

Implementación de un sistema de visión artificial basado en YOLO para detección de calidad de ruedas de autos

Rodríguez D. Axel
 Universidad Autónoma de Tamaulipas
 a2243018003@alumnos.uat.edu.mx

Resumen— En este trabajo se presenta la implementación de un modelo de YOLO para la detección del estado de las ruedas de un automóvil. Se diseñó y entrenó un modelo que permite identificar las ruedas y clasificarlas en dos estados: ruedas en buen estado y ruedas pinchadas, esto utilizando imágenes capturadas por una cámara y recolectadas de internet. Los resultados con los datos de testeó mostraron una precisión mAP50 del 98%. Posteriormente el sistema fue evaluado en diferentes escenarios para verificar su precisión y efectividad. Los resultados muestran un desempeño confiable del modelo, con una aceptable precisión en la detección en condiciones adversas, lo cual lo hace adecuado para aplicaciones de mantenimiento y sistemas de vigilancia de vehículos.

Abstract In this paper, the implementation of a YOLO model for detecting the condition of a car's wheels is presented. A model was designed and trained to identify wheels and classify them into two states: wheels in good condition and flat tires, using images captured by a camera and collected from the internet. The results with the test data showed an mAP50 accuracy of 98%. Subsequently, the system was evaluated in different scenarios to verify its precision and effectiveness. The results demonstrate a reliable performance of the model, with acceptable accuracy in adverse conditions, making it suitable for maintenance applications and vehicle surveillance systems.

I. INTRODUCCIÓN

La seguridad y el mantenimiento de vehículos son aspectos críticos en la industria automotriz y en tráfico y flujo vehicular. Las ruedas juegan un papel fundamental en el rendimiento y la seguridad de un automóvil, ya que su desgaste o daño puede derivar en accidentes graves.

En la actualidad, la mayoría de los defectos de los neumáticos se detectan mediante imágenes de rayos X de los neumáticos con una máquina de prueba de rayos X. El método de identificación y clasificación de los defectos de los neumáticos mediante observación manual es ineficiente y tiene diferentes criterios de evaluación [1].

La visión por computadora es una disciplina que permite a las máquinas "mirar". Hace que las computadoras y otros dispositivos tomen el lugar de los ojos humanos para la detección de objetos y el procesamiento adicional de imágenes. El algoritmo de detección de objetivos basado en el aprendizaje profundo puede extraer automáticamente características de la imagen, retener más detalles que otros tipos de métodos y tiene una gran portabilidad.

II. ESTADO DEL ARTE

A. Visión Artificial en la Industria Automotriz

La visión artificial ha sido ampliamente adoptada en la industria automotriz para tareas de inspección y control de calidad, proporcionando soluciones automatizadas que mejoran la precisión y eficiencia en la detección de defectos. Los sistemas de visión artificial permiten la automatización de procesos, reduciendo la intervención humana y mejorando la consistencia en la inspección de productos [1].

B. Arquitectura YOLO11 (You Only Look Once)

YOLO es una de las arquitecturas de detección de objetos más eficientes en términos de velocidad y precisión. Su estructura de red convolucional unificada permite la detección de objetos en tiempo real, lo cual es esencial para aplicaciones industriales donde el tiempo de respuesta es crítico [2].

C. Aplicaciones de YOLO en Inspección de Calidad

Diversos estudios han implementado la arquitectura YOLO para tareas de inspección de calidad en entornos industriales, incluyendo la detección de defectos en productos y componentes automotrices. Estos estudios destacan la capacidad de YOLO para adaptarse a diversas condiciones de iluminación y variabilidad de los objetos inspeccionados [2].

D. Retos en la Implementación de Sistemas de Visión Basados en YOLO

La implementación de sistemas de visión basados en YOLO enfrenta varios desafíos, como la necesidad de gran capacidad de cómputo para el entrenamiento del modelo y la dificultad para manejar datos no balanceados en la detección de defectos menores. Sin embargo, los avances en hardware y técnicas de optimización han permitido el despliegue de modelos más ligeros que pueden ser implementados en dispositivos de bajo consumo [3].

D. Transferencia de Aprendizaje en Modelos YOLO

La transferencia de aprendizaje permite utilizar modelos preentrenados en conjuntos de datos genéricos para adaptar rápidamente las capacidades de detección de YOLO a conjuntos de datos específicos de la industria automotriz, como los defectos en ruedas. Esto reduce significativamente el tiempo y recursos de entrenamiento, permitiendo una implementación rápida y eficaz en líneas de producción [3].

III. METODOLOGÍA

Para garantizar un proceso eficiente de diseño, entrenamiento, y evaluación del modelo de YOLO, la metodología se dividió en 5 partes principales.

A. Recolección de datos

La primera fase consistió en la recopilación de imágenes de ruedas de automóviles en diferentes condiciones. Se buscó crear un conjunto de datos variado que incluyera imágenes de ruedas en buen estado y ruedas desinfladas de diferentes modelos de carros. Las imágenes se obtuvieron de dos maneras principales:

Captura Directa: Se utilizaron cámaras convencionales para capturar imágenes de ruedas en diferentes ambientes y condiciones de iluminación, simulando entornos reales donde el sistema podría ser implementado.



Fig. 1. Ejemplo de foto de ruedas de carro por captura directa.

Busquedas de Internet: Se recolectaron imágenes de ruedas de automóviles a través de internet, utilizando palabras clave específicas en motores de búsqueda. Se filtraron imágenes irrelevantes y se seleccionaron aquellas que mostraban claramente el vehículo completo con sus ruedas en los estados de interés.



Fig. 2. Ejemplo de foto de ruedas de carro por búsquedas de internet.

Una vez recolectadas suficientes imágenes, se etiquetaron manualmente utilizando la plataforma Roboflow. Esta fase permitió contar con un dataset robusto que posteriormente fue dividido en conjuntos de entrenamiento (70%), validación (20%) y prueba (10%).

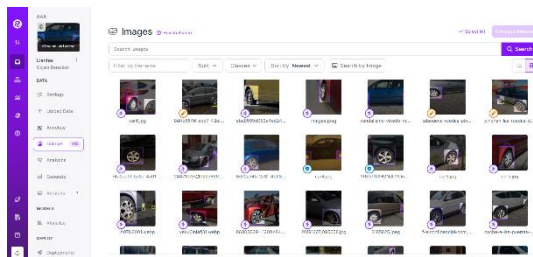


Fig. 3. Etiquetado de imágenes en la plataforma Roboflow.

B. Preprocesamiento de imágenes

El preprocesamiento de las imágenes fue una etapa fundamental para garantizar la calidad de los datos de entrada al modelo. La plataforma Roboflow también permite realizar estos ajustes, se realizaron los siguientes pasos:

Redimensionamiento: Todas las imágenes fueron redimensionadas a 640x640 píxeles para ajustarse al tamaño de entrada requerido por la arquitectura YOLO11.

Aumento de Datos (Data Augmentation): Para mejorar la capacidad de generalización del modelo y evitar el sobreajuste, se aplicaron técnicas de aumento de datos, tales como rotación, cambio de brillo y contraste, adición de ruido, y recorte aleatorio. Estas técnicas permitieron incrementar la diversidad del dataset de entrenamiento sin necesidad de recolectar nuevas imágenes.

C. Entrenamiento del modelo YOLO11

Se utilizó la arquitectura YOLO11 debido a su capacidad para realizar detección de objetos en tiempo real, manteniendo un equilibrio entre precisión y velocidad, además de ser la arquitectura más reciente y optimizada. El proceso de entrenamiento se realizó utilizando un equipo de cómputo con un procesador Intel CORE i5 de 7ma generación, 16 GB de ram, disco duro de estado sólido de 960 GB, y una tarjeta gráfica dedicada Nvidia GeForce gtx 1050 con el sistema operativo Windows 10, asegurándose de tener los recursos para acelerar los cálculos y optimizar el tiempo de entrenamiento. Las principales configuraciones del entrenamiento fueron las siguientes:

Parámetros de Entrenamiento: Se entrenó el modelo yolov11n.pt descargado del repositorio oficial de YOLO durante 100 épocas, con un parámetro imgSize para especificar el tamaño de las imágenes de entrenamiento (640x640).

Validación durante el Entrenamiento: Se realizó una validación continua con el conjunto de validación, monitoreando métricas como la precisión promedio (mAP) y la pérdida total para detectar sobreajuste.

D. Evaluación del modelo

Una vez completado el entrenamiento, el modelo fue evaluado utilizando el conjunto de prueba para medir su desempeño en términos de precisión, recall, y mAP50 (Mean Average Precision at IoU=0.50). Se generó una matriz de confusión para analizar los falsos positivos y falsos negativos, lo cual ayudó a identificar los casos en los que el modelo presentaba dificultades de detección con ayuda de varias fórmulas de métricas de evaluación.

IV. RESULTADOS Y CONCLUSIONES

Finalmente, el modelo fue optimizado utilizando la librería de Python ultralytics para la integración del modelo YOLO11 en un entorno de desarrollo sencillo para ser probado con imágenes ajenas al dataset original. Además, se probó la capacidad del sistema para procesar imágenes en tiempo real.

A partir de la matriz de confusión generada por el entrenamiento del modelo YOLO mostrado en la Tabla 1. se

evaluó el rendimiento final del modelo en la Tabla 2 para cada clase con las métricas “Recall”, “Precision”, “Accuracy” [4], “F1-score” que son factores importantes de evaluación para modelos de clasificación y definidas como se muestra a continuación:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Donde:

TP (True Positives): Número de predicciones correctas donde el modelo identificó correctamente una clase positiva (por ejemplo, identificó una rueda pinchada cuando realmente estaba pinchada).

TN (True Negatives): Número de predicciones correctas de la clase negativa (por ejemplo, identificó correctamente una rueda en buen estado).

FP (False Positives): Número de casos en los que el modelo identificó incorrectamente una clase positiva (por ejemplo, identificó una rueda pinchada cuando en realidad estaba en buen estado).

FN (False Negatives): Número de predicciones incorrectas donde el modelo no detectó una clase positiva que sí estaba presente.

Accuracy mide la proporción de predicciones correctas realizadas por el modelo en relación con el total de predicciones realizadas. Es una métrica global que muestra qué tan bien se desempeña el modelo en general. En este proyecto, la accuracy refleja cuántas ruedas fueron correctamente clasificadas como en buen estado o pinchadas sobre el total de imágenes analizadas. Sin embargo, en conjuntos de datos desbalanceados (por ejemplo, si hay muchas más ruedas en buen estado que pinchadas), la accuracy podría no ser una métrica completamente representativa del rendimiento del modelo.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

El recall, también conocido como sensibilidad o tasa de verdaderos positivos, mide la capacidad del modelo para detectar todos los casos positivos de una clase específica. En el contexto de la detección de ruedas, un recall alto indica que el modelo puede identificar la mayoría de las ruedas pinchadas. Es útil en aplicaciones donde es crucial detectar todos los casos positivos, aun si eso significa obtener algunos falsos positivos. Esto es particularmente importante para evitar fallos en sistemas de mantenimiento preventivo donde es preferible alertar sobre posibles problemas.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

La precisión (precision) indica la exactitud del modelo al identificar una clase positiva. Un valor alto de precisión significa que el modelo comete pocos errores al clasificar como positiva una muestra. En el contexto del proyecto, una alta precisión implica que cuando el modelo detecta una rueda pinchada, es probable que efectivamente esté pinchada. Es especialmente importante cuando se busca minimizar el número de falsas alarmas en el sistema, lo que es útil en escenarios donde

los recursos de mantenimiento son limitados y se desea evitar inspecciones innecesarias.

$$F1\ score = \frac{2 \times Precision \times Recall}{Recall + Precision} \quad (4)$$

El F1-score es la media armónica entre la precisión y el recall, y proporciona un balance entre ambas métricas. Es especialmente útil cuando existe un desbalance entre las clases positivas y negativas, ya que combina tanto la capacidad del modelo de detectar positivos como la exactitud de esas detecciones. En el contexto de la detección de ruedas, un F1-score alto indica que el modelo tiene un buen equilibrio entre no perder casos de ruedas pinchadas (recall) y ser preciso al hacer las detecciones (precision). Esto lo hace ideal para evaluar el rendimiento global del modelo cuando se busca un balance entre la sensibilidad y la precisión de las detecciones.

$$mAP = \frac{1}{|classes|} \sum_{c \in classes} \frac{\#TP(c)}{\#TP(c) + \#FP(c)} \quad (5)$$

Existen otras características como el mAP50 (Mean Average Precision) que es una métrica proporcionada por el entrenamiento de YOLO e indica la precisión media calculada con un umbral de intersección sobre unión (IoU) de 0,5. Es una medida de la precisión del modelo considerando sólo las detecciones "fáciles".

TABLA 1
MATRIZ DE CONFUSIÓN

		Valores Reales	
		Rueda Buena	Rueda Pinchada
Valor Predicción	Rueda Buena	33	1
	Rueda Pinchada	0	13

TABLA 2
RESULTADOS DE MÉTRICAS DE EVALUACIÓN PARA EL MODELO

0	RUEDA BUENA	RUEDA PINCHADA
RECALL	.9706	1
PRECISION	1	.9286
ACCURACY	.9787	.9787
F1-SCORE	.9851	.9620
MAP50	.9809	
MAP50-95	.6400	

Para verificar la efectividad del modelo entrenado en condiciones reales, se diseñaron pruebas en diferentes escenarios:

Entornos de Iluminación: Se probaron imágenes de ruedas tomadas en diferentes condiciones de iluminación, como luz natural, sombra y luz artificial, para evaluar la capacidad del modelo de adaptarse a variaciones en la iluminación.



Fig. 4. Pruebas del modelo entrenado con diversas detecciones.

Ángulos de Captura: Se grabó un video donde se obtienen perspectivas de ruedas desde distintos ángulos para analizar la precisión del modelo cuando la perspectiva de la rueda no es frontal.



Fig. 5. Pruebas del modelo entrenado en video.

REFERENCIAS

- [1] Saleh, R.A.A., Konyar, M.Z., Kaplan, K., and Ertunc, H.M.: 'Tire Defect Detection Model Using Machine Learning'. Proc. 2022 2nd International Conference on Emerging Smart Technologies and Applications (eSmarTA)2022 pp. Pages
- [2] Lwakatare, L.E., Raj, A., Crnkovic, I., Bosch, J., and Olsson, H.H.: 'Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions', Information and Software Technology, 2020, 127
- [3] Jing, J., Zhuo, D., Zhang, H., Liang, Y., and Zheng, M.: 'Fabric defect detection using the improved YOLOv3 model', Journal of Engineered Fibers and Fabrics, 2020, 15
- [4] Filoteo-Razo, J.D., Elizondo-Leal, J.C., Martinez-Angulo, J.R., Barron-Zambrano, J.H., Díaz-Manriquez, A., Saldivar-Alonso, V.P., Estudillo-Ayala, J.M., and Rojas-Laguna, R.: 'Non-Invasive Optoelectronic System for Color-Change Detection in Oranges to Predict Ripening by Using Artificial Neural Networks', IEEE Photonics Journal, 2023, 15, (5), pp. 1-10