

**Desarrollo de una Herramienta Inteligente para la Captura, Selección y Visualización
Automatizada de Imágenes Georreferenciadas en la Inspección de Vías Urbanas**

Juan Pablo Correa Sandoval

Cristian Acosta

Asesor

Duber Martínez Torres

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Sociales Artes y Humanidades ECSAH

Ingeniería Electrónica

2026

Resumen

Este trabajo presenta el desarrollo de una herramienta informática inteligente y modular cuyo objetivo es automatizar los procesos de captura, selección y visualización de imágenes georreferenciadas destinadas a la inspección de vías urbanas. Dicha herramienta facilita la generación de conjuntos de datos (datasets) optimizados, fundamentales para el entrenamiento y la evaluación de modelos de inteligencia artificial enfocados en la detección de defectos en vías vehiculares.

La metodología implementada articula varios procesos clave: una aplicación móvil captura simultáneamente video y coordenadas GPS, asegurando la correspondencia temporal y espacial de los datos. Posteriormente, el sistema segmenta el video en clips basados en un umbral de desplazamiento geográfico, logrando una notable reducción promedio del 96% en el volumen de fotogramas al eliminar redundancia. De cada clip, se selecciona automáticamente el frame más representativo utilizando un puntaje ponderado que combina métricas objetivas de calidad visual como nitidez (Laplaciano), entropía y la puntuación de calidad percibida HyperIQA.

Finalmente, los frames seleccionados son procesados utilizando el modelo Deep Learning NAFNet para corregir la distorsión por movimiento, lo que resultó en una mejora de la calidad visual promedio del 14,14% respecto al clip original. Los resultados experimentales validaron la efectividad de la herramienta, demostrando que esta mejora visual se traduce en un aumento absoluto del 20,93% en la tasa de detección de defectos por el modelo YOLOv12, además de reducir falsos positivos. La herramienta final proporciona una interfaz de usuario para la gestión de rutas, la visualización interactiva de trayectos sobre mapas y la evaluación visual de los resultados de detección, estableciéndose como una solución robusta para la generación estandarizada de datasets de alta calidad y la optimización del proceso de inspección vial. El

trabajo se desarrolla en el marco del proyecto institucional de investigación PGI3902ECBTI2024 de la UNAD, lo cual refuerza su pertinencia y articulación con las líneas de investigación de la Escuela de Ciencias Básicas, Tecnología e Ingeniería (ECBTI).

Palabras clave: *Detección, Georeferenciación, procesamiento, imágenes, dataset.*

Abstract

This work presents the development of an intelligent and modular computer tool aimed at automating the capture, selection, and visualization processes of georeferenced images intended for urban road inspection. This tool facilitates the generation of optimized datasets, which are essential for training and evaluating artificial intelligence models focused on detecting defects in vehicular roadways.

The implemented methodology articulates several key processes: a mobile application simultaneously captures video and GPS coordinates, ensuring the temporal and spatial correspondence of the data. Subsequently, the system segments the video into clips based on a geographical displacement threshold, achieving a notable average reduction of 96% in the volume of frames by eliminating redundancy. From each clip, the most representative frame is automatically selected using a weighted score that combines objective visual quality metrics such as sharpness (Laplacian), entropy, and the perceived quality score HyperIQA.

Finally, the selected frames are processed using the Deep Learning model NAFNet to correct motion blur, resulting in an average visual quality improvement of 14.14% compared to the original clip. Experimental results validated the tool's effectiveness, demonstrating that this visual improvement translates into an absolute increase of 20.93% in the defect detection rate by the YOLOv12 model, in addition to reduce false positives. The final tool provides a user interface for route management, interactive visualization of routes on maps, and visual evaluation of detection results, establishing itself as a robust solution for the standardized generation of high-quality datasets and the optimization of the road inspection process. The work is developed within the framework of the UNAD institutional research project PGI3902ECBTI2024, which reinforces its relevance and alignment with the research lines of the

School of Basic Sciences, Technology, and Engineering (ECBTI).

Keywords: *Detection, Georeferencing, Processing, Images, Dataset.*

Tabla De Contenido

Introducción	10
Planteamiento del Problema	12
Justificación	14
Objetivos.....	15
Objetivo General.....	15
Objetivos Específicos	15
Marco Conceptual y Teórico	16
Visión por Computador y Aplicaciones Viales	16
Aprendizaje Profundo y Detección de Objeto	17
Calidad de Imagen y Reducción de Borrosidad.....	17
Georreferenciación y Sistemas GPS.....	18
Diseño metodológico.....	20
Enfoque y Tipo de Investigación.....	20
Estructura Funcional y Flujo Metodológico del Sistema Propuesto.....	20
Captura Sincronizada de Video y GPS.....	22
Segmentación en Clips por Umbral de Desplazamiento	23
Selección Automática de Frames	25
Corrección de Distorsión por Movimiento	26
Etapa de Pruebas y Validación.....	27
Implementación de la Arquitectura Propuesta.....	28
Aplicación para Adquisición de Datos	29
Interfaz para la Visualización y Gestión de Datos Geoespaciales y Visuales.....	30
Pruebas Experimentales y Resultados Obtenidos.....	39

	7
Descripción del Entorno de Desarrollo.....	39
Visualización Geográfica.....	39
Reducción de Datos.....	40
Mejora Visual de las Imágenes.....	41
Impacto Sobre la Detección Automática de Defectos.....	43
Discusión de los Resultados.....	47
Conclusiones y Trabajos Futuros.....	49
Referencias Bibliográficas.....	51

Lista De Tablas

Tabla 1	<i>Reducción Alcanzada después del Proceso de Selección Automática.....</i>	41
Tabla 2	<i>Desempeño de Detección con y sin Corrección de Distorsión.....</i>	45
Tabla 3	<i>Promedio del Tiempo de Procesamiento por Imagen en Cada Etapa.....</i>	47

Lista de Figuras

Figura 1	<i>Diagrama de Bloques Metodológico.....</i>	21
Figura 2	<i>Diagrama de Flujo de la Aplicación.....</i>	23
Figura 3	<i>Interfaz de la Aplicación Android.....</i>	29
Figura 4	<i>Interfaz de Visualización y Gestión de Datos Geoespaciales y Visuales</i>	31
Figura 5	<i>Previsualización de Recorrido Cargado en el Mapa</i>	32
Figura 6	<i>Visualización Multimedia del Recorrido.....</i>	33
Figura 7	<i>Barra de Progreso de Extracción de Frames Representativos</i>	34
Figura 8	<i>Ventana Emergente para la Selección de Frames.....</i>	35
Figura 9	<i>Visualización de los Puntos Marcados en el Mapa.....</i>	36
Figura 10	<i>Ventana Emergente para el Procesamiento Mediante el Modelo YOLO.....</i>	37
Figura 11	<i>Diagrama de Flujo Operacional.....</i>	38
Figura 12	<i>Visualización de las Rutas Registradas durante la Etapa de Captura de Datos.....</i>	40
Figura 13	<i>Ejemplos de la Corrección de Distorsión por Movimiento.....</i>	42
Figura 14	<i>Comparación de Evaluación de Calidad</i>	43
Figura 15	<i>Ejemplos de Detección Automática de Defectos</i>	46

Introducción

La infraestructura vial urbana es un activo crítico que requiere un mantenimiento constante y eficiente para garantizar la seguridad y movilidad. Tradicionalmente, la inspección de vías se ha basado en métodos manuales, que son inherentemente subjetivos, lentos y costosos, resultando en una baja estandarización de los datos recopilados. En la actualidad, las tecnologías de Visión por Computador y Aprendizaje Profundo (Deep Learning) ofrecen un camino prometedor hacia la automatización de esta tarea, permitiendo la detección y clasificación objetiva de defectos viales. No obstante, la implementación exitosa de estos modelos requiere la existencia de vastos conjuntos de datos (datasets) de imágenes georreferenciadas y de alta calidad, una necesidad que a menudo no se satisface con los recursos locales.

Para abordar esta problemática, el presente trabajo desarrolla una Herramienta Inteligente y Modular para la Captura, Selección y Visualización Automatizada de Imágenes Georreferenciadas en la Inspección de Vías Urbanas. El objetivo primordial de esta herramienta es optimizar drásticamente el flujo de trabajo de recolección de datos mediante la automatización, asegurando que solo las imágenes más nítidas y representativas sean retenidas. Esto se logra a través de un proceso que incluye la captura sincronizada de video y GPS, la segmentación geográfica para la eliminación de la redundancia (con una reducción de fotogramas superior al 96%), la selección automática de frames clave basada en métricas de calidad visual (nitidez, entropía, HyperIQA) y la corrección de distorsión por movimiento mediante el modelo NAFNet.

Los resultados de este desarrollo demuestran que el procesamiento y la optimización de las imágenes impactan directamente en el rendimiento de los modelos de detección; la corrección de la borrosidad, por ejemplo, incrementó la tasa de detección del modelo YOLOv12 en un

20,93% absoluto. La herramienta final ofrece una solución robusta para la estandarización de los datasets y la mejora del proceso de inspección, proporcionando una interfaz de usuario para la visualización de rutas y la gestión de resultados. El proyecto se enmarca dentro del proyecto institucional de investigación PGI3902ECBTI2024 de la Universidad Nacional Abierta y a Distancia (UNAD), lo cual subraya su pertinencia y alineación con las líneas de investigación de la Escuela de Ciencias Básicas, Tecnología e Ingeniería (ECBTI).

El resto del documento está estructurado de la siguiente manera: se presenta el Planteamiento del Problema y la Justificación del proyecto, seguidos por los Objetivos; el Marco Conceptual y Teórico sienta las bases de las tecnologías utilizadas; la Metodología detalla el diseño y desarrollo de la herramienta; finalmente, se exponen los Resultados, las Conclusiones y las Recomendaciones derivadas del trabajo.

Planteamiento del Problema

En el contexto actual de crecimiento urbano acelerado y deterioro progresivo de la infraestructura vial, la evaluación oportuna, precisa y sistemática del estado de las vías se ha convertido en una necesidad crítica para garantizar la seguridad vial, optimizar la movilidad y planificar adecuadamente las labores de mantenimiento. Las entidades responsables de la gestión del espacio público enfrentan múltiples retos debido a que los métodos tradicionales de inspección suelen ser manuales, poco sistemáticos, dependientes de observaciones subjetivas y difíciles de escalar, lo que limita la capacidad para generar diagnósticos confiables y actualizados.

El uso de tecnologías de visión por computador representa una alternativa prometedora para automatizar la inspección vial. Sin embargo, uno de los principales obstáculos para la implementación efectiva de estas soluciones radica en la disponibilidad y calidad de los datasets de imágenes empleados para entrenar y validar los modelos de detección. Estos datasets suelen generarse a partir de videos capturados desde vehículos en movimiento, pero la selección de los frames adecuados generalmente se realiza de forma totalmente manual, lo que implica procesos lentos, costosos y con alto grado de subjetividad. Además, los videos pueden incluir imágenes redundantes, distorsionadas por movimiento, con iluminación inadecuada o con baja nitidez, lo que afecta negativamente la utilidad del conjunto de datos generado.

Otro aspecto importante es que muchas de estas imágenes no están georreferenciadas o no presentan una correspondencia clara entre la información visual y geoespacial. Esta desalineación entre los datos visuales y las coordenadas geográficas limita la posibilidad de mapear con precisión la ubicación de los defectos detectados, impidiendo su uso efectivo para

aplicaciones como la generación de alertas, análisis estadísticos o planeación de mantenimiento focalizado.

Actualmente existe carencia de herramientas que automaticen de forma integral la adquisición de datos, la selección inteligente de imágenes representativas, basada en criterios objetivos como la calidad visual, el nivel de solapamiento espacial entre frames, la georreferenciación y la visualización estructurada de resultados obtenidos por modelos de detección. Esta carencia dificulta tanto la generación de datasets estandarizados como la evaluación comparativa de distintos algoritmos en condiciones reales.

Ante esta necesidad se propone diseñar e implementar una plataforma tecnológica que permita automatizar la captura, evaluación, selección y visualización georreferenciada de imágenes obtenidas desde recorridos urbanos en vehículo, contribuyendo así a la construcción de datasets útiles para la experimentación científica y al desarrollo de modelos de visión por computador orientados a la inspección de pavimentos urbanos.

Pregunta de investigación

¿Cómo diseñar una herramienta que automatice la captura, selección georreferenciada y visualización de imágenes viales para crear datasets útiles en la detección de defectos por visión por computador?

Justificación

La herramienta propuesta aporta soluciones a varios problemas actuales en la inspección de infraestructura vial mediante tecnología de bajo costo, reutilizable y escalable. En primer lugar, permite capturar datos georreferenciados de manera sincronizada desde dispositivos móviles comunes, lo cual facilita su replicabilidad.

Además, facilita la generación de nuevos datasets experimentales, clave para entrenar modelos de detección de defectos basados en visión por computador. Esta funcionalidad es especialmente relevante para el avance de la investigación en contextos locales, donde escasean repositorios de imágenes adecuados.

La herramienta también permitirá registrar la ubicación geoespacial de defectos detectados, generando información potencialmente útil tanto para advertir a conductores, como para respaldar decisiones sobre mantenimiento vial basado en la información estadística suministrada.

Por último, la propuesta está alineada con el proyecto de investigación institucional titulado “Caracterización de desempeño de modelos Deep Learning para la evaluación del estado de vías vehiculares bajo condiciones nocturnas” (código: PGI3902ECBTI2024), lo que refuerza su pertinencia académica y su valor como insumo para investigaciones presentes y futuras en la UNAD.

Objetivos

Objetivo General

Desarrollar una herramienta informática modular para la adquisición, procesamiento y visualización de imágenes georreferenciadas de vías urbanas, orientada a facilitar la generación de datasets optimizados y la evaluación visual de resultados de modelos de detección de defectos en pavimento.

Objetivos Específicos

Implementar un módulo de captura que integre la grabación de video y la sincronización con datos de geolocalización.

Implementar en la interfaz de usuario un módulo que permita visualizar sobre un mapa los trayectos georreferenciados asociados a los vídeos registrados.

Implementar un módulo de selección automática de frames clave basado en métricas de calidad visual y control de solapamiento espacial.

Aplicar técnicas de mejora de imágenes a los frames seleccionados, incluyendo normalización de iluminación, contraste y reducción de distorsión por movimiento.

Incorporar un módulo de visualización de resultados de modelos Deep Learning pre-entrenados para detección de defectos sobre imágenes o videos de prueba.

Marco Conceptual y Teórico

La inspección automatizada de infraestructuras viales, como parte de sistemas de visión por computador, se fundamenta en varios campos teóricos: visión computacional, aprendizaje profundo, calidad de imagen, georreferenciación y procesamiento digital avanzado. Cada uno aporta herramientas críticas para la detección precisa y eficiente de defectos en pavimentos urbanos.

Visión por Computador y Aplicaciones Viales

La visión por computador es una disciplina de la inteligencia artificial que permite a los sistemas informáticos percibir, procesar y comprender información visual del entorno, con el objetivo de tomar decisiones o ejecutar acciones basadas en dicha interpretación (Ettalibi et al., 2024). Esta área combina técnicas de análisis de imágenes, reconocimiento de patrones y aprendizaje automático, y ha tenido un crecimiento exponencial gracias al avance de los algoritmos de aprendizaje profundo y al aumento de la capacidad de cómputo disponible (Shaolin et al., 2020).

En el ámbito de la infraestructura vial, la visión por computador ha permitido transformar procesos tradicionalmente manuales, lentos y subjetivos, como la inspección del estado del pavimento, en procedimientos automáticos, sistemáticos y reproducibles. Mediante técnicas de detección de objetos, segmentación semántica y clasificación, es posible identificar defectos comunes en las vías, tales como baches, grietas longitudinales, fisuras transversales, desprendimientos o irregularidades superficiales, con un alto grado de precisión (Ma et al., 2022; Guan et al., 2023).

Estos sistemas, al analizar imágenes capturadas desde vehículos o drones, no solo permiten reducir significativamente los costos operativos asociados al monitoreo vial, sino que

también amplían la cobertura territorial y la frecuencia con la que se pueden recolectar datos actualizados. Además, al eliminar la subjetividad inherente a la observación humana, se mejora la consistencia y la trazabilidad de los diagnósticos sobre el estado de la infraestructura (Askarzadeh et al., 2025).

La adopción de estas tecnologías no solo representa un avance técnico, sino que también abre nuevas oportunidades para la integración de soluciones inteligentes en el mantenimiento preventivo de vías, la gestión de recursos urbanos y la construcción de ciudades más seguras y sostenibles.

Aprendizaje Profundo y Detección de Objeto

El deep learning se centra en arquitecturas neuronales profundas capaces de aprender representaciones jerárquicas de datos visuales. Modelos como YOLOv8, Mask R-CNN y variantes enriquecidas con transformadores han mostrado un rendimiento sobresaliente en detección de baches y grietas en tiempo real (Wang et al., 2025; Nafaa *et al.*, 2024; Men *et al.*, 2024).

En el contexto de la visión por computador aplicada a la inspección vial, una práctica común consiste en utilizar modelos previamente entrenados en grandes datasets para resolver tareas similares en nuevos contextos. Esta técnica, conocida como transferencia de aprendizaje, permite ahorrar tiempo de entrenamiento y aprovechar características ya aprendidas por modelos como YOLO o Mask R-CNN (Bhavana et al., Xu *et al.*, 2022). En particular, esto es útil cuando no se cuenta inicialmente con un volumen amplio de datos locales o anotados.

Calidad de Imagen y Reducción de Borrosidad

La fidelidad visual de las imágenes empleadas en tareas de inspección vial es un factor determinante en el rendimiento de los modelos de visión por computador. Imágenes capturadas

desde vehículos en movimiento están expuestas a diversas Notas de distorsión, entre las que se incluyen la borrosidad por movimiento (motion blur), la iluminación inconsistente por sombras, reflejos o variaciones del entorno. Estas alteraciones reducen la capacidad del modelo para detectar detalles finos, afectando negativamente la precisión de la segmentación y la localización de defectos (Arya *et al.*, 2024).

Para mitigar estos efectos, se aplican técnicas objetivas de evaluación de calidad visual que permiten seleccionar, de manera automática, los frames más adecuados para conformar un dataset robusto. Entre las métricas tradicionales se destacan la varianza del operador Laplaciano, utilizada para estimar la nitidez, y el Índice de Similitud Estructural (SSIM), que evalúa la pérdida de información estructural entre imágenes. En años recientes, se han incorporado modelos más avanzados como MUSIQ (Multi-scale Image Quality), basado en transformadores, que permite estimar la calidad percibida desde múltiples escalas espaciales de forma más coherente con la percepción humana (Ke *et al.*, 2021).

Además de la evaluación, es posible aplicar procesos de mejora como el deblur (eliminación de desenfoque) tales como NAFNet (Esaulov & Esfahani, 2025), la normalización de histograma (para ajustar la iluminación) y técnicas de superresolución por redes generativas adversariales como ESRGAN, que incrementan el detalle visual en imágenes de baja calidad (Yuan *et al.*, 2022). Estas mejoras no solo optimizan la entrada al modelo, sino que también elevan la utilidad de las imágenes para tareas de inspección manual.

Georreferenciación y Sistemas GPS

La georreferenciación permite asociar coordenadas (latitud/longitud) a cada imagen o frame, facilitando su ubicación precisa en un entorno urbano. Esto es vital para almacenar el estado de la vía, generar alertas y realizar análisis espaciales y temporales eficaces.

El uso de receptores GPS integrados en teléfonos móviles o dispositivos embebidos permite registrar de forma continua coordenadas sincronizadas con la captura de video o imágenes, generando un registro cronológico de la ruta. Esta sincronización temporal-espacial es crucial para garantizar la trazabilidad y reproducibilidad de los datos

Una vez adquirida la información georreferenciada, su visualización sobre mapas digitales es un recurso clave para facilitar la interpretación por parte de usuarios. Para esto, existen diversas librerías de software que permiten integrar de forma eficiente coordenadas geográficas con visualizaciones interactivas. En el ecosistema Python, herramientas como Folium (basada en Leaflet.js), geopandas (para análisis geoespacial con estructuras tipo DataFrame), gmap3 (basada en Google Maps) y Plotly Express ofrecen interfaces accesibles para representar rutas, puntos de interés o resultados de detección sobre mapas reales. Estas librerías permiten incluir capas personalizadas, leyendas, imágenes vinculadas y resultados del modelo, enriqueciendo la experiencia de análisis.

Diseño metodológico

Enfoque y Tipo de Investigación

Este trabajo se enmarca en un enfoque metodológico de tipo aplicado y de carácter tecnológico, orientado al desarrollo de una herramienta software para la solución de una necesidad específica en el ámbito de la visión por computador. La metodología adoptada es de naturaleza exploratoria y experimental, en tanto que se busca diseñar, implementar y validar un sistema informático basado en datos capturados en recorridos urbanos desde un vehículo en movimiento. El proceso de desarrollo se organizó por etapas modulares que se integran progresivamente, permitiendo la verificación de cada componente del sistema de forma independiente y en conjunto.

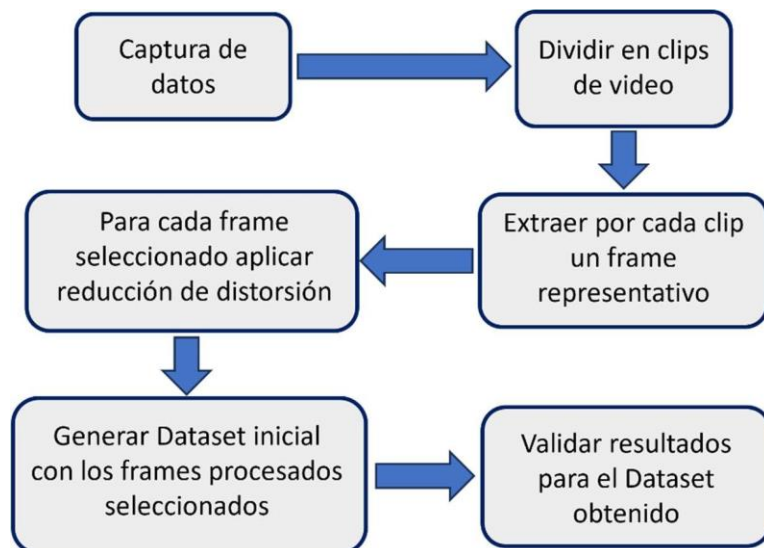
A continuación, se describen los principales módulos funcionales:

Estructura Funcional y Flujo Metodológico del Sistema Propuesto

La plataforma integral propuesta tiene como objetivo la adquisición y optimización de información geoespacial y visual orientada al análisis del entorno vial. Su concepción parte de la necesidad de unificar en un mismo sistema los procesos de captura, procesamiento y mejora de datos, asegurando así la generación de información precisa, coherente y de buena calidad. El diseño adopta una arquitectura modular, lo que permite distribuir las tareas entre componentes especializados que trabajan juntos de manera coordinada para asegurar eficiencia, escalabilidad y solidez operativa.

Figura 1

Diagrama de Bloques Metodológico



Nota. Elaboración propia

La estructura funcional del sistema se desarrolla mediante un flujo de trabajo automatizado que combina la captura sincronizada de video y coordenadas GPS, la organización espacial de los recorridos y la generación de un conjunto de datos optimizado para análisis futuros. Este flujo asegura la concordancia temporal y espacial entre los datos visuales y geográficos, de modo que cada registro pueda asociarse con precisión a un punto determinado del trayecto. Este principio es la piedra angular del sistema y permite que las etapas posteriores de segmentación, evaluación de calidad y mejora visual se realicen dentro de un marco de datos coherente y rastreado.

En la Figura 1 se presenta el proceso propuesto para la generación automatizada del dataset se compone de seis fases claramente definidas. En primer lugar, se realiza la captura sincronizada de datos, combinando video y georreferenciación obtenida desde un vehículo en

movimiento. Posteriormente, el material registrado se divide en clips de video utilizando criterios espaciales para evitar redundancia visual. Para cada clip, se selecciona automáticamente un frame representativo con base en métricas objetivas de calidad. A cada uno de estos frames seleccionados se le aplica un procedimiento de reducción de distorsión, orientado a mitigar efectos de desenfoque por movimiento y mejorar la nitidez general de la imagen. Con los frames ya procesados, se genera un dataset inicial consistente y depurado. Finalmente, este dataset es sometido a una etapa de validación con el fin de verificar su calidad, coherencia espacial y pertinencia para tareas posteriores de inspección vial mediante visión por computador.

Captura Sincronizada de Video y GPS

La etapa de captura contempla la adquisición simultánea de video e información de posicionamiento geográfico mediante una aplicación móvil desarrollada para este propósito. El sistema combina el uso de la cámara y el sensor GPS del dispositivo, de modo que cada fotograma queda asociado a sus respectivas coordenadas, garantizando la coherencia temporal y espacial de los datos.

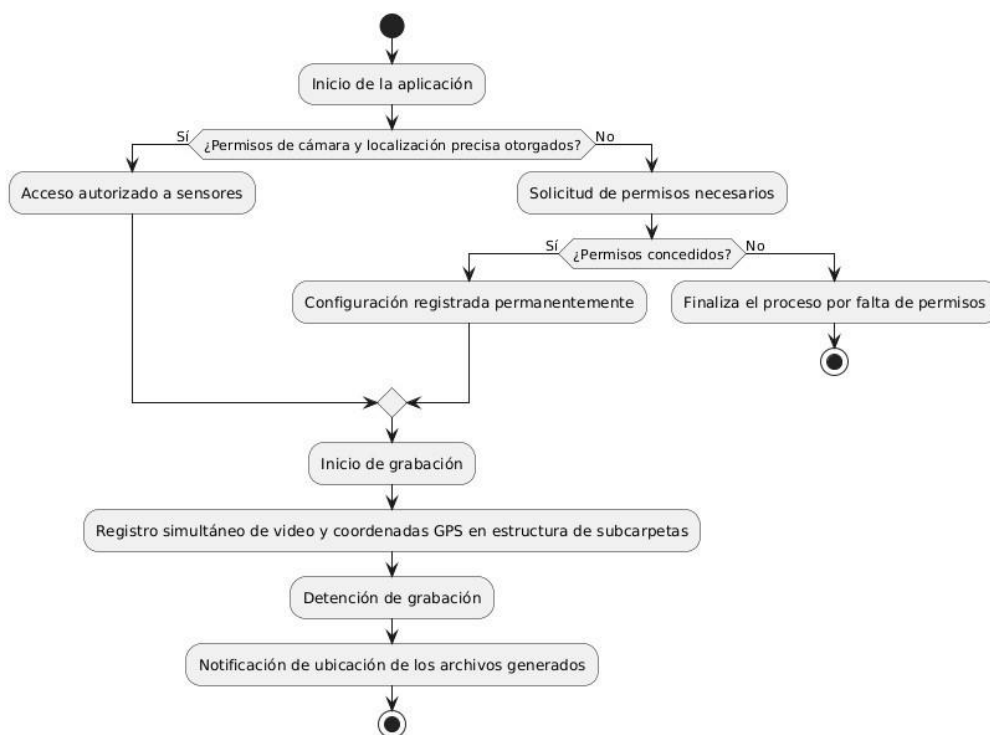
Para el correcto funcionamiento, se requiere otorgar permisos de acceso a la cámara y a la localización precisa, los cuales se configuran una única vez al inicio del proceso. La interfaz de la aplicación se plantea con un diseño simple e intuitivo, priorizando la operatividad en campo y reduciendo la posibilidad de errores durante la recolección.

Los datos adquiridos se guardan de manera ordenada, dividiendo los archivos de video de los registros de localización para conservar la trazabilidad y hacer más sencillo su procesamiento posterior dentro del flujo general del sistema. De igual manera, la estructura de almacenamiento tiene en cuenta la creación de subcarpetas con nombres descriptivos para cada recorrido, lo que facilita una clasificación más precisa y un acceso más rápido a la información obtenida. En la

figura 2 se muestra el diagrama de flujo operativo para l aplicación desarrollada en Android para la captura de datos.

Figura 2

Diagrama de Flujo de la Aplicación



Nota. Elaboración propia

Segmentación en Clips por Umbral de Desplazamiento

Dado que durante la captura de video se registra información visual de forma continua, es posible que múltiples fotogramas consecutivos representen una misma porción del entorno cuando el vehículo se desplaza lentamente o se detiene. Esta redundancia de información puede afectar la eficiencia del procesamiento y almacenamiento de los datos. Para mitigar este efecto, se implementa un proceso de segmentación del video en clips definidos por un umbral de distancia de desplazamiento preestablecido.

Para estimar el tamaño de los clips se debe establecer la velocidad de desplazamiento del

vehículo, la cual se obtiene a partir de los datos de georeferenciación obtenidos junto con la traza de tiempo.

Sean

- **V** : la velocidad de desplazamiento del vehículo en (m/s)
- **D** : el umbral de distancia en metros para reducir la redundancia en la captura.
- **T_m** : el periodo de muestreo que define el tamaño del clip de video
- **V_c** : la velocidad de captura de la cámara (fps).

El tiempo requerido para alcanzar el mínimo desplazamiento establecido por el umbral de distancia se obtiene de la ecuación (1).

$$T_m = \frac{D(m)}{V\left(\frac{m}{s}\right)} \quad (1)$$

La compresión o reducción (R) de *frames* para almacenamiento y procesamiento estará dada por:

$$R = 100 \left(1 - \frac{1}{T_m V_c} \right) \quad (2)$$

El fundamento de esta metodología es la relación entre el tiempo que dura el clip y la distancia recorrida, lo que significa que la duración de cada sección está directamente relacionada con la velocidad del vehículo. Cuando el sistema detecta, a partir de los datos GPS, que el desplazamiento ha alcanzado el umbral de un metro, se considera completado un tramo de captura y se genera un nuevo clip de video. Por lo tanto, los videos se dividen en segmentos (Clips) que representan áreas específicas del trayecto, manteniendo así una correspondencia precisa entre la información visual y el recorrido geográfico.

Esta etapa garantiza que cada clip cubra una sección única de la vía, manteniendo la

continuidad espacial sin dejar tramos significativos sin inspeccionar mientras reduce la redundancia de datos.

Selección Automática de Frames

El objetivo de la fase de selección automática de frames es determinar, entre cada clip generado, qué imagen representa mejor la calidad visual del segmento. Para ello, se considera la aplicación de métricas objetivas que permiten evaluar distintos aspectos de la nitidez, la información y la percepción visual de cada fotograma. Para el trabajo desarrollado se emplearon tres indicadores independientes:

- **Sharpness** (nitidez) que mide el nivel de detalle y enfoque de la imagen.

Es calculado a partir de una variante del filtro Laplaciano aplicada a cada *frame*.

- **Entropy** (Entropía) obtenida con la entropía de Shannon del canal de luminancia, la cual refleja la cantidad de información contenida en la imagen.

- **HyperIQA** modelo *Deep learning* que estima la calidad percibida de manera más cercana a la evaluación humana. Este modelo presenta la ventaja de no requerir una imagen de referencia adicional.

Cada métrica se normaliza al rango [0,1] y se combina en un puntaje ponderado según la siguiente distribución:

$$Score = 0.4xSharpness + 0.2 * Entropy + 0.4xHyperIQA \quad (3)$$

Los pesos se asignaron considerando la importancia relativa de cada métrica para la inspección vial: Sharpness y HyperIQA capturan la nitidez y la calidad perceptual de la imagen, aspectos fundamentales para la detección precisa de defectos, mientras que Entropy aporta información complementaria sobre el contenido visual. Los valores de ponderación se determinaron manualmente mediante un análisis cualitativo basado en la inspección visual de

imágenes de prueba.

La combinación de estas tres medidas permite obtener un análisis balanceado que incluye tanto parámetros objetivos como subjetivos.

Cada frame dentro de un clip recibe un puntaje calculado a partir de dichas métricas por medio de la ecuación (3), y aquel con la puntuación más alta se designa como el frame representativo del segmento. Este procedimiento busca reducir la redundancia visual y optimizar la cantidad de información relevante que se conserva, asegurando que los resultados sean consistentes y adecuados para el posterior procesamiento y visualización.

Corrección de Distorsión por Movimiento

Dada la naturaleza de la metodología de adquisición de datos desde un vehículo, que implica la captura de video en movimiento, es común que parte de los fotogramas presenten desenfoque o pérdida de detalle debido a vibraciones o cambios bruscos en la cámara. En esta fase, los frames seleccionados tras la evaluación de calidad son sometidos a el proceso de mejora visual mediante la aplicación de un modelo *Deep Learning* pre-entrenado denominado *NAFNet* descargado del repositorio oficial de *BasicSR*.

Este modelo se encuentra en el estado del arte y ha demostrado un desempeño sólido en la restauración de imágenes afectadas por desenfoque por movimiento. Al integrar esta etapa dentro del flujo general del sistema como un paso intermedio entre la selección automática de frames y las fases de análisis o representación final, se logra garantizar una calidad visual más estable y adecuada para las siguientes fases del proyecto. De esta forma, los resultados obtenidos reflejan de manera más precisa las condiciones reales de los trayectos y contribuyen a un análisis más confiable.

Etapa de Pruebas y Validación

En esta etapa se describen los mecanismos utilizados para verificar el funcionamiento adecuado de cada uno de los componentes desarrollados. Se presenta el proceso de integración de las distintas fases dentro de la interfaz de usuario implementada, así como su modo de operación. Finalmente, se exponen los resultados obtenidos al aplicar la herramienta en un caso práctico, con el propósito de evidenciar tanto las ventajas de la metodología propuesta en la reducción de datos a procesar como su impacto en el desempeño de la detección automática de defectos.

Implementación de la Arquitectura Propuesta

Para garantizar la usabilidad práctica de la herramienta desarrollada, se implementó una interfaz de usuario que permite ejecutar de forma automática las diferentes etapas de la metodología propuesta. Dicha interfaz fue desarrollada en Python, haciendo uso de los siguientes recursos:

- PyTorch (2.8.0+cu128): Ejecución del modelo HyperIQA, NafNET y redes definidas en *models.py*.
- Torchvision (0.23.0+cu128): Transformaciones y normalización de imágenes.
- NumPy (2.2.6): Cálculos numéricos y manejo de arreglos multidimensionales.
- OpenCV (4.12.0.88): Lectura, escritura y procesamiento de video e imágenes; cálculo de nitidez mediante Laplaciano.
- Pillow (11.3.0): Conversión y manipulación de imágenes entre formatos.
- Scikit-Image (0.25.2): Cálculo de entropía mediante *Shannon entropy*.
- PyQt5 (5.15.11): Creación de la interfaz gráfica (ventanas, botones, diálogos).
- PyQtWebEngine (5.15.7 / 5.15.2): Visualización embebida de mapas HTML.
- Folium (0.20.0): Generación de mapas interactivos con marcadores y recorridos.
- Geopy (2.4.1): Cálculo de distancias y velocidades geodésicas.
- B branca, Jinja2, xyzservices: Dependencias para renderizado de mapas.
- ImageIO: Soporte de entrada/salida de imágenes y video.
- QThread (PyQt5): Procesamiento en segundo plano sin bloquear la GUI.
- Ultralytics(8.3.227): Necesaria para utilizar modelos YoloV12
- Formatos:
 - Video: *.webm*, *.avi* (códecs *XVID*, *mp4v*).

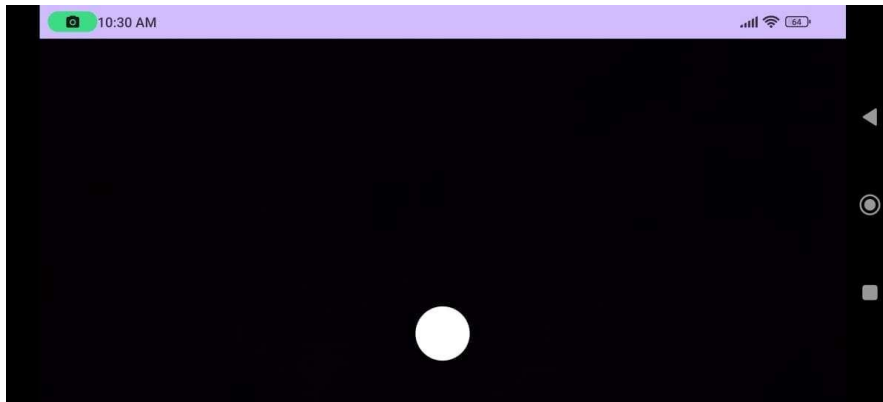
- Imágenes: *.jpg*, *.webp*.
- Datos: *.txt*, *.json*, *.html*.
- Técnicas IQA: Combinación de métricas de nitidez, contraste RMS, entropía y puntuación HyperIQA.

A continuación, se describe cada módulo de la interfaz, incluyendo la interfaz desarrollada en Android para la captura de datos.

Aplicación para Adquisición de Datos

Figura 3

Interfaz de la Aplicación Android



Nota. Elaboración propia

En la Figura 3 se presenta la interfaz principal de la aplicación diseñada específicamente para realizar la captura sincronizada de video y datos GPS en tiempo real. El primer momento en el que se inicia la aplicación el sistema solicita al usuario la autorización para acceder a la cámara y a la ubicación precisa del dispositivo, permisos que deben ser otorgados una sola vez y que son esenciales para garantizar la correcta vinculación entre la información visual y la geoespacial. La precisión de la localización es un requisito fundamental, ya que permite asociar cada fotograma de video con sus coordenadas geográficas exactas, asegurando así la fidelidad de

los datos recolectados.

El sistema tiene un solo botón central, cuyo propósito es comenzar y finalizar la grabación del video, siguiendo el comportamiento habitual de la cámara del dispositivo. Esta selección de diseño tiene como objetivo disminuir la complejidad operativa y minimizar las equivocaciones durante el proceso de captura en campo.

En relación con el almacenamiento de datos, la aplicación implementa una estructura jerárquica cuidadosamente diseñada dentro del directorio `Android/data/com.example.cameraFr`, con el propósito de mantener un control organizado y trazable de las sesiones de registro. En este directorio se generan dos carpetas principales: *Movies* y *Documents*. La primera contiene los archivos de video producidos en formato `.mp4`, mientras que la segunda almacena los archivos de texto `.txt` que registran las coordenadas geográficas obtenidas durante la grabación.

Para cada sesión de captura, el sistema crea de manera automática una subcarpeta identificada con el nombre del recorrido, lo que permite una correspondencia directa entre los videos y los datos de localización.

Interfaz para la Visualización y Gestión de Datos Geoespaciales y Visuales

La interfaz gráfica desarrollada para PC, representa el componente de interacción directa entre el usuario y el sistema de análisis geográfico y visual. Su diseño se centra en la integración de los procesos para capturar, procesar y visualizar datos dentro de un ambiente funcional y unificado. Con esta interfaz se puede monitorear cada etapa del flujo operativo, desde que se cargan los datos hasta que se muestran los resultados finales. Esto simplifica la gestión total de las rutas registradas y de los componentes visuales vinculados.

Figura 4

Interfaz de Visualización y Gestión de Datos Geoespaciales y Visuales



Nota. Elaboración propia

En la parte superior se dispone una barra de herramientas que concentra las funciones principales del sistema. Entre ellas se encuentran la importación de recorridos, la actualización del entorno de trabajo, la extracción de frames y la opción de limpiar los registros cargados. Cada componente de esta barra está conectado con una fase concreta del flujo de procesamiento, lo que facilita la preservación de un orden lógico y funcional en la interfaz.

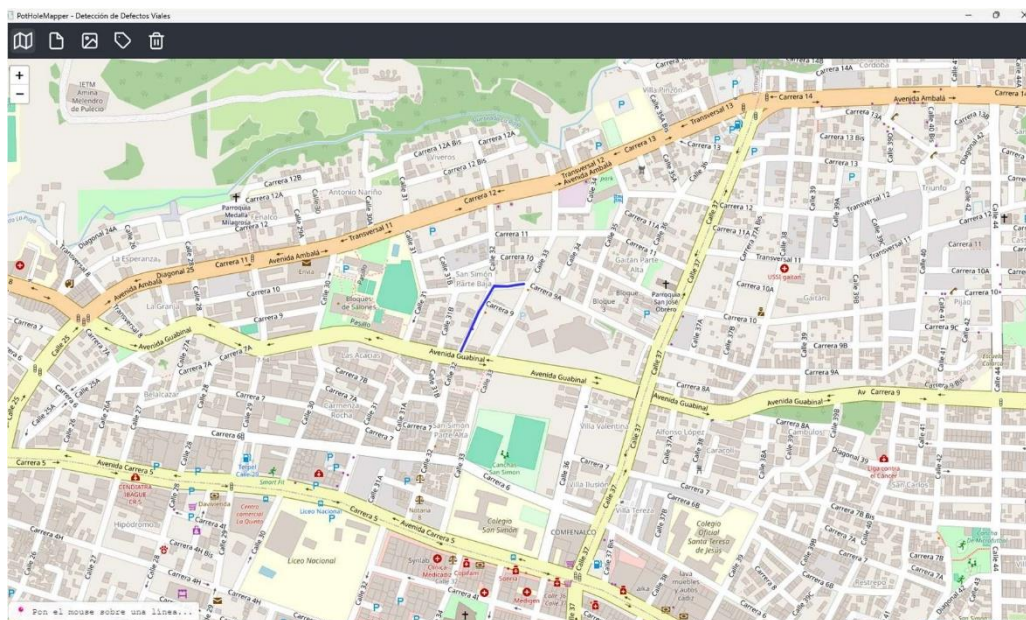
El mapa interactivo, que tiene la responsabilidad de mostrar visualmente los datos geoespaciales procesados, constituye la parte central del sistema. El sistema se mantiene en espera hasta que el usuario lleva a cabo la importación de los datos, en su estado inicial. Para comenzar el procedimiento, el usuario escoge la opción de "Importar datos" en la barra de herramientas superior. Cuando se realiza esto, aparece un cuadro de diálogo que facilita la selección simultánea del archivo de video y el archivo de coordenadas (.txt) que se produjeron en la fase de captura.

Una vez confirmada la selección, el sistema carga automáticamente ambos archivos, generando la representación del recorrido en el mapa. En esta visualización, se representan

la trayectoria que muestra los segmentos de datos registrados y procesados, estableciendo de este modo la correspondencia espacial del trayecto realizado (véase la Figura 5).

Figura 5

Previsualización de Recorrido Cargado en el Mapa

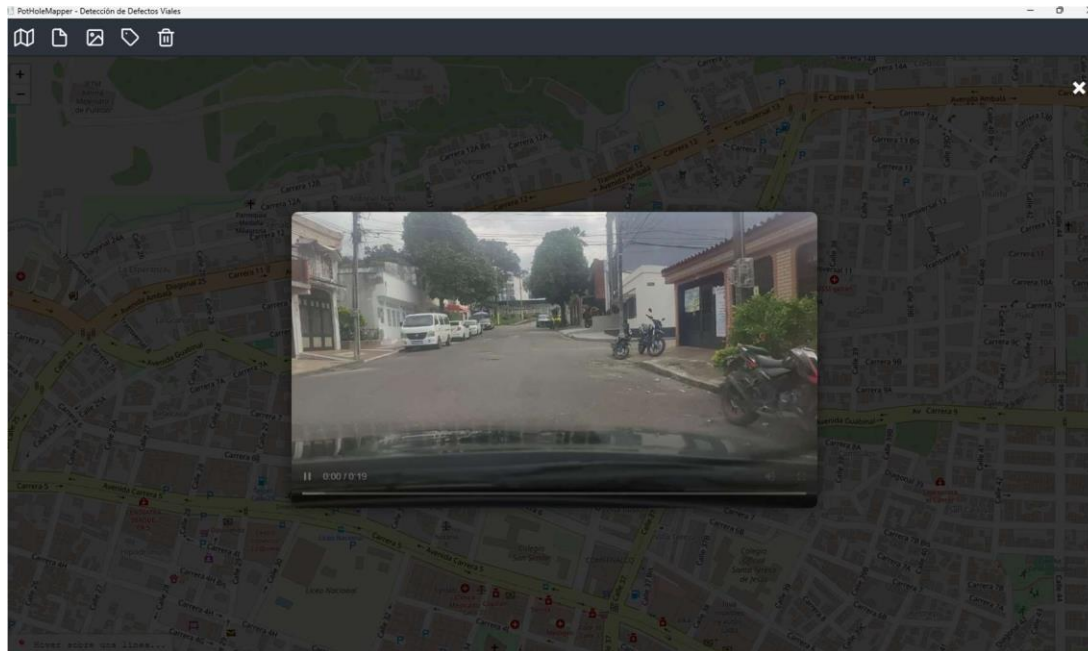


Nota. Elaboración propia

Tras la carga del trayecto, la interfaz habilita la función de previsualización multimedia asociada al recorrido realizado. Al seleccionar la línea de la trayectoria, se despliega una ventana emergente que reproduce el video correspondiente (Figura 6), lo que permite verificar de forma simultánea la coherencia entre el desplazamiento georreferenciado y la información visual capturada durante el registro de datos.

Figura 6

Visualización Multimedia del Recorrido



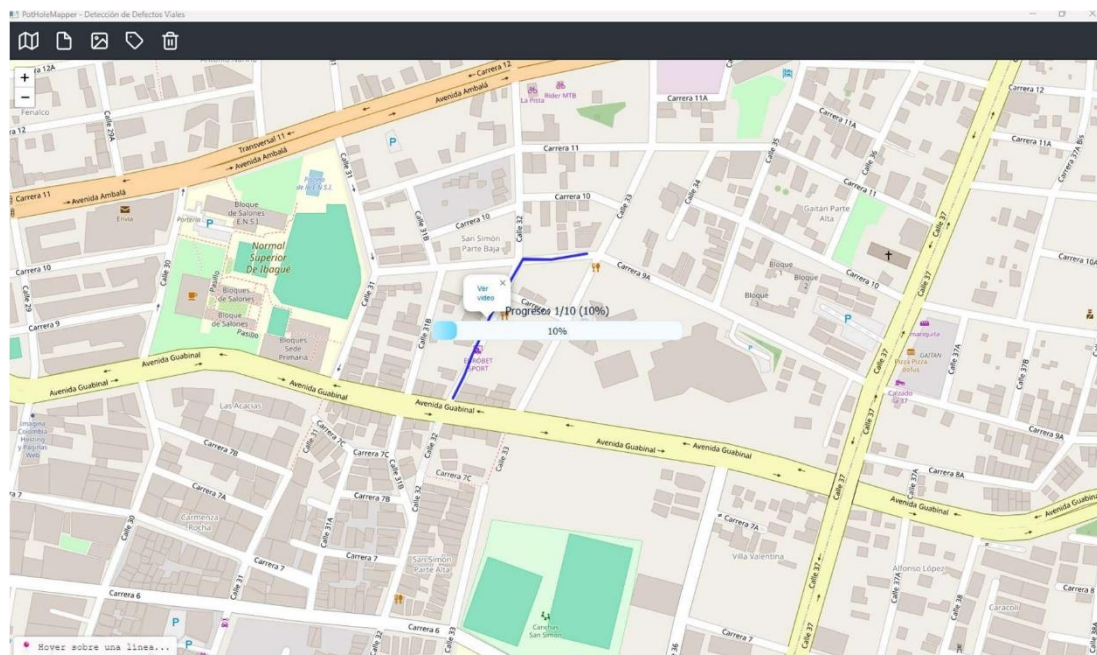
Nota. Elaboración propia

Una vez validada la correspondencia entre la trayectoria y el video, el usuario puede iniciar la extracción de frames representativos mediante la herramienta correspondiente en la barra superior. Al seleccionar esta opción, el sistema ejecuta un proceso automático que analiza el video cargado y selecciona los fotogramas más relevantes según criterios de calidad, iluminación y estabilidad visual.

Durante este procedimiento, la interfaz muestra una barra de progreso (Figura 7) que indica el estado de la extracción y permite al usuario monitorear la evolución del proceso en tiempo real. Este componente visual contribuye a mantener la trazabilidad de la operación y garantiza que el usuario conozca en qué etapa del análisis se encuentra el sistema.

Figura 7

Barra de Progreso de Extracción de Frames Representativos

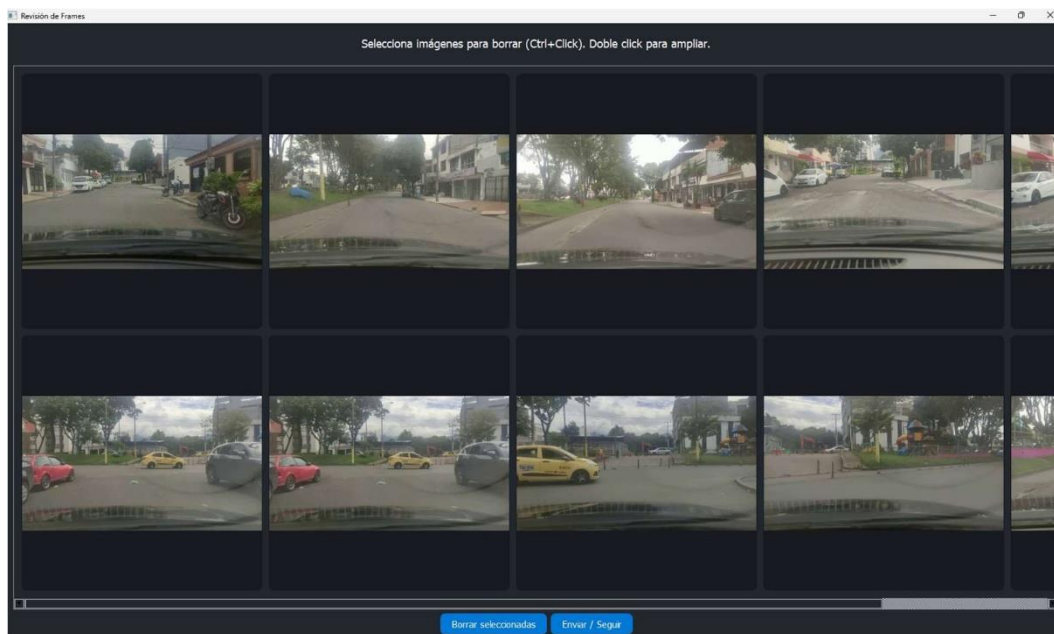


Nota. Elaboración propia

Una vez finalizada la extracción, el sistema despliega la ventana emergente de selección de frames (Figura 7), en la cual se presentan los fotogramas más representativos obtenidos de cada clip segmentado. Esta interfaz permite al usuario revisar visualmente los resultados del proceso automático y determinar, de forma informada, qué imágenes mantener o eliminar. Por lo tanto, actúa como un espacio de control y revisión de calidad. Este sistema asegura un balance entre la intervención humana y la automatización, garantizando que los datos visuales elegidos se ajusten no solo a criterios computacionales, sino también a la valoración contextual del operador.

Figura 8

Ventana Emergente para la Selección de Frames



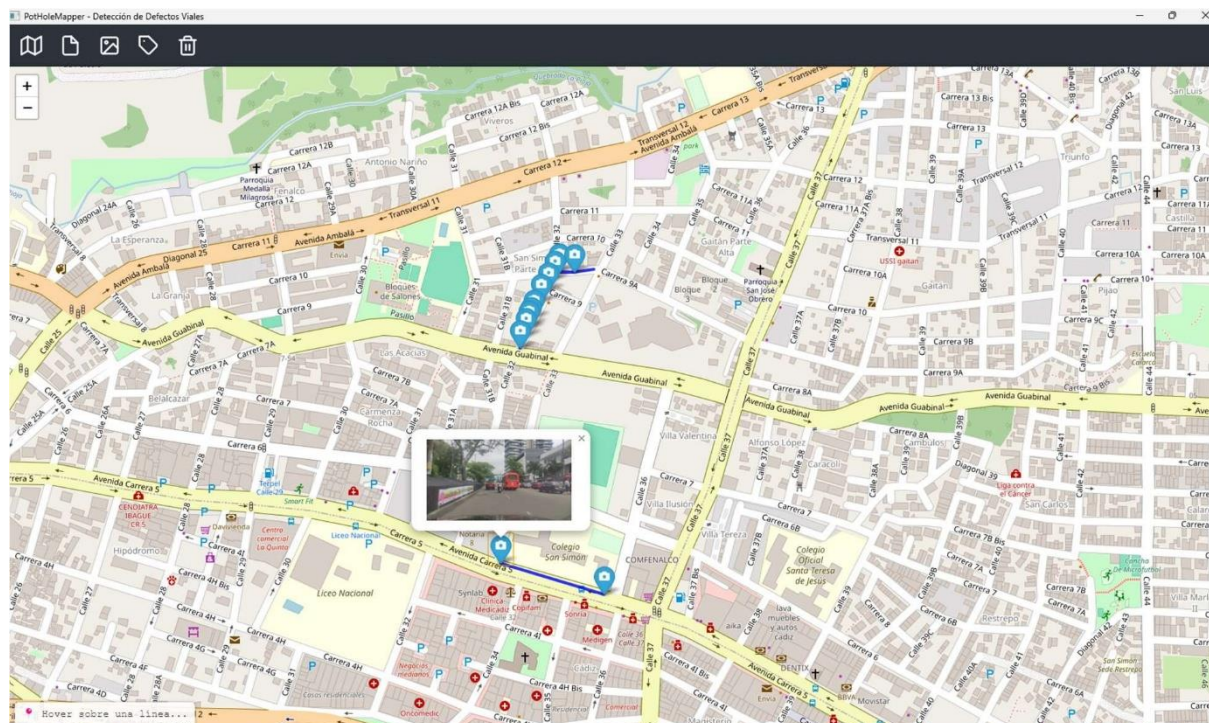
Nota. Elaboración propia

Cuando se termina la validación manual, los fotogramas que han sido aprobados se mandan directamente al módulo de mejora visual NAFNet. Allí, se llevan a cabo el procedimiento de restauración con el objetivo de eliminar el desenfoque por movimiento y recuperar detalles finos que se hayan perdido en la captura.

Después del proceso de mejora visual, las imágenes optimizadas se incorporan al contexto de la interfaz. En este lugar, cada punto que aparece en el mapa simboliza un fotograma procesado y elegido (Figura 9). Esta disposición espacial posibilita una interpretación exacta del trayecto, haciendo más fácil la correlación entre los segmentos obtenidos y su localización geográfica real. Cada punto funciona como un nodo de información visual, desde el que se puede desplegar una vista previa de la imagen asociada y, si es necesario, ampliar su visualización para examinarla con más detalle.

Figura 9

Visualización de los Puntos Marcados en el Mapa



Nota. Elaboración propia

Adicionalmente, en la esquina inferior izquierda de la interfaz visual se incluye una leyenda informativa que presenta las coordenadas geográficas (latitud y longitud) y velocidad en ese tramo del lugar donde el cursor está ubicado.

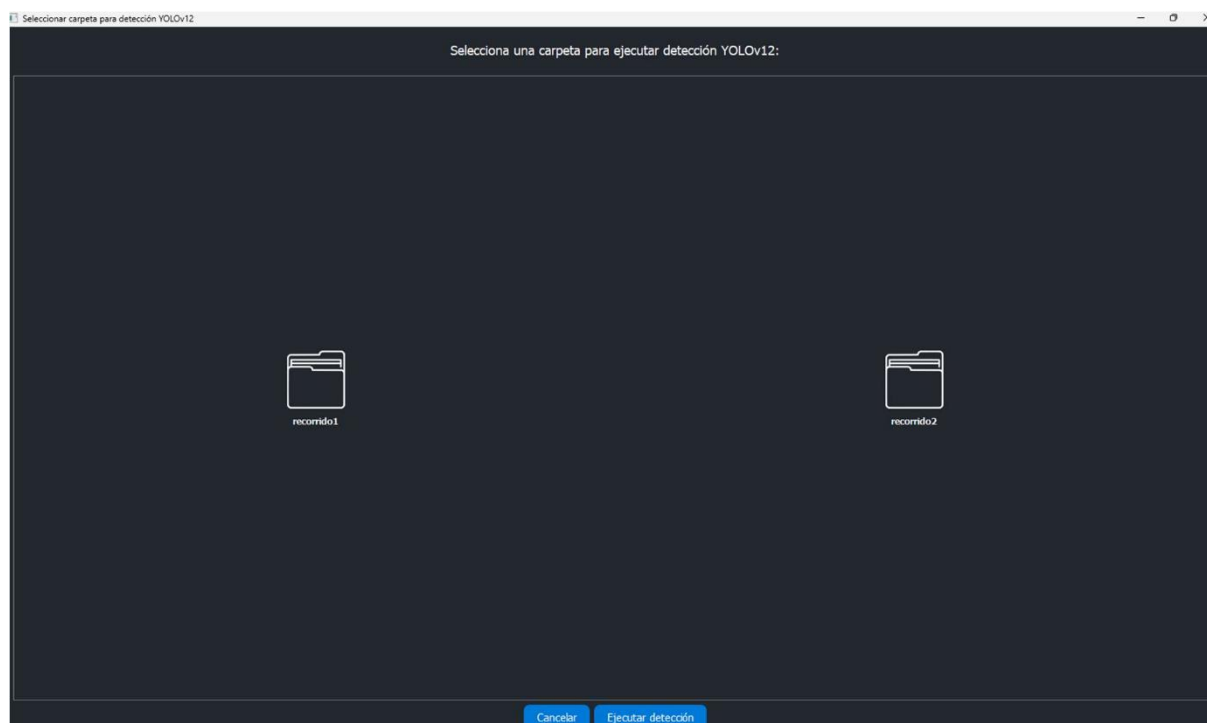
En una etapa posterior, el sistema incorpora un módulo destinado al análisis automatizado de los defectos viales presentes en las imágenes ya procesadas. Esta funcionalidad permite la integración de modelos de detección preentrenados basados en YOLO, brindando la posibilidad de comparar distintos modelos y determinar cuál se ajusta con mayor precisión a las características específicas del entorno de las imágenes.

En la barra de herramientas de la interfaz principal se dispone una opción que permite cargar el modelo de detección previamente entrenado, al seleccionarlo, el sistema despliega una

ventana emergente que organiza los distintos recorridos registrados en carpetas independientes, cada una correspondiente a un conjunto de imágenes optimizadas tras el proceso de mejora visual con NAFNet (Figura 10). Una vez que se ha determinado el trayecto a examinar, el modelo lleva a cabo la identificación automática de defectos viales en las imágenes elegidas, creando nuevas versiones anotadas que sustituyen a las previas.

Figura 10

Ventana Emergente para el Procesamiento Mediante el Modelo YOLO.

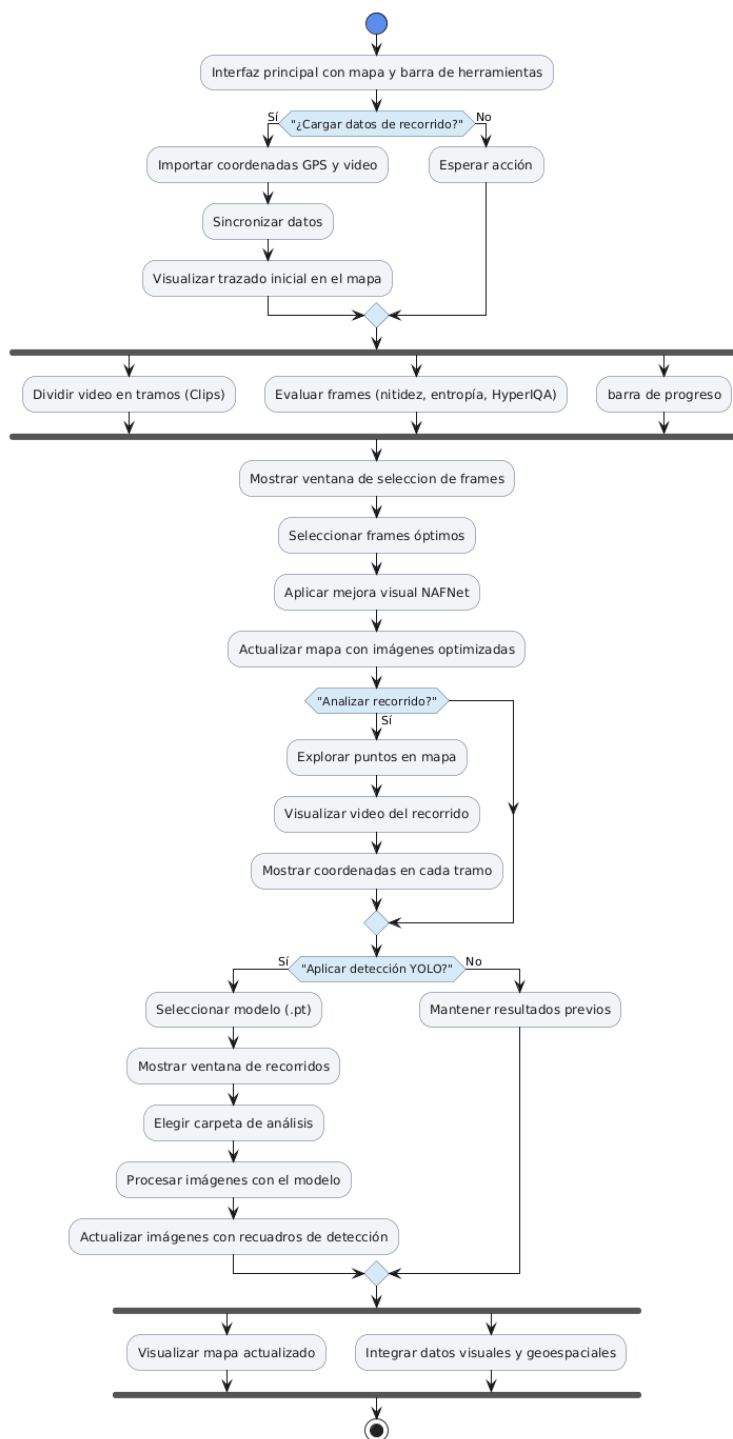


Nota. Elaboración propia

En la Figura 11 se muestra el diagrama de flujo que resume la operatividad de la plataforma desarrollada.

Figura 11

Diagrama de Flujo Operacional



Nota. Elaboración propia

Pruebas Experimentales y Resultados Obtenidos

Con el objetivo de evaluar el funcionamiento y la usabilidad de la implementación desarrollada, se llevó a cabo un caso práctico completo orientado a la generación de un dataset de prueba para la evaluación de defectos en vías vehiculares. En esta sección se describen los detalles específicos correspondientes a la metodología propuesta y se presentan los resultados obtenidos en cada etapa del proceso. Es importante mencionar que estos resultados fueron presentados por los autores de este trabajo en el *19 Congreso Colombiano de Computación*.

Descripción del Entorno de Desarrollo

El entorno de desarrollo empleado para la implementación del sistema se configuró sobre una estación de trabajo con un procesador AMD Ryzen 5 5600X, 16 GB de memoria RAM y una unidad de almacenamiento SSD M.2 de 500 GB. Adicionalmente, para la ejecución de los modelos de aprendizaje profundo y las tareas de mejora visual, se utilizó una GPU NVIDIA GeForce RTX 3060.

El sistema fue desarrollado en un ambiente basado en Python y PyTorch como marco principal para implementar y ejecutar los modelos de aprendizaje profundo. El rendimiento estable durante la implementación de todo el sistema fue asegurado por este entorno, que permitió una integración eficaz de los procesos de optimización.

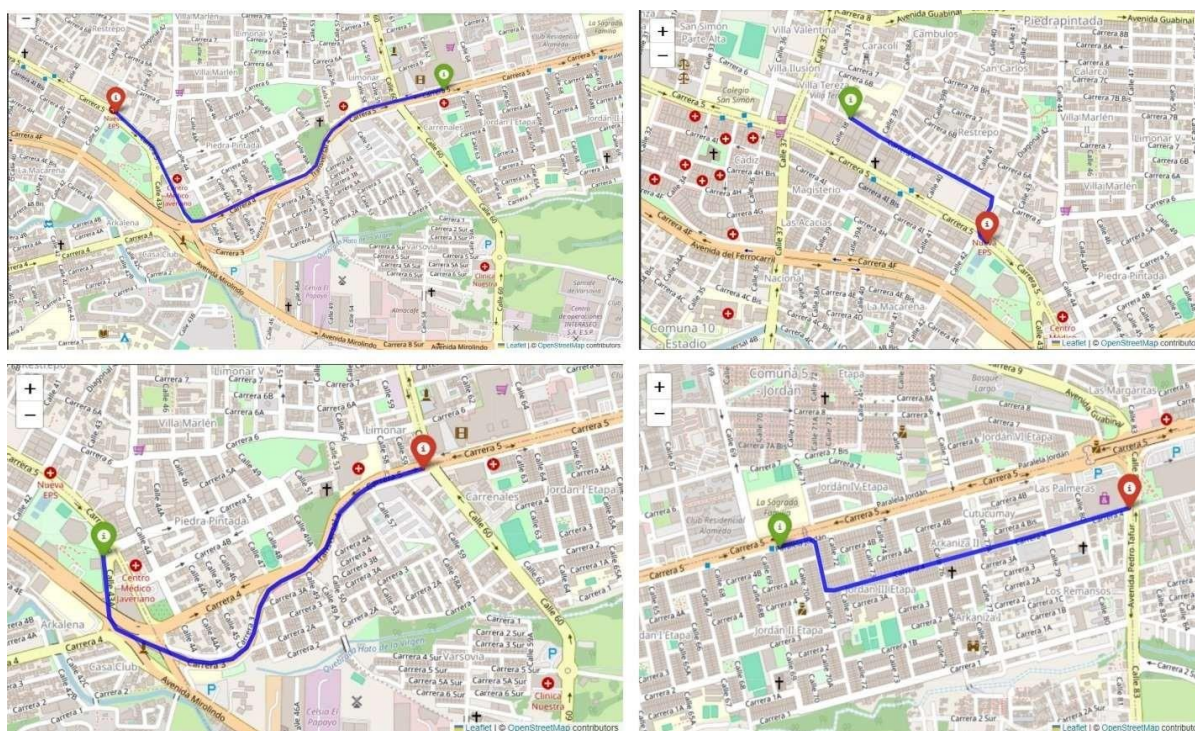
Visualización Geográfica

Para validar el desempeño del sistema propuesto, se realizaron pruebas en cuatro rutas urbanas con duraciones entre 7 y 10 minutos. Las rutas incluyeron segmentos mixtos, como superficies pavimentadas, baches visibles y áreas previamente reparadas. Este tipo de escenario permitió evaluar la capacidad del sistema para operar bajo condiciones reales de inspección vial.

Como parte del proceso automatizado, los datos de georreferenciación recopilados durante cada ruta fueron utilizados para generar una representación espacial precisa de la trayectoria. La Figura 12 muestra la visualización generada sobre un mapa correspondiente a las cuatro rutas registradas. Esta funcionalidad permite asociar cada imagen seleccionada con su ubicación exacta, lo que facilita futuros análisis geospaciales y el reporte automatizado de anomalías en la vía.

Figura 12

Visualización de las Rutas Registradas durante la Etapa de Captura de Datos.



Nota. Elaboración propia

Reducción de Datos

Uno de los objetivos principales del sistema es reducir la cantidad de fotogramas redundantes sin comprometer la cobertura espacial. Los resultados muestran una disminución significativa en el volumen total de imágenes procesadas. La Tabla 1 presenta un resumen que

compara el número de fotogramas originales con el número de fotogramas seleccionados como representativos después de aplicar el proceso de segmentación y selección automática basado en calidad visual. Estos resultados indican que, en promedio, más del 96% de los fotogramas originales son descartados, lo que representa una mejora sustancial en términos de eficiencia de almacenamiento y postprocesamiento..

Tabla 1

Reducción Alcanzada Después del Proceso de Selección Automática

Trayecto	No Frames originales	No Frames seleccionado	Reducción
1	38100	1560	95.91%
2	23400	630	97.30%
3	29400	1307	95.55%
4	38400	1153	97.00%

Mejora Visual de las Imágenes

Además de la reducción del dataset, se llevó a cabo una evaluación cuantitativa para analizar la mejora en la calidad visual de las imágenes seleccionadas. Para ello, se calcularon métricas objetivas de calidad de imagen antes y después del proceso de selección y corrección utilizando el modelo NAFNet.

La Figura 13 presenta ejemplos de fotogramas afectados por distorsión por movimiento, junto con sus versiones corregidas tras aplicar NAFNet. Se observa una mejora clara en la nitidez y en la definición de los detalles.

Figura 13

Ejemplos de la Corrección de Distorsión por Movimiento



Nota. En el lado izquierdo se muestran dos fotogramas con distorsión por movimiento. En el lado derecho, las imágenes correspondientes después de aplicar NAFNet para reducir dicha distorsión. Elaboración propia

La Figura 14 muestra tres curvas que ilustran el comportamiento del indicador de calidad visual a lo largo de las diferentes etapas del proceso:

- **Curva roja:** valor promedio de calidad visual en cada clip original.
- **Curva azul:** calidad de los fotogramas seleccionados automáticamente.
- **Curva verde:** calidad de los fotogramas después de aplicar el modelo NAFNet.

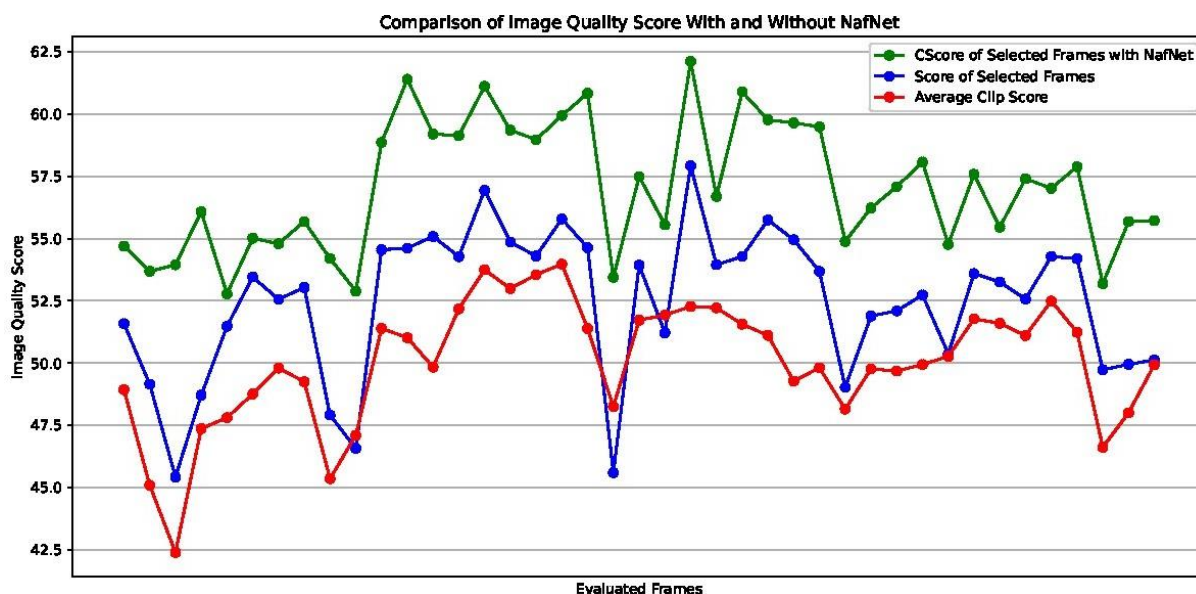
Los porcentajes promedio de mejora obtenidos fueron los siguientes:

- Imagen seleccionada vs. promedio del clip: **+4.86%**. Imagen con NAFNet vs. promedio del clip: **+14.14%**.
- Imagen con NAFNet vs. imagen seleccionada sin procesar: **+9.28%**.

Estos valores demuestran la efectividad tanto del algoritmo de selección como del modelo de Deep Learning en mejorar la calidad perceptual y cuantitativa de las imágenes.

Figura 14

Comparación de Evaluación de Calidad



Nota. El eje vertical representa el valor del indicador de calidad visual. El eje horizontal muestra, en rojo, el valor promedio de las imágenes dentro de cada clip; en azul, el valor de los fotogramas seleccionados; y en verde, el valor de las imágenes procesadas para eliminar la distorsión por movimiento. Elaboración propia

Impacto Sobre la Detección Automática de Defectos

Para evaluar la efectividad de la metodología propuesta para la detección automatizada de defectos viales, se realizaron experimentos adicionales utilizando videos capturados desde un vehículo en movimiento bajo condiciones nocturnas y con baja iluminación. A partir de estas

grabaciones, se generó un conjunto de datos siguiendo la canalización propuesta, compuesto por 681 imágenes que contienen 1.098 defectos de interés anotados. Este conjunto de datos se dividió en tres subconjuntos: 70% para entrenamiento, 10% para validación y 20% para pruebas. Para el conjunto de prueba, se seleccionaron específicamente imágenes con el mayor grado de distorsión por movimiento, con el fin de evaluar el impacto de la etapa de corrección.

Inicialmente, el modelo YOLOv12 fue entrenado utilizando las imágenes originales sin procesar y evaluado sobre el conjunto de prueba. En una segunda etapa, se creó un conjunto de datos duplicado con la misma estructura e imágenes que el original, pero con todos los fotogramas preprocesados mediante el modelo NAFNet para la corrección de distorsión por movimiento. Luego, el modelo fue nuevamente entrenado con este conjunto de datos procesado, y el rendimiento sobre el conjunto de prueba se comparó con el escenario sin procesamiento, lo que permitió cuantificar el impacto de la mejora visual en la detección de defectos viales.

La Tabla 2 resume los resultados obtenidos sobre el subconjunto de prueba, considerando un Ground Truth de 229 defectos anotados. En el primer escenario, utilizando las imágenes originales sin preprocesamiento, el modelo YOLOv12 detectó correctamente 153 defectos (66.80%). En el segundo escenario, donde las mismas imágenes fueron procesadas con el modelo NAFNet para la corrección de distorsión por movimiento, el número de defectos detectados correctamente aumentó a 200 (87.73%). Esto representa una mejora absoluta de 20.93%, destacando la efectividad de incorporar técnicas de mejora visual en la etapa de preprocesamiento para la detección automatizada de defectos en carreteras. En la figura 15 se muestran algunos ejemplos de defectos que fueron exitosamente detectados después de la etapa de mejoramiento de imágenes.

Tabla 2*Desempeño de Detección con y sin Corrección de Distorsión*

Condición	Tasa de Detección (%)	Verdaderos Positivos	Falsos Positivos
Sin Preprocesamiento	66.80	153/229	2
Mejorada con NafNet	87.73	200/229	0

Es importante destacar que en ambos escenarios se observó una tasa muy baja de falsos positivos. En la primera condición, utilizando las imágenes originales sin preprocesamiento, solo se reportaron dos detecciones incorrectas en el conjunto de prueba. En contraste, cuando se aplicó la corrección de distorsión por movimiento con NAFNet, no se registraron falsos positivos.

Además, aunque se generaron algunas cajas delimitadoras en regiones no incluidas en las anotaciones originales, la inspección manual confirmó que correspondían a defectos reales de la vía, lo cual indica la capacidad del modelo para detectar anomalías previamente no etiquetadas.

Figura 15*Ejemplos de Detección Automática de Defectos*

Nota. Al lado Izquierdo: imágenes sin corrección de distorsión. Derecha: imágenes después de aplicar la corrección de distorsión. Elaboración propia

El tiempo promedio de procesamiento por imagen fue analizado para las principales etapas del flujo propuesto, tal como se resume en la Tabla 3. El cálculo del puntaje de calidad requirió 0,0254 segundos por imagen, y el proceso de inferencia para la detección de defectos utilizando YOLOv12 tuvo un promedio de 7,3 milisegundos por imagen. Ambas etapas demuestran un desempeño adecuado para una implementación en tiempo real o casi en tiempo real. En contraste, la corrección de distorsión basada en NAFNet, con un promedio de 1,76 segundos por imagen, representa la etapa computacionalmente más exigente, lo que la hace más apropiada para procesamiento offline en lugar de ejecución en tiempo real.

Tabla 3

Promedio del Tiempo de Procesamiento por Imagen en Cada Etapa.

Etapa	Tiempo promedio por imagen (seg)
Computación del factor de calidad	0.0254
Corrección distorsión con NafNet	1.76
Inferencia detección de defectos YOLOv12	0.0073

Discusión de los Resultados

Los resultados demuestran la efectividad del sistema propuesto para generar, de manera automática, datasets de imágenes urbanas de alta calidad destinados a la inspección de vías mediante visión por computador. La metodología implementada reduce de forma significativa el número de fotogramas procesados, manteniendo una adecuada cobertura espacial y mejorando sustancialmente la calidad visual de las imágenes seleccionadas.

La etapa de segmentación del video basada en la distancia geográfica mostró un buen desempeño al evitar la redundancia visual sin comprometer la continuidad espacial. Al establecer un umbral de 1 metro entre clips consecutivos, se logró una reducción superior al 96% en el volumen total de datos. Esto representa una ventaja operativa notable, ya que disminuye los requerimientos de almacenamiento y el tiempo necesario para tareas posteriores como la anotación o el entrenamiento de modelos de aprendizaje automático.

El proceso de selección del fotograma representativo, basado en una combinación ponderada de métricas objetivas de calidad visual (nitidez, entropía y HyperIQA), resultó eficaz para identificar automáticamente la imagen más clara dentro de cada clip. En promedio, la imagen seleccionada mostró una mejora del 4.86% respecto al valor promedio del clip, lo que

valida la capacidad del algoritmo para discriminar entre imágenes borrosas o de bajo contraste y aquellas más adecuadas para su análisis posterior. La aplicación del modelo de Deep Learning NAFNet para la corrección de distorsión por movimiento fue clave para mejorar la calidad final del dataset generado. Tras el procesamiento, se observó una mejora del 14.14% respecto al promedio del clip y de 9.28% en comparación con la imagen original. Estos resultados destacan la importancia de incorporar técnicas avanzadas de restauración de imágenes en escenarios reales donde el movimiento del vehículo puede afectar la nitidez de los fotogramas capturados.

Desde una perspectiva práctica, el sistema propuesto reduce de manera significativa la intervención manual en la construcción del dataset, minimiza errores subjetivos y aumenta la consistencia de los datos utilizados para entrenar modelos de detección de defectos viales. Además, la integración del proceso de georreferenciación permite asociar cada imagen con su ubicación geográfica precisa, facilitando futuras aplicaciones en sistemas automatizados de reporte de anomalías en la vía.

Adicionalmente, se realizó una evaluación experimental para analizar el impacto de la corrección de distorsión por movimiento en el desempeño de la detección de defectos viales. Usando un conjunto de prueba con 229 defectos anotados, la tasa de detección aumentó de 66.80% (153 detecciones correctas) con imágenes sin procesar a 87.73% (200 detecciones correctas) tras aplicar el preprocesamiento basado en NAFNet, lo que representa una mejora absoluta de 20.93%. Asimismo, los falsos positivos se redujeron de dos a cero. Estos hallazgos confirman que la incorporación de técnicas de mejora visual en la etapa de preprocesamiento no solo optimiza la calidad de las imágenes, sino que también tiene un efecto directo y medible sobre el rendimiento de los modelos de detección subsiguientes, especialmente en condiciones desafiantes como desenfoque por movimiento y baja iluminación.

Conclusiones y Trabajos Futuros

La metodología propuesta ha demostrado ser efectiva para automatizar la generación de conjuntos de datos de imágenes urbanas de alta calidad, destinados al entrenamiento y evaluación de modelos de visión por computador para la inspección vial. El sistema integra técnicas de procesamiento de video, análisis de datos georreferenciados y modelos avanzados de Deep Learning, logrando una reducción superior al 95% en el volumen total de fotogramas, manteniendo al mismo tiempo una adecuada cobertura espacial y mejorando la calidad visual de las imágenes seleccionadas.

El uso combinado de métricas objetivas de calidad visual (nitidez, entropía y HyperIQA) permitió identificar los fotogramas más representativos dentro de cada clip, garantizando consistencia y minimizando la subjetividad típicamente asociada a los procesos manuales. Además, la aplicación del modelo NAFNet fue clave para corregir distorsiones relacionadas con el movimiento, obteniendo una mejora promedio en la calidad visual del 14,14% en comparación con el promedio del clip.

Más allá de la mejora en la calidad visual, la efectividad de este paso de preprocesamiento fue confirmada mediante un experimento adicional de detección, donde la aplicación de la corrección de movimiento condujo a un aumento absoluto del 20,93% en la tasa de detección y eliminó completamente los falsos positivos. Esto demuestra que la mejora de imágenes no solo beneficia la calidad del conjunto de datos, sino que también se traduce en ganancias tangibles en el rendimiento de los modelos posteriores de detección de defectos.

Estos resultados validan la viabilidad del sistema como una herramienta para estandarizar y optimizar la construcción de conjuntos de datos utilizados en el desarrollo de soluciones tecnológicas para la detección automatizada de defectos viales. Al automatizar toda la línea de

procesamiento, desde la captura del video hasta la selección y mejora de fotogramas, el enfoque propuesto reduce significativamente el tiempo y el esfuerzo humano requerido, disminuye los costos operativos y aumenta la trazabilidad de los datos utilizados.

Se proponen varias líneas de mejora para trabajos futuros. En primer lugar, podría explorarse la integración de filtros semánticos mediante la incorporación de módulos de detección preliminar de características específicas, como baches o grietas, con el fin de priorizar la selección de fotogramas relevantes. En segundo lugar, se podrían implementar estrategias de etiquetado semiautomático basadas en modelos preentrenados para acelerar el proceso de anotación de conjuntos de datos destinados al entrenamiento de redes neuronales. Finalmente, podría realizarse un análisis de sensibilidad para optimizar la distribución de pesos entre las métricas que componen el puntaje de calidad de imagen, garantizando un proceso de selección de fotogramas más robusto y eficaz.

Referencias Bibliográficas

- Arya, Deeksha, et al. "RDD2022: A multi-national image dataset for automatic road damage detection." *Geoscience Data Journal* 11.4 (2024): 846-862.
- Askarzadeh, T., Bridgelall, R., & Tolliver, D. D. (2025). Drones for road condition monitoring: applications and benefits. *Journal of Transportation Engineering, Part B: Pavements*, 151(1), 04024055.
- Bhavana, N., Kodabagi, M. M., Kumar, B. M., Ajay, P., Muthukumaran, N., & Ahilan, A. (2024). POT YOLO: Real Time Road Potholes Detection Using Edge Segmentation Based YOLO v8 Network. *IEEE Sensors Journal*, 24(15).
- Esaulov, V., & Esfahani, M. M. (2025). A Comparative Study of NAFNet Baselines for Image Restoration. *arXiv preprint arXiv:2506.19845*.
- Ettalibi, A., Elouadi, A., & Mansour, A. (2024). AI and computer vision-based real-time quality control: a review of industrial applications. *Procedia Computer Science*, 231, 212-220.
- Guan, S., Liu, H., Pourreza, H. R., & Mahyar, H. (2023). Deep learning approaches in pavement distress identification: A review. *arXiv preprint arXiv:2308.00828*.
- Ke, J., Wang, Q., Wang, Y., Milanfar, P., & Yang, F. (2021). Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 5148-5157).
- Ma, N., Fan, J., Wang, W., Wu, J., Jiang, Y., Xie, L., & Fan, R. (2022). Computer vision for road imaging and pothole detection: a state-of-the-art review of systems and algorithms. *Transportation safety and Environment*, 4(4), tdac026.
- Selvia Nafaa, H. Essam, K. Ashour, et al.: Automated Pavement Cracks Detection and Classification Using Deep Learning. *arXiv* (2024) [2024.07674]

- Su, Shaolin, et al. "Blindly assess image quality in the wild guided by a self-adaptive hyper network." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
- T. Men, B. Wang, N. Zhang, Q. Sun: Pavement Defect Detection Method Based on Deep Learning. In: Proc. CENet 2023, LNEE 1126, pp. 494–500 (2024)
- Wang, N., Gao, Z., Fang, H., Hu, H., Li, B., & Wang, C. (2025). Deep Learning Network Model for Road Pavement Damage Detection Based on YOLOv8s-GES. Transportation Research Record, 03611981241312224.
- Xu, X., Zhao, M., Shi, P., Ren, R., He, X., Wei, X., & Yang, H. (2022). Crack detection and comparison study based on faster R-CNN and mask R-CNN. Sensors, 22(3), 1215.
- Yuan, B., Sun, Z., Pei, L., Li, W., Ding, M., & Hao, X. (2022). Super-resolution reconstruction method of pavement crack images based on an improved generative adversarial network. Sensors, 22(23), 9092.