Aircraft detection in aerial imagery based on YOLO architectures

Vita Kashtan[†], Yevhen Radionov[†] and Volodymyr Hnatushenko[*,†]

Dnipro University of Technology, Dmytra Yavornytskoho Ave 19, Dnipro, 49005, Ukraine

Abstract

The study is devoted to determining the most efficient YOLO-based architecture for the task of aircraft detection in high-resolution aerial imagery. A comparative analysis was conducted across YOLO models v8 through v11 under three experimental conditions: using pre-trained (raw) models, fine-tuning the models on a domain-specific dataset, and fine-tuning models to a dataset enhanced through a proposed image preprocessing method. The evaluation considered both accuracy and inference performance metrics. The proposed methodology reduced the false negative rate from 19.5% to 3.2% at a confidence threshold of 0.75, underscoring its effectiveness in enhancing target visibility under challenging imaging conditions such as low contrast or background clutter.

Keywords

machine learning, aircraft detection, object detection, optical image preprocessing, YOLO

1. Introduction

Accurately detecting objects in imagery represents a cornerstone of contemporary computer vision, with far-reaching implications for surveillance, defense, and autonomous systems [1]. In aerospace, precisely identifying aircraft in aerial imagery is fundamental to performing mission-critical operations such as air traffic control, border surveillance, and search-and-rescue efforts. Aircraft detection in high-resolution remote sensing imagery introduces specific challenges requiring dedicated evaluation strategies [2]. Unlike general-purpose datasets, aerial imagery often contains densely populated scenes, varying viewing angles, and non-uniform illumination conditions, all of which can degrade the performance of detectors trained on conventional benchmarks.

Recent advancements in deep learning-based object detection models have significantly improved the capabilities of visual recognition systems. Among these, the You Only Look Once (YOLO) family of architectures has emerged as a leading approach due to its superior real-time inference performance and high detection accuracy. The YOLO framework has undergone continuous evolution—from the original YOLOv1 to the most recent YOLOv11—with each version introducing architectural refinements and achieving improved results across widely used benchmarks such as COCO and PASCAL VOC [3]. However, the direct applicability of these models to domain-specific tasks, such as aircraft detection in remote sensing imagery, remains non-trivial. It is primarily due to the unique characteristics of aerial data, including significant variations in object scale, visually complex and cluttered backgrounds, low spatial resolution, and partial occlusions. These factors can affect the detection accuracy of different YOLO versions and must be carefully considered when selecting a model for operational use in aerospace contexts.

✉ vitalionkaa@gmail.com (V. Kashtan); eugene.radionov@gmail.com (Y. Radionov); vvgnat@ukr.net (V. Hnatushenko)

ⓘD 0000-0002-0395-5895 (V. Kashtan); 0009-0002-2839-7161 (Y. Radionov); 0000-0003-3140-3788 (V. Hnatushenko)

2. Related works

Aircraft detection in aerial and satellite imagery has received growing attention in remote sensing and computer vision, mainly due to its strategic significance in civilian and defense-related contexts. Traditional approaches relied on handcrafted features and classical machine learning techniques, such as support vector machines, Haar-like features, or HOG descriptors. However, these methods typically struggle to generalize across varying imaging conditions and object appearances.

With the emergence of deep learning, convolutional neural networks have revolutionized object detection, enabling the automatic extraction of hierarchical features directly from raw image data. Region-based detectors, such as Faster R-CNN, and single-shot detectors, such as SSD and YOLO, have demonstrated superior accuracy and speed across diverse visual tasks. Among these, YOLO models have gained traction due to their unified detection pipeline, which allows for real-time inference without compromising accuracy. Numerous studies have employed different YOLO versions for object detection tasks, demonstrating their effectiveness in terms of speed and accuracy. In aircraft detection, YOLO-based models have been adopted to address challenges such as cluttered backgrounds, occlusion, and visually similar non-target objects.

In the study [3], the authors compared the YOLOv7, YOLOv8, and RT-DETR models for military aircraft detection tasks. Their findings indicate that YOLOv8 achieved the highest mAP at 94.0%, outperforming YOLOv7 and RT-DETR, which achieved mAP values of 90.2% and 92.7%, respectively. However, RT-DETR demonstrated better performance in terms of Recall, reaching 90.4%, compared to 88.1% for YOLOv8 and 82.7% for YOLOv7.

U-YOLO is an enhanced detection architecture [4] for vehicle detection in high-resolution remote sensing imagery. The proposed model demonstrated improved detection accuracy, yielding an increase ranging from 4.94% to 6.89% relative to the baseline YOLO model, depending on the specific dataset used. Furthermore, compared to conventional object detection frameworks such as RFBNet, M2Det, and SSD300, the U-YOLO model achieved performance gains ranging from 6.84% to 12.41%. To improve fine-grained aircraft recognition quality, authors proposed an efficient detection model called FGA-YOLO [5]. Extensive experimental evaluations conducted on MAR20 and FAIR1M aircraft recognition datasets demonstrate that the proposed method enhances the accuracy of fine-grained aircraft classification. YOLO-based models can be adapted and optimized for use with Synthetic Aperture Radar (SAR) data. In the paper [6], the authors presented YOLO-SAATD (SAR Airport and Aircraft Target Detector), specifically designed to detect airport facilities and aircraft targets in SAR data efficiently. The proposed model demonstrated notable performance improvements, achieving a 1–2% increase in mAP50 and a 15% enhancement in detection frame rate when evaluated on the SAR-AIRPort-1.0 and SAR-AirCraft-1.0 benchmark datasets.

Techniques and innovations developed for ship detection in satellite imagery can be adapted for aircraft detection tasks, owing to the similar operational conditions and challenges encountered in satellite-based object detection, such as small target size, low contrast, complex backgrounds, and varying illumination. In the study [7], authors achieved an accuracy of more than 84% in semantic segmentations for ships on satellite imagery using a two-step approach: building a classifier based on XCeption and using a baseline U-Net model with Resnet18 as an encoder for exact segmentation.

Despite the availability of a wide range of advanced technologies and object detection models, aircraft detection in high-resolution satellite imagery continues to present significant challenges, including low contrast between aircraft and background, small object size, complex scene composition, and visually similar non-target objects. While existing YOLO-based methods have demonstrated promising speed and detection accuracy results, they often rely on high-quality image inputs and perform suboptimally under adverse visual conditions.

The aim of this research is to determine the most effective YOLO-based architecture for the task of aircraft detection in high-resolution aerial imagery. To this end, a systematic comparative analysis involves four recent YOLO versions: YOLOv8 through YOLOv11—evaluated across key performance metrics, including detection accuracy, inference speed, and robustness under diverse imaging conditions. Beyond baseline assessment, the study introduces an enhanced preprocessing pipeline

integrating histogram equalization and bilateral filtering. This preprocessing strategy aims to improve image contrast and edge preservation, thereby enhancing the overall quality of input data and potentially increasing detection performance across all evaluated models.

3. Proposed methodology

To enhance the confidence and reliability of aircraft detection in remote sensing imagery, we propose the methodology shown in Figure 1.
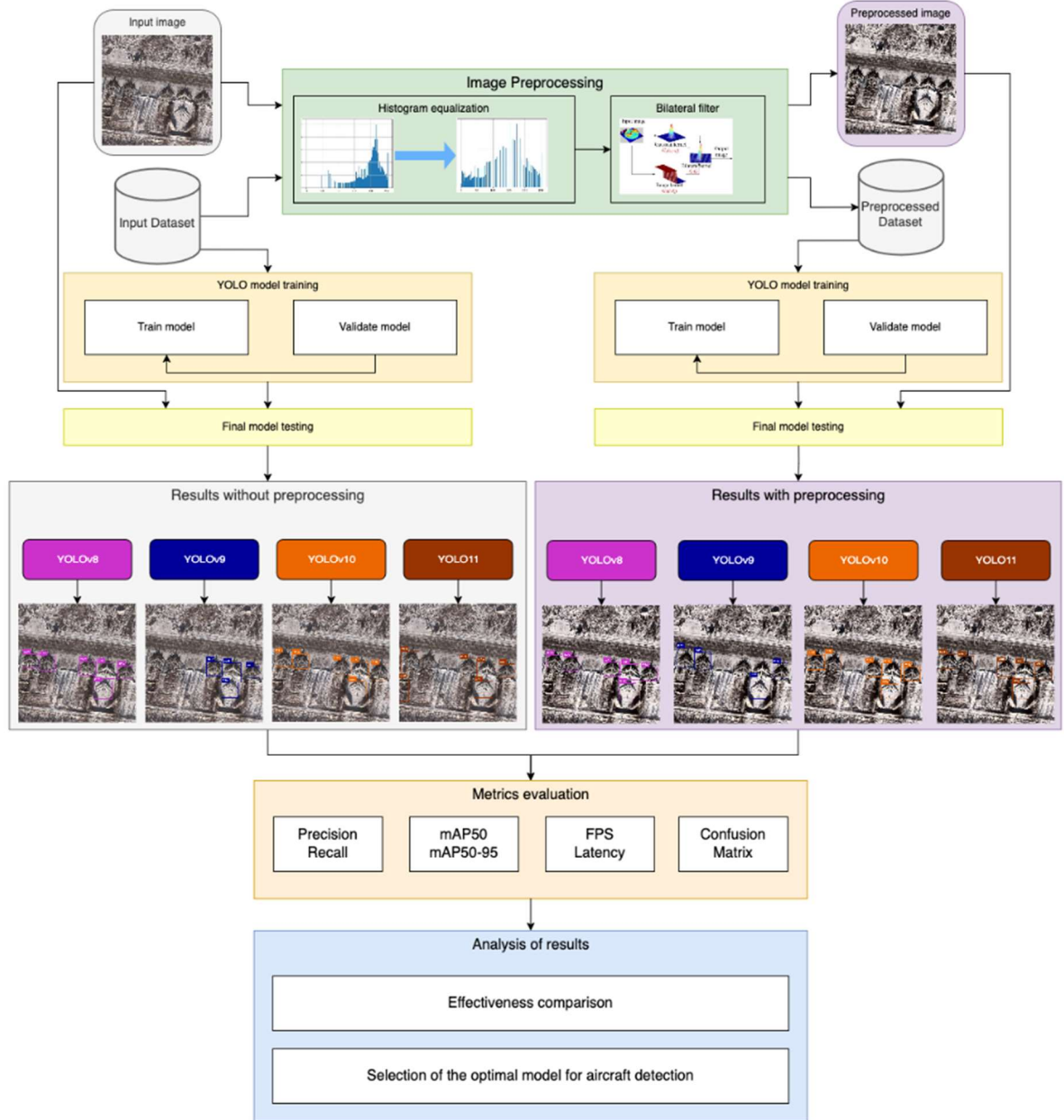

Figure 1: Diagram of the proposed methodology.

The proposed methodology is structured into six sequential stages.

The first stage consists of digital image and dataset acquisition. In this study we used the Military Aircraft Recognition dataset [8] that is publicly available on the Kaggle platform. The dataset contains 3842 high-resolution remote sensing images containing a total of 22341 annotated aircraft instances with 20 distinct aircraft types. The images in the dataset have a wide range of visual conditions, including variations in lighting, contrast, and image clarity. The dataset has been divided into three parts: 70% for training, 20% for validation and 10% for testing.

The second stage of the proposed methodology involves applying preprocessing techniques specifically designed to enhance the visual quality of the input data, thereby improving object detection performance, particularly in low-contrast or noisy conditions [9]. Preprocessing constitutes a component of the overall approach and comprises two primary operations: histogram equalization and bilateral filtering. This combination aims to increase the visibility of target objects by enhancing contrast while simultaneously reducing background noise through edge-preserving smoothing, ultimately facilitating more accurate detection results [10, 11]. The initial preprocessing stage entails applying histogram equalization to each input image to enhance global contrast. This technique expands the dynamic range of intensity values, thereby improving the separability of salient features and facilitating more effective execution of subsequent tasks such as object detection and classification. By redistributing pixel intensity values across the entire available dynamic range, histogram equalization enhances the contrast between aircraft and background elements, especially in low-contrast scenes where aircraft may otherwise be difficult to discern. This method is well-established in image processing because it can improve feature visibility by modifying the intensity distribution such that the resulting histogram approximates a uniform distribution [12]. The transformation function utilized in histogram equalization is derived from the original image histogram's cumulative distribution function (CDF). Following histogram equalization, bilateral filtering, as defined in Equation 1, is applied to reduce image noise while preserving important edge information. Bilateral filtering is a non-linear, edge-preserving, noise-reducing smoothing technique extensively utilized in computer vision and image processing [13]. Unlike conventional linear filters, such as Gaussian blur, which uniformly smooth the image and tend to degrade edge sharpness, the bilateral filter incorporates both spatial proximity and intensity similarity to compute the filtered pixel value. It reduces noise without compromising critical structural details such as edges and contours.

$$I^{filtered}(x) = \frac{1}{W_p} \sum_{x_i \in \Omega} I(x_i) f_r(\|I(x_i) - I(x)\|) g_s(\|x_i - x\|) \,,$$

$$W_p = \sum_{x_i \in \Omega} f_r(\|I(x_i) - I(x)\|) g_s(\|x_i - x\|) \,,$$

(1)

where $I^{filtered}$ is the filtered image, I is the original input image, x are the coordinates of the current pixel to be filtered, Ω is the window centered in x, so $x^i \in \Omega$ in another pixel, f_r is the range kernel for smoothing differences in intensities, g_s is the spatial (or domain) kernel for smoothing differences in coordinates [14].

Figure 2 provides a visual illustration of the qualitative improvements achieved through the proposed preprocessing approach, which combines histogram equalization and bilateral filtering applied to raw aerial imagery. Figure 2a depicts the original image, which typically exhibits common challenges characteristic of remote sensing data, including suboptimal contrast, haze, and subtle illumination variations. These factors can obscure fine details of aircraft and reduce edge distinctness, particularly against complex or shadowed backgrounds. In contrast, Figure 2b shows the image after histogram equalization, demonstrating a marked enhancement in overall contrast. This global contrast adjustment redistributes intensity values across the full dynamic range, rendering previously faint features more discernible. However, while histogram equalization improves contrast, it may also amplify noise or introduce undesirable artifacts. Figure 2c presents the result after bilateral filtering, which effectively smooths homogeneous regions while preserving critical edge and structural information. This edge-preserving smoothing is essential for aircraft detection, as it sharpens aircraft contours without blurring their distinct shapes or causing them to merge with the background. Collectively, the combined preprocessing steps yield a perceptibly clearer and more defined aerial scene, wherein aircraft outlines become more prominent and distinguishable from the surrounding environment. Its enhanced visual clarity constitutes a foundational improvement that may contribute to better detection performance by YOLO models in subsequent processing stages.
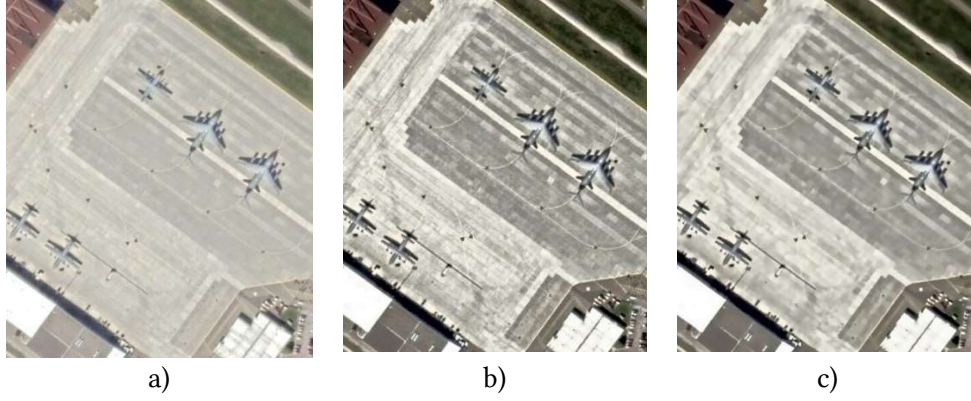
Figure 2: Remote sensing image: a) original; b) histogram equalized; c) bilateral filter.

The third stage involves training various YOLO models on original and preprocessed datasets. Since its inception, the YOLO object detection framework has undergone substantial architectural evolution, with successive versions introducing innovations to enhance detection accuracy, computational efficiency, and task versatility. YOLOv8, released by Ultralytics in 2023, represents a significant advancement by adopting an anchor-free detection paradigm, simplifying the architecture, and improving performance in detecting small-scale objects. It supports a unified framework for multiple vision tasks, including object detection, instance segmentation, and classification, and extends applicability to object tracking, pose estimation, and oriented bounding boxes [15].

YOLOv9 (2024) further introduces the Programmable Gradient Information mechanism and the Generalized Efficient Layer Aggregation Network, addressing information bottlenecks and improving lightweight model performance without sacrificing inference speed [16]. YOLOv10, also from 2024, implements a non-maximum suppression (NMS)-free training paradigm that enables end-to-end deployment and reduces computational costs. Additional architectural enhancements, such as spatial-channel decoupled down sampling and large-kernel convolutions, contribute to increased accuracy and efficiency, while a dual assignment strategy lowers latency for real-time detection [17].

The latest iteration, YOLOv11, continues this trajectory by integrating novel components, including the C3k2 block, Spatial Pyramid Pooling-Fast, and a Convolutional Block with Parallel Spatial Attention. These advances substantially improve feature extraction and object localization, enhancing performance across various vision tasks such as oriented object detection and instance-level segmentation [18, 19].

The fourth stage evaluates model performance using a comprehensive set of metrics. The metrics computed include Precision, Recall, mAP50, and mAP50-95 to quantify detection accuracy. In addition to these accuracy-oriented measures, efficiency metrics such as Frames Per Second and Latency were assessed to characterize the models' real-time inference capability and response time, respectively. Confusion matrices were constructed to provide deeper insights into classification performance, enabling detailed visualization of true positives, false positives, false negatives, and true negatives. The evaluation also includes a comparative analysis of models trained on preprocessed versus non-preprocessed datasets, thereby elucidating the impact of the preprocessing pipeline on detection performance, particularly in challenging conditions characterized by low contrast or image noise.

Finally, the optimal YOLO model architecture was selected based on a thorough trade-off analysis considering detection accuracy, inference speed, and robustness to low-contrast imagery. The chosen model represents the best overall compromise between Precision, Recall, computational efficiency, and reliability, thus constituting the most effective approach for accurate and efficient aircraft detection in high-resolution remote sensing imagery.

4. Experiment

4.1. Description of the experiment

The experiment aimed to evaluate and compare the performance of four YOLO model versions – YOLOv8, YOLOv9, YOLOv10, and YOLOv11 – on the task of aircraft detection using aerial imagery and compare them in the following states: not fine-tuned, fine-tuned, fine-tuned on the dataset with the proposed method applied. To conduct the experiment, the software was created in the Python programming language, using Ultralytics YOLO library for model training. The experiment was conducted on the following setup: CPU AMD Ryzen 5600X, GPU: MSI RTX 3080 Gaming X Trio, RAM: 32GB.

4.2. Metrics

The mean Average Precision (mAP) metric is employed for comprehensive assessment as YOLO-based models evaluate performance based on object classification and localization accuracy. The mAP metric integrates Precision and Recall over a range of confidence thresholds and is computed at varying Intersections over Union (IoU) thresholds. These thresholds define the minimum overlap required between the predicted and ground truth bounding boxes for a detection to be considered a True Positive. Varying the IoU threshold affects the classification of detections as True Positives or False Positives, thereby influencing the resulting Precision and Recall values. Precision is defined as the ratio of True Positive detections to the total number of positive detections made by the model at a given IoU threshold and is computed according to Equation 2:

$$Precision = \frac{TP}{TP + FP}, \tag{2}$$

where *TP* is true positive, *FP* is false positive.

As defined by Equation 3, Recall represents the proportion of True Positive detections identified by the model relative to the total number of ground truth instances. It quantifies the model's ability to detect all relevant objects at a given threshold correctly:

$$Recall = \frac{TP}{TP + FN}, \tag{3}$$

where *TP* is true positive, *FN* is false negative.

The general formulation for calculating mAP is provided in Equation 4:

$$mAP = \frac{1}{N} \sum_{n=1}^{N} AP_i, \tag{4}$$

where *N* is the total number of object classes, and AP_i denotes the Average Precision for the *i*-th class.

This metric reflects the meaning of the precision-recall integrals computed for each class, providing a single detection accuracy measure across multiple categories.

Specifying the IoU thresholds used in mAP calculations with numeric suffixes is customary. For instance, mAP50-95 indicates that the mAP was computed by averaging AP values over IoU thresholds ranging from 0.50 to 0.95 in increments of 0.05. In contrast, mAP50 refers to the AP value calculated at a single IoU threshold of 0.5.

The Average Precision (AP) for a given object class is computed by first ranking all predicted bounding boxes in descending order according to their confidence scores. Following this, Precision and Recall values are calculated at each detection threshold. The AP is then obtained by computing the area under the precision-recall (P–R) curve, which is estimated through numerical integration over all Recall levels. This process is formally expressed in Equation 5:

$$AP = \int_{r=0}^{1} p(r)dr \, , \tag{5}$$

where $p(r)$ represents the *Precision* as a function of *Recall.*

To assess the inference performance of each model, two key metrics were employed: Frames Per Second (*FPS*) and *Latency. FPS* measures the number of image frames that a model can process within one second, providing an indicator of real-time processing capability. In contrast, *Latency* quantifies the time required by the model to process a single image, typically expressed in milliseconds. Together, these metrics offer a comprehensive evaluation of the computational efficiency and responsiveness of each YOLO variant during inference.

In the fields of machine learning and computer vision, the confusion matrix is a fundamental tool for evaluating the performance of classification models [20]. It provides a comprehensive visualization of classification outcomes by quantifying the number of correct and incorrect predictions across all categories. Specifically, it reports the count of true positives, which are correctly identified positive instances, and true negatives, which are correctly identified negative instances. False positives correspond to negative samples that are incorrectly classified as positive, while false negatives denote positive samples that are mistakenly classified as negative. By capturing these four key components, the confusion matrix facilitates a detailed assessment of a model's predictive accuracy and error distribution.

5. Results and discussion

The performance of the non-fine-tuned YOLO models on the aircraft detection task, shown in Table 1, demonstrates strong baseline capabilities. Specifically, the models achieved high Precision values ranging from 0.90 to 0.96 and Recall values between 0.89 and 0.94. The mAP50 scores exceeded 0.94 across all models, indicating effective localization of aircraft in most cases. The models exhibited competitive performance in inference efficiency, with frame rates between 75 and 87 FPS and Latency values ranging from 5.67 to 7.92 milliseconds per image, depending on the model.

However, the mAP50-95 metric provides a more comprehensive evaluation of detection accuracy across varying IoU thresholds — remained comparatively low, ranging from 0.65 to 0.69. It suggests that while the models identify object presence and location well, their ability to generalize across stricter localization criteria without fine-tuning is limited.

According to Table 2, the experimental results indicate minimal variation in detection accuracy across the evaluated fine-tuned YOLO models when applied to the aircraft detection task. Specifically, the *mAP50* remains nearly identical across all versions, suggesting comparable performance in terms of basic object localization.

Table 1
Performance Metrics of not Fine-Tuned YOLO Models

Model	Parameters	Precision	Recall	mAP50	mAP50-95	FPS	Latency (ms/img)
YOLOv8m	25.8M	0.904415	0.894466	0.947656	0.664185	86.29	6.33
YOLOv9m	20.1M	0.958055	0.929658	0.978657	0.675673	75.83	7.92
YOLOv10m	16.5M	**0.963751**	**0.944347**	**0.984384**	**0.689463**	80.04	7.02
YOLOv11m	20.1M	0.955961	0.930024	0.980038	0.654377	**87.85**	**5.67**

Table 2
Performance Metrics of Fine-Tuned YOLO Models

Model	Parameters	Precision	Recall	mAP50	mAP50-95	FPS	Latency (ms/img)
YOLOv8m	25.8M	0.994690	0.989565	0.994677	0.799413	86.31	6.34

YOLOv9m	20.1M	**0.995392**	0.990256	**0.994679**	0.800972	56.76	10.93
YOLOv10m	16.5M	0.994222	0.984797	0.994462	0.806911	80.60	7.62
YOLOv11m	20.1M	0.994453	**0.991707**	0.994639	**0.812434**	**87.82**	**5.66**

In terms of inference efficiency, YOLOv11m outperformed all other models. It achieved the lowest *Latency* at 5.66 milliseconds per image and the highest throughput with 87.82 *FPS*. The next best-performing model in this regard was YOLOv8m, with a *Latency* of 6.34 milliseconds per image and 86.31 *FPS*. Visual comparison charts for performance metrics are shown in Figure .
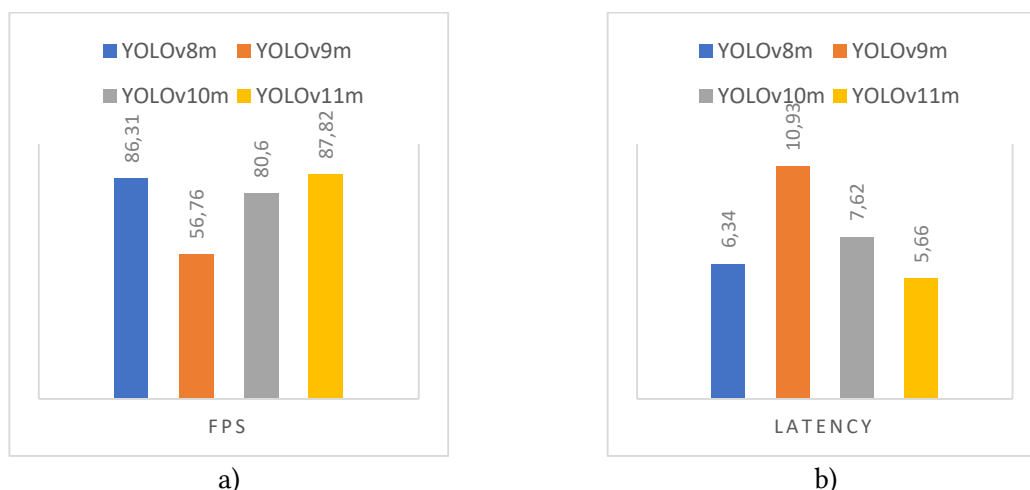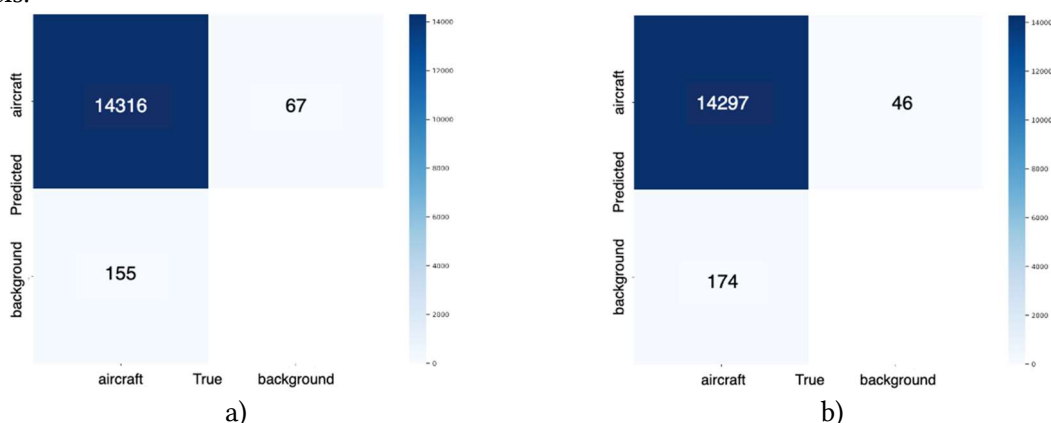

Figure 3: Fine-tuned models metrics comparison: a) FPS; b) latency.

An additional aspect considered in the evaluation was the number of trainable parameters directly impacting model size and deployment feasibility. The models exhibited noticeable differences: YOLOv8m contains 25.8 million parameters, YOLOv9m and YOLOv11m each have 20.1 million, while YOLOv10m is the most lightweight, with only 16.5 million parameters. YOLOv10m maintains competitive performance despite its lower parameter count, suggesting an efficient architectural design. Confusion matrices for the evaluated fine-tuned YOLO models are presented in Figure 4, computed using a confidence threshold of 0.25 and an IoU threshold of 0.45. These results demonstrate that YOLOv11m offers improved accuracy under stricter evaluation criteria and superior computational efficiency while maintaining average trainable parameter numbers across models.
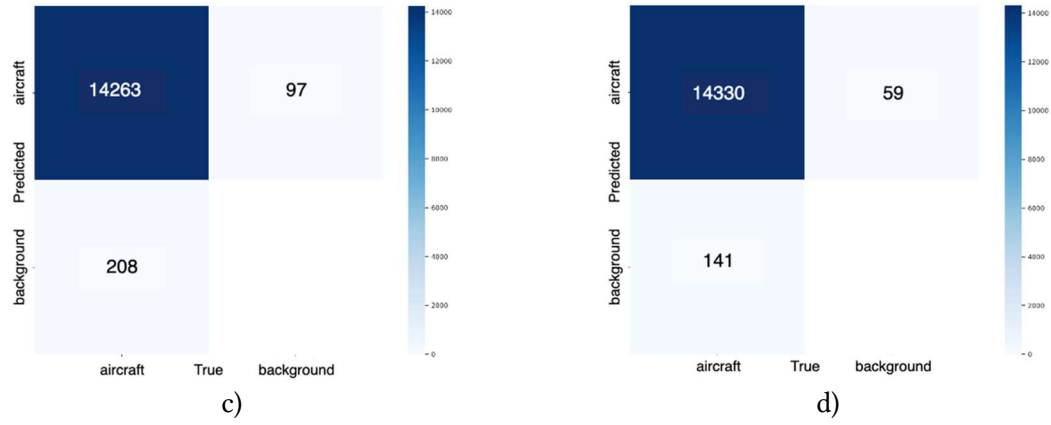
Figure 4: Confusion matrix for fine-tuned models: a) YOLOv8m; b) YOLOv9m; c) YOLOv10m; d) YOLOv11m.

A visual comparison of detection outputs from the evaluated YOLO models is presented in Figure 5, using a test image characterized by moderate contrast between aircraft and background. Notably, the plane in the scene exhibits color tones similar to the surrounding environment, which poses a challenge for object-background discrimination. The image also contains multiple buildings of varying shapes and sizes, potentially leading to false positive detections due to their structural similarity to aircraft.

All models demonstrate high confidence scores for correctly detected aircraft, particularly larger objects. However, performance degrades for smaller aircraft, with reduced confidence levels and, in some cases, missed detections. Specifically, the confidence scores for a small aircraft in the scene were 0.78 for YOLOv8, 0.68 for YOLOv9, and 0.47 for YOLOv10. YOLOv11 failed to detect the small aircraft, indicating a limitation in its sensitivity to small-scale targets under low-contrast conditions.
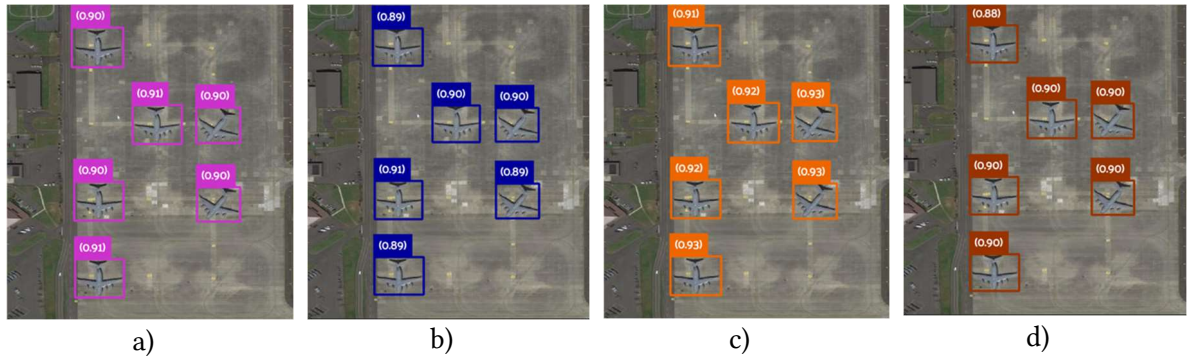

Figure 5: Fine-tuned models aircraft recognition comparison on the image fragment with moderate contrast: a) YOLOv8m; b) YOLOv9m; c) YOLOv10m; d) YOLOv11m.

The models demonstrated moderate detection performance in another test image, which features low contrast between the aircraft and the background and contains six aircraft (Figure 6). All models successfully detected at least four aircraft, though with varying levels of confidence and accuracy.

YOLOv8 detected all six aircraft. However, two detections were associated with low confidence scores of 0.26 and 0.67, respectively. YOLOv9 correctly identified four aircraft, one with a confidence score of 0.63. YOLOv10 achieved the most consistent performance in this scenario, detecting all six aircraft with confidence scores exceeding 0.70. YOLOv11 detected three aircraft with high confidence values (greater than 0.84), identified a fourth aircraft with a lower confidence of 0.44, failed to detect one aircraft, and produced one false positive detection with a confidence score of 0.42.
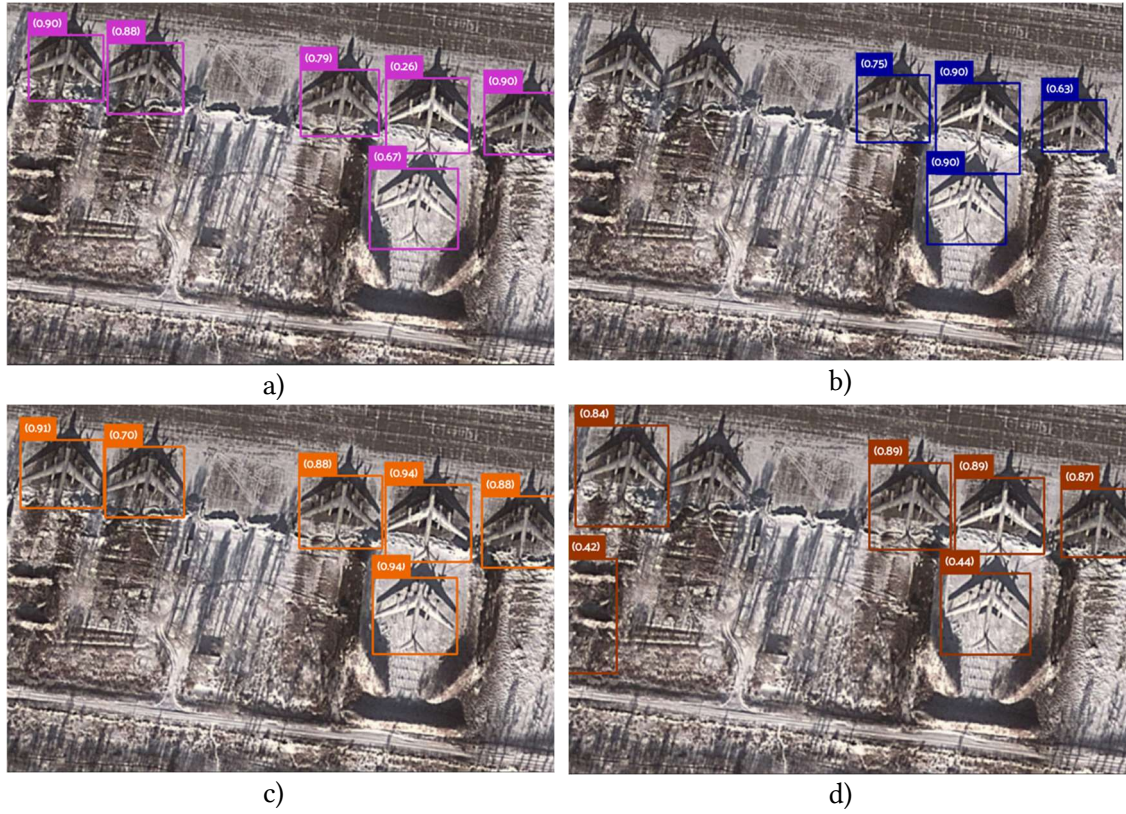
Figure 6: Fine-tuned models aircraft recognition comparison on the test image with low contrast: a) YOLOv8m; b) YOLOv9m; c) YOLOv10m; d) YOLOv11m.

Table 3 presents the performance metrics of the fine-tuned YOLO models evaluated on the dataset preprocessed using the proposed method. The results indicate that Precision, Recall, and mAP50 values remain comparable to those obtained from the same models fine-tuned on the original, non-preprocessed dataset (as shown in Table 2). Similarly, the inference speed metrics, namely FPS and Latency, exhibit negligible differences, confirming that the preprocessing pipeline does not introduce computational overhead.

Table 3

Performance Metrics of Fine-Tuned YOLO Models on Preprocessed Images

Model	Parameters	Precision	Recall	mAP50	mAP50-95	FPS	Latency (ms/img)
YOLOv8m	25.8M	0.992688	0.985180	0.994364	0.782252	86.45	6.30
YOLOv9m	20.1M	0.993128	**0.988755**	**0.994658**	0.785548	75.79	7.94
YOLOv10m	16.5M	0.963750	0.944347	0.984384	0.689463	86.24	6.96
YOLOv11m	20.1M	**0.993315**	0.985812	0.994526	**0.786554**	**88.12**	**6.21**

Although a slight decrease is observed in the mAP50-95 metric, indicating a modest reduction in bounding box localization accuracy under stricter IoU thresholds, this is offset by an improvement in the confidence levels of the detected objects.

As shown in Figure 7, the proposed technology significantly reduces the number of false negative detection results on higher confidence levels. From 2364 (19.5%) to 449 (3.2%).
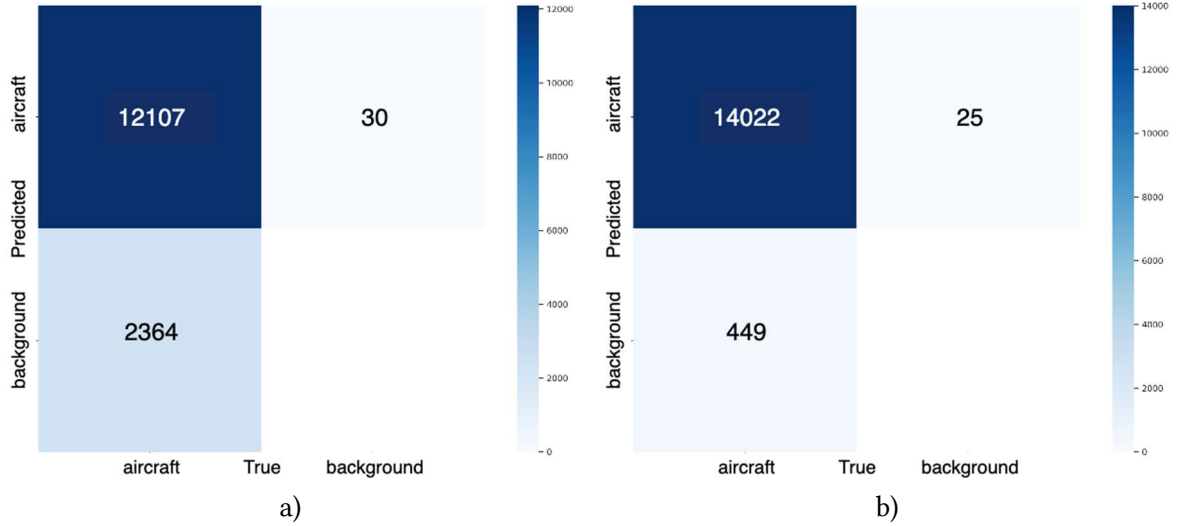
Figure 7: YOLOv11m confusion matrix with confidence ≥ 0.75: a) original dataset; b) dataset with the proposed preprocessing method.

All evaluated YOLO models demonstrated consistently high Precision, with the lowest recorded value of 0.964 for YOLOv10 and the highest of 0.993 for YOLOv11. Similarly, Recall values remained robust across models, ranging from 0.944 (YOLOv10) to 0.989 (YOLOv9). The mAP50 exhibited a comparable trend, with values spanning from 0.984 (YOLOv10) to 0.995 (YOLOv9). For the stricter mAP50-95 metric, YOLOv10 again yielded the lowest score at 0.689, while YOLOv11 achieved the highest at 0.787.

Regarding performance, YOLOv9 was identified as the slowest model, with a frame rate of 75.79 FPS and Latency of 7.94 ms/image. YOLOv8, YOLOv10, and YOLOv11 demonstrated comparable performance results, with YOLOv11 emerging as the most efficient, delivering 88.12 FPS and 6.21 ms/image latency.

In conclusion, based on accuracy and performance metrics, YOLOv11 is the most effective architecture for aircraft detection in high-resolution remote sensing imagery. However, YOLOv10 and YOLOv8 also yielded competitive results, and due to their widespread adoption and the availability of extensive modifications and enhancements made by other researchers, these models may provide practical benefits in scenarios where adaptability and extensibility are prioritized.

6. Conclusions

This study conducted a systematic comparative evaluation of four recent YOLO architectures (YOLOv8 through YOLOv11) to identify the most effective solution for aircraft detection in high-resolution satellite imagery. To enhance detection performance under challenging visual conditions, the proposed methodology incorporated a preprocessing stage combining histogram equalization and bilateral filtering. This approach was specifically designed to improve global contrast and suppress background noise while preserving critical edge details, thereby enhancing the quality of the input data provided to the detection models.

This study conducted a systematic comparative evaluation of four recent YOLO architectures (YOLOv8 through YOLOv11) to identify the most effective solution for aircraft detection in high-resolution satellite imagery. To enhance detection performance under challenging visual conditions, the proposed methodology incorporated a preprocessing stage combining histogram equalization and bilateral filtering. This approach was specifically designed to improve global contrast and suppress background noise while preserving critical edge details, thereby enhancing the quality of the input data provided to the detection models. The effectiveness of this preprocessing strategy was validated through a comparative analysis of model performance on original versus preprocessed datasets. Notably, it resulted in a substantial reduction in the false negative rate—from 19.5% to 3.2% at a confidence threshold of 0.75—demonstrating its significant contribution to improved object

visibility in low-contrast or cluttered scenes. Among the evaluated models, YOLOv11 demonstrated the best overall performance. It achieved the highest Precision (0.993), the second-highest Recall (0.9858), and the best mAP50-95 (0.7866) score, reflecting strong performance in both localization and classification tasks. Furthermore, with a frame rate of 88.12 FPS and latency of 6.21 ms/image, YOLOv11 meets the requirements for real-time inference, making it highly suitable for operational use. The results confirm that integrating a preprocessing stage with the YOLOv11 architecture provides a robust and efficient framework for accurate aircraft detection in high-resolution aerial imagery, particularly under suboptimal imaging conditions.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] W. Lu, C. Niu, C. Lan, W. Liu, S. Wang, J. Yu, T. Hu. "High-Quality Object Detection Method for UAV Images Based on Improved DINO and Masked Image Modeling". Remote Sens. 2023, 15, 4740. https://doi.org/10.3390/rs15194740.

[2] J. Terven, D.-M. Córdova-Esparza, J.-A. Romero-González. "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS". Mach. Learn. Knowl. Extr. 2023, 5, 1680-1716. doi: 10.3390/make5040083

[3] F. Şengül, K. Adem. "Detection of Military Aircraft Using YOLO and Transformer-Based Object Detection Models in Complex Environments". Bilişim Teknolojileri Dergisi 18, 2025, 85–97. doi: 10.17671/gazibtd.1549034.

[4] D. Guo, Y. Wang, S. Zhu, X. Li. "A Vehicle Detection Method Based on an Improved U-YOLO Network for High-Resolution Remote-Sensing Images". Sustainability, 15, 2023. doi: 10.3390/su151310397.

[5] J. Wu, F. Zhao, G. Yao, and Z. Jin. "FGA-YOLO: A one-stage and high-precision detector designed for fine-grained aircraft recognition". Neurocomputing. 618, 2025. doi: 10.1016/j.neucom.2024.129067.

[6] D. Ma, Z. Lu, Z. Dai, Y. Wei, L. Yang, H. Hu, W. Zhang, D. Zhang. "YOLO-SAATD: An efficient SAR airport and aircraft targets detector". Visual Informatics 9, 2025. doi: 10.1016/j.visinf.2025.100240.

[7] D. Hordiiuk, I. Oliinyk, V. Hnatushenko, K. Maksymov. "Semantic Segmentation for Ships Detection from Satellite Imagery". 2019 IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO), Kyiv, Ukraine, 2019, pp. 454–457. doi: 10.1109/ELNANO.2019.8783822.

[8] Military Aircraft Recognition dataset. 2022. URL: https://www.kaggle.com/datasets/khlaifiabilel/military-aircraft-recognition-dataset.

[9] V.J. Kashtan, V.V. Hnatushenko, Y.I. Shedlovska. "Processing technology of multispectral remote sensing images". 2017 IEEE International Young Scientists Forum on Applied Physics and Engineering (YSF), Lviv, Ukraine, 2017, pp. 355–358. doi: 10.1109/YSF.2017.8126647.

[10] A. Sarinova, A. Neftissov, Y. Rzayeva, K. Alimzhan, et al. "Development of aerospace images preliminary processing method for subsequent recognition and identification of various objects". Scientific Journal of Astana IT University. 2024, 96-106. doi:10.37943/18BIAC9844.

[11] V. Kashtan, V. Hnatushenko. "Computer Technology of High Resolution Satellite Image Processing Based on Packet Wavelet Transform". Conflict Management in Global Information Networks (CMiGIN 2019), Lviv, Ukraine, 2019, pp. 370–380. URL: https://nbn-resolving.org/urn:nbn:de:0074-2588-5.

[12] O. Patel, Y. P. S. Maravi, S. Sharma. "A Comparative Study of Histogram Equalization Based Image Enhancement Techniques for Brightness Preservation and Contrast Enhancement". SIPIJ4, 2013, pp. 11–25. doi: 10.5121/sipij.2013.4502.

[13] G. Papari, N. Idowu, T. Varslot. "Fast bilateral filtering for denoising large 3D images". IEEE Transactions on Image Processing, vol. 26, no. 1, 2017, pp. 251–261.

[14] R. Gavaskar, K. Chaudhury. "Fast Adaptive Bilateral Filtering". IEEE Transactions on Image Processing, vol. 28, no. 2, pp. 779-790, Feb. 2019, doi: 10.1109/TIP.2018.2871597.

[15] Y. Chen, X. Yuan, J. Wang, R. Wu, X. Li, Q. Hou. "YOLO-MS: Rethinking Multi-Scale Representation Learning for Real-time Object Detection". IEEE Transactions on Pattern Analysis and Machine Intelligence 47, pp. 4240–4252, 2023. doi: 10.48550/ARXIV.2308.05480.

[16] C.-Y. Wang, I.-H. Yeh, H.-Y. M. Liao, YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information, in: Computer Vision – ECCV 2024, Springer Cham, 2024, pp.1–12. doi: 10.1007/978-3-031-72751-1_1.

[17] Y. Li, W. Leong, H. Zhang, YOLOv10-Based Real-Time Pedestrian Detection for Autonomous Vehicles, in: 2024 IEEE 8th International Conference on Signal and Image Processing Applications (ICSIPA), Kuala Lumpur, Malaysia, 2024, pp. 1–6. doi: 10.1109/ICSIPA62061.2024.10686546.

[18] R. Khanam, M. Hussain, YOLOv11: An Overview of the Key Architectural Enhancements, 2024. doi: 10.48550/arXiv.2410.17725.

[19] L. He, Y. Zhou, L. Liu, W. Cao, J. Ma, Research on object detection and recognition in remote sensing images based on YOLOv11, Scientific Reports 15, 2025. doi: 10.1038/s41598-025-96314-x.

[20] V. Hnatushenko, D. Mozgovoy, V. Vasyliev, Accuracy evaluation of automated object recognition using multispectral aerial images and neural network, in: Tenth International Conference on Digital Image Processing (ICDIP 2018), Shanghai, China, SPIE, 2018, p. 72. doi: 10.1117/12.2502905.