International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4399-4404
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0646

Performance Comparison of Yolo Versions for Small Object Detection on UAV Images

V Nivashini[1], Dr. G Rajesh[2]
[1]Research Scholar, Dept. of IT, Anna University, MIT Campus, Chennai, Tamil Nadu, India.
[2]Associate Professor, Dept. of IT, Anna University, MIT Campus, Chennai, Tamil Nadu, India.
Emails: *nivashinivenkatachalapathy@gmail.com[1], raajiimegce@gmail.com[2]*

Abstract
With the rapid advancement in YOLO (You Only Look Once) versions, object detection has become a crucial task. In this article, the performance of three state-of-the-art YOLO versions: YOLOv8, YOLOv10, and YOLOv11 is analyzed. Using the VisDrone dataset, which contains 10 classes; these models were evaluated on metrics including precision, recall, F1 score, training time, and inference time. Based on the experiments, YOLOv8 achieved higher precision, recall, F1 score, and mAP compared to YOLOv10 and YOLOv11. On the other hand, YOLOv11 demonstrated reduced training time and inference time. Additionally, the performance of these detectors was analyzed on the Jetson Nano edge platform. YOLOv8 achieved the highest precision, recall, F1 score, and mAP on edge platforms as well. These experimental results provide valuable insights into selecting the most suitable YOLO version for specific object detection tasks and can guide further optimization efforts.

Keywords: *Precision, Recall, F1 score, mAP, Jetson Nano*

1. Introduction
In recent years, Unmanned Aerial Vehicles (UAVs) have been used in various applications, such as surveillance, agriculture, remote sensing, disaster management, and rescue [1], [2]. Object detection in UAV images poses significant challenges due to several factors. The distance between the object and the UAV causes scaling complexities. Target objects are often densely accumulated, making detection more complicated. Furthermore, images captured by UAVs often have low resolution, which hinders effective object detection. For instance, when an image has dimensions smaller than 32x32 pixels, it becomes challenging even for humans to identify the target. Objects in images with resolutions below 32x32 pixels are categorized as tiny objects. Many traditional algorithms struggle to detect such small objects due to these difficulties [3], [4]. Therefore, improving and optimizing conventional algorithms to extract crucial features related to small objects in aerial images is a critical problem [5], [6]. The main contributions of this work are:

- The performance of different YOLO models is evaluated using the VisDrone dataset, providing insights into their effectiveness for UAV-based detection tasks. Various metrics and comparison criteria are used to benchmark the models. This analysis ensures that the models are suitable for real-world applications.
- Pretrained YOLO models are optimized for compatibility with the Jetson Nano, ensuring smooth deployment on edge devices. The optimization process involves fine- tuning model parameters to enhance performance without sacrificing accuracy. This enables efficient operation on the limited resources of the Jetson Nano.
- The performance of the optimized models is compared on the Jetson Nano, demonstrating promising results that are comparable to those obtained using a local PC with a GPU. The evaluation showcases the Jetson Nano's potential for edge-based detection tasks. This confirms that the Nano can deliver high performance in UAV-based rescue applications.

This paper is structured as follows: Section II presents the literature review. Section III explains the methodology, including the dataset description and evaluation metrics. Section IV discusses the results and analysis. Finally, Section V presents the conclusion.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4399-4404
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0646

2. Literature Review

Recent advancements in object detection have extensively utilized deep learning architectures, particularly the YOLO series, for real-time applications in UAVs. YOLOv8, an enhanced version of the YOLO framework, has demonstrated superior performance in detecting a wide range of objects in complex environments, including aerial views [7]. In contrast, MFFCI–YOLOv8 offers a lightweight design that incorporates multiscale feature fusion and context information, which enhances detection performance in remote sensing applications [8]. Additionally, the study on the ”Improved Deformable Convolution Method for Aircraft Object Detection in Flight Based on Feature Separation in Remote Sensing Images” presents a novel approach to aircraft detection, leveraging deformable convolution techniques to address the challenges of handling complex aerial imagery [9]. Despite these advancements, challenges remain in optimizing these models for edge devices with limited resources, such as the Jetson Nano, while ensuring high detection accuracy. The MSFE-YOLO approach aims to improve YOLOv8's efficiency by integrating innovative techniques that better manage aerial perspectives, resulting in enhanced detection speed and precision for UAV- based applications. The VisDrone dataset, which consists of UAV captured images, has proven highly beneficial for various computer vision tasks, such as object detection. However, small object detection using the VisDrone dataset presents significant challenges due to the low resolution of the images and the dense distribution of target objects. Reference [10] introduced a large-scale UAV-captured dataset, designed for tasks like object detection and tracking, which is applicable across various environmental conditions. The study also emphasized the challenges encountered during the dataset's collection. A proposed improvement to the YOLOv8 algorithm for object detection was presented in [11], where the authors modified the original YOLOv8 architecture by replacing the detection head with a Convolutional Block Attention Module (CBAM) spatial attention mechanism, resulting in an 11% increase in detection accuracy. Additionally, a cross-domain fusion attention mechanism and a feature fusion model were integrated into YOLOv8, further enhancing multi-scale detection accuracy. As a result, [12] achieved a 39.2% mAP@0.5 on the VisDrone2019 dataset. Furthermore, [13] reviewed existing datasets that contain aerial images and highlighted their applications in classification, segmentation, detection, and tracking, shown in Figure 1.

3. Methodology

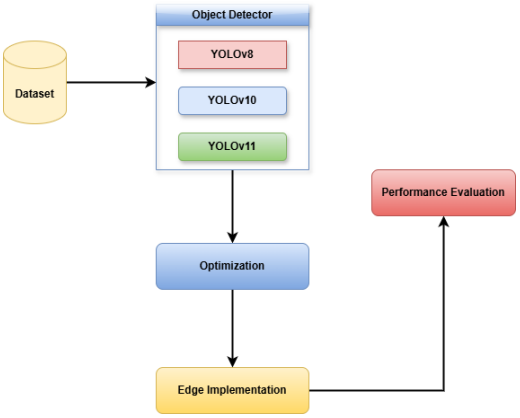

Figure 1 Pipeline for UAV Object Detection and Edge Deployment

The process begins with the preparation of the VisDrone dataset, which is split into training, validation, and test sets, followed by pre-processing and augmentation. Three YOLO models (YOLOv8, YOLOv9, YOLOv10) are trained using the training set and fine-tuned on the validation set, adjusting hyperparameters to optimize accuracy, precision, and recall. The fine-tuned models are then evaluated on the test set, com- paring performance metrics. These models are benchmarked on a discrete PC for inference speed and efficiency, and then optimized for edge deployment through techniques such as quantization and pruning to reduce model size and improve performance. Finally, the optimized models are deployed on a Jetson Nano running L4T 32.7.6, and their performance is evaluated.

3.1. Dataset Description

The VisDrone dataset [11], developed by the AISKYEYE team at Tianjin University in China, is a comprehensive benchmark for UAV-based object detection. It is collected from multiple drone platforms across various real-world scenarios, incorporating diverse weather and lighting conditions. The dataset is manually annotated with

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4399-4404
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0646

over 2.6 million bounding boxes, covering 10 distinct classes: pedestrian, people, bicycle, car, van, truck, tricycle, awning-tricycle, bus, and motor. Table 1 outlines the division of the dataset into training, validation, and test sets, providing a balanced distribution of data for effective model training and evaluation. This rich annotation enables accurate object detection and supports the development of robust models for UAV-based applications.

3.2. Experimental Setup

In this work, the performance of YOLOv8, YOLOv10, and YOLOv11 object detection on the VisDrone dataset are analyzed by setting the hyperparameters shown in TABLE 2. Both the training and validation performance are compared using precision, recall, and mAP. The goal of training these models with the same hyperparameters is to ensure a precise and effective comparison of the performance of YOLOv8, YOLOv9, and YOLOv10 models.

Table 1 Division of the VisDrone Dataset into Training, Validation and Testing Sets

CATEGORY	NO. OF IMAGES	PERCENTAGE
Training	6,469	75%
Validation	547	6%
Testing	1,610	19%
Total	8,626	100%

Table 2 Values of Hyperparameters

Hyperparameters	Values
Input Image size	640*640
Optimizer	AdamW
Epoch	50
Batch size	16
Learning rate	0.01
Momentum	0.937
Decay	0.005

In the experimental setup, the AdamW optimizer is employed, a variant of the Adam algorithm that decouples weight decay from the gradient update, enhancing regularization and improving training performance. The models are trained for 50 epochs with a batch size of 16 and an image size of 640x640. The learning rate is set to 0.01, a value known to facilitate stable and relatively fast convergence by controlling the magnitude of weight updates. The momentum is set to 0.937, which helps speed up convergence by incorporating past gradients, reducing oscillations in the optimization process. Additionally, a weight decay of 0.0005 is applied to penalize large weights, promoting generalization and reducing the risk of overfitting, shown in Table 3, Figure 2.

Figure 2 Edge Implementation Setup

Table 3 System Specifications

Parameters	Jetson Nano
Type	Embedded GPU
Architecture	Maxwell
GPU	128-core
Memory	4GB

3.3. Evaluation Metrics

The evaluation metrics used for comparative analysis are precision, recall, F1 score, inference time, and training time. Precision measures the rate of true positive predictions among all positive predictions made by the model. Recall measures the rate of true positive predictions among all actual positives. A high F1 score indicates a balanced distribution of classes. The overall performance of object detection can be evaluated by Mean Average Precision (mAP), which measures the average precision for all classes. Inference time is the amount of time the model takes to analyze new data and make predictions. Training time is the amount of time the model takes to train.

4. Results and Discussion

4.1. Discrete GPU-based Detection

From Figure 3 YOLOv8 achieves the highest precision, particularly excelling in Bicycle (0.641)

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4399-4404
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0646

and Car (0.679), demonstrating its superior detection accuracy. YOLOv10 shows competitive precision, with its best performance in Car (0.6), but overall precision remains lower than YOLOv8. YOLOv11 slightly outperforms YOLOv10 in certain classes, such as Car (0.62), but is generally less accurate than YOLOv8. YOLOv8 is the best choice for precision-critical tasks, while YOLOv10 and YOLOv11 offer acceptable accuracy with potentially better computational efficiency. This demonstrates a trade-off between detection accuracy and resource efficiency across the models. YOLOv8 achieves the highest recall across most classes, particularly excelling in Car (0.736) and Bi- cycle (0.596), showcasing its superior sensitivity in object detection. YOLOv10 and YOLOv11 perform comparably, with YOLOv11 slightly surpassing YOLOv10 in specific classes like Bicycle (0.567). Despite their lower recall, YOLOv10 and YOLOv11 may still be viable options for scenarios prioritizing computational efficiency. Overall, YOLOv8 is the most suitable model for tasks demanding high recall, which is illustrated in Figure 4.

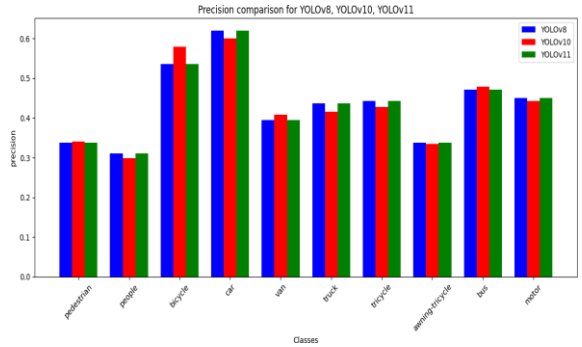

Figure 3 Precision Comparison Across YOLO Models

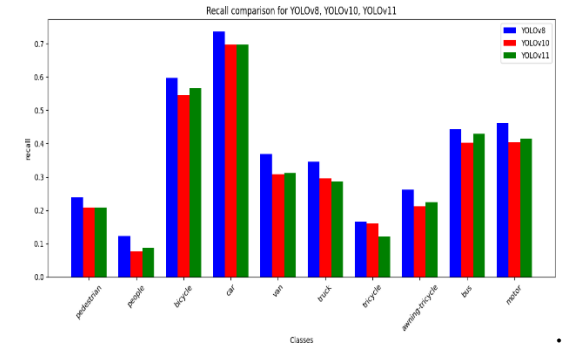

Figure 4 Recall Comparison Across YOLO Models

Based on the mAP (50-90) results, YOLOv8 outperforms both YOLOv10 and YOLOv11 across most classes, with higher values in precision and detection performance, particularly in challenging scenarios like 'pedestrian' and 'tricycle'. YOLOv10 shows the lowest mAP, indicating weaker performance in detecting objects across the board, particularly for smaller and more complex objects. YOLOv11 performs slightly better than YOLOv10, offering a balanced trade-off between accuracy and speed, but still falls short of YOLOv8's detection capability. Therefore, YOLOv8 is the most reliable for accuracy, as shown in figure 5, making it the top choice for tasks requiring high detection precision, Table 4.

Table 4 Comparison of Performance on VisDrone Dataset

Model	Precision	Recall	F1 score	mAP(50-90)
YOLOv8	48.8	37.4	42.4	22.8
YOLOv10	43.3	33.1	37.4	18.6
YOLOv11	43.6	33.5	37.8	19.2

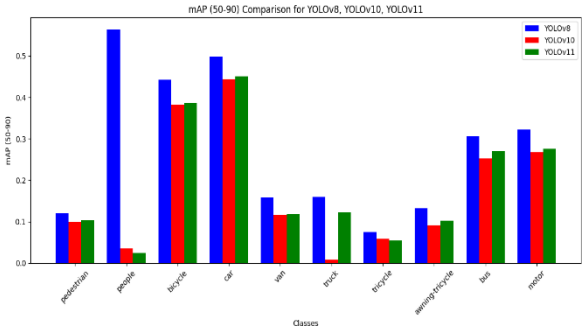

Figure 5 mAP (50-90) Comparison Across YOLO Models

Based on the comparison of YOLOv8, YOLOv10, and YOLOv11, YOLOv8 outperforms the others in terms of detection accuracy, as it has the highest precision, recall, F-1 score, and mAP (50-90), making it the most balanced model for object detection, as shown in Table 4 and 5. This highlights YOLOv8's superior ability to detect and classify objects with

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4399-4404
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0646

high precision while maintaining an efficient balance between detection performance and recall, shown in Table 5.

Table 5 Time Based Comparison

Model	Inference time (ms)	Training time (hrs)
YOLOv8	2.3	3.667
YOLOv10	3.3	3.275
YOLOv11	2.0	2.778

4.2. Edge Based Detection

UAV applications where computational resources are limited. When deploying the optimized YOLOv8 model on the Jetson Nano, the model demonstrates efficient object detection performance despite the constraints of edge hardware. Techniques such as model quantization, pruning, and conversion to TensorRT format significantly reduce the model size and inference time, making it well-suited for low-power, embedded platforms like the Jetson Nano. The results from edge-based detection show that YOLOv8 retains its high precision, recall, and F1-score, even with the reduced computational resources. This makes YOLOv8 an excellent choice for UAV-based rescue missions, where real-time decision-making is crucial. The combination of high detection accuracy and optimized inference time on the Jetson Nano ensures that UAVs can perform reliable and fast object detection even in resource- constrained environments. Figure 6 & 7 illustrates the detection result, confirming the effectiveness of YOLOv8 for real-time, edge-based detection in UAV systems [13-16].

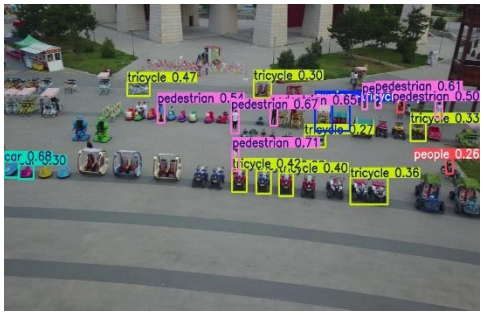

Figure 6 Visualization Result Input Image

Figure 7 Detected Output

Conclusion

The experimental results demonstrate a clear trade- off between detection performance and computational efficiency across YOLOv8, YOLOv10, and YOLOv11. YOLOv8 achieves the highest precision, recall, and F1-score, with improvements of up to 11 % in precision and 4.9 % in F1- score compared to YOLOv10 and YOLOv11. These metrics highlight YOLOv8's effectiveness in scenarios demanding high detection accuracy. On the other hand, YOLOv11 excels in computational efficiency, offering reduced training and inference times compared to YOLOv8 and YOLOv10. In this experiment the performance of detector models are compared with Jetson Nano platform and found that YOLOv8 achieved better performance than others. Future research can focus on optimizing YOLOv8 for better computational efficiency and high detection accuracy.

References

[1]. Roy, A., Bose, R., & Bhaduri, J. (2022). A fast accurate fine-grain object detection model based on YOLOv4 deep neural network. Neural Computing and Applications, 34(5), 3895–3921.

[2]. Wang, K., Liu, M., & Ye, Z. (2021). An advanced YOLOv3 method for small-scale road object detection. Applied Soft Computing, 112, Article 107846.

[3]. Tian, G., Liu, J., & Yang, W. (2021). A dual neural network for object detection in UAV images. Neurocomputing, 443, 292–301.

[4]. Rayhana, R., Xiao, G., & Liu, Z. (2020). Internet of Things empowered smart greenhouse farming. IEEE Journal of Radio Frequency Identification, 4, 195–211.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 12 December 2025
Page No: 4399-4404
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0646

[5]. Liu, Q., Qi, X., Liu, S., Cheng, X., Ke, X., & Wang, F. (2022). Application of lightweight digital twin system in intelligent transportation. IEEE Journal of Radio Frequency Identification, 6, 729–732.

[6]. Zhu, G., Zhu, F., Wang, Z., Xiong, G., & Tian, B. (2023). Small target detection algorithm based on multi-target detection head and attention mechanism. In Proceedings of the IEEE 3rd International Conference on Digital Twins and Parallel Intelligence (DTPI) (pp. 1–6).

[7]. Qi, S., Song, X., Shang, T., Hu, X., & Han, K. (2024). MSFE-YOLO: An improved YOLOv8 network for object detection on drone view. IEEE Geoscience and Remote Sensing Letters, 21, 1–5. doi:10.1109/LGRS.2024.3432536

[8]. Xu, S., Song, L., Yin, J., Chen, Q., Zhan, T., & Huang, W. (2024). MF-FCI–YOLOv8: A lightweight remote sensing object detection network based on multiscale features fusion and context information. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 19743–19755. doi:10.1109/JSTARS.2024.3474689

[9]. Yu, L., et al. (2024). Improved deformable convolution method for aircraft object detection in flight based on feature separation in remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 8313–8323. doi:10.1109/JSTARS.2024.3386696

[10]. Zhu, P., et al. (2022). Detection and tracking meet drones challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44, 7380–7399. doi:10.1109/TPAMI.2021.3119563

[11]. Shao, S., Yuan, W., Wang, Y.-P. E., & Zhou, Z. (2024). Research on small target detection for drone based on improved YOLOv5. Proceedings of the International Conference on Machine Learning and Computational Applications (ICMLCA), 489–492. doi:10.1109/ICMLCA63499.2024.10754503

[12]. He, Z., & Cao, L. (2024). SOD-YOLO: Small object detection network for UAV aerial images. IEEJ Transactions on Electrical and Electronic Engineering. doi:10.1002/tee.24195

[13]. [13]. Rahman, M. M., Siddique, S., Kamal, M., Hossain, R. R., & Gupta, K. D. (2024). UAV (Unmanned Aerial Vehicles): Diverse applications of UAV datasets in segmentation, classification, detection, and tracking. Preprints, 202411.0829.v1. doi:10.20944/preprints202411.0829.v1

[14]. Lou, H., et al. (2023). DC-YOLOv8: Small-size object detection algorithm based on camera sensor. Electronics, 12(10), 2323.

[15]. Wang, A., et al. (2024). YOLOv10: Real-time end-to-end object detection. Computer Vision and Pattern Recognition - Cornell University, 1–18.

[16]. Khanam, R., & Hussain, M. (2024). YOLOv11: An overview of the key architectural enhancements.