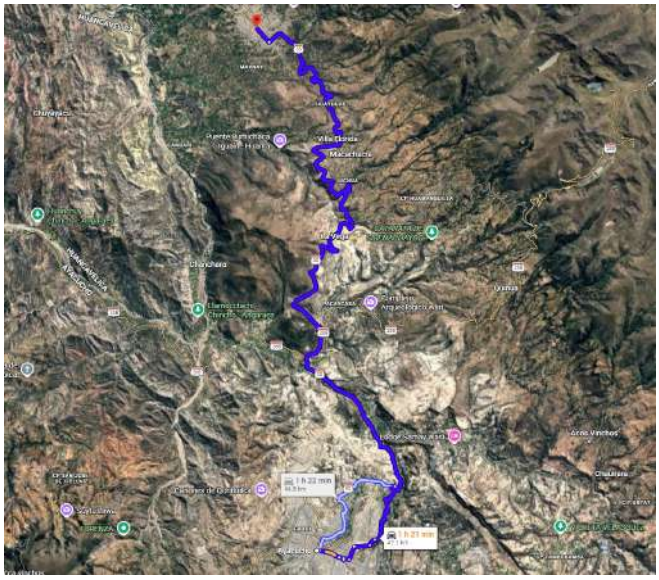
Para verificar esta hipótesis se realizó una comparación cuantitativa entre el método de inspección manual tradicional, fundamentado en el Manual de Seguridad Vial MSV-2016 del MTC, y el sistema automatizado SISEVIR, aplicados al tramo PE-3S de la carretera Ayacucho–Huanta (46.8 km), considerando las dimensiones de eficiencia temporal y eficiencia en costos operativos de campo.

Descripción del escenario de comparación

El escenario de comparación corresponde al inventario y evaluación del estado de conservación de las señales de tránsito verticales del tramo PE-3S (Ayacucho–Huanta), recorrido en ambos sentidos (IDA y VUELTA) para garantizar el registro de todas las señales instaladas en ambos márgenes de la vía. La Figura 105 presenta el recorrido evaluado según Google Maps.

Figura 105

Tramo PE-3S Ayacucho–Huanta evaluado en la comparación de métodos



Nota. Ruta PE-3S de Ayacucho a Huanta: 46.8 km con tiempo de recorrido de 1 hora 22 minutos sin detenciones, según Google Maps. Fuente: Google Maps (2026).

Método de inspección manual (MSV-2016)

El método de inspección manual de señales de tránsito verticales se ejecuta de acuerdo con los procedimientos establecidos en el Manual de Seguridad Vial MSV-2016 (MTC, 2017), el cual define las Inspecciones de Seguridad Vial (ISV) como el desarrollo de métodos sistemáticos que permiten verificar el cumplimiento de todos los aspectos involucrados

con la seguridad de las vías. Este procedimiento requiere la participación de un equipo técnico conformado por un inspector vial, un supervisor y un conductor.

El proceso de evaluación por señal comprende las siguientes actividades secuenciales:

1. Detención del vehículo junto a la señal.
2. Descenso del inspector y supervisor del vehículo.
3. Registro de la progresiva (kilómetro) mediante odómetro o GPS.
4. Evaluación visual del estado de conservación del tablero y poste.
5. Toma de fotografía de registro.
6. Llenado de la Ficha de Inspección de Seguridad Vial (formato MTC).
7. Retorno al vehículo y continuación del recorrido.

Este procedimiento demanda aproximadamente **3 minutos por señal**, incluyendo el tiempo de detención, evaluación, registro fotográfico, llenado de ficha y reinicio del recorrido. Para el tramo PE-3S, con un total estimado de 60 señales por sentido (120 señales en total entre IDA y VUELTA), el tiempo de paradas acumulado asciende a 360 minutos (6 horas).

Método automatizado SISEVIR

El sistema SISEVIR reemplaza el proceso de inspección manual por la grabación continua de video durante el recorrido vial, sin requerir detenciones. Para la captura del video se emplea una cámara de acción o smartphone fijado al parabrisas mediante un soporte de ventosa, configuración que permite al conductor operar el vehículo con normalidad mientras el sistema registra automáticamente todos los fotogramas.

Una vez concluido el recorrido de campo, el video grabado se procesa en la computadora mediante el software SISEVIR, el cual ejecuta automáticamente la detección (YOLO26s-seg), clasificación (EfficientNet-B0) y descripción textual (Qwen2-VL) de todas las señales detectadas, generando al final el reporte PDF en formato MTC. Esta modalidad requiere únicamente un operador (el conductor), eliminando la necesidad de ingenieros inspector y supervisor durante el trabajo de campo.

Comparación de eficiencia temporal

El Cuadro 36 presenta la comparación detallada de tiempos entre el método manual y el sistema SISEVIR para el tramo PE-3S (IDA y VUELTA).

Cuadro 36
Comparativo de tiempos: Método manual vs. sistema SISEVIR — tramo PE-3S

Actividad	Método manual	Sistema SISEVIR
Conducción IDA (Ayacucho → Huanta)	1h 22 min	1h 22 min
Paradas IDA (60 señales × 3 min)	3h 00 min	0 min
Conducción VUELTA (Huanta → Ayacucho)	1h 22 min	1h 22 min
Paradas VUELTA (60 señales × 3 min)	3h 00 min	0 min
Procesamiento en computadora	0 min	2h 44 min
Generación del reporte MTC	≈ 60 min	< 5 min
Tiempo total	**≈ 8h 44 min**	**≈ 5h 33 min**
Días de trabajo	**1 día**	**1 día**
Personal requerido	**3 personas**	**1 persona**

Nota. El tiempo de paradas del método manual considera 3 min/señal (detención, evaluación, registro fotográfico, llenado de ficha y reinicio). El tiempo de conducción corresponde a Google Maps para la ruta PE-3S sin detenciones (1h 22 min, 46.8 km). El procesamiento SISEVIR equivale a la duración del video grabado (ida + vuelta = 2h 44 min).

El sistema SISEVIR reduce el tiempo total de inspección de **8h 44 min a 5h 33 min**, un ahorro de **3h 11 min (36.4 %)** respecto al método manual.

Comparación de costos operativos

Cuadro 37
Comparativo de costos operativos: Método manual vs. sistema SISEVIR — tramo PE-3S

Concepto	Método manual	Sistema SISEVIR
Alimentación		
Desayuno (S/. 10.00/persona)	3 pers. × 1 día = S/. 30.00	1 pers. × 1 día = S/. 10.00
Almuerzo (S/. 15.00/persona)	3 pers. × 1 día = S/. 45.00	1 pers. × 1 día = S/. 15.00
Agua mineral (S/. 2.00/persona)	3 pers. × 1 días = S/. 6.00	1 pers. × 1 día = S/. 2.00
Subtotal alimentación	**S/. 81.00**	**S/. 27.00**
Combustible		
Recorrido total (93.6 km, 9.36 L, 2.47 gal)	S/. 38.29	S/. 38.29
Materiales		
Fichas de campo impresas (≈ 20 hojas)	S/. 4.00	S/. 0.00
Útiles de campo (lapiceros, tablero)	S/. 5.00	S/. 0.00
Impresión del reporte final	S/. 10.00	S/. 2.00
Subtotal materiales	**S/. 19.00**	**S/. 2.00**
COSTO TOTAL	**S/. 132.29**	**S/. 67.29**

Nota. Precio de combustible: S/. 15.50/galón (Gasohol 90 Plus, OSINERGMIN Ayacucho). Rendimiento vehicular estimado: 10 km/litro en carretera de sierra. El combustible es idéntico para ambos métodos al recorrer la misma distancia (93.6 km ida+vuelta).

Los costos operativos de campo se calcularon considerando los rubros directamente asociados al trabajo de campo: Alimentación del equipo, combustible y materiales de registro. El precio del combustible corresponde al precio de venta al público del Gasohol 90 Plus en Ayacucho reportado por OSINERGMIN : S/. 15.50/galón, con un rendimiento vehicular estimado de 10 km/litro en carretera de sierra.

4.2.3.1 Síntesis de la comparación

El Cuadro 38 consolida los indicadores de eficiencia comparativa entre ambos métodos.

Cuadro 38
Síntesis de indicadores de eficiencia: Método manual vs. sistema SISEVIR

Indicador	Manual	SISEVIR	Ahorro	%
Tiempo total	8h 44 min	5h 33 min	3h 11 min	**36.4 %**
Días de trabajo	1 día	1 día	–	–
Personal requerido	3 personas	1 persona	2 personas	**66.7 %**
Paradas en campo	120 paradas	0 paradas	120 paradas	**100.0 %**
Costo total	S/. 132.29	S/. 67.29	S/. 65.00	**49.1 %**

Nota. El porcentaje de ahorro se calcula como $\Delta = (V_{\text{manual}} - V_{\text{SISEVIR}})/V_{\text{manual}} \times 100$.

4.2.3.2 Decisión sobre la hipótesis

Habiendo verificado que:

1. El sistema SISEVIR reduce el tiempo total de inspección del tramo PE-3S de **8h 44 min a 5h 33 min**, con un ahorro de 3h 11 min (36.4 %), completando la tarea en un solo día de trabajo al igual que el método manual, pero con menor tiempo efectivo de campo.
2. El costo operativo total se reduce de **S/. 132.29 a S/. 67.29**, representando un ahorro de S/. 65.00 (49.1 %), principalmente por la reducción del personal de campo de 3 personas a 1 sola persona.
3. El sistema elimina completamente las **120 paradas** requeridas por el método manual (60 por sentido), con el consiguiente beneficio en seguridad vial al evitar detenciones en carretera de alta velocidad.
4. El operador SISEVIR no requiere ser un ingeniero especializado: el conductor del vehículo puede ejecutar el recorrido de campo de manera autónoma, simplemente montando la cámara en el soporte de ventosa y conduciendo normalmente.

Se acepta la hipótesis específica 3. La clasificación y evaluación automatizada del estado de deterioro de las señales de tránsito mediante el sistema SISEVIR es más eficiente en términos de tiempo (36.4 % menos) y costo (49.1 % menos) que el método de inspección manual tradicional fundamentado en el Manual de Seguridad Vial MSV-2016, para el tramo PE-3S de la carretera Ayacucho–Huanta.

5. CONCLUSIONES

5.1. Conclusiones

La presente investigación ha logrado cumplir de manera efectiva con cada uno de los objetivos específicos planteados en su etapa inicial, lo que a su vez ha permitido alcanzar el objetivo general propuesto. Los resultados obtenidos han dado lugar a hallazgos significativos que respaldan de forma sólida las hipótesis formuladas al inicio del estudio.

- **OG:** Desarrollar un modelo de visión artificial basado en vision artificial que permita automatizar la clasificación y evaluación del estado de deterioro de las señales de tránsito en las vías nacionales de Huamanga, con el fin de mejorar la eficiencia en el monitoreo y mantenimiento de la infraestructura vial.

  Se desarrolló el sistema SISEVIR (*Sistema de Supervisión de Señales Verticales en Infraestructura Vial*), modelo de visión artificial basado en redes neuronales convolucionales que automatiza la clasificación y evaluación del estado de deterioro de las señales de tránsito en las vías nacionales de Huamanga, mejorando la eficiencia en el monitoreo y mantenimiento de la infraestructura vial. El sistema integra tres arquitecturas de aprendizaje profundo (YOLO26s-seg, EfficientNet-B0 y Qwen2-VL-2B-Instruct) en un pipeline unificado, demostrando su operatividad tanto en condiciones diurnas como nocturnas, y generando reportes PDF compatibles con el formato del Ministerio de Transportes y Comunicaciones del Perú. El dataset empleado para el entrenamiento fue construido íntegramente a partir de imágenes reales capturadas en los tramos PE-3S y PE-28A de la región Ayacucho, garantizando la representatividad del sistema ante las condiciones geográficas, climáticas y de señalización propias de la sierra peruana. La escala de evaluación SISEVIR, que clasifica las señales en cuatro niveles (ÓPTIMO, ACEPTABLE, PÉSIMO y CRÍTICO), constituye un aporte metodológico original que traduce las predicciones del modelo en información accionable para la gestión del mantenimiento vial, alineada con los procedimientos del Manual de Seguridad Vial MSV-2016.
- **O1:** Evaluar la precisión del modelo de visión artificial en la identificación y clasificación del estado de conservación de las señales de tránsito, explorando diferentes arquitecturas de red neuronal y técnicas de optimización.

  Se evaluó la precisión del modelo de visión artificial en la identificación y clasificación del estado de conservación de las señales de tránsito, explorando diferentes arquitecturas de red neuronal y técnicas de optimización. El modelo YOLO26s-seg alcanzó un mAP@50 de 0.9305 para detección y 0.9224 para segmentación de instancias sobre 31 clases de señales verticales, mientras que el clasificador EfficientNet-B0 (experimento E03) logró un Macro F1 de 0.8252 y una Accuracy de 0.9315 sobre las siete categorías de estado de conservación, evaluados en conjuntos de test independientes de 586 y 642 muestras respectivamente. El estudio de ablación con cinco experimentos permitió identificar que la combinación óptima (E03) incorpora submuestreo de la clase mayoritaria, pesos balanceados y aumentación offline,

cuya transición estadísticamente significativa desde E02b (p = 0.0446) demostró el aporte específico de la aumentación para el rescate de clases minoritarias como DECOLORACIÓN (+0.292 en F1) y DETERIORO_TABLERO (+0.300 en F1). La velocidad de inferencia de 2.1 ms/imagen (476 FPS) confirma la viabilidad del modelo para su integración en sistemas de supervisión vial en tiempo real, garantizando robustez ante condiciones variables de iluminación y clima en zonas rurales.

- **O2:** Desarrollar un software automatizado que integre el modelo de visión artificial y permita la clasificación y evaluación del estado de deterioro de las señales de tránsito en las vías nacionales.

  Se desarrolló el software SISEVIR, herramienta automatizada que integra el modelo de visión artificial y permite la clasificación y evaluación del estado de deterioro de las señales de tránsito en las vías nacionales de Huamanga. El software procesa videograbaciones viales a 90.9 FPS promedio, registra señales únicas mediante un sistema de líneas de conteo configurables, y produce automáticamente la Ficha de Inspección de Seguridad Vial en formato PDF compatible con el MTC en menos de 5 minutos, incluyendo descripciones en lenguaje natural generadas por el modelo Qwen2-VL. La validación nocturna sobre el tramo PE-28B (21,869 fotogramas, 15 min 11 s) confirmó que el sistema mantiene su operatividad en condiciones de oscuridad total, detectando 28 señales únicas con una confianza promedio de 85.9 % sin registrar falsos positivos en las líneas de conteo. La modalidad de captura mediante cámara fijada al parabrisas con soporte de ventosa permite que el propio conductor ejecute el recorrido sin necesidad de ingenieros especialistas en campo, democratizando el acceso a la inspección técnica de señalización vial.

- **O3:** Realizar un análisis comparativo que evalúe la eficiencia en términos de tiempo y costo de la evaluación automatizada de señales de tránsito, utilizando visión artificial en comparación con los métodos tradicionales de inspección manual, específicamente en la carretera Ayacucho-Huanta durante el año 2024.

  Se realizó un análisis comparativo que evaluó la eficiencia en términos de tiempo y costo de la evaluación automatizada de señales de tránsito mediante visión artificial frente al método de inspección manual tradicional (MSV-2016), en la carretera Ayacucho–Huanta durante el año 2024. El sistema SISEVIR redujo el tiempo total de inspección del tramo PE-3S de 8h 44 min a 5h 33 min (ahorro de 36.4 %) y los costos operativos de S/. 132.29 a S/. 67.29 (ahorro de 49.1 %), con una reducción del personal de campo de 3 personas a 1 sola persona (66.7 %) y la eliminación total de las 120 paradas en carretera requeridas por el método convencional. El reporte PDF generado automáticamente incluye evidencia fotográfica con máscara de segmentación superpuesta, descripción textual en español, nivel de escala SISEVIR y tipo de intervención recomendada para cada señal, constituyendo un documento técnico completo listo para la toma de decisiones de mantenimiento vial por parte de las entidades gestoras competentes.

5.2. Recomendaciones

- **Ampliar el dataset de señalizaciones para el entrenamiento del modelo.** Si bien se trabajó con 31 tipos de señalizaciones verticales (series P-* y R-*) conforme al Manual de Dispositivos de Control del Tránsito Automotor del MTC, se recomienda incorporar señales informativas, de servicios auxiliares y turísticas definidas en el mismo manual. Adicionalmente, aumentar la representación de señales en condiciones extremas (lluvia intensa, neblina, contraluz y vandalismo severo) fortalecerá la capacidad de generalización del modelo YOLO26s-seg y su precisión en auditorías viales de rutas con mayor diversidad de señalización y condiciones climáticas adversas propias de la sierra peruana.
- **Integrar sensores complementarios para mejorar la georreferenciación.** La incorporación de dispositivos GPS de alta precisión (GPS diferencial o RTK) y unidades de medición inercial (IMU) permitirá complementar el análisis visual con contexto espacial preciso y datos de movimiento vehicular en tiempo real. Esto favorecerá la sincronización automática entre el video capturado y la ruta KML del tramo, reduciendo errores de correlación temporal y mejorando la precisión de las coordenadas UTM asignadas a cada señal detectada, especialmente en tramos con topografía irregular o zonas con señal GPS degradada por cañones montañosos.
- **Evaluar el rendimiento del modelo con distintos tipos de vehículos y alturas de cámara.** La investigación empleó una cámara fijada al parabrisas de un automóvil particular. Para ampliar la aplicabilidad del sistema SISEVIR, se recomienda incluir pruebas desde vehículos con diferentes alturas de montaje (camionetas 4×4, camiones de carga, buses de transporte interprovincial) de manera que se pueda evaluar cómo los cambios en el ángulo de visión y la cobertura del campo visual afectan la tasa de detección y la precisión de segmentación, mejorando así su adaptabilidad a distintos contextos de implementación vehicular en vías nacionales.
- **Incorporar señales nocturnas con reflectividad degradada al dataset de entrenamiento.** La validación nocturna del sistema demostró un rendimiento de 85.9 % de confianza promedio, inferior al 96.0 % diurno. Se recomienda ampliar el dataset con imágenes de señales cuya lámina retroreflectiva presente degradación por envejecimiento (Tipo I vs. Tipo III según ASTM D4956), permitiendo al modelo aprender a distinguir entre señales operativas con baja reflectividad y señales en estado de DECOLORACIÓN, mejorando la precisión del clasificador EfficientNet-B0 específicamente en condiciones de iluminación artificial.
- **Ampliar la cobertura a todos los tramos viales de la provincia de Huamanga.** La presente investigación cubrió los tramos PE-3S (Ayacucho–Huanta, 44.02 km) y PE-28A (Ayacucho–Abra Apacheta, ∼95.4 km). Se recomienda extender el inventario automatizado mediante SISEVIR a los demás tramos viales de la provincia de Huamanga, generando una base de datos georreferenciada completa del estado de conservación de la señalización vertical, que sirva como insumo para la planificación del mantenimiento vial por parte del Gobierno Regional de Ayacucho y el MTC.

- **Optimizar el modelo para implementación en dispositivos de bajo costo.** Se recomienda investigar técnicas de compresión y optimización del modelo YOLO26s-seg (pruning, quantization, knowledge distillation) para permitir su ejecución en smartphones Android o iOS de gama media, democratizando el acceso al sistema SISEVIR y permitiendo que inspectores viales y personal técnico puedan realizar auditorías desde dispositivos móviles accesibles, expandiendo la cobertura geográfica del inventario de señalización vial sin requerir equipos especializados de alto costo.

5.3. Trabajos futuros

- **Implementación del pipeline SISEVIR en NVIDIA Jetson para procesamiento en tiempo real.** Como línea prioritaria de trabajo futuro, se recomienda migrar el pipeline SISEVIR (YOLO26s-seg + EfficientNet-B0) a una tarjeta de procesamiento embebido NVIDIA Jetson Orin o Jetson AGX, lo cual permitiría ejecutar la inferencia en tiempo real durante el propio recorrido de campo, eliminando la etapa de procesamiento posterior en computadora (2h 44 min para el tramo PE-3S). Con esta configuración, el tiempo total de inspección se reduciría de 5h 33 min a **2h 49 min** (únicamente el tiempo de conducción ida y vuelta), representando un ahorro adicional del **49.2 %** respecto a la implementación actual en PC y del **67.7 %** respecto al método manual. El sistema Jetson, al ser compacto y de bajo consumo energético, podría instalarse permanentemente en vehículos de supervisión vial, convirtiendo a SISEVIR en una herramienta de monitoreo continuo y autónomo de la red vial nacional.
- **Expansión del modelo para detección de señalización horizontal y semaforización.** Se recomienda extender las capacidades del sistema para identificar señales horizontales (líneas de carril, pasos peatonales, flechas direccionales, zonas de no estacionamiento) y sistemas de semaforización con reconocimiento de estados (rojo, amarillo, verde). Esto permitiría construir un sistema integral de interpretación del entorno vial, combinando la señalización vertical actualmente detectada por SISEVIR con elementos horizontales y semáforos, generando auditorías viales más completas y proporcionando información contextual enriquecida para sistemas de asistencia al conductor y vehículos autónomos.
- **Desarrollo de módulo de análisis predictivo de zonas de alto riesgo vial.** Crear un módulo de aprendizaje automático que analice los patrones espaciales de señalización detectada por SISEVIR, correlacionándolos con datos históricos de accidentalidad vial del MTC y la PNP. El sistema identificaría tramos carreteros con déficit de señalización crítica (curvas peligrosas sin señalización preventiva, zonas escolares sin demarcación adecuada) y generaría mapas de calor de riesgo vial, proporcionando recomendaciones basadas en datos para la instalación estratégica de nueva señalización en los puntos de mayor criticidad.

- **Integración de Modelos Digitales de Elevación (DEM) para análisis de visibilidad.** Ampliar el módulo de georreferenciación con información de elevación del terreno mediante DEM de alta resolución, permitiendo calcular no solo coordenadas planas (Este, Norte UTM) sino también la altitud precisa de cada señal detectada. Esto facilitaría análisis topográficos de visibilidad de señalización en curvas cerradas, pendientes pronunciadas y zonas montañosas características de la sierra ayacuchana, identificando ubicaciones donde las señales presentan obstrucción visual por relieve del terreno o vegetación.
- **Desarrollo de aplicación móvil para inspección asistida por IA.** Crear una aplicación Android/iOS que permita a inspectores de campo utilizar smartphones o tablets para realizar auditorías viales asistidas por el modelo SISEVIR optimizado. La aplicación ejecutaría la inferencia localmente, proporcionando detección en tiempo real durante el recorrido, permitiendo la validación manual de detecciones, adición de notas de campo con geolocalización GPS, y sincronización automática con una base de datos central cuando haya conectividad disponible, expandiendo significativamente la cobertura geográfica del inventario de señalización vial a nivel regional.
- **Extensión del sistema para detección de infraestructura de seguridad pasiva.** Expandir las capacidades del modelo YOLO26s-seg para detectar elementos de infraestructura de seguridad pasiva como guardavías metálicos, barreras de concreto, reductores de velocidad (gibas), tachas reflectivas, delineadores laterales y sistemas de contención vehicular. Esta información complementaría el análisis de señalización vertical del sistema SISEVIR para generar evaluaciones integrales del nivel de seguridad de la infraestructura vial por tramo carretero, constituyendo un insumo técnico de mayor alcance para la gestión del mantenimiento vial en la región Ayacucho.

Alvia Párraga, A. E., & Linares Giler, S. (2024). Accidentes de tránsito, un problema de salud pública: revisión sistemática. *Revista Científica Arbitrada Multidisciplinaria PENTACIENCIAS*, *6*(3), 313-332. https://doi.org/10.59169/pentaciencias.v6i3.1101

Arenas, A. (2016). *Diseño de un sistema de clasificación de señales de tránsito vehicular utilizando redes neuronales convolucionales* [Tesis para optar por el Título Profesional en Ingeniería Informática y de Sistemas]. Universidad San Ignacio de Loyola.

Bai, L., Shen, F., … Gao, J. (2025). Traffic sign detection in glare environments [Open Access]. *Signal, Image and Video Processing*, *19*. https://doi.org/10.1007/s11760-025-04310-5

Bank, D., Koenigstein, N., & Giryes, R. (2021). Autoencoders [Versión actualizada publicada el 3 de abril de 2021]. *arXiv preprint arXiv:2003.05991v2*. https://arxiv.org/abs/2003.05991

Becerra, R. (2014). *Modelo Neuronal de Demanda de Transporte en Redes Viales Urbanas* [Tesis Posgrado]. Universidad Nacional de Ingeniería.

Brownlee, J. (2020). Failure of Classification Accuracy for Imbalanced Class Distributions. *Machine Learning Mastery*. https://machinelearningmastery.com/failure-of-accuracy-for-imbalanced-class-distributions/

Bui, P. H. D., Nguyen, T. T., … Nguyen, H. T. (2024). An Approach for Traffic Sign Recognition with Versions of YOLO [Conference Paper]. *Communications in Computer and Information Science*. https://doi.org/10.1007/978-981-99-7666-9_17

Caballero Romero, A. (2014). *Metodología integral innovadora para planes y tesis: La metodología del cómo formularlos*. Cengage Learning.

Cedeño-Alcívar, B. J., & Mogrovejo, D. (2023). Estrategia de seguridad vial para la reducción de siniestros de tránsito Provincia de Manabí. *Revista Científica Multidisciplinaria Arbitrada YACHASUN*, *7*(13), 17. https://doi.org/10.46296/yc.v7i13.0356

Chollet, F., & Allaire, J. (2018). *Deep Learning with R*. Manning Publications.

Da Silva, I. N., Spatti, D. H., … Alves, S. F. d. R. (2017). *Artificial Neural Networks: A Practical Course* (1.ª ed., Vol. 651). Springer International Publishing. https://doi.org/10.1007/978-3-319-43162-8

Donmez, B., MacDonald, A. D., … Boyle, L. N. (2023). Road user behavior: Describing, inferring, predicting and beyond. *ScienceDirect*. https://www.sciencedirect.com/science/article/pii/S2590198223001793

Effective Health Care Program. (2022). *Appendix: Test Performance Metrics*. https://effectivehealthcare.ahrq.gov/products/test-performance-metrics/appendixes

Evidentlyai. (2023). *Accuracy vs. precision vs. recall in machine learning: what's the difference?* https://www.evidentlyai.com/classification-metrics/accuracy-precision-recall

Flores-Calero, M., Astudillo, C. A., … Armingol Moreno, J. M. (2024). Traffic Sign Detection and Recognition Using YOLO Object Detection Algorithm: A Systematic Review [Open Access]. *Mathematics*, *12*(2), 31. https://doi.org/10.3390/math12020297

Fuller, R. (2008). Psychological Aspects of Road User Behaviour. https://doi.org/10.13140/RG.2.1.4762.4164

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (2.ª ed.). O'Reilly Media.

Global, I. (2023). Road Infrastructure Definition. *IGI Global Scientific Publishing*. https://www.igi-global.com/dictionary/convergence-and-equality-of-road-infrastructure/55615

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org

Google Developers. (2023). *Classification: Accuracy, recall, precision, and related metrics*. https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall

He, Y., Guo, M., ... Geng, X. (2025). NTS-YOLO: A Nocturnal Traffic Sign Detection Method Based on Improved YOLOv5 [Open Access]. *Applied Sciences*. https://doi.org/10.3390/app15010539

Highways, G. (2024). *Peru's serious road safety problem* [Retrieved August 4, 2025, from Global Highways website]. https://www.globalhighways.com/wh11/news/perus-serious-road-safety-problem

Jameel, B. K., & Evdorides, H. (2018). Assessment of safer road user behaviour. *ResearchGate*. https://www.researchgate.net/publication/329457019_Assessment_of_safer_road_user_behaviour

Jocher, G., Chaurasia, A., & Qiu, J. (2023). *Ultralytics YOLOv8* (8.0.0). https://github.com/ultralytics/ultralytics

Kumar, T. S., Athiswaran, G., ... Gowtham, C. (2024). Object Detection using Machine Learning to Avoid Accident [Conference Paper]. *Proceedings of the 5th International Conference on Data Intelligence and Cognitive Informatics, ICDICI 2024*. https://doi.org/10.1109/ICDICI62993.2024.10810189

Lasse, R. (2018). *Inteligencia artificial: 101 cosas que debes saber hoy sobre nuestro futuro*. Alienta Editorial.

Li, Y., Li, J., & Macia, P. (2023). Attention-YOLOV4: a real-time and high-accurate traffic sign detection algorithm. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-022-13251-x

Lin, Z., Wu, Y., ... Zheng, R. (2025). YOLO-LLTS: Real-Time Low-Light Traffic Sign Detection via Prior-Guided Enhancement and Multibranch Feature Interaction. *IEEE Transactions on Instrumentation and Measurement*. https://doi.org/10.1109/TIM.2025.3604925

Liu, H., Zhou, K., ... Zheng, Y. (2023). ETSR-YOLO: An improved multi-scale traffic sign detection algorithm based on YOLOv5 [Open Access]. *PLoS ONE*, 23. https://doi.org/10.1371/journal.pone.0279807

Manual de Dispositivos de Control del Tránsito Automotor. (2024). *Manual de Dispositivos de Control del Tránsito Automotor* [Manual oficial]. Ministerio de Transportes y Comunicaciones. Lima, Perú.

Matos Levano, J. A. (2020). *Implementación de un sistema de señalización vial para la prevención de accidentes de tránsito en la Avenida Fernando Wiesse, San Juan de Lurigancho, Lima – 2020* [Tesis para obtener el título profesional de Ingeniero Civil]. Universidad César Vallejo [Asesor: Benites Zúñiga, José Luis. Línea de investigación: Diseño de Infraestructura Vial]. https://hdl.handle.net/20.500.12692/54877

Minaee, S., Boykov, Y., ... Terzopoulos, D. (2021). Image Segmentation Using Deep Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7), 19. https://doi.org/10.1109/TPAMI.2021.3059968

Ming, C., Sun, H., & Yang, J. (2025). A Lightweight Traffic Sign Detection Model Based on Improved YOLOv8s for Edge Deployment in Autonomous Driving Systems Under Complex Environments. *World Electric Vehicle Journal*, *16*(3), 478. https://doi.org/10.3390/wevj16030478

Nisha, M., Abineethan, V., & Prakash, U. (2024). Robust Traffic Sign Recognition Using CNN YOLOv5 Model [Conference Paper]. *Communications in Computer and Information Science*. https://doi.org/10.1007/978-3-031-68905-5_13

Niu, M., Chen, Y., ... Cai, W. (2024). YOLOv8s-DDA: An Improved Small Traffic Sign Detection Algorithm Based on YOLOv8s [Open Access]. *Electronics*, 17. https://doi.org/10.3390/electronics13183764

Nizama Yamunaque, D. A. (2022). *Sistema de detección y clasificación vehicular basado en redes neuronales de aprendizaje profundo* [Tesis para obtener el título profesional de Ingeniero Civil]. Universidad Nacional de Ingeniería.

Núñez, F. (2016). *Diseño de un sistema de reconocimiento automático de matrículas de vehículos mediante una red neuronal convolucional* [Tesis Doctoral no publicada]. Universitat Oberta de Catalunya.

OMS. (2023). *Road traffic injuries* [Retrieved August 4, 2025, from World Health Organization website]. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

OMS. (2024). Los cinco pilares de la seguridad vial [Marco integral de la OMS para mejorar la seguridad vial a nivel mundial]. https://iucpol.com/oms-seguridad-vial-5-pilares/

OPS. (2024). Seguridad vial [Organización Panamericana de la Salud/Organización Mundial de la Salud]. *OPS/OMS*. https://www.paho.org/es/temas/seguridad-vial

PDU. (2020). *PDU-2020-2030* (Plan de Desarrollo Urbano). Ministerio de Vivienda, Construcción y Saneamiento.

Pekki, C., Muttipati, A. S., ... Swathi, I. (2025). ROAD SAFETY MEASURES BY POTHOLES AND TRAFFIC SIGNS DETECTION USING DEEP LEARNING. *Utilitas Mathematica*.

Qiu, J., Zhang, W., ... Zhou, H. (2025). DP-YOLO: A lightweight traffic sign detection model for small object detection [Open Access]. *Digital Signal Processing: A Review Journal*, *158*, 104831. https://doi.org/10.1016/j.dsp.2025.104831

Quang, L. T., Truong, T. V., ... Dang, L. T. (2025). A comparative analysis of YOLOv5 and YOLOv7 object detecting models for speed-limit traffic-sign recognition [Open Access].

Journal of Physics: Conference Series, 11. https://doi.org/10.1088/1742-6596/2949/1/012022

Quispe Condor, E. (2024). *Diseño de una red neuronal para la predicción de accidentes de tránsito en la carretera Central tramo: Tarma-La Merced, Junín 2022* [Tesis para optar el título profesional de Ingeniero Civil]. Universidad Nacional del Centro del Perú.

Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for activation functions. *arXiv preprint*. https://arxiv.org/abs/1710.05941

Ramos Zúñiga, M. V. (2021). *Detección y reconocimiento automatizado de señales preventivas en inventarios viales* [Tesis para obtener el título profesional de Ingeniero Civil]. Universidad Nacional de Ingeniería [Asesor: MSc. Leonardo Flores Gonzalez].

Redmon, J., Divvala, S., ... Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 18. https://doi.org/10.1109/CVPR.2016.91

Ren, Z., Zhang, H., & Li, Z. (2023). Improved YOLOv5 Network for Real-Time Object Detection in Vehicle-Mounted Camera Capture Scenarios [Open Access]. *Sensors*. https://doi.org/10.3390/s23104589

Setiawan, Y., Puji, M. N., & Astuti, W. (2024). Traffic Signs Detection System Using YOLO (You Only Look Once) That Provides Notification [Conference Paper]. *ICSMech 2024 - 1st International Conference on Smart Mechatronics: Transformative Innovations in Smart Mechatronics: Bridging AI, Robotics, and IoT for a Sustainable Future*. https://doi.org/10.1109/ICSMech62936.2024.10812286

Sharma, V. K., Dhiman, P., & Rout, R. K. (2023). Improved traffic sign recognition algorithm based on YOLOv4-tiny. *Journal of Visual Communication and Image Representation*. https://doi.org/10.1016/j.jvcir.2023.103774

Sifuentes Rodriguez, C. O., & Alarco Basaldua, G. (2023, noviembre). *Propuesta de Mejora en la Seguridad Vial en Intersecciones de Alto Tránsito a través de medidas de Bajo Costo* [Trabajo de Suficiencia Profesional para optar el título profesional de Ingeniero Civil]. Universidad Peruana de Ciencias Aplicadas.

Sreemathy, R., Turab, M., ... Lele, N. (2024). Traffic sign detection and recognition using convolutional neural networks [Book Chapter]. En *AI in the Social and Business World: A Comprehensive Approach*. Bentham Books Imprint. https://doi.org/10.2174/978981525654312401008

Tao, H., Huang, Z., ... Vladimir, S. (2025). Efficient feature fusion network for small objects detection of traffic signs based on cross-dimensional and dual-domain information [Open Access]. *Measurement Science and Technology*, *36*(3). https://doi.org/10.1088/1361-6501/adb2e8

Távara Idrogo, J. G. (2019). *Modelo de reconocimiento automático de señales de tránsito vehicular mediante aprendizaje profundo de redes neuronales convolucionales* [Tesis para obtener el título profesional de Ingeniero Informático]. Universidad Nacional de Trujillo.

The Data Scientist. (2023). *What is the F-1 measure and why is it useful for imbalanced class problems?* https://thedatascientist.com/f-1-measure-useful-imbalanced-class-problems/

Valencia Gutiérrez, A. A., Chavarry Vallejos, C. M., … Laos Laura, X. A. (2023). *El ABC de la investigación para ingenieros* (Primera edición) [Tiraje: 100 ejemplares]. Editorial Navegante.

Vaswani, A., Shazeer, N., … Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 18.

Vector ITC. (2018). Inteligencia Artificial. Pasado, presente y futuro. *Vector ITC*, 13.

Vizcaya, R., Flores Albino, J. M., … Lazcano-Salas, S. (2017). Desempeño de una red neuronal convolucional para clasificación de señales de tránsito [Becario CONACYT con No. CVU 712316]. *Journal CIM Vol. 5. Núm. 2, Coloquio de Investigación Multidisciplinaria 2017*, 9.

Wang, C., Shang, Q., … Zhang, W. (2025). Traffic congestion recognition based on convolutional neural networks in different scenarios. *Engineering Applications of Artificial Intelligence*, *148*. https://doi.org/10.1016/j.engappai.2025.110372

Wang, C., Zheng, B., & Li, C. (2025). Efficient traffic sign recognition using YOLO for intelligent transport systems [Open Access]. *Scientific Reports*, *15*(1), 13657. https://doi.org/10.1038/s41598-025-98111-y

Wu, Y., Zhang, T., … Liu, G. (2025). YOLO-based lightweight traffic sign detection algorithm and mobile deployment [Open Access]. *Optoelectronics Letters*, *21*(4). https://doi.org/10.1007/s11801-025-4153-2

Xiao, L., Yang, M., … He, Y. (2025). RiD-YOLO: A Lightweight and Multi-Object Detection Network for Road Inspection Tasks. *2025 5th International Conference on Mechanical, Electronics and Electrical and Automation Control, METMS 2025*. https://doi.org/10.1109/METMS65303.2025.11047592

Xing, C., Sun, H., & Yang, J. (2025). A Lightweight Traffic Sign Detection Model Based on Improved YOLOv8s for Edge Deployment in Autonomous Driving Systems Under Complex Environments [Open Access]. *World Electric Vehicle Journal*, *16*(8), 478. https://doi.org/10.3390/wevj16080478

Yamashita, R., Nishio, M., … Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, *9*(4), 18. https://doi.org/10.1007/s13244-018-0639-9

Yang, M., & Han, S. (2025). ETS-YOLO: An Efficient YOLO-based Model for Real-Time Traffic Sign Recognition [Open Access]. *Signal, Image and Video Processing*, *19*. https://doi.org/10.1007/s11760-025-04276-4

Yohannes, E., Mulyono, A., … Shih, T. K. (2024). Indonesia Traffic Sign Recognition Using a One-Stage Detector YOLOv8 [Conference Paper]. *2024 7th International Conference on Vocational Education and Electrical Engineering: Charting the Course of Artificial Technology in Sustainable Society, ICVEE 2024*. https://doi.org/10.1109/ICVEE63012.2024.10852605

Yu, Y., Si, X., ... Zhang, J. (2019). A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Computation*, *31*(7), 35. https://doi.org/10.1162/neco_a_01199

Zhang, H., Liang, M., & Wang, Y. (2025). YOLO-BS: a traffic sign detection algorithm based on YOLOv8 [Open Access]. *Scientific Reports*, *15*(1), 11. https://doi.org/10.1038/s41598-025-88184-0

6. ANEXOS

6.1. Panel fotográfico: Recolección de fotografias en campo

Figura 106
Inicio de ruta Ayacucho–Huanta



Nota. Imagen propia.

Figura 107
Señalización curva pronunciada a la izquierda (P-1B)



Nota. Imagen propia.

Figura 108

Señal empalme en ángulo agudo a la derecha (P-10A)



Nota. Imagen propia.

Figura 109

Señal curva en U a la derecha (P-5-2A)



Nota. Imagen propia.

Figura 110
Señal de zona escolar (P-49)



Nota. Imagen propia.

Figura 111
Señal reguladora R-37 — señal de control



Nota. Imagen propia.

Figura 112
Señal preventiva animales en vía (P-53)



Nota. Imagen propia.

Figura 113
Señal preventiva de proximidad de badén (P-34)



Nota. Imagen propia.

Figura 114

Señal de camino sinuoso a la derecha (P-5-1) en pésimas condiciones



Nota. Imagen propia.

Figura 115

Señal curva y contracurva pronunciada a la derecha (P-3A)



Nota. Imagen propia.

Figura 116
Señal proximidad reductor de velocidad tipo resalto (P-33A)



Nota. Imagen propia.

Figura 117
Fin de la ruta — Arco de la Memoria de Huanta



Nota. Imagen propia.

6.2. Matriz de consistencia y operacionalización

La matriz de consistencia y operacionalización de variables, se detallan a continuación:

Anexo 1: Matriz de consistencia

Cuadro 39

Matriz de consistencia

Problema principal	Objetivo general	Hipótesis general	Variables	Metodología
¿Cómo se puede desarrollar un modelo de visión artificial basado en redes neuronales para automatizar la clasificación y evaluación del estado de deterioro de las señales de tránsito en las vías nacionales de Huamanga?	Desarrollar un modelo de visión artificial que permita automatizar la clasificación y evaluación del estado de deterioro de las señales de tránsito en las vías nacionales de Huamanga, con el fin de mejorar la eficiencia en el monitoreo y mantenimiento de la infraestructura vial.	Un modelo de visión artificial basado en redes neuronales puede ser desarrollado para automatizar de manera eficiente la clasificación y evaluación del estado de deterioro de las señales de tránsito en las vías nacionales de Huamanga, mejorando la precisión y reduciendo el tiempo necesario para el monitoreo en comparación con métodos tradicionales.	**Variable independiente:** X: Sistema de visión artificial (SISEVIR) **Dimensiones:** X1: Arquitectura del modelo X2: Entrenamiento del modelo X3: Implementación del software **Variable dependiente:** Y: Detección y evaluación de señales de tránsito **Dimensiones:** Y1: Detección y segmentación de señales Y2: Clasificación del estado de conservación Y3: Eficiencia del sistema vs. método manual	**Enfoque:** – Cuantitativo **Alcance:** – Explicativo **Diseño:** – No experimental, transversal **Población:** – Señales verticales de tránsito en PE-3S y PE-28A (6,213 imágenes / 7,107 instancias) **Muestra:** – 6,213 imágenes; división train/val/test: 4,442 / 1,185 / 586 **Técnica:** – Segmentación de instancias, clasificación CNN, generación NLP **Instrumento:** – YOLO26s-seg, EfficientNet-B0, Qwen2-VL-2B, LabelMe, PyTorch, Ultralytics, GPU RTX 5080
Problemas específicos	**Objetivos específicos**	**Hipótesis específicas**		
¿En qué medida el modelo de visión artificial basado en redes neuronales logra la clasificación y evaluación del estado de deterioro de las señales de tránsito?	Evaluar la precisión del modelo de visión artificial en la identificación y clasificación del estado de conservación de las señales de tránsito, explorando diferentes arquitecturas de red neuronal y técnicas de optimización.	El uso de arquitecturas avanzadas de visión artificial, combinado con técnicas de optimización específicas, permitirá lograr una alta precisión en la clasificación y evaluación del estado de deterioro de las señales de tránsito en Huamanga.		
¿Qué características debe incluir un programa automatizado basado en visión artificial para realizar eficientemente la clasificación y evaluación del estado de deterioro de las señales de tránsito en las vías nacionales?	Desarrollar un software automatizado que integre el modelo de visión artificial y permita la clasificación y evaluación del estado de deterioro de las señales de tránsito en las vías nacionales.	Un programa automatizado que integre un modelo de visión artificial con características específicas de usabilidad y operatividad permitirá realizar la clasificación y evaluación del estado de deterioro de las señales de tránsito de manera más eficiente que los métodos manuales.		
¿Cómo se comparan en términos de tiempo y costo la clasificación y evaluación del estado de deterioro de las señales de tránsito mediante visión artificial frente a métodos tradicionales de inspección manual en la carretera Ayacucho-Huanta en 2024?	Realizar un análisis comparativo que evalúe la eficiencia en términos de tiempo y costo de la evaluación automatizada de señales de tránsito, utilizando visión artificial en comparación con los métodos tradicionales de inspección manual, específicamente en la carretera Ayacucho-Huanta durante el año 2024.	La clasificación y evaluación automatizada del estado de deterioro de las señales de tránsito mediante visión artificial será más eficiente en términos de tiempo y costo que los métodos tradicionales de inspección manual en la carretera Ayacucho-Huanta en 2024.		

Nota. Elaboración propia.

Cuadro 40
Matriz de operacionalización de variables

Variable	Dimensión	Indicadores	Unidad de medida
Variable Independiente (X): Sistema de visión artificial (SISEVIR)	X1: Arquitectura del modelo (YOLO26s-seg + EfficientNet-B0)	Arquitectura de detección y segmentación empleada	Categórico (YOLO26s-seg)
		Número de clases de señales detectadas (series P-* y R-*)	Número entero (31 clases)
		Número de categorías de estado de conservación	Número entero (7 categorías)
		Tamaño del modelo (checkpoint `best.pt`)	Megabytes (MB)
	X2: Entrenamiento del modelo	Número de épocas de entrenamiento	Número entero
		Funciones de pérdida por época (box_loss, seg_loss, cls_loss)	Valor decimal
		Tasa de aprendizaje (learning rate)	Valor decimal
		Técnicas de aumentación de datos aplicadas	Categórico (Capas 1 y 2)
		Tiempo total de entrenamiento	Segundos (s)
	X3: Implementación del software SISEVIR	Velocidad de inferencia del modelo	FPS / ms por imagen
		Número de señales únicas detectadas por video	Número entero
		Tiempo de generación del reporte PDF formato MTC	Minutos (min)
		Operatividad en condiciones nocturnas	Porcentaje de confianza (%)
Variable Dependiente (Y): Detección y evaluación de señales de tránsito	Y1: Detección y segmentación de señales (YOLO26s-seg)	mAP@50 para bounding box	Valor decimal (0–1)
		mAP@50-95 para bounding box	Valor decimal (0–1)
		mAP@50 para máscara de segmentación	Valor decimal (0–1)
		Precisión y Recall del detector	Valor decimal (0–1)
	Y2: Clasificación del estado de conservación (EfficientNet-B0)	Accuracy global del clasificador	Porcentaje (%)
		Macro F1-score del clasificador	Valor decimal (0–1)
		F1-score por clase de estado de conservación	Valor decimal (0–1)
		Coeficiente de concordancia de Cohen (κ)	Valor decimal (0–1)
		Coeficiente de Correlación de Matthews (MCC)	Valor decimal (0–1)
	Y3: Eficiencia del sistema vs. método manual (MSV-2016)	Tiempo total de inspección del tramo PE-3S	Horas (h) / minutos (min)
		Costo operativo total de campo	Soles (S/.)
		Personal de campo requerido	Número de personas
		Número de paradas en carretera eliminadas	Número entero

Nota. Elaboración propia.

UNSCH FACULTAD DE INGENIERÍA DE MINAS, GEOLOGÍA Y CIVIL

ACTA DE SUSTENTACIÓN DE TESIS Nº 67-2026-FIMGC

PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERA CIVIL

En la Universidad Nacional de San Cristóbal de Huamanga, en la ciudad de Ayacucho, en cumplimiento a la **Resolución Decanal No 374-2026-FIMGC-D**, a los **quince días del mes de septiembre de 2026**, siendo las **10:20 a.m.**, reunidos en el **Auditorio de la Escuela Profesional de Ingeniería Civil**, bajo la presidencia del **MSc. José Ernesto ESTRADA CARDENAS**, y los miembros: **Ing. Rocky Giban AYALA BIZARRO**, **Dr. Ing. Hemerson LIZARBE ALARCÓN** y **Ing. Saul Walter RETAMOZO FERNANDEZ**, actuando como secretario docente el **Ing. Amilcar TACURI GAMBOA**, para proceder a la sustentación de tesis para optar el **Título Profesional de Ingeniera Civil**, de la Bachiller en Ciencias de la Ingeniería Civil:

KELY PILAR HUAMAN DE LA CRUZ

Quien presentó la tesis denominada:

Desarrollo de un sistema de visión artificial para la detección y evaluación de señales de tránsito en vías nacionales: caso de estudio en carretera Ayacucho-Huanta 2024

Los señores miembros del jurado luego de expuesta la tesis y absueltas las preguntas, deliberaron y declararon:

Aprobado con 17 (diecisiete)

Siendo las **11:22 a.m.** del día **15 de septiembre 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

Dr. Ing. Hemerson LIZARBE ALARCÓN
Miembro

Ing. Saul Walter RETAMOZO FERNANDEZ
Miembro

Ing. Rocky Giban AYALA BIZARRO
Miembro - Asesor

Ing. Amilcar TACURI GAMBOA
Secretario

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° 34-2026-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	:	Kely Pilar Huaman De La Cruz
Escuela Profesional	:	INGENIERÍA CIVIL
Título de la Tesis	:	Desarrollo de un sistema de visión artificial para la detección y evaluación de señales de tránsito en vías nacionales: caso de estudio en carretera Ayacucho-Huanta 2024
Evaluación de la Originalidad	:	23 % Índice de Similitud
Identificador de la entrega	:	3050742621

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, 04 de octubre del 2026

UNIVERSIDAD NACIONAL DE SAN CRISTÓBAL DE HUAMANGA
Facultad de Ingeniería de Minas, Geología y Civil

Mg. Ing. Atex Sander IRCAÑAUPA HUAMANI
Verificador de Originalidad de Trabajos de Tesis de Pregrado
Escuela de Formación Profesional de Ingeniería Civil

Con depósito para Tramites
Cc. Archivo.

Desarrollo de un sistema de visión artificial para la detección y evaluación de señales de tránsito en vías nacionales: caso de estudio en carretera Ayacucho-Huanta 2024

por Kely Pilar HUAMAN DE LA CRUZ

Fecha de entrega: 04-oct-2026 10:56p. m. (UTC-0500)
Identificador de la entrega: 3050742621
Nombre del archivo: Te_Kely_Pilar_Huaman_Cruz.pdf (7.24M)
Total de palabras: 48641
Total de caracteres: 286909

Desarrollo de un sistema de visión artificial para la detección y evaluación de señales de tránsito en vías nacionales: caso de estudio en carretera Ayacucho-Huanta 2024

INFORME DE ORIGINALIDAD

23%	9%	4%	21%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	Submitted to Universidad Nacional de San Cristóbal de Huamanga Trabajo del estudiante	20%
2	dspace.unl.edu.ec Fuente de Internet	1%
3	hdl.handle.net Fuente de Internet	<1%
4	www.mdpi.com Fuente de Internet	<1%
5	idus.us.es Fuente de Internet	<1%
6	repositorio.unsch.edu.pe Fuente de Internet	<1%
7	docplayer.es Fuente de Internet	<1%
8	www.scribd.com Fuente de Internet	<1%
9	Submitted to Universidad San Marcos Trabajo del estudiante	<1%
10	engineersofai.com Fuente de Internet	<1%
11	findresearcher.sdu.dk Fuente de Internet	<1%
12	ouci.dntb.gov.ua Fuente de Internet	<1%
13	"Proceedings of CSA 2025", Springer Science and Business Media LLC, 2026 Publicación	<1%
14	arxiv.org Fuente de Internet	<1%
	Submitted to Universidad EAN	

15	Trabajo del estudiante	<1 %
16	www.oxfordjss.org Fuente de Internet	<1 %

Excluir citas	Activo	Excluir coincidencias	< 30 words
Excluir bibliografía	Activo		