

Análisis comparativo de modelos YOLO para la detección de defectos en infraestructura vehicular

Comparative analysis of YOLO models for defect detection in vehicular infrastructure

Duber Martínez Torres¹

Jairo Antonio Muñoz Arango²

Universidad Nacional Abierta y a Distancia, Colombia

Resumen

Este trabajo presenta un estudio preliminar que compara el desempeño del modelo YOLOv11 con algunas de sus versiones anteriores en la detección de defectos en vías vehiculares. Se empleó el conjunto de datos UDTIRI, compuesto por 1.000 imágenes etiquetadas, evaluando las versiones más ligeras de cada modelo bajo las mismas condiciones experimentales. Los resultados demuestran que YOLOv11 logra un mejor equilibrio entre precisión, sensibilidad y eficiencia. Este estudio subraya la importancia de futuras investigaciones para evaluar modelos con un mayor número de parámetros y explorar el impacto de técnicas como el preprocesamiento de datos y el ajuste de hiperparámetros, con el objetivo de mejorar la detección de defectos y optimizar el uso de estos modelos en el monitoreo de infraestructura vial.

Palabras clave: detección de defectos, modelos YOLO, aprendizaje profundo.

Abstract

YOLOv11 model with some of its earlier versions in the specific task of detecting road defects. The UDTIRI dataset, consisting of 1,000 labeled images, was used, and the lightweight versions of each model were evaluated under the same experimental conditions. The results show that YOLOv11 achieves a better balance between precision, recall, and

¹ <https://orcid.org/0000-0002-0061-7153> / duber.martinez@unad.edu.co

² <https://orcid.org/0000-0001-5264-7453> / jairo.munoz@unad.edu.co

efficiency. This study highlights the importance of future research to assess models with a larger number of parameters and to explore the impact of techniques such as data preprocessing and hyperparameter tuning, aiming to enhance defect detection and optimize the application of these models for monitoring road infrastructure.

Keywords: Defect detection, YOLO models. deep learning.

1. Introducción

El deterioro de las vías vehiculares es una problemática común en muchas ciudades del mundo y especialmente en Colombia. La tecnología de visión por computador se presenta como una herramienta eficaz para automatizar la inspección de pavimentos, donde las arquitecturas basadas en *deep learning* se han consolidado como una herramienta dominante en el análisis automático de imágenes (Ayman & Fakhr, 2023). En los últimos años se ha popularizado el uso de arquitecturas inicialmente destinadas para la identificación de objetos para que mediante un proceso de sintonización permitan detectar los defectos de interés. Uno de los modelos más utilizados con este enfoque son los basados en arquitecturas YOLO (Redmon *et al.*, 2016). La familia de redes YOLO en sus diferentes versiones, se ha constituido en una de las arquitecturas más populares para la identificación y localización de objetos en la imagen. Esto gracias a sus altas prestaciones de desempeño y la capacidad de detección de múltiples elementos en la imagen a una alta velocidad de procesamiento. Esta arquitectura ha sido explorada para la detección de anomalías en el pavimento con buenos resultados (Jeong & Kim, 2022; Wan *et al.*, 2022). Recientemente se presentó a la comunidad el último modelo de esta familia, denominado YOLOv11 (Ultralytics, n.d.), el cual promete mejoras en precisión y rendimiento respecto a sus predecesores. En este trabajo se presenta un estudio preliminar que permita comparar el desempeño del modelo YOLOv11 respecto a versiones previas en el ámbito específico de la detección de defectos en vías vehiculares.

2. Metodología

Datos de experimentación: para este trabajo se utilizó el conjunto de datos público UDTIRI (Guo *et al.*, 2024), ampliamente reconocido como benchmark. Este dataset está compuesto por 1.000 imágenes RGB etiquetadas y capturadas en diversos escenarios. Se dividió en tres

subconjuntos: 600 imágenes para entrenamiento, 100 imágenes para validación y 300 imágenes para prueba.

Modelos evaluados: la experimentación se realizó utilizando cuatro modelos de la familia YOLO, incluyendo el modelo de interés YOLOv11 y las tres versiones anteriores: YOLOv10, YOLOv9 y YOLOv8. Para cada uno de estos modelos, se emplearon las arquitecturas equivalentes con el menor número de parámetros, es decir, YOLOv11n, YOLOv10n, YOLOv9t y YOLOv8n, respectivamente. Dado que el objetivo de este estudio preliminar es comparar el desempeño del modelo más reciente con sus predecesores, todos los modelos fueron evaluados bajo condiciones similares y en su forma básica, sin aplicar preprocesamiento a las imágenes de entrada, aumento de datos o ajuste de hiperparámetros.

Todos los experimentos fueron implementados en PyTorch y ejecutados en una GPU Tesla T4, 15GB bajo el entorno de Google Colab. Cada modelo fue entrenado durante 500 épocas con detención temprana.

Métricas de evaluación: para comparar el desempeño del modelo YOLOv11 respecto a versiones previas se utilizaron como métricas, un conjunto de curvas que permiten establecer el equilibrio entre la capacidad de cada modelo de detectar los huecos reales presentes en la escena reduciendo a su vez la generación de falsas detecciones. Estas curvas denominadas en inglés como *Precision-Confidence*, *Recall-Confidence*, *Precision-Recall* y *F1-Confidence*, muestran las variaciones y relaciones entre la precisión y la sensibilidad (*recall*) al cambiar el umbral de confianza.

3. Resultados y discusión

Con el fin de identificar qué modelo logra el mejor rendimiento según las métricas analizadas, en la Figura 1 se combinan diversas curvas que permiten evaluar no solo valores puntuales, sino también el comportamiento y la estabilidad de cada modelo a través de varios umbrales. A partir de esta evaluación, se puede determinar el equilibrio entre la capacidad de cada modelo para detectar la mayor cantidad de verdaderos defectos sin clasificar incorrectamente áreas sanas como defectos.

Se observa que los cuatro modelos analizados tienen un comportamiento similar, con una precisión que cae considerablemente para umbrales de confianza inferiores a 0.2 dejando de detectar un importante número de defectos. Por otra parte, a partir de umbrales de confianza superiores a 0.8 la sensibilidad o *recall* comienza a decrementar

rápidamente con la generación de un alto número de falsos positivos. Aunque las diferencias no son demasiado grandes, en las cuatro gráficas se observa que el modelo YOLOv11 (curva roja), presenta en forma consistente, el mejor balance permitiendo un mejor ajuste entre la precisión y la sensibilidad (*Recall*) con un valor de umbral de confianza cercano a 0.7.

En la Tabla 1 se presentan los indicadores de complejidad computacional, expresados en GFLOPS (giga operaciones de punto flotante por segundo), así como la velocidad de procesamiento, medida por el tiempo de inferencia. En relación con los modelos y datos utilizados, se encontró que el modelo YOLOv11n tiene la menor complejidad computacional y es el segundo más rápido, superado únicamente por el modelo YOLOv8n.

Tabla 1. Indicadores de rendimiento

Modelo	GFLOPS	Tiempo inferencia (ms)
YOLOv11n	6.3	3.73
YOLOv10n	8.2	4.92
YOLOv9t	7.6	8.23
YOLOv8n	8.1	3.24

En la Figura 2 se muestran ejemplos de predicciones realizadas por el modelo YOLOv11 en algunas de las imágenes del conjunto de datos de prueba.

Alcance y limitaciones de este estudio: los resultados presentados en este documento cumplen con el objetivo planteado en este estudio preliminar: demostrar que el último modelo de la familia de detectores YOLO disponible hasta la fecha muestra ventajas respecto a sus predecesores en la tarea específica de detección de huecos en vías vehiculares. En investigaciones futuras, se planea ampliar el estudio para abordar aspectos como la inclusión de modelos más complejos de la misma familia, en sus versiones *small*, *medium* y *large*. Además, se evaluará el rendimiento con conjuntos de datos más extensos y se analizará el efecto del preprocesamiento de imágenes, el aumento de datos, y la optimización de los hiperparámetros de cada modelo.

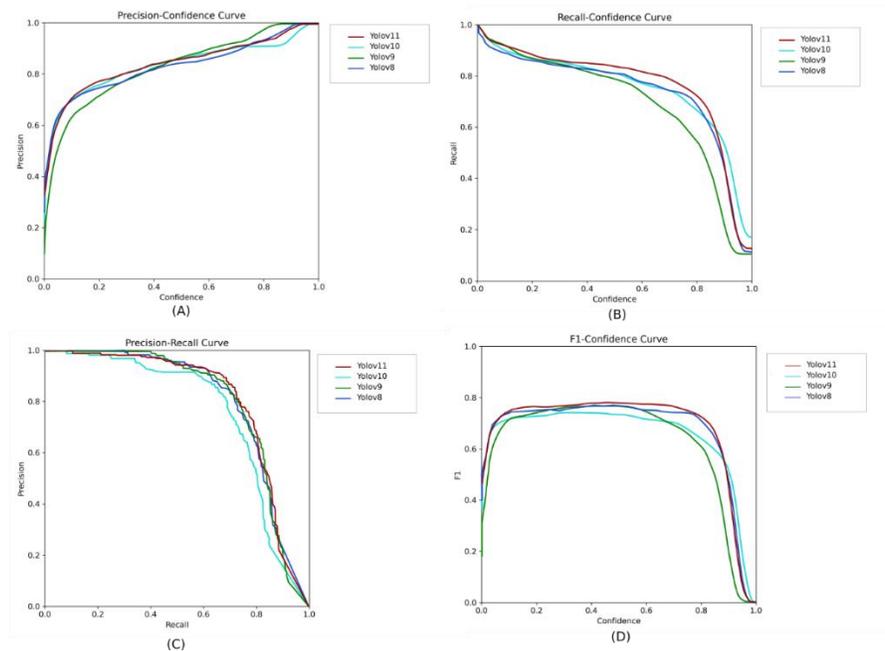


Figura 1. Curvas comparativas de desempeño. (A) Curvas Precision-Confidence, (B) Recall-Confidence, (C) Precision-Recall, (D) F1-Confidence.

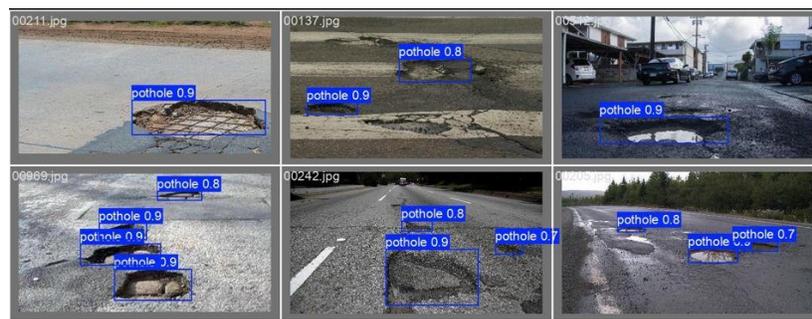


Figura 2. Ejemplo de predicciones realizadas por el modelo.

4. Conclusiones

En este estudio preliminar, se ha comparado el desempeño del modelo YOLOv11 con sus versiones anteriores, YOLOv10, YOLOv9 y YOLOv8, en el contexto específico de la detección de defectos en vías vehiculares. Las evaluaciones realizadas utilizando diversas métricas basadas en curvas de precisión y sensibilidad demuestran que YOLOv11 alcanza un equilibrio superior en términos de precisión y *recall*, en comparación con sus predecesores. Este mejor desempeño es consistente en todas las métricas analizadas, en adición se comprueba que, para este caso de estudio,

YOLOv11 destaca por ser más eficiente en velocidad de inferencia que YOLOv10 y YOLOv9.

En futuros trabajos, se contempla ampliar la investigación incorporando modelos más complejos de la familia YOLO y aplicar técnicas avanzadas de optimización para maximizar el rendimiento. De igual forma, ampliar el conjunto de datos de experimentación, así como explorar técnicas de preprocesamiento y aumento artificial de datos que puedan incrementar el desempeño de los modelos.

Referencias

- Ayman, H., & Fakhr, M. W. (2023). Recent computer vision applications for pavement distress and condition assessment. *Automation in Construction, 146*, 104664.
- Guo, S., Li, J., Feng, Y., Zhou, D., Zhang, D., Chen, C., ... & Fan, R. (2024). UDTIRI: An online open-source intelligent road inspection benchmark suite. *IEEE Transactions on Intelligent Transportation Systems*.
- Jeong, D., & Kim, J. (2022). Road Damage Detection using YOLO with Image Tiling about Multi-source Images. *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*.
<https://doi.org/10.1109/BigData55660.2022.10020282>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779–788.
- Ultralytics. (n.d.). YOLOv11 models. *Ultralytics Documentation*.
<https://docs.ultralytics.com/models/yolo11/>
- Wan, F., Sun, C., He, H., Lei, G., Xu, L., & Xiao, T. (2022). YOLO-LRDD: a lightweight method for road damage detection based on improved YOLOv5s. *Eurasip Journal on Advances in Signal Processing, 2022*(1). <https://doi.org/10.1186/s13634-022-00931-x>