High-Precision Multi-Class Object Detection Using Fine-Tuned YOLOv11 Architecture: A Case Study on Airborne Vehicles

Nasser S. Albalawi
Department of Computer Sciences-Faculty of Computing and Information Technology,
Northern Border University, Rafha, Saudi Arabia

***Abstract*—The widespread adoption of airborne vehicles, including drones and UAVs, has brought significant advancements to fields such as surveillance, logistics, and disaster response. Despite these benefits, their increasing use poses substantial challenges for real-time detection and classification, particularly in multi-class scenarios where precision and scalability are essential. This paper proposes a high-performance detection framework based on YOLOv11, specifically tailored for identifying airborne vehicles. YOLOv11 integrates innovative features, such as anchor-free detection and enhanced attention mechanisms, to deliver superior accuracy and speed. The proposed framework is tested on a comprehensive airborne vehicle dataset featuring diverse conditions, including variations in altitude, occlusion, and environmental factors. Experimental results demonstrate that the fine-tuned YOLOv11 model exceeds the performance of existing models. Additionally, its ability to operate in real-time makes it ideal for critical applications like air traffic management and security monitoring.**

***Keywords*—Airborne vehicles; YOLOv11; object detection; surveillance**

I. Introduction

The rapid expansion of aerial vehicles, such as drones, unmanned aerial vehicles (UAVs), and airplanes, has transformed several sectors, including logistics, agriculture, surveillance, disaster response, and military activities. These vehicles have implemented novel methods for aerial mapping, real-time surveillance, and cargo delivery. Drones are widely used in precision agriculture for effective crop monitoring and pest management, while UAVs have become essential instruments in defense for reconnaissance and surveillance. Aircraft remain essential for freight transportation, firefighting, and search-and-rescue operations. Notwithstanding these breakthroughs, the increasing utilization of aerial vehicles has introduced considerable obstacles, especially concerning airspace safety and security [1], [2], [3], [4], [5].

Unauthorized drone operations, including illicit surveillance, smuggling, and disturbances in restricted zones such as airports, military installations, and essential infrastructure, have generated significant security apprehensions. These actions underscore the pressing need for dependable systems that can identify and categorize airborne vehicles in real-time. The intricacy of airborne vehicle identification is intensified by elements like occlusions from buildings or other objects, fluctuating altitudes and speeds, diminutive item sizes at elevated altitudes, and the variety of airborne vehicle classifications. Conventional detection techniques, including radar, acoustic sensors, and optical systems, often encounter constraints regarding precision and scalability. Radar systems, while proficient in monitoring bigger aircraft, may have difficulties with tiny drones because to their reduced radar cross-sections. Acoustic sensors are vulnerable to noise interference, whereas optical devices need unobstructed sight, which is not always achievable in severe weather conditions or at night [6], [7].

Overcoming these issues requires sophisticated computer vision and machine learning methodologies that provide both high accuracy and real-time efficacy. Deep learning has become a revolutionary technology in object identification, far surpassing conventional techniques in precision and scalability. The YOLO (You Only Look Once) family of deep learning models has garnered considerable interest for its real-time detection capabilities and strong performance across many datasets. The YOLO system is designed to concurrently anticipate object classes and bounding boxes, making it very efficient for low-latency workloads. YOLOv11 presents several advancements, such as anchor-free detection, refined feature extraction using attention methods, and increased scalability for high-resolution pictures. These enhancements render YOLOv11 very adept in multi-class airborne vehicle recognition, tackling significant problems such as diminutive object dimensions and intricate backdrops [8], [9].

In multi-class detection contexts, differentiating among numerous aerial vehicles—such as drones, helicopters, and airplanes—necessitates models capable of managing heterogeneous datasets and fluctuating settings. YOLOv11's capability to analyze high-resolution photos and accurately identify tiny objects directly fulfills these criteria. Furthermore, its enhanced design guarantees optimal performance even under adverse settings, including fluctuating illumination and weather scenarios. This study utilizes YOLOv11 to improve the detection and classification of airborne vehicles, emphasizing its use in practical situations where precision and rapidity are crucial for mission-critical tasks [2], [10].

This study presents many significant contributions:

- Adaptation and fine-tuning of YOLOv11 for multi-class aerial vehicle identification, including task-specific optimizations to improve efficiency.
- Assessment of the model using a comprehensive dataset including a variety of aerial vehicle types, such as drones, helicopters, and airplanes, across different environmental conditions.

- Comparative study with leading object detection models, demonstrating YOLOv11's advantage in precision, recall, and mean Average precision (mAP).

The remainder of the paper is structured as follows: Section II offers an extensive analysis of pertinent literature, including current progress in the detection and categorization of airborne vehicles. Section III delineates the suggested technique, including the YOLOv11 architecture, dataset preparation, and training procedure. Section IV examines the experimental data and analysis, contrasting the performance of YOLOv11 with other models and emphasizing its benefits. Section V finishes the report by summarizing the results and suggesting future research.

II. Related Work

Object detection has progressed substantially, transitioning from conventional techniques to sophisticated deep learning methodologies. Initial methodologies, such Haar cascades and Histogram of Oriented Gradients (HOG), depended on manually created features and traditional machine learning methods. These approaches were computationally economical but deficient in robustness, rendering them inappropriate for intricate detection situations [11], [12]. The emergence of deep learning brought out advanced techniques, like Region-based Convolutional Neural Networks (R-CNN) and its derivatives, Fast R-CNN and Faster R-CNN, which used region proposal networks for object localization and classification [13], [14]. Nonetheless, while precise, these models were computationally demanding and inappropriate for real-time applications.

Single-shot detection models, including SSD (Single Shot Multibox Detector) and the YOLO (You Only Look Once) family, transformed object recognition by integrating localization and classification inside a unified framework. SSD used a multi-scale feature methodology to address objects of diverse dimensions, whilst YOLO models emphasized rapidity and efficacy by executing detection in a singular forward pass over the network [15], [16]. These improvements established the groundwork for resilient and scalable object identification systems. Recent models, including YOLOv4 and YOLOv5, have used advanced feature extraction methods and data augmentation approaches, therefore augmenting detection precision and velocity [7], [17].

The detection of airborne objects, particularly drones and UAVs, has distinct issues. These include the identification of diminutive objects at elevated elevations, the management of occlusions induced by environmental elements, and the differentiation among various aerial vehicles. Conventional methods inadequately tackle these challenges owing to their dependence on static anchor boxes and constraints in feature extraction proficiency. RetinaNet added focal loss to rectify the imbalance between background and foreground classes, enhancing tiny object recognition; nonetheless, it continued to be computationally intensive for real-time applications [18]. Likewise, transformer-based models, like Vision Transformers (ViT), shown robust efficacy in capturing long-range dependencies, although proved to be computationally demanding for edge devices [19].

Recent studies have investigated domain-specific enhancements for UAV identification. Ma et al. [20] introduced a hybrid methodology that integrates radar and image data, showcasing enhanced classification precision for drones in low-visibility environments. Zhang et al. [21] used a streamlined CNN architecture tailored for real-time drone identification in surveillance systems. Furthermore, Hossain et al. [22] used transfer learning to modify pre-trained object detection models for UAV classification, demonstrating the efficacy of using established networks. Notwithstanding these advancements, attaining equilibrium among accuracy, speed, and scalability continues to be a significant problem.

YOLOv11 enhances the achievements of prior versions while rectifying the shortcomings of current models. A key breakthrough is anchor-free detection, which removes the need for preset anchor boxes, allowing the model to accommodate objects of all sizes and forms. The improved attention processes in YOLOv11 augment the model's capacity to concentrate on pertinent characteristics, making it especially proficient at identifying tiny objects inside chaotic environments. Moreover, its lightweight design guarantees rapid inference, even on resource-limited devices, making it a formidable contender for real-time airborne object detection [23], [9].

Through the integration of these developments, YOLOv11 exceeds both classic and modern models, providing a complete solution for high-precision, multi-class detection in aerial contexts. Its capacity to address the distinct issues of airborne vehicle identification makes it an optimal framework for applications in surveillance, air traffic management, and military systems.

III. Methodology

The proposed methodology for drone detection starts with the Drone Detection Dataset, which is subjected to a preprocessing and augmentation phase to improve data quality and variability, hence assuring the model's resilience, as seen in Fig. 1. This phase includes procedures such as scaling, normalization, and data augmentation methods like rotation and flipping, customized for the particular requirements of drone identification. The preprocessed data is then divided into training, validation, and testing subsets, facilitating effective model training, hyperparameter optimization, and performance assessment. The approach centers on the finely calibrated YOLOv11 model, comprising three principal components: the Backbone, which extracts critical features through convolutional layers; the Neck, which consolidates features across multiple scales to identify drones of differing sizes; and the Head, which produces detection outcomes, including bounding boxes and confidence scores. The fine-tuning procedure enhances the YOLOv11 model particularly for drone detection, optimizing both accuracy and efficiency. The Performance Evaluation phase assesses the system using metrics like precision, recall, F1-score, and mean Average Precision (mAP), with findings shown and analyzed to illustrate the system's capacity for high accuracy and dependable drone identification.

A. Fine-Tuned YOLOv11 Architecture

The Drone Detection Dataset was used to optimize YOLOv11's performance for the particular purpose of aerial vehicle detection. Fine-tuning is modifying a pre-trained model

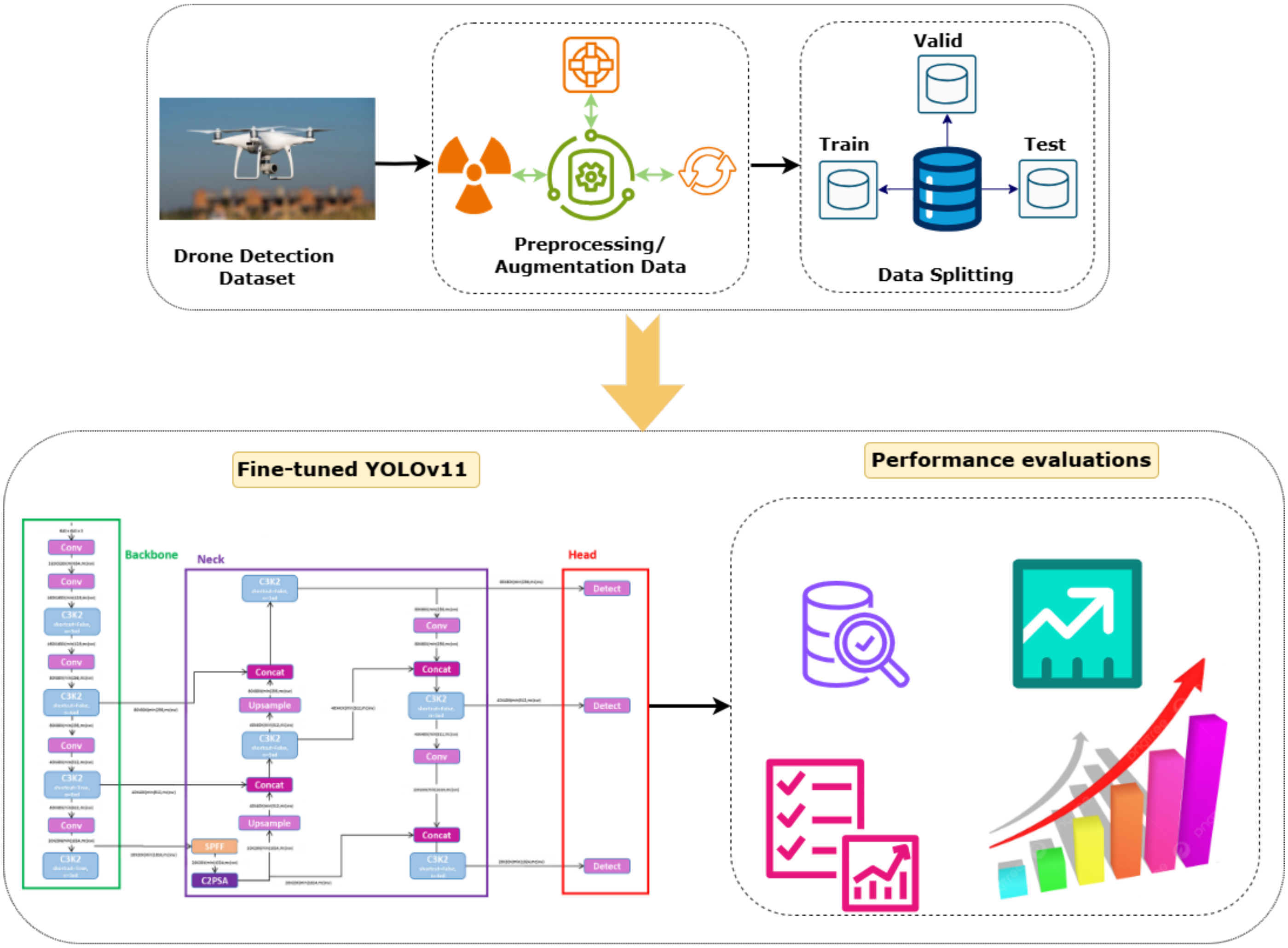


Fig. 1. Proposed approach-based fine-tuned YOLOv11.

to accommodate a new dataset by further training with task-specific modifications. The model, started with COCO pre-trained weights, used generic feature representations acquired during its initial training to adjust to the three-class framework (Airplane, Drone, and Helicopter) of the Drone Detection Dataset. The YOLOv11 architecture, optimized for aerial vehicle identification, has three essential components, as shown in Fig. 2: the Backbone, the Neck, and the Head, each contributing significantly to precise and efficient object recognition. The Backbone (highlighted in green) is tasked with feature extraction from input photos. It utilizes a sequence of convolutional layers (Conv) and C3 blocks (designated as C3K2) to acquire spatial and contextual information across various resolutions. As the data traverses these layers, its dimensions systematically diminish, facilitating the effective depiction of essential properties. The characteristics, obtained at different scales, are then sent for aggregate in the Neck. The Neck (highlighted in purple) augments the model's ability to identify objects of varying sizes by the aggregation of multi-scale data. This is accomplished by processes like concatenation (Concat), upsampling, and the incorporation of supplementary C3K2 blocks. The use of sophisticated elements such as SPFF (Spatial Pyramid Feature Fusion) and C2PSA (Cross-Scale Pairwise Self-Attention) enhances feature fusion across scales, hence augmenting localization and detection precision, especially for little objects such as drones. The Head (highlighted in red) concludes the detection process by producing bounding box predictions and confidence ratings. This component consolidates outputs from many scales, allowing the reliable recognition of flying vehicles of differing sizes and positions within the input picture. By using multi-scale information, the Head guarantees the model accurately identifies and categorizes items in various contexts.

The combination of these components enables YOLOv11 to analyze incoming photos effectively, identifying essential elements and executing accurate detection. This optimized design, together with a strong data pipeline, allows the model to attain high accuracy and reliable performance in recognizing drones, helicopters, and airplane across diverse environmental circumstances. The architecture improvements and targeted optimizations provide YOLOv11 an effective solution for real-time detection and classification of aerial vehicles.

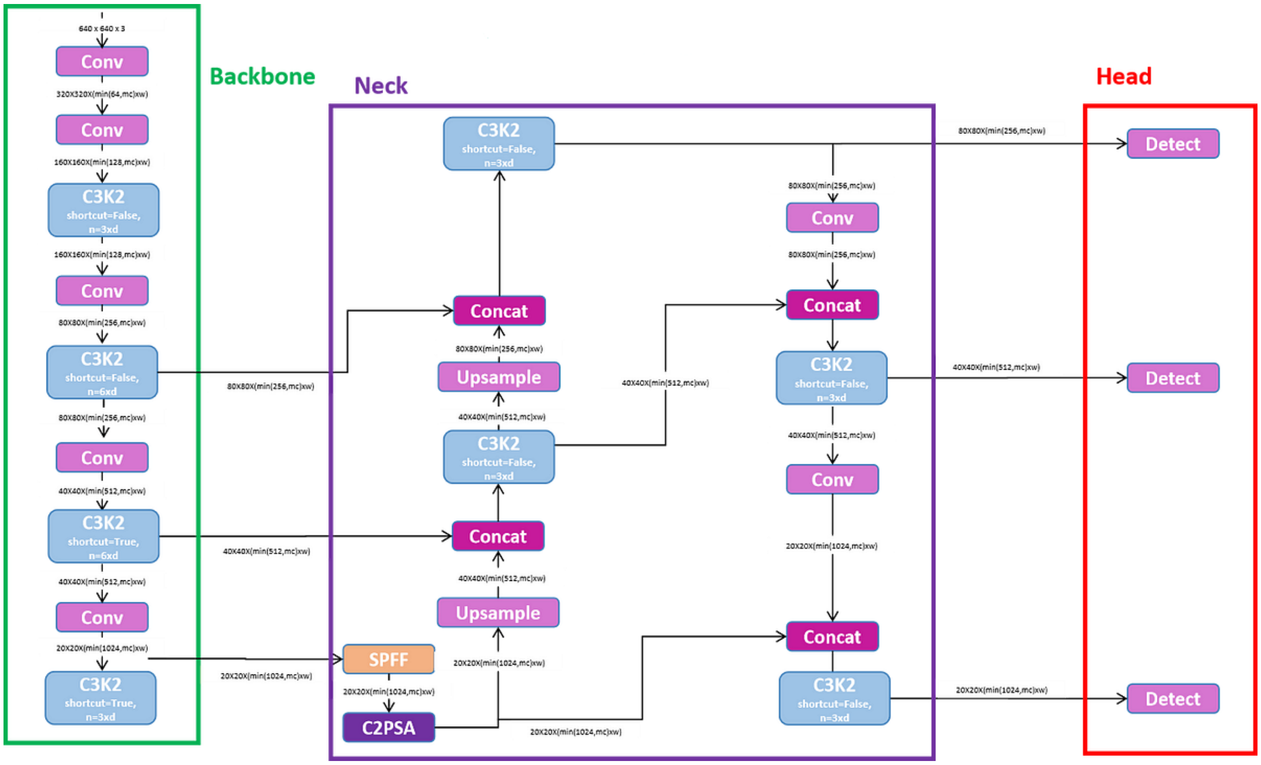


Fig. 2. Fine-tuned YOLOv11 architecture.

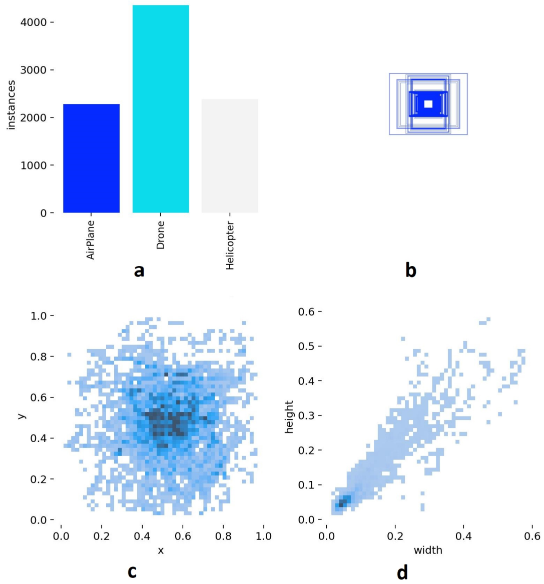


Fig. 3. Visualization of the Dataset. (a) Number of annotations per class. (b) Visualization of the location and size of each bounding box. (c) Statistical distribution of the positions of the bounding boxes. (d) Statistical distribution of the sizes of the bounding boxes.

B. Dataset Preparation

This work uses the Drone Detection Dataset obtained from Roboflow, which contains 11,998 images tagged with bounding boxes for three categories: Airplane, Drone, and Helicopter, as seen in Fig. 3. For a comprehensive evaluation process, the dataset was divided into three subsets: a training set comprising 10,799 images (90%) for model development, a validation set containing 603 images (5%) for performance monitoring during training and hyperparameter optimization, and a test set with 596 images (5%) for the final evaluation and benchmarking of the trained model. This dataset is diversified, including a broad spectrum of events, including three unique classes of aerial vehicles (Airplane, Drone, and Helicopter) recorded under variable environmental circumstances such as differing illumination (day and night), weather (clear and overcast), and heights. Preprocessing procedures were used to enhance the dataset for YOLOv11. All photos were downsized to 640×640 pixels with a stretch transformation to conform to YOLOv11's input specifications. Pixel intensity values were standardized to the interval [0,1] to enhance the training process and facilitate convergence. Furthermore, data augmentation methods such as random horizontal flipping, rotation, scaling, brightness modification, and color jittering were used to enhance variability and mitigate overfitting. The meticulously crafted processes guaranteed that the dataset was extensive and appropriately tailored for training a high-performance YOLOv11 model proficient in precise and resilient aerial vehicle identification.

C. Model Training and Optimization

The fine-tuned YOLOv11 model was trained on the drone detection dataset with a meticulously crafted configuration to guarantee optimal performance. A learning rate of 0.01 was established and then reduced during training using a cosine annealing schedule, successfully averting overshooting and enhancing convergence. A batch size of 32 was used to improve computational efficiency and ensure stable convergence, while the AdamW optimizer was utilized to integrate adaptive learning rate modifications with weight decay, hence improving generalization and training stability. Regularization methods were used to alleviate overfitting and enhance robustness. Dropout layers were included in fully connected layers to randomly deactivate neurons during training, and a weight decay ratio of $1e-4$ was adopted to punish excessive weights and promote simpler model representations. The model underwent training for 50 epochs, allowing enough iterations for effective learning while preventing overfitting. Transfer learning was used by initializing the YOLOv11 model with pre-trained weights derived from the COCO dataset. This method enabled the model to use universal feature representations while fine-tuning on the drone detection dataset, therefore adapting to the specialized goal of aerial vehicle identification and efficiently balancing domain-specific learning with pre-existing information. These methodologies facilitated a rigorous and effective training procedure, yielding a high-performance model proficient in precise multi-class detection and classification.

D. Evaluation Metrics

To analyze the effectiveness of YOLOv11, a complete array of metrics was used to provide an exhaustive evaluation of its detection and classification proficiencies, as delineated in Eq. 1, 2, 3, 4, and 5. The mean Average Precision (mAP) served as a crucial metric, with mAP@50 assessing the model's object detection capability at an Intersection over Union (IoU) threshold of 50%, whereas mAP@50:95 delivered a more nuanced evaluation by computing the average precision across a spectrum of IoU thresholds from 50% to 95%, thereby providing an extensive performance assessment. Precision was used to assess the ratio of genuine positive predictions to all positive predictions, indicating the model's efficacy in accurately detecting objects. Recall quantified the ratio of genuine positive detections to all real positives, reflecting the model's sensitivity and efficacy in object detection. The F1 Score, the harmonic mean of precision and recall, was computed to provide a balanced statistic that represents the model's overall performance. Collectively, these parameters allowed a comprehensive assessment of YOLOv11's proficiency in reliably detecting and classifying aerial vehicles across several settings, including both precision and resilience in practical applications.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{1}$$

$$\text{mAP} = \frac{1}{n}\sum_{i=1}^{n} \text{AP}_i \tag{2}$$

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \tag{3}$$

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \tag{4}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

IV. Experimental Results

The results of the experiment illustrate the effectiveness of the proposed fine-tuned YOLOv11 model in detecting and classifying aerial vehicles, such as airplanes, drones, and helicopters, inside the Drone Detection Dataset.

Fig. 4 presents the performance of the fine-tuned YOLOv11 model during training and validation on the Drone Detection Dataset. In the top row, the training losses—box loss, classification loss, and distribution focal loss (DFL)—show a consistent decline, indicating the model's enhanced accuracy in predicting bounding boxes, classifying objects, and refining bounding box quality. Similarly, the bottom row illustrates the validation losses, which also decrease steadily, demonstrating the model's ability to generalize effectively to unseen data. Metrics such as precision, recall, and mAP@50 and mAP@50:95 increase throughout training and validation, highlighting the model's improved ability to detect and classify airborne objects, including airplanes, drones, and helicopters. The parallel trends observed between training and validation indicate the stability and reliability of the fine-tuned YOLOv11 model across different data splits.

Fig. 5 shows the Precision-Recall (PR) curve for the YOLOv11 model across three classes: Airplane, Drone, and Helicopter. Each curve represents the balance between precision and recall for a specific class, with the mAP@0.5 (mean Average Precision at IoU 0.5) values annotated in the legend. The Airplane class achieves a high mAP of 0.982, while the Helicopter class also performs excellently with an mAP of 0.983. The Drone class shows a slightly lower performance with an mAP of 0.933. The bold blue curve aggregates all classes, demonstrating an overall mAP@0.5 of 0.966. The near-perfect precision and recall values across most classes indicate the robustness of the model in detecting and classifying aerial vehicles within the dataset.

Fig. 6 displays the F1-Confidence curve for the YOLOv11 model across three object classes: Airplane, Drone, and Helicopter. Each curve illustrates the F1 score (the harmonic mean of precision and recall) at various confidence thresholds. The Airplane and Helicopter classes achieve high F1 scores close to 0.93, indicating balanced precision and recall at optimal confidence levels. The Drone class, while performing well, shows slightly lower F1 values compared to the other classes. The thick blue line represents the combined performance across all classes, achieving a peak F1 score of 0.93 at a confidence threshold of 0.340. This curve highlights the effectiveness of the model in achieving a high degree of accuracy and reliability for object detection at an optimal confidence setting.

Fig. 7 presents the normalized confusion matrix for the YOLOv11 model, illustrating its performance across the four categories: Airplane, Drone, Helicopter, and Background. Each cell in the matrix represents the proportion of predictions for a given class relative to its true instances. The diagonal cells indicate correct predictions, with high values of 0.97 for Airplane, 0.94 for Drone, and 0.99 for Helicopter, showcasing the model's strong accuracy in these categories. Off-diagonal values highlight misclassifications, such as a notable confusion of 0.19 where some Airplanes are misclassified as Drones and 0.10 where some Helicopters are misclassified as Background. The matrix underscores the model's overall reliability while

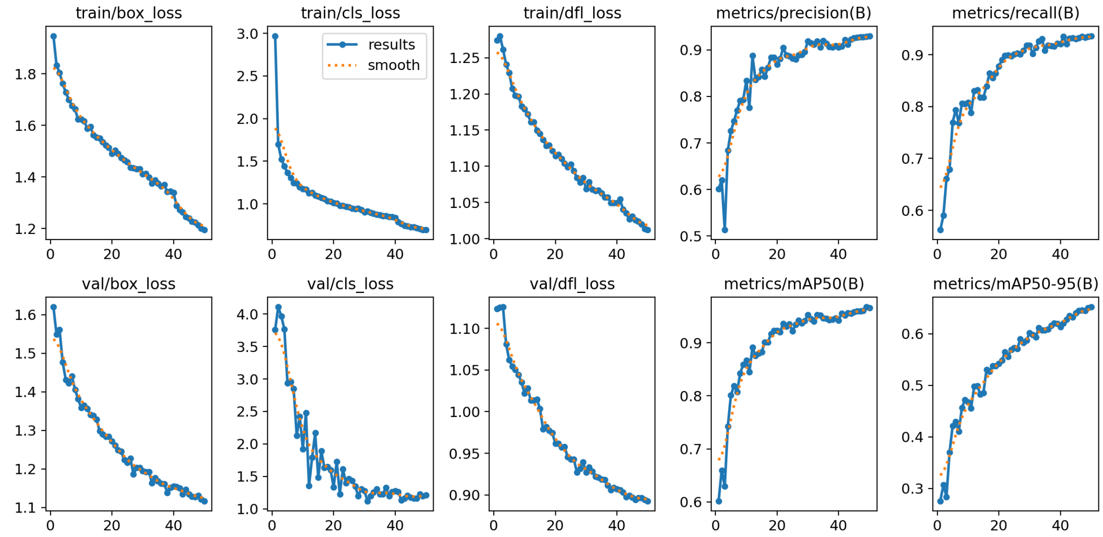


Fig. 4. Training and validation performance metrics.

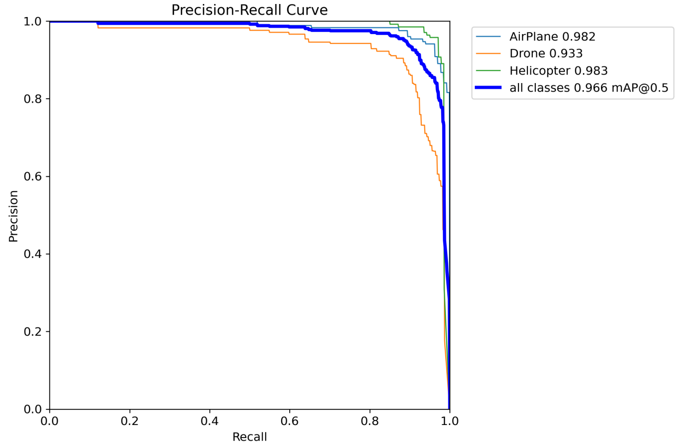


Fig. 5. PR Curve for YOLOv11 on drone detection dataset.

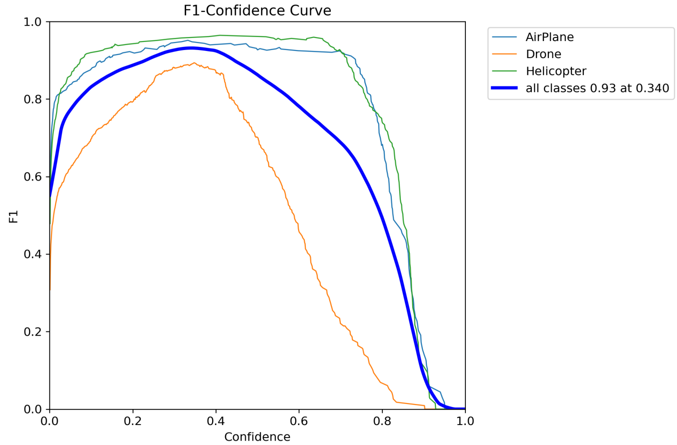


Fig. 6. F1-Confidence curve for YOLOv11 on drone detection dataset.

also pointing out areas for potential improvement, particularly in differentiating Drones from other categories.

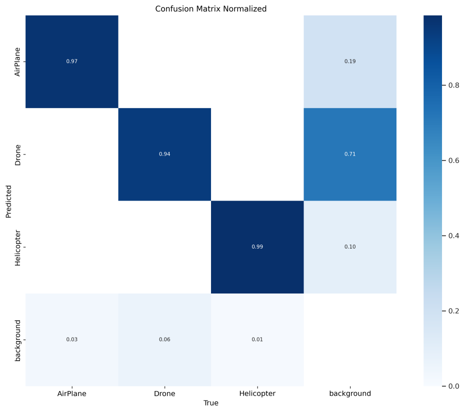
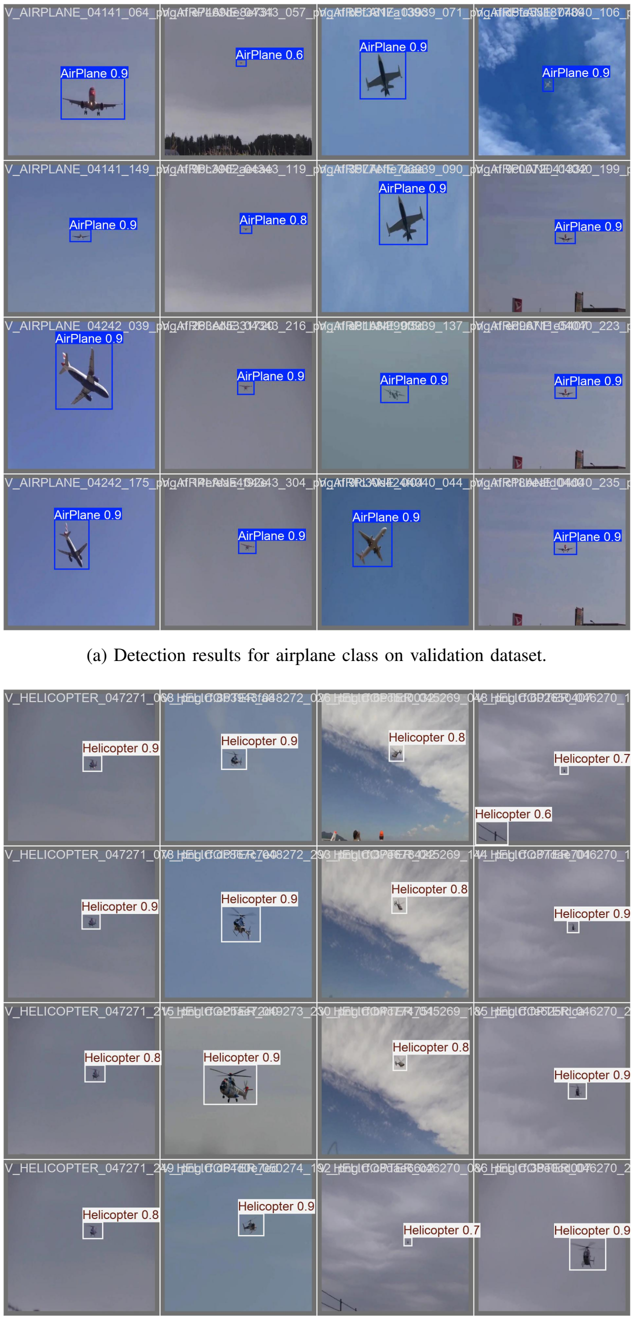


Fig. 7. Normalized confusion matrix for YOLOv11 on drone detection dataset.

Fig. 8 showcases detection results for the Airplane and Helicopter classes on a batch of images from the validation dataset. Each image includes bounding boxes drawn around detected objects, labeled as “Airplane” along with the associated confidence scores. The confidence values range from 0.6 to 0.9, reflecting the model’s confidence in the accuracy of its predictions. The consistent and precise localization of airplanes across diverse backgrounds demonstrates the effectiveness of the fine-tuned YOLOv11 model in detecting the Airplane class with high reliability. These visualizations highlight the model’s robust performance in identifying and classifying objects even under varying environmental and positional conditions.

A. Comparative Study

Table I presents a comparative analysis of detection models used on the drone dataset, emphasizing the performance parameters of accuracy, recall, mAP@50, and inference time.



(a) Detection results for airplane class on validation dataset.

(b) Detection results for helicopter class on validation dataset.

Fig. 8. Prediction results on validation dataset.

The findings from [24] indicate a precision of 0.91, a recall of 0.89, and a mAP@50 of 0.93; nevertheless, the inference time remains unreported. Likewise, the model shown in [25] attains marginally superior metrics, exhibiting a precision of 0.94, a recall of 0.92, and a mAP@50 of 0.94. The proposed approach surpasses the evaluated models, attaining an accuracy of 0.94, a recall of 0.943, and a mAP@50 of 0.966. Moreover, it has an inference time of about 1.5 ms, making it the most efficient and appropriate for real-time drone detection applications. These results emphasize the efficacy and feasibility of the suggested method, integrating high detection accuracy with rapid processing speed.

TABLE I. Comparison of Detection Models

Model	Precision	Recall	mAP@50	Inference Time (ms)
YOLOv4 [24]	0.91	0.89	0.93	—
YOLOv5 [25]	0.94	0.92	0.94	—
Proposed Approach	0.94	0.943	0.966	1.5

V. Conclusion

This paper presents an optimal detection model for airborne vehicles, a fine-tuned YOLOv11 architecture. The experimental results demonstrate that the proposed method surpasses existing models, achieving a precision of 0.94, a recall of 0.943, and an mAP@50 of 0.966, with an inference time of only 1.5 ms. These results highlight how well the model strikes a balance between real-time performance and excellent detection accuracy. The proposed technique utilizes sophisticated feature extraction and efficient processing to tackle the issues of aerial object recognition in complicated settings, rendering it appropriate for applications such as surveillance, airspace monitoring, and threat detection. Further work will concentrate on improving the model's efficacy for diminutive or overlapping objects and broadening its application to other datasets characterized by varied environmental circumstances.

Acknowledgment

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFR-2025-1260-01"

References

[1] S.-W. Roh and J.-W. Lim, "Drone detection and classification using deep learning," *Sensors*, vol. 21, no. 9, p. 3002, 2021.

[2] A. Sharma and R. Mittal, "Drone detection and identification in the rf spectrum using a machine learning approach," *IEEE Access*, vol. 9, pp. 96 856–96 867, 2021.

[3] N. Al-lQubaydhi, A. Alenezi, T. Alanazi, A. Senyor, N. Alanezi, B. Alotaibi, M. Alotaibi, A. Razaque, and S. Hariri, "Deep learning for unmanned aerial vehicles detection: A review," *Computer Science Review*, vol. 51, p. 100614, 2024.

[4] D. Ojdanić, C. Naverschnigg, A. Sinn, D. Zelinskyi, and G. Schitter, "Parallel architecture for low latency uav detection and tracking using robotic telescopes," *IEEE Transactions on Aerospace and Electronic Systems*, 2024.

[5] D. Aouladhadj, E. Kpre, V. Deniau, A. Kharchouf, C. Gransart, and C. Gaquière, "Drone detection and tracking using rf identification signals," *Sensors*, vol. 23, no. 17, p. 7650, 2023.

[6] A. Mohan and R. Smith, "Deep learning for drone detection and tracking," *Pattern Recognition Letters*, vol. 131, pp. 123–129, 2020.

[7] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," pp. 779–788, 2016.

[9] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," *arXiv preprint arXiv:2107.08430*, 2021.

[10] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," pp. 886–893, 2005.

[12] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," vol. 1, pp. I–I, 2001.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, 2014.

[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," *European conference on computer vision*, pp. 21–37, 2016.

[16] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[17] G. Jocher, "Ultralytics yolov5: cutting-edge object detection at real-time speeds," *GitHub Repository*, 2021.

[18] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988, 2017.

[19] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, and X. Zhai, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2021.

[20] J. Ma and Z. Zhou, "Detection of drones using hybrid radar and vision-based system," *Sensors*, vol. 20, no. 10, p. 2930, 2020.

[21] M. Zhang and X. Li, "Drone detection and tracking using lightweight cnns in surveillance systems," *Computer Vision Applications*, vol. 11, pp. 42–55, 2021.

[22] A. Hossain and S. Ahmed, "Detection and classification of uavs using transfer learning with deep neural networks," *Neural Computing and Applications*, vol. 33, pp. 12 345–12 360, 2021.

[23] Z. T. Wang and W. Sun, "Fcos: Fully convolutional one-stage object detection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, pp. 987–1003, 2022.

[24] L. Tan, X. Lv, X. Lian, and G. Wang, "Yolov4_drone: Uav image target detection based on an improved yolov4 algorithm," *Computers & Electrical Engineering*, vol. 93, p. 107261, 2021.

[25] N. Al-Qubaydhi, A. Alenezi, T. Alanazi, A. Senyor, N. Alanezi, B. Alotaibi, M. Alotaibi, A. Razaque, A. A. Abdelhamid, and A. Alotaibi, "Detection of unauthorized unmanned aerial vehicles using yolov5 and transfer learning," *Electronics*, vol. 11, no. 17, p. 2669, 2022.