Evaluación de Detectores de Objetos Generales para Vigilancia Aérea

La vigilancia es crucial para lograr la seguridad y eficiencia en una ciudad. Los sistemas de vigilancia tradicionales requieren de intervención humana constante y son propensos a errores. Este estudio compara detectores generales de objetos destacados en la literatura de vigilancia por imágenes aéreas asistida por visión artificial, evaluando su rendimiento en el dataset DOTA. Los resultados identifican los detectores más eficientes para distintas áreas de vigilancia, obteniendo información valiosa para agencias de seguridad interesadas en mejorar sus sistemas con inteligencia artificial (Martinez-Escobar et al., 2025).

Introducción

Los sistemas de vigilancia tradicionales dependen altamente de personal de seguridad que esté constantemente monitoreando áreas de interés; no obstante, la mayoría del tiempo los sistemas de vigilancia tienen puntos ciegos, son ineficientes o costosos (Alfara et al., 2020; Martinez-Escobar et al., 2025). Estos sistemas pueden ser asistidos con técnicas de visión artificial y aprendizaje profundo; sin embargo, estas incorporaciones presentan serios retos, tales como los cambios en perspectivas, orientación, formas de los objetos, variaciones significativas en la resolución de entrada u oclusión de objetos.

Este estudio busca comparar el rendimiento de diversos detectores de objetos generales para imágenes aéreas, permitiendo descubrir cuáles son mejores para campos específicos de vigilancia. Se utilizó el Dataset of Object deTection of Aerial images (DOTA) como benchmark estándar para evaluar los modelos, el cual consiste enteramente en imágenes satelitales. Este artículo muestra la evaluación de modelos de detección de objetos del estado del arte, explorando sus dificultades prácticas, tales como altas resoluciones y recursos limitados.

Metodología

Selección del dataset

DOTA (Dataset of Object Detection for Aerial Images) contiene 15 categorías de objetos en 2,806 imágenes, con un total de 400,000 instancias (Ding et al., 2021). El dataset es un gran benchmark debido a la diversidad de objetos en cuanto a tamaños, así como a la dimensión de sus imágenes, que va desde los 800×800 píxeles hasta los 20,000×20,000. Para el entrenamiento, validación y pruebas de los modelos, se conservaron las divisiones planteadas por los autores de DOTA.

Selección de los modelos

Para la selección de modelos, se buscaron detectores en tiempo real que tuvieran fuertes capacidades de generalización, así como modelos con arquitecturas nuevas basadas en otras ya conocidas. Entre los modelos seleccionados se encuentran las versiones de YOLO v5, v8, v9 y NAS; los modelos Roboflow Object Detection Fast 2.0 y 3.0; y FasterRCNN.

Evaluación de rendimiento

El rendimiento de los modelos se basó en tres métricas: precisión, sensibilidad y mAP50 %. La precisión es la razón entre el número total de positivos detectados y el número total de positivos verdaderos más el número total de falsos positivos (1). La sensibilidad es la razón de positivos verdaderos entre la suma de positivos verdaderos y falsos negativos (2). El mAP50 % es la precisión media promedio en un intervalo de confianza del 50 % (3).}

$$(1)\ precision = \frac{Positivos\ Verdaderos}{Positivos\ Verdaderos + Falsos\ Positivos}$$

$$(2)\ sensibilidad = \frac{Positivos\ Verdaderos}{Positivos\ Verdaderos + Falsos\ Negativos}$$

$$(3)\ mAP50\% = \frac{1}{n}\sum_{n=1}^{N} AP_n$$

Resultados

Evaluación de rendimiento

La tabla 1 muestra los resultados generales para los detectores de objetos. Se puede observar que el modelo con mejor mAP fue Roboflow 3.0 (33.8 %), el de mayor precisión fue YOLOv8 (72 %) y el de mayor sensibilidad fue Roboflow 2.0 (31.6 %).

Modelo	mAP50 %	Precisión (%)	Sensibilidad (%)
Faster RCNN	21.1	48.8	19.5
YOLOv5	28.4	63.2	27
YOLOv8	32.2	**72**	28.5
YOLOv9	33	68	29
YOLO NAS-S	26.8	70.5	23.2
Roboflow 2.0	33.6	66.1	**31.6**
Roboflow 3.0	**33.8**	69.6	30.7

Tabla 1. Métricas generales para cada modelo

Detección de objetos pequeños

Los objetos pequeños son los más difíciles de encontrar en imágenes aéreas, y en la evaluación particular de los modelos hay una clara tendencia hacia un menor rendimiento en clases de objetos más pequeños, especialmente aquellos que se encontraban en imágenes grandes y que, en consecuencia, fueron afectados por la pérdida de detalles debido a la reducción de escala. En la figura 1, se puede observar cómo los cambios de resolución afectan de manera significativa a dos modelos diferentes. Roboflow 2.0 (figura 1a), con un intervalo de confianza del 50 %, detecta con mayor precisión que su contraparte YOLO-NAS Small (figura 1b), el cual fue entrenado con menor resolución.

Figura 1ª. Roboflow 2.0 con un interval de confianza del 50%

Figura 1b. YOLO NAS Small con un intervalo de confianza del 50%

Figura 1. Imagen P0060 del dataset de validación de DOTA. Las cajas rojas son predicciones de piscinas, las cajas moradas son detecciones de vehículos pequeños y las cajas azules son detecciones de vehículos grandes.

Distribución de instancias en el dataset

En el dataset hay algunas clases que están considerablemente menos representadas que otras; tal es el caso de *container-crane* y *bridge*, las cuales tuvieron un mal rendimiento en todos los modelos debido a la escasez de instancias dentro del conjunto de datos. De igual manera, hay ciertas clases que se encontraban en imágenes de menor resolución, haciendo que se vieran menos afectadas por los efectos de la reducción de escala, como lo es la clase *tennis-court*, un objeto de gran tamaño que se encontraba, en promedio, en imágenes de resolución más baja que otras clases similares como *soccer-ball-field*.

Posibles aplicaciones y retos a futuro

En vigilancia urbana, donde la vigilancia se utiliza para la identificación de incidentes, vehículos robados o accidentes automovilísticos, es necesario disminuir la cantidad de falsos positivos para evitar desperdiciar recursos en problemas inexistentes. Por ello, modelos como YOLOv8 y YOLO-NAS, con una precisión alta, se

desempeñarían mejor en estas tareas. Para estudios futuros, es importante optimizar las técnicas de preprocesamiento para imágenes aéreas, con el fin de evitar que se vean afectadas por la disminución de resolución.

Referencias:

- Alfara, A. U. A., Ivanov, N. S., & Ivanenko, V. G. (2020). Research on the vulnerabilities of urban video surveillance systems. *2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, 222–224. https://doi.org/10.1109/EIConRus49466.2020.9039519
- Ding, J., Xue, N., Xia, G.-S., Bai, X., Yang, W., Yang, M., Belongie, S., Luo, J., Datcu, M., Pelillo, M., & Zhang, L. (2021). Object detection in aerial images: A large-scale benchmark and challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1. https://doi.org/10.1109/TPAMI.2021.3117983
- Martinez-Escobar, D., Caso-Benitez, M., Mila-Ceron, A., & Guzman, Z. (2025). Assessing general object detectors for aerial surveillance. In S. Nesmachnow & L. Hernández Callejo (Eds.), *Smart cities* (pp. 77–88). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-85324-1_6

Sobre los autores:

Diego Iván Martínez Escobar

Estudiante de Ingeniería en Sistemas Computacionales en la Universidad de las Américas Puebla. Explore Intern en Microsoft y miembro del Programa de Honores.

Contacto: diego.martinezer@udlap.mx

Zobeida J. Guzmán Zavaleta

Doctora y Maestra en Ciencias con Especialidad en Ciencias Computacionales por el Instituto Nacional de Astrofísica, Óptica y Electrónica. Licenciada en Ciencias de la Computación por la Benemérita Universidad Autónoma de Puebla.

Contacto: zobeida.guzman@udlap.mx