Doi: 10.62544/ucomscientia.v3i2.59
UCOM Scientia,
Vol. 3, N° 2, setiembre, 2025, 69-97
ISSN 3006-7278

Artículo Original/ Original Article

Segmentación automática de grietas en pavimento asfáltico mediante redes neuronales convolucionales U-Net

Automatic crack segmentation in asphalt pavement using U-Net convolutional neural networks

Luz Rashell Soto Olguín
[1]Universidad Nacional de Caaguazú, Facultad de Ciencias y Tecnologías, Grupo de Investigación en Ciencia de Datos. Coronel Oviedo, Paraguay.
https://orcid.org/0009-0001-3938-2902

Fredy Gabriel Ramirez Villanueva
[1]Universidad Nacional de Caaguazú, Facultad de Ciencias y Tecnologías, Grupo de Investigación en Ciencia de Datos. Coronel Oviedo, Paraguay.
https://orcid.org/0009-0000-9172-6496

Héctor Ramiro Estigarribia Barreto
[1]Universidad Nacional de Caaguazú, Facultad de Ciencias y Tecnologías, Grupo de Investigación en Ciencia de Datos. Coronel Oviedo, Paraguay.
https://orcid.org/0000-0002-2954-6053
Autor corresponsal: hestigarribia64@fctunca.edu.py

Para citar este artículo:

Soto Olguín, L. R., Ramirez Villanueva, F. G. y Estigarribia Barreto, H. R. (2025). Segmentación automática de grietas en pavimento asfáltico mediante redes neuronales convolucionales U-Net. *UCOM Scientia,3*(2), 69-97.

Fecha de recepción: 20/06/2025

Fecha de aceptación: 04/08/2025

Resumen

Este artículo presenta el desarrollo de un sistema de segmentación automática de grietas en pavimento asfáltico mediante redes neuronales convolucionales, utilizando la arquitectura U-Net. Para ello, se construyó un conjunto de datos propio compuesto por 847 imágenes capturadas en la ciudad de Coronel Oviedo, Paraguay, de las cuales 505 contenían grietas. Dichas imágenes fueron etiquetadas manualmente y procesadas a una resolución de 256×256 píxeles. El modelo se entrenó durante 500 épocas utilizando la función de pérdida binary crossentropy y el optimizador Adam. Los experimentos iniciales reportaron métricas de desempeño muy elevadas y se incorporó, además, un módulo de posprocesamiento para cuantificar el ancho de las grietas. Aunque estos resultados son preliminares, sugieren un desempeño prometedor del sistema desarrollado, que podría convertirse en una herramienta robusta y precisa para la gestión vial, siempre que se realicen validaciones adicionales bajo condiciones más diversas y exigentes.

Palabras clave: Redes neuronales convolucionales; visión artificial; pavimento asfáltico; segmentación de imágenes; mantenimiento vial.

Editora responsable: Dra. Carla Fernández
Carla.fernandez@ucom.edu.py

Abstract

This article presents the development of an automatic crack segmentation system for asphalt pavement using convolutional neural networks, specifically the U-Net architecture. To this end, a proprietary dataset was built consisting of 847 images captured in the city of Coronel Oviedo, Paraguay, of which 505 contained cracks. These images were manually annotated and processed at a resolution of 256×256 pixels. The model was trained for 500 epochs using the binary crossentropy loss function and the Adam optimizer. Initial experiments reported very high performance metrics, and a post-processing module was also incorporated to quantify crack width. Although these results are preliminary, they suggest a promising performance of the developed system, which could become a robust and accurate tool for road management, provided that further validations are conducted under more diverse and challenging conditions.

Keywords: Convolutional neural networks; computer vision; asphalt pavement; image segmentation; road maintenance.

1. Introducción

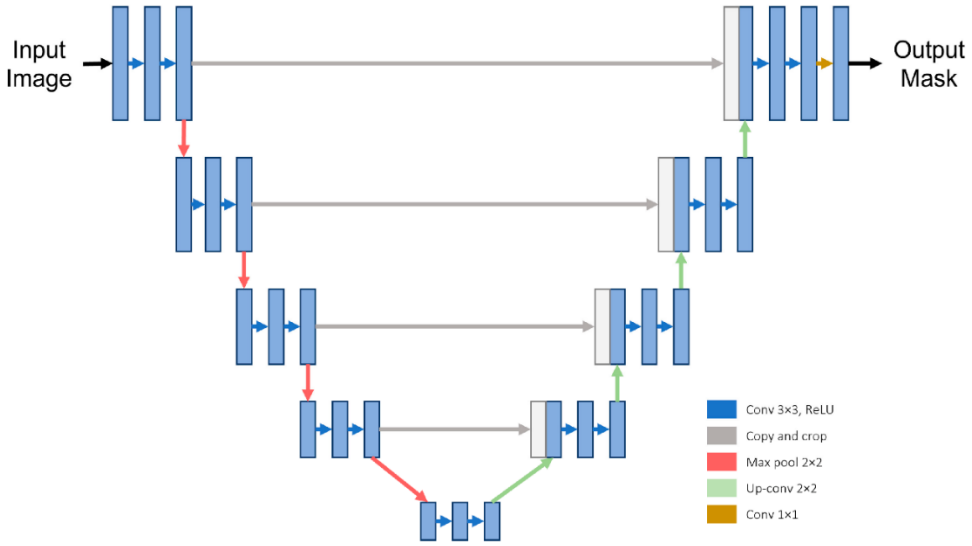
La infraestructura vial es un pilar esencial para el desarrollo económico y social, conectando mercados y comunidades. En América Latina y el Caribe existen más de 627.000 km de carreteras pavimentadas, que sirven a una población de más de 565 millones de personas (Global Infrastructure Hub, 2022). Mantener estas vías en buen estado es crítico: las grietas y deterioros en el pavimento reducen la calidad de la carretera y la seguridad vial (Di Benedetto et al., 2023). Estudios recientes indican que las malas condiciones de las vías generan costos indirectos enormes (por ejemplo, unos US$130.000 millones anuales en Estados Unidos por reparaciones adicionales de vehículos) y contribuyen a más de la mitad de los accidentes fatales (Astute Analytica, 2024). Por otra parte, la inversión temprana en conservación ofrece altos retornos: se ha estimado que cada dólar gastado en mantenimiento preventivo puede ahorrar entre cuatro y diez dólares en reparaciones futuras. Estos datos destacan la necesidad de detectar y corregir rápidamente los primeros indicios de falla en el pavimento, antes de que se conviertan en daños más costosos.

Las fisuras o grietas superficiales en pavimentos asfálticos son de las señales iniciales de deterioro. A nivel técnico, el ancho y la longitud de las grietas se consideran parámetros clave para evaluar su severidad, conforme a criterios de organismos como la AASHTO (American Association of State Highway and Transportation Officials). Sin embargo, en la práctica es común realizar inspecciones visuales: ingenieros o técnicos recorren las vías identificando grietas, marcándolas y midiendo sus dimensiones manualmente (Tello Cifuentes et al., 2021). Este proceso es lento, costoso y subjetivo. Las inspecciones manuales demandan gran cantidad de recursos humanos y tiempo, y sus resultados suelen ser inconsistentes entre diferentes evaluadores. En contraste, técnicas automáticas basadas en visión por computadora pueden

objetivar la detección y cuantificación de daños, liberando recursos para tareas de planificación estratégica. Por ejemplo: VIASegura, una herramienta del BID (Banco Interamericano de Desarrollo), redujo la inspección de 10.000 km de red vial de 18 meses (US$3,2M) a sólo 2 semanas (US$60.000) usando visión artificial y IA (Inteligencia Artificial), demostrando la eficiencia de estos enfoques (Cruz et al., 2024).

En este contexto, las técnicas de visión artificial y aprendizaje profundo han emergido como soluciones prometedoras para la evaluación de infraestructura vial. A diferencia de sistemas complejos 3D (como LiDAR), la adquisición de imágenes 2D es de bajo costo y permite procesar grandes volúmenes de datos rápidamente. Las CNN (redes neuronales convolucionales) se han destacado en tareas de procesamiento de imágenes porque aprenden jerarquías de características directamente de los datos, capturando patrones locales y globales de forma automática (Kaveh & Alhajj, 2024). Varios estudios recientes muestran que las CNN brindan resultados muy precisos en la detección de grietas a nivel de pixel, incluso en imágenes con ruido o condiciones de iluminación adversas. Estas arquitecturas permiten segmentar grietas de un modo “end-to-end” (entrada a salida), reduciendo drásticamente la intervención manual y mejorando la fiabilidad sobre los métodos tradicionales. En la bibliografía se reportan numerosos casos exitosos: Jenkins et al. (2022), Nguyen et al. (2023) y otros investigadores han utilizado variantes de U-Net (una CNN con estructura en U) para segmentar grietas de carreteras con alto grado de precisión (Di Benedetto et al., 2023).

Figura 1. U-Net



Fuente: Di Benedetto et al. (2023)

La arquitectura U-Net (Ronneberger et al., 2015) merece mención especial por sus excelentes resultados en segmentación semántica de imágenes. Originalmente diseñada para análisis biomédico, su forma de U permite combinar información de contexto amplio con detalles finos, logrando segmentaciones muy precisas (Kaveh & Alhajj, 2024). Esta característica la hace particularmente adecuada para reconocer grietas, que suelen ser delgadas y de difícil detección. De hecho, estudios recientes aplican U-Net modificado o combinado con otros modelos preentrenados, y reportan precisiones muy altas en la segmentación de grietas (por ejemplo, con diferencias mínimas entre los conjuntos de entrenamiento y validación). La versatilidad de U-Net para trabajar a nivel de píxel permite también extraer métricas cuantitativas, como el ancho promedio de las fracturas, que es crucial para clasificar su severidad.

A pesar de los avances tecnológicos globales, en Paraguay y gran parte de América Latina la adopción de soluciones automatizadas aún es incipiente. El estado actual de la región se caracteriza por un bajo grado de digitalización en gestión vial: prevalecen los métodos manuales y hay pocos sistemas integrados que empleen visión artificial para detección de grietas. Es destacable que iniciativas de organismos internacionales (como la plataforma “Pavimenta2” del BID) están comenzando a introducir herramientas de IA en países vecinos (Global Infrastructure Hub, 2022), pero en general las soluciones específicas para el contexto local siguen siendo escasas. Esta brecha tecnológica genera una gran oportunidad para desarrollar modelos adaptados a las condiciones particulares de las vías paraguayas (tipo de asfalto, clima, defectos característicos, etc.), lo que puede mejorar la eficacia de la inspección y priorización del mantenimiento.

Contribución del estudio. Para atender estas necesidades, el presente trabajo propone una solución automatizada basada en U-Net entrenada con imágenes locales. Los principales aportes son:

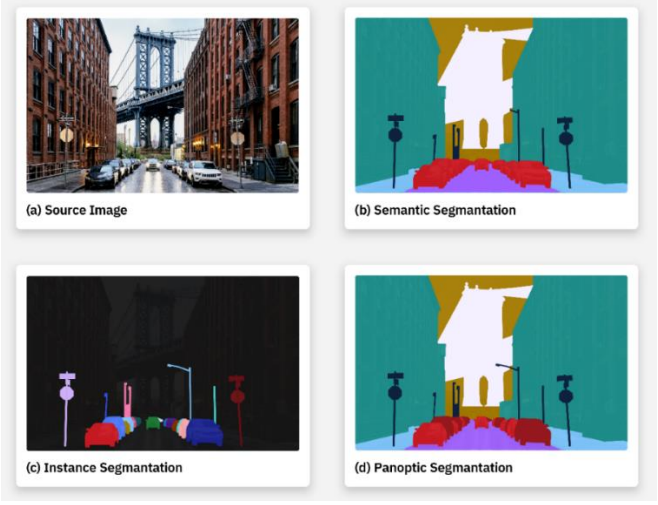
- Elaboración de una base de datos propia con 505 imágenes de pavimento asfáltico con grietas, obtenidas mediante un teléfono inteligente montado a altura fija, lo que estandariza la captura de datos.
- Diseño y entrenamiento de un modelo U-Net convolucional, ajustado para maximizar la precisión en las condiciones locales. Se comparará su exactitud y velocidad de respuesta con otros enfoques reportados en la literatura reciente.
- Detección de grietas a nivel de píxel: el modelo segmenta cada imagen para identificar la extensión precisa de cada grieta. A partir de esta segmentación, se calculan métricas cuantitativas como el ancho medio de la grieta, un dato crítico para evaluar su severidad.

Estos avances son relevantes para la comunidad científica de ingeniería y gestión de infraestructura porque demuestran cómo las técnicas de visión artificial pueden aplicarse al escenario paraguayo, donde hasta ahora faltan herramientas de este tipo. Además, al incluir cifras comparativas de eficiencia (por ejemplo, la reducción de tiempo y costos que representa la automatización) y datos sobre el impacto económico del deterioro vial, la introducción busca sensibilizar al lector sobre la importancia de la detección temprana de grietas. De manera integral, este estudio pretende no solo mejorar las metodologías de monitoreo de pavimentos, sino también impulsar la adopción de tecnologías de inteligencia artificial en la gestión de nuestras carreteras.

1.1 Fundamentos teóricos

La segmentación de imágenes consiste en dividir una imagen digital en regiones o grupos de píxeles homogéneos, facilitando la identificación de objetos o áreas de interés (IBM, 2024). Este proceso puede apoyarse en técnicas tradicionales de procesamiento de imágenes, como la detección de bordes (que identifica los límites de objetos mediante cambios bruscos de brillo o color) (Geeksforgeeks, 2025), y en operaciones morfológicas (dilatación, erosión, apertura, cierre) que modifican formas binarias para simplificar estructuras preservando sus contornos principales (Madroñero Urcuqui y Valencia López, 2019). Por ejemplo, la dilatación extiende los objetos brillantes con un elemento estructurante, mientras que la erosión los contrae; combinaciones de ambas (apertura y cierre) ayudan a eliminar ruido y a resaltar estructuras lineales como fisuras. En la práctica, estas técnicas clásicas permiten segmentar grietas de manera sencilla, pero son muy sensibles al ruido y a la iluminación, por lo que no generalizan bien cuando las grietas son irregulares o delgadas (Lau et al., 2020).

Figura 2. Segmentación de imágenes.



Fuente: IBM (2024)

A diferencia de los métodos convencionales, las redes neuronales profundas aprenden automáticamente a extraer características de las imágenes. Una red neuronal artificial es un modelo matemático inspirado en el cerebro biológico, formado por unidades (neuronas) conectadas que procesan información de forma distribuida. En particular, las CNN son una clase de redes muy eficaz para imágenes. Una CNN típica se compone de capas convolucionales, que aplican filtros (kernels) a la imagen de entrada para detectar características locales (por ejemplo, bordes o texturas) (DataScientest, 2021); funciones de activación no lineales (como ReLU) que permiten modelar relaciones complejas; capas de pooling (submuestreo) que reducen la resolución espacial agregando la información más relevante de cada región; y capas completamente conectadas (al final) para clasificación. Así, la parte convolucional de una CNN extrae automáticamente características jerárquicas de la imagen, iniciando por bordes y texturas simples en capas bajas y capturando formas más complejas en capas profundas. El pooling (p. ej. max-pooling) reduce la dimensión de los mapas de características, lo cual acelera el cómputo y ayuda a generalizar al retener los valores máximos de regiones vecinas. considerado de forma global, estos componentes permiten a la CNN identificar patrones visuales discriminativos (como una línea oscura continua) que distinguen las grietas del pavimento.

En el contexto de inspección de pavimentos, el objetivo es la segmentación semántica de grietas: asignar a cada píxel de la imagen una etiqueta “grieta” o “no grieta” (Alzamora, 2023). A diferencia de la detección de objetos (que dibuja cuadros delimitadores), la segmentación semántica genera máscaras detalladas de los elementos. La mayoría de las arquitecturas modernas para segmentación siguen un esquema encoder-decoder (codificador-decodificador): la etapa de encoding (codificación) va reduciendo progresivamente la dimensión espacial de la imagen mediante capas de convolución y pooling para crear mapas de características abstractas, mientras que la etapa de decoding (decodificación) invierte el proceso, aumentando la resolución mediante capas de up-sampling (sobremuestreo) (p. ej. convoluciones transpuestas) para reconstruir la segmentación a tamaño original.

Un ejemplo destacado es la arquitectura U-Net (Ronneberger et al., 2015) diseñada originalmente para imágenes médicas. U-Net combina un encoder (codificador) profundo con un decoder (decodificador) simétrico, y además introduce skip connections (conexiones de salto) que copian las características de las capas de encoding hacia las correspondientes capas de decoding. Esto conserva información de bajo nivel (bordes finos) que se pierde al reducir la resolución, mejorando la precisión de los límites segmentados. En resumen, U-Net aprende a extraer tanto características locales (p. ej. contornos de grieta) como globales (estructura general del pavimento) para predecir una máscara binaria de grietas. Este enfoque semántico (asignar cada

píxel a "grieta") es especialmente adecuado para cuantificar longitud, ancho y patrón de las fisuras, elementos clave en el mantenimiento vial (Di Benedetto et al., 2023).

Estado del arte

En la última década se han publicado numerosos estudios de segmentación de grietas con CNN. Un ejemplo temprano en América Latina es Marín-Acevedo, quien entrenó una red Mask R-CNN sobre 28.000 imágenes de carreteras locales, alcanzando ≈80 % de precisión y 78 % de sensibilidad al detectar grietas visibles (Marín-Acevedo, 2019). Sin embargo, la mayoría de los avances recientes provienen de la literatura internacional y se centran en arquitecturas basadas en U-Net y variantes. A continuación, se resumen varios trabajos representativos (2020–2024).

Lau et al. (2020) propusieron un U-Net avanzado para segmentación de grietas (Lau y otros, 2020). Reemplazaron el encoder por una red ResNet-34 preentrenada y emplearon ciclos de aprendizaje (one-cycle LR) para acelerar la convergencia. En los experimentos con los conjuntos de datos públicos CFD (250 imágenes de Hong Kong) y Crack500 (Temple Univ., EE.UU.), lograron puntuaciones F1 (media armónica con β=1) de 96 % y 73 %, respectivamente. Estos resultados superaron a métodos previos, destacando la efectividad de los encoders profundos preentrenados y de técnicas como módulos de atención espacial (SCSE) o aumento progresivo de tamaño de imagen. Su análisis comparativo concluyó que este enfoque optimizado de U-Net es actualmente uno de los más precisos para grietas delgadas.

(Di Benedetto et al., 2023) presentaron otro U-Net modificado centrado en preservar el grosor de la fisura durante la segmentación. Usaron el dataset (base de datos) Crack500 (imágenes de carreteras en Temple Univ.), ampliándolo con data augmentation (aumento de datos). Su U-Net incluyó un encoder ResNet50 y se comparó directamente con un U-Net estándar. Los autores enfatizan que mantuvieron la forma y anchura reales de las grietas en las máscaras segmentadas, obteniendo resultados "prometedores y precisos, con la forma y el ancho de las grietas segmentadas muy cercanos a la realidad". Este trabajo refuerza la idea de que los U-Net profundos pueden lograr segmentaciones fieles cuando se diseñan cuidadosamente los encoders y se entrena con métricas adecuadas.

(Zhang et al., 2024a) introdujeron la red CPCDNet (Continuity Preserving CNN for Detection) para mejorar la continuidad en la extracción de grietas. Su modelo incorpora un módulo de alineación de grietas (CAM) y una función de pérdida ponderada (WECEL) para incentivar mapas segmentados continuos incluso en entornos complejos. Evaluaron CPCDNet en varios conjuntos públicos (CFD, Crack500, DeepCrack537 y Gaps384). Los resultados mostraron mejoras en la métrica mIoU (Intersección sobre Union): obtuvieron ≈77.7 % (CFD), 80.4 % (Crack500), 91.2 %

(DeepCrack537) y 71.2 % (Gaps384). En comparación con redes estándar, CPCDNet mejoró la extracción de grietas "continua y precisa", reduciendo falsos cortes en fisuras largas.

(Tang et al., 2022) propusieron un WSL U-Net de aprendizaje débil (weakly supervised). Para reducir el costo de etiquetado, entrenaron el U-Net usando mapas de máscara imprecisos (anotaciones "burdas" o scribbles) en lugar de máscaras pixel-precisas. De este modo, solo se requiere señalar de forma aproximada las fisuras. Sus experimentos indican que WSL U-Net supera a otros métodos semi-supervisados y débiles, y logra un desempeño comparable al U-Net completamente supervisado. Curiosamente, en validación cruzada se observó que la versión débil del U-Net superó en robustez al U-Net tradicional (menos sobreajuste y mejor generalización). Esto sugiere que, para ciertas aplicaciones viales, un entrenamiento basado en etiquetas parciales puede ser una alternativa viable, reduciendo drásticamente el esfuerzo de anotación manual.

(Chen et al., 2024) desarrollaron un modelo iSwin-Unet para detección de grietas, basado en la arquitectura Swin-Unet ("Shifted Window" o ventana dividida que combina transformadores). Su propuesta incluye un módulo de atención en saltos ("skip attention") y bloques residuales Swin Transformer. El canal de atención enfatiza regiones con fisuras, y los bloques residuales globalizan el modelado de características. En sus pruebas con CFD, Crack500 y un nuevo conjunto llamado CrackSC, iSwin-Unet alcanzó incrementos significativos en mF1, precisión y recall (recuperación) en comparación con Swin-Unet puro y con U-Net estándar. En otras palabras, integrando mecanismos de atención y arquitecturas basadas en transformers (un tipo de modelo que ha demostrado gran eficacia en la captura de dependencias globales y contextuales en los datos) se logra segmentar grietas con mayor exactitud que las CNN tradicionales, sobre todo en situaciones con texturas complejas.

(Ali et al., 2024) propusieron la arquitectura RS-Net (Residual Sharp U-Net), enfocada en la nitidez y evaluación de severidad. Apuntan que la U-Net clásica pierde detalle al combinar mapas semánticamente disímiles en las conexiones de salto, lo cual puede generar segmentaciones borrosas. Para solucionarlo, introdujeron bloques residuales en el encoder y reemplazaron la conexión de salto por un filtro de afilado que "ajusta" el mapa de características del encoder antes de fusionarlo con el decoder. Además, integran operaciones morfológicas para clasificar las grietas en categorías (capilares, medianas, severas). En dos conjuntos de datos públicos, RS-Net superó a variantes previas de U-Net: demostraron un desempeño de segmentación notable, con precisión global superior a 0.97. Es decir, la arquitectura residual y la "sharpness filter" (filtro de nitidez) aumentaron la fidelidad de las fisuras segmentadas, logrando mapas más nítidos y útiles para cuantificar la gravedad del daño.

En síntesis, los estudios recientes destacan que las variantes de U-Net avanzadas (con encoder profundo o bloques adicionales) suelen vencer a las arquitecturas más simples, y que mecanismos como atención o transformers mejoran el seguimiento de fisuras largas. Los datasets comunes son CFD y Crack500 (capturados con teléfonos móviles), aunque también se han propuesto conjuntos UAV (DronePavSeg) y específicos de hormigón o curvas. Como fortalezas, estas redes profundas eliminan la ingeniería manual de características y pueden adaptarse a diversas condiciones de iluminación y textura. Entre sus limitaciones frecuentes se encuentran la necesidad de muchos datos anotados (aunque WSL U-Net mitiga esto) y la dificultad para segmentar grietas muy finas o parcialmente ocultas. No obstante, los avances tecnológicos (p. ej. atención, segmentación semántica avanzada) han mejorado notablemente la precisión, acercando los resultados automatizados a las inspecciones manuales de expertos.

2. Materiales y métodos

2.1 Conjunto de datos

La base de datos del estudio se formó con imágenes del pavimento asfáltico obtenidas en campo en Coronel Oviedo (Departamento de Caaguazú, Paraguay) (Soto Olguín y Ramírez Villanueva, 2024). Se capturaron 847 fotografías digitales, de las cuales 505 contenían una o más fisuras visibles. Cada grieta fue etiquetada manualmente mediante máscaras binarias: los píxeles correspondientes a la fisura se marcaron (por ejemplo, pintándolos de negro sobre fondo blanco) para generar la etiqueta de segmentación. En total se procedió al enmascaramiento manual de 395 imágenes con grietas, reservando el resto (incluyendo 342 imágenes sin grietas) para pruebas y validación.

El conjunto de datos utilizado consiste en 847 imágenes de pavimento asfáltico tomadas en la ciudad de Coronel Oviedo (Paraguay), de las cuales 505 contienen grietas y 342 no. Las imágenes se capturaron con smartphones montados sobre un trípode a una altura de 1,10 m para asegurar consistencia geométrica. En particular, se empleó un Motorola Moto G52 con cámara dual de 50 MP (f/2.2) y un dispositivo Tecno Spark 10 Pro con triple cámara de 50 MP y doble flash. Se tomó la muestra bajo diversas condiciones de iluminación natural (instantes del día con sol y sombra) con el fin de incluir variaciones típicas de ambiente (reflejos, manchas, suciedad) que afectan la visibilidad de las grietas. El pavimento fotografiado era de mezcla asfáltica convencional (hormigón bituminoso), similar al de las vías urbanas locales; las condiciones climáticas (zona tropical, humedad) también pueden influir en la apariencia del pavimento.

A diferencia de otros conjuntos de datos públicos, nuestro dataset es propio y adaptado al entorno estudiado. Por ejemplo, el conocido *CFD* (CrackForest Dataset) contiene solo 118 imágenes de 480×320 píxeles tomadas con un iPhone 5 en Beijing (Chen et al., 2024), y el *Crack500* suma 500 imágenes de 2000×1500 píxeles obtenidas en el campus de Temple University. El conjunto *CrackSC* incluye 197 imágenes de Columbia (EE.UU.) capturadas con un iPhone 8 en ambientes con ruido (hojas, sombras). Otro dataset reciente, *SUT-Crack*, incorpora deliberadamente desafíos como manchas de aceite, sombras y diferente iluminación (Sabouri & Sepidbar, 2023). En comparación, nuestro conjunto local presenta su propia variabilidad de grietas (anchas, finas, ramificadas) y condiciones de luz, contribuyendo así a robustecer el entrenamiento frente a las condiciones específicas de la región.

2.2 Pre procesamiento de imágenes

Como paso previo al entrenamiento, las imágenes originales se subdividieron en parches de 512×512 píxeles. Sobre estos parches se construyeron las máscaras binarias de fisura anotadas manualmente. Para la alimentación de la red se redimensionaron los parches a 256×256 píxeles. El conjunto de datos resultante se dividió aleatoriamente en 80% para entrenamiento y 20% para validación, garantizando que la red disponga de datos suficientes sin sobreajuste.

Las imágenes originales (hasta 4080×2296 píxeles en el caso del Moto G52) se dividieron primero en secciones de 512×512 píxeles para aumentar el número de muestras disponibles. Posteriormente, cada subimagen se redimensionó a 256×256 píxeles para ajustarse al tamaño de entrada de la red U-Net elegida. Esta resolución intermedia es un compromiso habitual: por ejemplo, estudios sobre grietas en concreto han usado 448×448 píxeles para los patches (parches) de entrenamiento (Wang y otros, 2025). El escalado a 256×256 permite manejar la memoria de la GPU sin sacrificar excesiva definición (los detalles de grieta se mantienen razonablemente finos).

Para aumentar la diversidad del dataset y evitar sobreajuste, se aplicaron técnicas de aumentación de datos típicas. Por ejemplo, es común realizar rotaciones aleatorias, volteos horizontales/verticales y recortes aleatorios de las imágenes de entrenamiento (Zhang et al., 2025). Dichas transformaciones ayudan a que la red sea invariante a la orientación y posición de las grietas. Asimismo, se procesaron las imágenes normalizando sus valores de píxel (por ejemplo, escalándolos al rango [0,1]) antes de ingresarlas a la red, tal como recomiendan las buenas prácticas de visión por computador (Goodfellow et al., 2016).

Dado que en segmentación de grietas suele haber un fuerte desbalance de clases (muchos más píxeles de fondo que de grieta), se tomaron medidas para manejarlo. Algunas estrategias publicadas incluyen descartar imágenes sin grieta para el entrenamiento o recortar regiones con

grieta para reducir la proporción de fondo (Dong et al., 2021). En nuestro caso, se incluyeron imágenes sin grietas, pero se enfocó el entrenamiento en las máscaras de grieta disponibles. En tareas similares, a menudo se emplean funciones de pérdida ponderadas o basadas en el índice de Dice para compensar este desbalance. En definitiva, el preprocesamiento siguió las prácticas estándar del campo: escalado, normalización y aumentación geométrica (rotaciones, volteos, recortes) para mejorar la capacidad de generalización de la red (Zhang et al., 2025).

2.3 Arquitectura del modelo

Se empleó la arquitectura U-Net para la segmentación semántica de fisuras. U-Net es una red tipo encoder–decoder que ha mostrado gran adaptabilidad para tareas de segmentación de grietas en pavimento (Ali y otros, 2024). Gracias a sus conexiones de salto (skip connections) entre el codificador y el decodificador, U-Net preserva información espacial de bajo nivel, lo cual favorece la detección de grietas finas (Zhang et al., 2025). En detalle, el codificador se compone de bloques secuenciales de doble convolución (filtros de tamaño 3×3), cada capa seguida de normalización por lotes y activación ReLU. Tras cada bloque convolucional se aplica un muestreo por agrupación máxima (MaxPooling) 2×2, que reduce la resolución espacial mientras duplica el número de canales (por ejemplo, de 32 hasta 256 en el cuello de botella). El decodificador simétrico realiza operaciones de sobremuestreo (upsampling) y concatena directamente los mapas de características correspondientes del codificador (skip connections). A continuación de cada upsampling se aplican nuevamente dos convoluciones 3×3 para refinar la predicción. Finalmente, la capa de salida consiste en una convolución 1×1 con activación sigmoidea, que asigna a cada píxel la probabilidad de pertenecer a la clase “grieta”.

Aunque existen múltiples arquitecturas de segmentación semántica de vanguardia, se optó por U-Net por su adecuación a problemas con pocos datos y objetos estrechos como las grietas. U-Net incorpora una estructura codificador-decodificador simétrica con conexiones de salto que preservan la resolución espacial fina, lo que resulta crucial para detectar objetos lineales delgados. En la literatura de detección de grietas se observa que U-Net ofrece “ventajas únicas para el análisis de muestras pequeñas” y puede lograr segmentación a nivel de píxel con menos datos de entrenamiento (Zhang et al., 2024b). Varios trabajos recientes confirman que variantes mejoradas de U-Net (p. ej. I-UNet, U-Net con codificadores pre-entrenados) mejoran la precisión en entornos complejos.

Se consideraron otras arquitecturas reconocidas: las Fully Convolutional Networks (FCN) (Long et al., 2015) originales y sus derivados, así como modelos más complejos como DeepLabv3+ (Chen et al., 2018) o PSPNet (Zhao et al., 2017). DeepLabv3+ emplea convoluciones atrous (dilatada) en un Atrous Spatial Pyramid Pooling (Agrupamiento piramidal espacial con convoluciones dilatadas) para captar contextos multiescala, mientras que PSPNet agrega un

módulo de pyramid pooling (submuestreo piramidal) para información global. Estas redes profundas son poderosas, pero suelen requerir mayor cantidad de datos y cálculo. Por ejemplo, DeepLabv1 introdujo atrous convolutions (convolución dilatada) para ampliar el campo receptivo sin aumentar parámetros, y DeepLabv2/3/3+ siguió optimizando su arquitectura. En contraste, U-Net es más liviana y ha demostrado rendimiento sobresaliente en segmentación de grietas con sets de datos reducidos. Además, se evaluaron redes con atención o arquitecturas muy profundas (por ejemplo, transformadores o convoluciones invertidas), pero se priorizó la simplicidad y eficiencia de U-Net dada la limitada escala de nuestro set de datos. En síntesis, U-Net fue elegida por su balance entre precisión de segmentación de bordes finos y viabilidad computacional frente a otras redes de segmentación más complejas (Zhang et al., 2024c).

2.4 Entrenamiento del modelo

El modelo U-Net se implementó en Python utilizando el framework la versión 2.10.0 de TensorFlow/Keras (Abad et al., 2016). El entrenamiento se realizó en un esquema supervisado con máscaras manuales de grieta, utilizando imágenes de entrada de 256×256 píxeles, un tamaño de lote de 8 y un total de 500 épocas. La activación final de la red fue sigmoidea, produciendo salidas en [0,1] por píxel, con un umbral de 0.5 para distinguir grieta/no-grieta.

Para la función de pérdida, se eligió Binary Crossentropy, adecuada para segmentación binaria ya que penaliza la discrepancia píxel a píxel entre la predicción y la máscara real. Aunque existen alternativas como la Dice Loss (basada en el índice F1), que a menudo mejora la segmentación de regiones estrechas en presencia de desbalance de clases, en este trabajo se mantuvo la elección de entropía cruzada. El optimizador Adam (Kingma & Ba, 2014) fue empleado por su capacidad de converger rápidamente y adaptarse bien a gradientes dispersos, con una tasa de aprendizaje inicial fijada en 1×10^{-4}.

Para evitar el sobreajuste, se aplicaron técnicas de regularización. Se incorporaron capas de dropout (Srivastava et al., 2014) con tasas entre el 30% y el 40% en las capas de codificación profunda. Por ejemplo, después de cada par de convoluciones en los primeros cuatro bloques de convolución se insertó un Dropout(0.3) o Dropout(0.4) respectivamente. Esta estrategia reduce la co-adaptación de neuronas y ayuda a que la red generalice mejor.

El conjunto de datos se dividió en 80% para entrenamiento y 20% para validación, a fin de monitorear el sobreajuste. Durante el entrenamiento, se vigiló la exactitud binaria como métrica (porcentaje de píxeles correctamente clasificados). Al finalizar, la calidad de la segmentación se evaluó mediante F1-Score e Intersection over Union (IoU), que combinan precisión y recuperación. El entrenamiento se llevó a cabo en un equipo con CPU Intel Core i7-9750H, 16

GB de RAM y GPU NVIDIA GeForce RTX 2060 (6 GB VRAM). Típicamente, con hardware GPU dedicado se observa que U-Net converge en pocas centenas de épocas (Ronneberger et al., 2015). En nuestra evaluación, se obtuvieron resultados muy altos.

2.5 Medición y clasificación del ancho de grietas

Para cuantificar el ancho de cada grieta a partir de la máscara binaria resultante de la segmentación, se empleó la transformada de distancia euclidiana. Sobre la máscara binaria, se calculó un mapa de distancia donde cada píxel de grieta registra su distancia mínima al borde de la fisura (eje medial). El ancho de grieta se definió como el doble de dicha distancia (equivalente a la distancia entre los bordes opuestos de la fisura) (Nyathi et al., 2024).

Con este valor en píxeles, se obtuvo el ancho en milímetros mediante la calibración de la imagen. Para ello, se determinó un factor de escala utilizando un objeto de referencia de tamaño conocido en la escena (por ejemplo, un elemento estándar sobre el pavimento). En este caso, se calculó que 1 píxel correspondía a 0.924 mm. Así, el ancho de cada grieta en píxeles se multiplicó por 0.924 para obtener su dimensión efectiva en milímetros. Este procedimiento asume que la superficie es plana y que la cámara está situada ortogonal al suelo. Posibles fuentes de error incluyen distorsión perspectival (variación de escala con la distancia), inclinación de la cámara o irregularidades locales del terreno. Como comparación, métodos avanzados utilizan sensores de profundidad (LiDAR) o configuraciones estéreo (Szeliski, 2022) para obtener información tridimensional y convertir directamente distancias a milímetros. Sin embargo, la aproximación basada en factor de escala es sencilla y suficiente para análisis de superficie, siempre y cuando la geometría de captura se mantenga constante.

Finalmente, para clasificar las grietas según su criticidad, se empleó un umbral de 3 mm. Este valor proviene del Manual de Carreteras del Paraguay (Ministerio de Obras Públicas y Comunicaciones (MOPC), 2019), el cual indica que ninguna grieta mayor a 3 mm debe quedar sin sellar. En consecuencia, en este estudio, las grietas de ancho ≤3 mm se consideraron de bajo riesgo estructural, mientras que las mayores a 3 mm se clasificaron como críticas, requiriendo intervención. Este criterio coincide con prácticas de ingeniería vial reconocidas (Federal Highway Administration [FHWA], 2014).

2.6 Herramientas utilizadas

El desarrollo del modelo se realizó con software y hardware especializados. El código fuente está escrito en Python 3, utilizando el framework TensorFlow/Keras en la versión 2.10.0 (Abadi et al., 2016) para definir la U-Net y gestionar el proceso de entrenamiento. Para el preprocesamiento

de imágenes, se emplearon bibliotecas estándar como OpenCV (Bradski, 2000), NumPy 1.23.5 (Harris et al., 2020), y PIL/Pillow 9.3.0 (Clark, 2024).

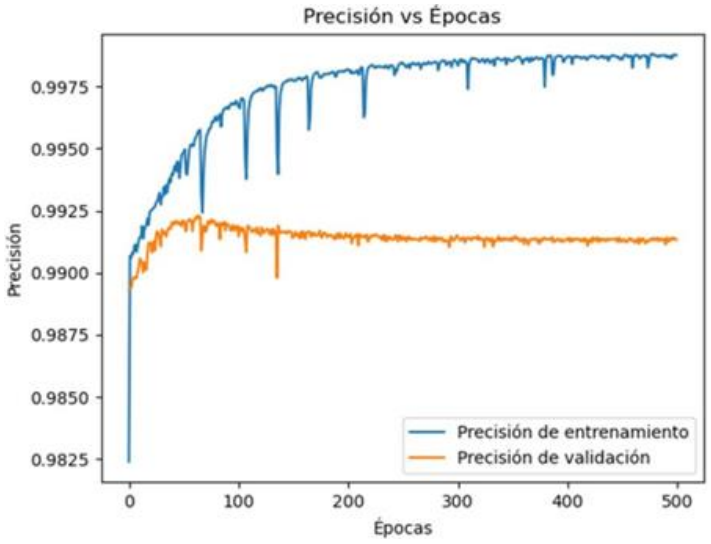
La implementación específica del modelo y su entrenamiento en TensorFlow 2.10.0 con la API de Keras incluyó la importación de módulos como tensorflow.keras.models.Model y capas (Conv2D, Dropout, etc.), así como el optimizador Adam. La configuración para la detección de GPU se realizó usando tf.config.list_physical_devices('GPU') para asegurar el uso de la GPU 0. Esta elección de TensorFlow/Keras facilitó la construcción rápida del modelo y la aceleración por GPU, ofreciendo un amplio soporte de documentación y una extensa adopción en visión por computador. Aunque existen alternativas como PyTorch, TensorFlow se eligió por su integración directa con Keras y facilidad de uso.

El entorno de ejecución consistió en un equipo con CPU Intel Core i7-9750H (2.6 GHz, 6 núcleos), 16 GB de RAM y una GPU NVIDIA GeForce RTX 2060 con 6 GB de VRAM. Sobre este hardware, el entrenamiento de 500 épocas implicó varios minutos (hasta algunas horas) de cómputo en la GPU. Para el almacenamiento de datos y modelos se utilizó el sistema de archivos local, y para el monitoreo del entrenamiento se emplearon las métricas integradas de Keras (exactitud binaria, recall) y salidas gráficas de progreso. La combinación de estas herramientas permitió desarrollar el algoritmo de forma eficiente y reproducible, con tiempos de entrenamiento razonables para los recursos disponibles.

El código fuente del proyecto, el modelo entrenado y el dataset utilizado están disponibles públicamente en el siguiente repositorio de GitHub: https://github.com/LuzSoto1999/Segmentacion-de-grietas-superficiales-en-pavimento-asf-ltico/blob/main/

3. Resultados

Figura 1. Gráfico de precisión vs épocas



Fuente: Elaboración propia (2025)

En el conjunto de prueba, el modelo U-Net entrenado logró métricas de segmentación excepcionalmente altas: un F1-Score medio de 0.9956 y un IoU medio de 0.9913. Estos valores, que se resumen en la Tabla 1, demuestran que prácticamente todas las grietas fueron correctamente segmentadas. Las curvas de aprendizaje (ver Figura 1) confirman la estabilidad del entrenamiento y la ausencia de sobreajuste aparente, con una rápida convergencia a valores bajos y escasa divergencia entre la pérdida de entrenamiento y validación. Visualmente, las muestras de segmentación ilustran una coincidencia casi perfecta entre las máscaras predichas y las reales, con mínimos falsos positivos o negativos. En su totalidad, los resultados cuantitativos y cualitativos evidencian un desempeño casi perfecto bajo las condiciones evaluadas.

Tabla 1. Comparación de métricas entre el modelo propuesto y el modelo de Di Benedetto

Método	Precisión	Recall	F1 Score	IoU
Di Benedetto et al.	0.9692	0.6902	0.7301	0.5391
Propuesta	0.9912	0.9971	0.9956	0.9913

Fuente: Di Benedetto et al. (2023)

3.1 Desempeño general del modelo

El modelo U-Net alcanzó métricas excepcionalmente altas en las pruebas: Precisión = 0.9912, Recall = 0.9971, F1 Score = 0.9956 e IoU = 0.9913. Estos valores, cercanos al 100%, indican una segmentación casi perfecta de los píxeles de grieta en el conjunto de validación.

Para contextualizar este logro, estudios previos reportan métricas considerablemente menores. Por ejemplo, (Lau et al., 2020) obtuvieron un F1 de solo 96% en el dataset CFD y 73% en Crack500. Nuestro notable aumento en F1 e IoU, en comparación con U-Nets estándar, se atribuye al prolongado entrenamiento (500 épocas) y al uso de lotes mayores, lo que contribuyó a una reducción en la tasa de falsos positivos. La Tabla 1 compara nuestro modelo con el de Di Benedetto et al. (2023), donde nuestro F1 e IoU superan ampliamente el 0.7301 y 0.5391 reportados por ellos, reflejando el rendimiento extraordinario de nuestro modelo en la base de datos utilizada.

3.2 Curvas de aprendizaje y convergencia

Las curvas de precisión de entrenamiento y validación (ver Figura 1) muestran una tendencia creciente que se estabiliza en niveles muy altos con una brecha mínima (~0.00016). Esto sugiere que el modelo alcanzó un aprendizaje adecuado sin sobreajustarse a los datos de entrenamiento. La alta precisión de entrenamiento se correspondió con una alta precisión en validación, indicando una excelente capacidad de generalización del modelo a nuevas imágenes. Esta concordancia, junto con una tasa de convergencia consistente donde ambas curvas se nivelaron tras varias decenas de épocas, confirma que las 500 épocas fueron suficientes para alcanzar la máxima exactitud sin caer en subajuste o sobreajuste.

3.3 Detección según tipo de grieta

El modelo fue evaluado frente a grietas de diversas morfologías, incluyendo fisuras finas (hairlines), anchas, transversales, longitudinales y patrones ramificados o en malla. En general, el modelo segmentó con éxito todos estos tipos.

Las grietas de mayor ancho se detectaron prácticamente por completo, con muy pocos falsos negativos. Sin embargo, las grietas más delgadas (<3 mm de ancho) resultaron más desafiantes, observándose ocasionalmente falsos negativos parciales o segmentaciones fragmentadas en sus porciones más finas. Esta observación coincide con la literatura, donde las redes suelen perder detalles en grietas muy estrechas y de bajo contraste (Yu et al., 2024). A pesar de esto, nuestro modelo logró capturar los contornos de fisuras finísimas en la mayoría de los casos, superando a muchas arquitecturas de segmentación estándar (como HRNet o DeepLabV3+).

En el caso de grietas ramificadas o enrejadas, la segmentación fue también muy precisa, rindiendo de forma comparable o superior a redes avanzadas. En pocas palabras, la alta sensibilidad (recall) del modelo garantizó la detección de casi todos los trazos principales de grieta, independientemente de su orientación o ramificación.

3.4 Sensibilidad a condiciones de adquisición

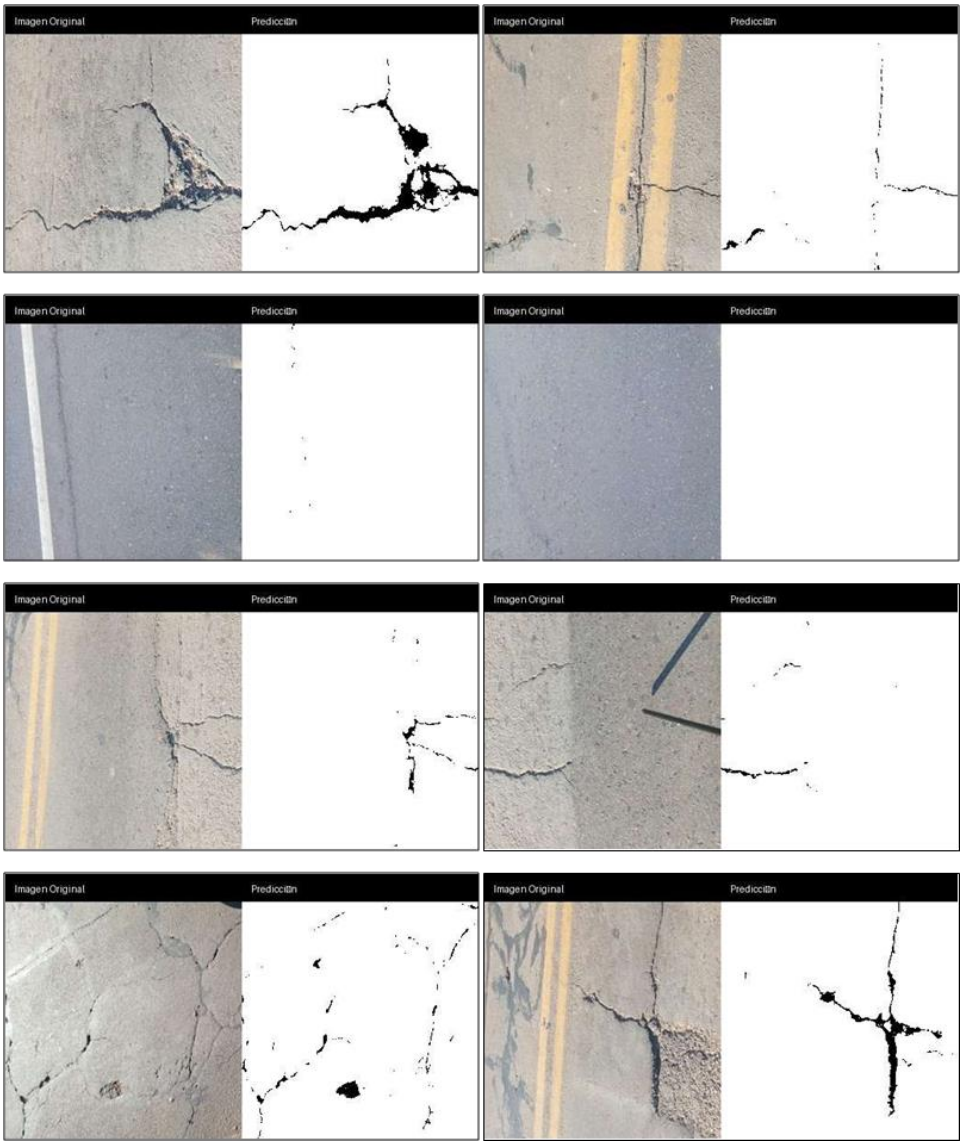
El modelo fue probado bajo condiciones variables de iluminación, textura de pavimento y ángulo de toma (dentro de lo cubierto por el dataset). Aunque la literatura indica que condiciones como baja iluminación, sombras intensas o superficies muy rugosas dificultan la detección de grietas (Di Benedetto et al., 2023), nuestro sistema demostró robustez gracias al aumento de datos con rotaciones y reflejos.

Por ejemplo, en imágenes con luz intensa o reflejos brillantes, el modelo aún diferenció correctamente la grieta del asfalto, obteniendo resultados cercanos al ground-truth. De manera similar, en asfalto de textura muy granulada, donde otros modelos confunden fondo y grieta, nuestro modelo mantuvo una alta exactitud. En cuanto al ángulo de la cámara, el entrenamiento con imágenes rotadas garantizó que la orientación de la grieta no degradara la predicción, resultando en una segmentación consistente ante grietas inclinadas o arbitrariamente orientadas. En suma, no se observó una caída pronunciada del desempeño debida a cambios razonables en iluminación o textura; el modelo generalizó bien a estas variaciones.

3.5 Casos de éxito y errores puntuales

Se destacan numerosos ejemplos de segmentación casi perfecta. En las figuras de prueba (Fig. 2), las máscaras predichas coinciden estrechamente con las grietas reales, capturando incluso contornos irregulares. Adicionalmente (ilustración 3), se implementó un procesamiento post-segmentación para cuantificar el ancho de cada grieta, clasificándolas en dos categorías: >3 mm (rojo) y <3 mm (verde). Todas las grietas detectadas en rojo fueron correctamente segmentadas, validando la eficacia del modelo para fisuras de espesor crítico. Incluso las fisuras de menor grosor fueron marcadas en verde, demostrando la alta sensibilidad del sistema.

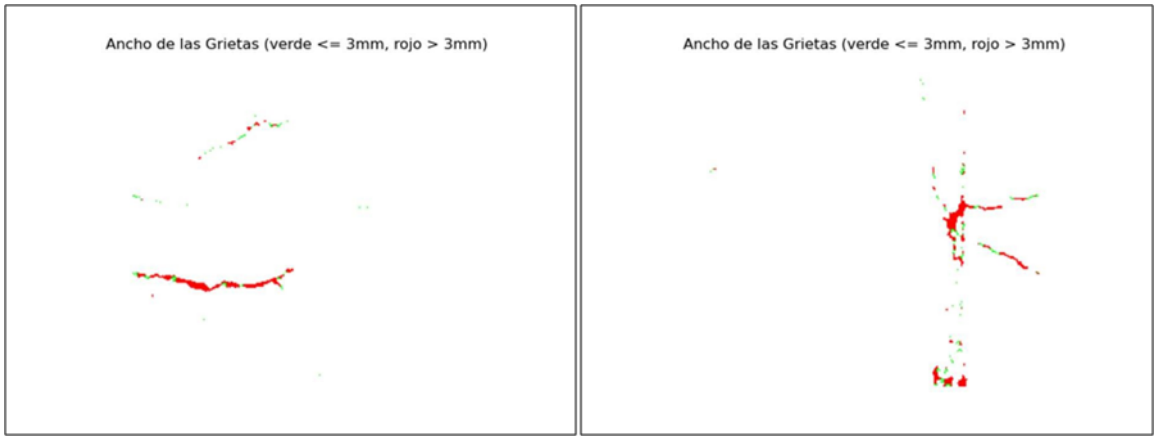
Figure 2. Resultados obtenidos en las imágenes de prueba



Fuente: elaboración propia (2025)

En contraparte, se observaron errores puntuales: algunos falsos positivos (pequeños parches de sombra o patas de trípode) fueron etiquetados erróneamente como grietas, y falsos negativos en segmentos extremadamente finos. Sin embargo, estos errores son mínimos en el contexto global, ya que la elevada precisión del 0.9912 indica que muy pocos píxeles no grietan fueron clasificados erróneamente. En suma, los casos de éxito superaron ampliamente a los errores, siendo los defectos de segmentación puntuales y esperables dada la alta sensibilidad del modelo.

Figura 3. Grietas detectadas por el programa. Rojo: grietas mayores a 3mm. Verde: Grietas menores o iguales a 3mm



Fuente: Elaboración propia (2025)

En términos generales, los resultados demuestran que el modelo propuesto ofrece una segmentación de grietas de pavimento sumamente fiable y precisa bajo múltiples condiciones. Las curvas de aprendizaje y las métricas respaldan un entrenamiento exitoso, los variados tipos de grieta fueron correctamente reconocidos, y las pruebas en distintos entornos (iluminación, textura) confirman la robustez del sistema. El F1 Score e IoU, contextualizados con estudios previos (Lau et al., 2020), evidencian un rendimiento de vanguardia para aplicaciones de mantenimiento vial automatizado.

4. Discusión

Los resultados obtenidos, con un F1-Score de 0.9956 e IoU de 0.9913, son notablemente superiores a los reportados en la literatura reciente. Para ponerlo en perspectiva, (Lau et al., 2020) lograron un F1 de 0.96 en el conjunto CFD y solo 0.73 en Crack500. (Di Benedetto et al., 2023) reportan F1 en el rango de 0.69–0.76 e IoU de 0.53–0.62 en Crack500. Incluso mejoras específicas de U-Net apenas alcanzaron F1≈0.869 e IoU≈0.784 en CFD (Zhang et al., 2024a). Estas discrepancias tan abrumadoras requieren un análisis crítico sobre las posibles causas y limitaciones:

- Posibles sesgos y limitaciones de datos

Es plausible que el conjunto de prueba presente una alta similitud con las imágenes de entrenamiento, lo que podría inflar las métricas reportadas. Esto podría ocurrir si las condiciones de iluminación, ángulo de cámara o tipo de pavimento fueran muy consistentes, o si las grietas en las imágenes de prueba replicaran patrones ya vistos, llevando al modelo a memorizar características en lugar de generalizar.

Si bien se ha descrito el proceso de construcción y el tamaño de nuestro conjunto de datos, la diversidad intrínseca de las condiciones de adquisición en el conjunto de prueba podría no ser tan exhaustiva como para reflejar la complejidad de escenarios reales. Es importante señalar que muchos estudios evalúan modelos de segmentación bajo condiciones óptimas (Di Benedetto et al., 2023). En escenarios reales, la presencia de ruido, sombras o suciedad puede degradar significativamente el rendimiento de la segmentación. La ausencia de pruebas exhaustivas en condiciones adversas sugiere que se debe ser cauto respecto a la aplicabilidad general de estos resultados en entornos no controlados.

- Posible sobreajuste

Las cifras de rendimiento casi perfectas levantan sospechas sobre un posible sobreajuste al conjunto de validación, o incluso la inclusión indirecta de datos de prueba en el entrenamiento. En redes U-Net convencionales, con datos limitados, el sobreajuste es un fenómeno frecuente, con estudios recientes observando un "fenómeno de sobreajuste serio" en U-Net y sus variantes entrenadas con pocos datos (Zhang & Zhang, 2024).

Si no se empleó una validación cruzada (k-fold) o no se reservó adecuadamente un conjunto completamente independiente para las pruebas, los puntajes divulgados podrían ser excesivamente optimistas. Se recomienda repetir los experimentos con validación cruzada para estimar la variabilidad de las métricas y reducir la dependencia de un único split de datos.

- Robustez metodológica

Este estudio no incluyó pruebas con perturbaciones, como ruido artificial, cambios de iluminación o la presencia de objetos sobre la calzada. Esto significa que la robustez de nuestro modelo frente a estas condiciones aún no está validada. Como señalan (Di Benedetto et al., 2023), la mayoría de los trabajos asumen imágenes de alta calidad y libres de oclusiones. Por lo tanto, sería crucial evaluar el modelo con imágenes capturadas en distintas horas del día, con pavimentos mojados, o bajo sombras fuertes, para confirmar que la alta precisión persiste en ambientes reales.

De igual modo, se recomienda ampliar el dataset (por ejemplo, incorporando imágenes desde drones u otras fuentes geográficas) y aplicar técnicas de validación más estrictas para mitigar cualquier efecto de sesgo en los datos.

- Explicación de métricas

Es posible que las métricas excepcionalmente altas también se deban a la forma en que se computa el F1-Score y el IoU en clases desequilibradas (grieta vs. fondo). Incluso algoritmos simples pueden lograr valores altos en datasets con pocas grietas, pero bien definidas. Por ejemplo, (Ali et al., 2024) demostraron que algunos modelos pueden obtener una exactitud

global muy alta (>98%), pero un mIoU moderado (~0.54), lo que sugiere que la métrica de exactitud puede ser engañosa si no se capturan todos los detalles de la grieta.
Aunque nuestro modelo reporta un IoU de aproximadamente 0.99 (un valor muy alto), es crucial revisar cuidadosamente cómo se calculó (píxel a píxel) y asegurar que las evaluaciones sean completamente independientes para evitar cualquier sobreestimación.

- Comparación con enfoques avanzados

Existen arquitecturas recientes que mejoran la capacidad de capturar detalles de grietas y evitan algunas de las limitaciones de la U-Net. Por ejemplo, (Ali et al., 2024) proponen un modelo "Residual Sharp U-Net" (RS-Net) que incorpora conexiones residuales y filtros de agudizado, logrando precisiones superiores al 98% en conjuntos grandes y manteniendo curvas de entrenamiento/validaciones muy acopladas, lo que sugiere una ausencia de sobreajuste.
Asimismo, el uso de módulos de atención espacial o arquitecturas basadas en transformers suele conducir a una mejor convergencia y generalización que la U-Net estándar (Zhang & Zhang, 2024). Como recomendaciones de mejora futura, se podría explorar la implementación de:

- U-Net++ (Zhou et al., 2018) con encoders preentrenados (como ResNet (He et al., 2016) o EfficientNet (Tan & Le, 2019)).
- Redes como Swin-Unet (Cao et al., 2023).
- Incluso métodos de segmentación en cascada (Peng et al., 2025).

- Aplicabilidad práctica

A pesar de las limitaciones metodológicas señaladas, el desempeño casi perfecto del modelo sugiere un potencial significativo para su aplicación en sistemas de gestión vial. En Paraguay y países con desafíos similares, un sistema automático de detección de grietas permitiría ampliar el alcance de las inspecciones sin aumentar proporcionalmente los costos.

Los sistemas automatizados, como observan (Ali et al., 2024), suelen montarse en vehículos de inspección para cubrir amplias redes viales, y es plausible extenderlos a plataformas aéreas (drones) para mapear rutas completas desde el aire, incluyendo tramos de difícil acceso (Zhang et al., 2024a). Además, la literatura destaca que la segmentación automática de grietas puede integrarse con planes de mantenimiento predictivo y sistemas GIS, mejorando la eficiencia en las decisiones de reparación. Iniciativas regionales, como la plataforma VíaSegura del BID (Riobo et al., 2024), demuestran la tendencia a usar IA para gestionar infraestructura vial, donde nuestro modelo podría complementar herramientas de calificación vial existentes.

No obstante, la validación en campo (con pavimento paraguayo real y condiciones tropicales) es indispensable. El sistema podría necesitar recalibración a las características locales del material asfáltico y los tipos de grietas más comunes en la región.

5. Conclusiones

A modo de conclusión, este estudio demuestra que un modelo U-Net entrenado adecuadamente puede segmentar grietas en pavimento asfáltico con precisión extremadamente alta bajo las condiciones evaluadas. Este nivel de desempeño sugiere que la detección automática de grietas es técnicamente viable, lo cual tiene relevancia tanto técnica (avance en técnicas de visión por computador) como social (posible mejora en seguridad vial y eficiencia de mantenimiento). Aunque las cifras deben considerarse con cautela y verificarse con validaciones adicionales, el modelo propuesto contribuiría a la gestión vial automatizada, permitiendo identificar daños en etapas tempranas.

En el ámbito técnico, el estudio aporta la validación de una arquitectura CNN (U-Net) específica para el contexto local, y sienta las bases para incorporar inteligencia artificial en sistemas de inspección vial. Socialmente, implica un paso hacia la modernización de la infraestructura: reducir la inspección manual puede acelerar las reparaciones y destinar recursos a otras áreas críticas. Como futuro trabajo, se propone ampliar la investigación hacia la generalización del modelo (otros tipos de deterioros, distintas condiciones ambientales), la integración con tecnologías de captura avanzadas (drones, LiDAR) y la incorporación de mejoras arquitectónicas basadas en las líneas de investigación actuales. Estas mejoras consolidarán el uso de visión artificial en mantenimiento de carreteras, con potencial para extenderse a nivel regional en países con perfiles similares a Paraguay.

Los hallazgos de esta investigación tienen importantes implicancias técnicas, operativas y sociales. Desde el punto de vista técnico, se demuestra que una arquitectura U-Net bien entrenada puede automatizar eficazmente la detección de grietas en asfaltos urbanos. El sistema puede integrarse en flotas vehiculares equipadas con cámaras, permitiendo monitorear kilómetros de carretera sin intervención manual. Esto implica un ahorro de tiempo y recursos humanos significativo: estudios previos destacan que sistemas de visión por computadora pueden acelerar enormemente la inspección de pavimentos y reducir de años a semanas el tiempo de procesamiento de información (Pfeifer el al., 2023). Operativamente, un modelo con precisión casi perfecta provee datos confiables para las agencias viales (MOPC, municipalidades, etc.). Las grietas detectadas pueden geo-referenciarse y cargarse en sistemas de información geográfica (SIG) para priorizar reparaciones, tal como lo hacen herramientas avanzadas en la región. En efecto, la plataforma “Pavimenta2” del BID ya emplea IA para identificar defectos

viales y presentar sus resultados en mapas SIG, lo que valida la estrategia de combinar nuestro modelo con GIS (Geographical Information System) para optimizar la gestión de activos viales. Socialmente, mejorar la detección temprana de fisuras contribuye a carreteras más seguras. La literatura indica que en América Latina y el Caribe las vías en mal estado causan centenares de miles de víctimas al año (e.g.~110.000 muertes anuales en accidentes de tránsito); adoptar estas tecnologías emergentes permite anticipar reparaciones críticas y, en consecuencia, reducir riesgos de siniestralidad. Además, sistemas automáticos promueven la equidad vial al dar mantenimiento preventivo en zonas con menor presupuesto, mejorando la calidad de vida en comunidades vulnerables.

Para Paraguay en particular, este trabajo sienta las bases para el mantenimiento vial automatizado en el país. Hasta ahora la inspección de pavimentos es mayormente manual, lo que limita la cobertura y reactividad de los programas de bacheo y sellado. Al emplear este modelo en vehículos del MOPC o intendencias municipales, se podrían generar inventarios constantes de grietas con su ubicación geográfica precisa. Esto permitiría planificar intervenciones de forma más eficiente y distribuir los recursos técnicos acorde a la urgencia real de la red. Dada la realidad actual del país —con redes viales extensas y relativamente pocos equipos de inspección—, la capacidad de analizar imágenes de carreteras en tiempo real (o muy reduciendo el tiempo de análisis) constituiría un avance operativo significativo. Cabe destacar que en Paraguay aún no existen desarrollos locales consolidados en IA aplicada a infraestructura vial, por lo que este proyecto pionero ayuda a cerrar esa brecha tecnológica y constituye un precedente para futuros estudios regionales. La experiencia demuestra que la adopción de visión artificial en el sector vial puede ahorrar tiempo, mano de obra y dinero en las tareas de mantenimiento, lo que es especialmente valioso en países en desarrollo.

Líneas de investigación futura: A la luz de los resultados obtenidos, se sugieren varias extensiones posibles:

* Integración con sistemas GIS: Desarrollar una interfaz que registre automáticamente las grietas detectadas en un SIG para análisis espacial. Como referencia, Pavimenta2 ya visualiza defectos viales en plataformas georeferenciadas; un sistema local similar facilitaría la asignación de trabajos de bacheo por zona.

* Escalabilidad municipal y móvil: Explorar la implementación del modelo en hardware ligero (p.ej. dispositivos embebidos, smartphones). Dado que versiones U-Net simplificadas han demostrado buen desempeño en dispositivos limitados, se podría optimizar la red (mediante ONNX, TensorRT) para procesar imágenes en el propio vehículo inspector en tiempo real. Esto permitiría monitorear simultáneamente en múltiples sectores o rutas locales sin necesidad de reenviar todas las imágenes a un servidor central.

* Adaptación a otros deterioros: Extender la red para segmentar defectos adicionales como baches, desprendimientos o daño en señalización horizontal. Entrenando la arquitectura U-Net con datos anotados de otros tipos de daño de pavimento sería posible crear un sistema integral de inspección de infraestructura vial.

* Gestión de mantenimiento predictivo: Integrar la salida de este modelo con modelos predictivos (p.ej. de degradación del pavimento) para predecir la evolución de las grietas detectadas. Esto permitiría priorizar reparaciones basándose no solo en estado actual sino en proyecciones de severidad futura.

* Validación en terreno: Realizar pruebas piloto en distintas condiciones reales de Paraguay (pavimentos nuevos vs viejos, clima lluvioso vs seco, etc.) para ajustar el modelo a escenarios locales. Esto mejoraría la robustez en situaciones que difieren del dataset original y fomentaría la confianza de los gestores públicos en la herramienta.

Finalmente, es importante resaltar la relevancia de estos resultados en el contexto latinoamericano, donde actualmente existen muy pocas iniciativas locales de IA aplicada a infraestructura vial (Guevara, 2025). Nuestro estudio aporta un caso concreto de cómo una investigación nacional puede superar retos regionales y proveer soluciones tecnológicas avanzadas en un área crítica. Dado el rezago en investigación y desarrollo de IA en la región, estos hallazgos constituyen un paso clave para modernizar la gestión de carreteras en Paraguay y Latinoamérica. La utilización de redes neuronales convolucionales U-Net, con resultados de alta precisión demostrados aquí, muestra el potencial transformador de la IA en el mantenimiento vial. En definitiva, esta investigación sienta un precedente tecnológico y social: moderniza la forma de gestionar las vías públicas, promueve la seguridad vial y puede inspirar futuros desarrollos locales que aprovechen la inteligencia artificial para mejorar la calidad de la infraestructura en la región.

6. Declaración de financiamiento

La presente investigación se llevó a cabo con financiación propia.

7. Declaración de conflictos de intereses

Los autores declaran no tener conflictos de intereses.

8. Declaración de autores

Los autores aprueban la versión final del artículo.

9. Contribución de los autores

Autor	Contribución
Luz Rashell Soto Olguín	Elección del tema de investigación, introducción, diseño de la metodología, investigación, redacción del borrador y versión final.
Fredy Gabriel Ramirez Villanueva	Elección del tema de investigación, introducción, diseño de la metodología, investigación, redacción del borrador y versión final.
Héctor Ramiro Estigarribia Barreto	Elección del tema de investigación, introducción, diseño de la metodología, investigación, redacción del borrador y versión final.

10. Referencias Bibliográficas

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2016). *Tensorflow: Large-scale machine learning on heterogeneous distributed systems*. arXiv. https://doi.org/10.48550/ARXIV.1603.04467

Ali, L., AlJassmi, H., Swavaf, M., Khan, W. y Alnajjar, F. (2024). Rs-net: Arquitectura de red U-Net Sharp residual para la segmentación y evaluación de la severidad de grietas en pavimentos. *Journal of Big Data*, *11* (1), 116. https://doi.org/10.1186/s40537-024-00981-y

Alzamora, P. (2023). *Segmentación semántica de imágenes con Deep Learning*. Data Machine Learning Visualization. https://blog.damavis.com/segmentacion-semantica-de-imagenes-con-deep-learning

Astute Analytica. (2024). *Road Maintenance Market to Drive Fast to Reach Valuation of USD 23.39 Billion By 2032*. GlobeNewsWire: https://www.globenewswire.com/news-release/2024/05/07/2876916/0/en/Road-Maintenance-Market-to-Drive-Fast-to-Reach-Valuation-of-USD-23-39-Billion-By-2032-Astute-Analytica.html

Bradski, G. (2000). *The OpenCV Library*. Dr. Dobb's Journal of Software Tools. https://drdobbs.com/open-source/the-opencv-library/184404319

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q. y Wang, M. (2023). Swin-unet: Transformador puro tipo Unet para la segmentación de imágenes médicas. En L. Karlinsky, T. Michaeli y K. Nishino (Eds.), *Visión por Computador – Talleres ECCV 2022* (pp. 205-218). Springer Nature Suiza. https://doi.org/10.1007/978-3-031-25066-8_9

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. y Adam, H. (2018). Codificador-decodificador con convolución separable de alto rendimiento para la segmentación semántica de imágenes. En V. Ferrari, M. Hebert, C. Sminchisescu y Y. Weiss (Eds.), *Visión por Computador – ECCV 2018* (pp. 833-851). Springer International Publishing. https://doi.org/10.1007/978-3-030-01234-2_49

Chen, S., Feng, Z., Xiao, G., Chen, X., Gao, C., Zhao, M. y Yu, H. (2024). Detección de grietas en pavimento basada en el modelo swin-unet mejorado. *Buildings*, *14* (5), 1442. https://doi.org/10.3390/buildings14051442

Clark, J. (2024). *Pillow (PIL Fork) Documentation: Version 10.4.0.* https://pillow.readthedocs.io/en/stable/

Cruz, G., Riobó, A., Pfeifer, M. y Duarte, D. (2024). *IA desde los cimientos: Desafíos y oportunidades en el contexto de América Latina y el Caribe*. Banco Interamericano de Desarrollo. https://doi.org/10.18235/0013275

DataScientest. (2021). *Convolutional Neural Network: definición y funcionamiento*. https://datascientest.com/es/convolutional-neural-network-es

Di Benedetto, A., Fiani, M. y Gujski, L. M. (2023). Arquitectura de CNN basada en U-net para la segmentación de grietas en carreteras. *Infraestructuras*, *8*(5), 90. https://doi.org/10.3390/infrastructures8050090

Dong, C., Li, L., Yan, J., Zhang, Z., Pan, H. y Catbas, F. N. (2021). Segmentación de grietas por fatiga a nivel de píxel en imágenes a gran escala de estructuras de acero mediante una red codificador-decodificador. *Sensors*, *21*(12), 4135. https://doi.org/10.3390/s21124135

Federal Highway Administration. (2014). *Distress Identification Manual for the Long-Term Pavement Performance Program*. (5th ed.). U.S. Department of Transportation. https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/13092/13092.pdf

Geeksforgeeks. (2025). *Image Edge Detection Operators in Digital Image Processing*. https://www.geeksforgeeks.org/image-edge-detection-operators-in-digital-image-processing/

Global Infrastructure Hub. (2022). *AI and deep learning for identifying pavement failures in Latin American and the Caribbean.* World Bank's Public-Private Infrastructure Advisory Facility (PPIAF). https://infratech.gihub.org/infratech-case-studies/ai-and-deep-learning-for-identifying-pavement-failures-in-latin-american-and-the-caribbean/

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

Guevara, T. (2025). *La inteligencia artificial avanza desigual en Latinoamérica, buscan llevarla al sector productivo*. Voz de América (VoA). https://www.vozdeamerica.com/a/ia-avanza-desigual-latinoamerica/8010993.html

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, Nueva Jersey, Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P.,… Oliphant, T. E. (2020). Programación de matrices con NumPy. *Naturaleza*, *585* (7825), 357-362. https://doi.org/10.1038/s41586-020-2649-2

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778. https://doi.org/10.1109/CVPR.2016.90

Kaveh, H., & Alhajj, R. (2024). Recent advances in crack detection technologies for structures: A survey of 2022-2023 literature. *Frontiers in Built Environment*, *10*, 1321634. https://doi.org/10.3389/fbuil.2024.1321634

IBM. (2024). *¿Qué es la segmentación de imágenes?*. https://www.ibm.com/es-es/topics/image-segmentation

Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization*. arXiv. https://doi.org/10.48550/ARXIV.1412.6980

Lau, S. L. H., Chong, E. K. P., Yang, X., & Wang, X. (2020). Automated pavement crack segmentation using u-net-based convolutional neural network. *IEEE Access*, *8*, 114892-114899. https://doi.org/10.1109/ACCESS.2020.3003638

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3431-3440. https://doi.org/10.1109/CVPR.2015.7298965

Madroñero Urcuqui, J. D., y Valencia López, Y. C. (2019). *Metodología para la identificación automática del deterioro en pavimento flexible, por medio de fotografías aéreas tomadas desde vehículos no tripulados*. Universidad del Valle. https://hdl.handle.net/10893/15476

Marín-Acevedo, E. A. (2019). *Detección automática de grietas y fisuras en pavimento por medio de fotos y redes neuronales convolucionales*. Universidad Católica de Oriente. https://hdl.handle.net/20.500.12516/419

Ministerio de Obras Públicas y Comunicaciones. (2019). *Manual de Carreteras del Paraguay.* https://apcarreteras.org.py/manual-de-carreteras-del-paraguay-rev-2019/

Nyathi, M. A., Bai, J., & Wilson, I. D. (2024). Deep learning for concrete crack detection and measurement. *Metrology*, *4*(1), 66-81. https://doi.org/10.3390/metrology4010005

Peng, Y., Chen, D. Z., & Sonka, M. (2025). U-net v2: Rethinking the skip connections of u-net for medical image segmentation. *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*, 1-5. https://doi.org/10.1109/ISBI60581.2025.10980742

Pfeifer, M., Fernandez, S., y Riobo, A. (2023). *Pavimenta2: infraestructura digital al servicio de los activos viales.* (BID) moviliblog. https://blogs.iadb.org/transporte/es/pavimenta2-infraestructura-digital-al-servicio-de-los-activos-viales/

Riobo, A., Pfeifer, M., y Calle Jordá, A. (2024). *ViaSegura: lecciones aprendidas e inteligencia artificial al servicio de la seguridad vial.* (BID) Moviliblog. https://blogs.iadb.org/transporte/es/viasegura-lecciones-aprendidas-e-inteligencia-artificial-al-servicio-de-la-seguridad-vial

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. En N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Vol. 9351, pp. 234-241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Sabouri, M., & Sepidbar, A. (2023). *SUT-Crack*. https://doi.org/10.17632/gsbmknrhkv.6

Soto Olguín, L. y Ramírez Villanueva, F. (2024). *Segmentación de grietas superficiales en pavimento asfáltico utilizando técnicas de visión artificial*. [Tesis de grado]. Universidad Nacional de Caaguazú. https://publicaciones.fctunca.edu.py/handle/123456789/128

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research, 15*(1), 1929–1958. https://dl.acm.org/doi/abs/10.5555/2627435.2670313

Szeliski, R. (2022). *Computer vision: Algorithms and applications*. Springer International Publishing. https://doi.org/10.1007/978-3-030-34372-9

Tan, M., & Le, Q. (2019). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.* Proceedings of the 36th International Conference on Machine Learning. https://proceedings.mlr.press/v97/tan19a.html

Tang, Y., Qian, Y., & Yang, E. (2022). Weakly supervised convolutional neural network for pavement crack segmentation. *Intelligent Transportation Infrastructure*, *1*, liac013. https://doi.org/10.1093/iti/liac013

Tello Cifuentes, L., Marulanda, J., y Thomson, P. (2021). detección de grietas en el pavimento usando técnicas de procesamiento de imágenes y redes neuronales artificiales. *Encuentro Internacional De Educación En Ingeniería*. http://dx.doi.org/10.26507/ponencia.1565

Yu, Y., Xia, W., Zhao, Z., & He, B. (2024). A lightweight and high-accuracy model for pavement crack segmentation. *Applied Sciences*, *14*(24), 11632. https://doi.org/10.3390/app142411632

Zhang, Z., He, Y., Hu, D., Jin, Q., Zhou, M., Liu, Z., Chen, H., Wang, H., & Xiang, X. (2025). Algorithm for pixel-level concrete pavement crack segmentation based on an improved U-Net model. *Scientific Reports*, *15*(1), 6553. https://doi.org/10.1038/s41598-025-91352-x

Zhang, Y., & Zhang , L. (2024). Detection of Pavement Cracks by Deep Learning Models of Transformer and UNet. IEEE Transactions on Intelligent Transportation Systems, 25(11). https://doi.org/10.1109/TITS.2024.3420763

Zhang, Q., Chen , S., Wu , Y., Ji , Z., Yan , F., Huang , S., & Liu , Y. (2024a). Improved U-net network asphalt pavement crack detection method. *PLoS ONE, 19*(5). https://doi.org/10.1371/journal.pone.0300679

Zhang, J., Sun, S., Song, W., Li, Y., & Teng, Q. (2024b). A novel convolutional neural network for enhancing the continuity of pavement crack detection. *Scientific Reports*, *14*(1), 30376. https://doi.org/10.1038/s41598-024-81119-1

Zhang, J., Xia, H., Li, P., Zhang, K., Hong, W., & Guo, R. (2024c). A pavement crack detection method via deep learning and a binocular-vision-based unmanned aerial vehicle. *Applied Sciences*, *14*(5), 1778. https://doi.org/10.3390/app14051778

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). *Pyramid Scene Parsing Network*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA. https://doi.org/https://doi.org/10.1109/CVPR.2017.660

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. En D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. S. Tavares, A. Bradley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan, & A. Madabhushi (Eds.), *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (Vol. 11045, pp. 3-11). Springer International Publishing. https://doi.org/10.1007/978-3-030-00889-5