HiLL PUBLISHING

Advance in Information Technology and Computer Science, 2025, 2(1),14-18
https://www.hillpublisher.com/Journals/aitcs/
ISSN Online: 3066-3156

Deep Learning-based FPV Drone Target Detection

Junfei Chen[1,*], Bolin He[2]

[1]Immersion Academy, 4010 Barranca Parkway Suite 252, Irvine, California 92604, China.
[2]Wuhan Britain-China School, Wuhan Britain-China School (Qingyun Campus), No. 10 Gutian Ce Road, Qiaokou District, Wuhan 430034, China.

How to cite this paper: Junfei Chen, Bolin He. (2025) Deep Learning-based FPV Drone Target Detection. *Advance in Information Technology and Computer Science*, 2(1), 14-18.
DOI: 10.26855/aitcs.2025.06.003

Received: September 7, 2025
Accepted: September 11, 2025
Published: October 13, 2025

[*]**Corresponding author:** Junfei Chen, Immersion Academy, 4010 Barranca Parkway Suite 252, Irvine, California 92604, China.

Abstract

With the rapid development of unmanned aerial vehicle (UAV) technology and its expanding applications, potential security risks have become increasingly prominent. Some malicious actors exploit UAVs for unlawful activities. These UAVs are typically small in size, fast, and highly maneuverable, making them difficult to detect with traditional surveillance methods. Therefore, the exploration of more precise and efficient UAV detection methods carries considerable practical significance. In this study, based on real-world scenarios, first-person view (FPV) and racing drones were employed to collect image and video datasets containing multiple UAVs. The datasets were then annotated for training. Subsequently, the YOLOv11 object detection algorithm was applied for model training, enabling automatic detection and recognition of UAV positions within images. This approach provides a feasible technical pathway for UAV target detection and lays the groundwork for future studies on UAV monitoring in more complex environments.

Keywords

Object Detection; Aerial Targets; YOLOv11; FPV Drone Perspective

1. Introduction

In recent years, UAVs have developed rapidly and have gradually undertaken an increasing number of tasks in society. The recent proposal of a low-altitude economy has further accelerated their popularity. However, this rapid growth, coupled with the increasing accessibility of UAVs to the general public, has also led to a surge in incidents such as unauthorized flights ("black flights"), intrusions into restricted airspace (airports, nuclear facilities, and large-scale events), collisions with public and private assets, disruption of normal aviation operations, invasion of privacy, and even criminal activities including smuggling, espionage, and attacks. These occurrences pose serious threats to public safety, aviation security, critical infrastructure protection, and personal privacy [1]. Such UAVs typically fly at low altitudes, have small physical dimensions, and move at relatively low speeds. Traditional radar systems often struggle to detect them due to low-altitude multipath clutter and other interference. As a result, developing new detection methods has become highly necessary. The design of new detection techniques must also take cost efficiency into account, since excessively high costs would hinder the deployment of detection systems across large and no-fly zones. At the same time, these systems must deliver high resolution, ensuring the ability to distinguish UAVs from other low-altitude disturbances such as birds or balloons. Furthermore, given the diverse and often extreme weather conditions across the country, detectors must also be capable of operating reliably in all-weather and complex environments—further raising the challenges in developing such next-generation solutions.

2. Literature Review

2.1 Survey of Related Research Methods

In the areas of motion deblurring and small-object detection, numerous improved methods and technical approaches have been proposed. However, how to effectively reduce the false-alarm rate while maintaining high precision and,

at the same time, lowering algorithmic complexity to achieve efficient real-time monitoring remains an urgent challenge.

Existing UAV detection methods can broadly be categorized into traditional approaches and deep learning-based approaches. Traditional approaches generally follow two main ideas: (1) Extracting target features through feature engineering, and then employing a discriminant classifier to determine the target location. (2) Detecting moving objects in images first, and subsequently applying a generative classifier to decide whether the moving object is indeed the target. For instance, F. Lin [2] et al. proposed a UAV detection method based on Haar wavelets and AdaBoost. Flight experiments demonstrated that this method achieved satisfactory detection performance under relatively simple backgrounds, such as cloudy skies. In another study, F. Gökçe [3] et al. introduced a cascade approach that combined Haar-like features, Local Binary Patterns (LBP), and Histograms of Oriented Gradients (HOG). By integrating multiple features, this method improved detection robustness to some extent, but its processing speed was limited. Overall, traditional methods for UAV detection have achieved partial progress, yet their applicability is largely confined to scenarios where the target exhibits limited appearance variation or the background is relatively simple.

By contrast, deep learning-based methods have made remarkable breakthroughs in general object detection. To date, research on UAV visual detection based on deep learning remains relatively limited. For instance, some researchers have incorporated a motion compensation mechanism to detect aerial moving targets and subsequently employed a Convolutional Neural Network (CNN) to classify their features. While this approach demonstrated superior detection precision, it exhibited a strong dependence on high-precision measurements of camera motion parameters. In another study, J. James [4] combined SegNet with top-hat morphology to detect high-altitude aircraft. Although the method was able to identify large airplanes flying at an altitude of 2,800 m, its detection precision was only 13.4%. Moreover, several studies have explored the application of deep learning algorithms, such as YOLOv5, to UAV detection. Nevertheless, a systematic comparison and evaluation of the performance of different representative algorithms in UAV detection tasks remains absent.

2.2 Existing UAV Detection Datasets

At present, comprehensive datasets specifically designed for training deep learning algorithms in UAV detection are very limited. For instance, the dataset compiled by Rozantsev [5] comprises 20 video sequences, each containing about 4,000 grayscale frames with a resolution of 752 × 480. These images were captured by a camera mounted on another UAV in both indoor and outdoor environments, with the targets being UAVs in flight. The dataset compiled by Y. Chen [6] is composed of two subsets: the Public UAV Dataset, comprising 30 video sequences captured in both indoor and outdoor environments and involving various UAV models; and the USC UAV Dataset, consisting of 30 video clips of a single UAV target, recorded on the campus of the University of Southern California. In the USC dataset, most the samples feature clean or cloudy sky backgrounds, making the dataset relatively homogeneous compared with the dataset proposed in this paper. To enlarge the dataset, the authors of the USC dataset developed a model-based automatic data augmentation method, in which cropped UAV model images were pasted onto background images. Although this approach increases dataset size, X. Peng [7] has pointed out that models trained on such data may not achieve significant performance improvements merely through data expansion. Recently, V. Walter [8] introduced a new dataset called MIDGARD, which is characterized by diverse backgrounds and lighting conditions, and employs UltraViolet Direction And Ranging technology to achieve automatic annotation.

Through comparative analysis, our research group concluded that a vision-based detection approach best meets current needs. Its primary advantages lie in its relatively low cost, resilience against low-altitude multipath clutter and complex electromagnetic environments, as well as its compact equipment size, which facilitates deployment. To achieve target detection, however, the vision-based approach requires the integration of deep learning techniques. Accordingly, our group will carry out data collection, use the collected data to train models, and ultimately apply the trained models to perform UAV target detection tasks.

3. Algorithm Introduction

This study primarily employs the YOLOv11 algorithm to process the collected data. In the experiments, a deep learning model is trained to distinguish targets. The basic principle of the YOLO object detection model is that it divides the input image into an S × S grid. Each grid cell is assigned the task of predicting whether an object center lies within its boundaries. For those cells that contain a predicted object center, the model generates B candidate bounding boxes

together with their associated confidence scores. The bounding box with the highest confidence score is then selected as the final detection output for the target.

Figure 1. Prediction of Grid Cell Centers.

In this experiment, the YOLO deep learning framework is adopted, with the working principle of YOLOv11 illustrated as follows:

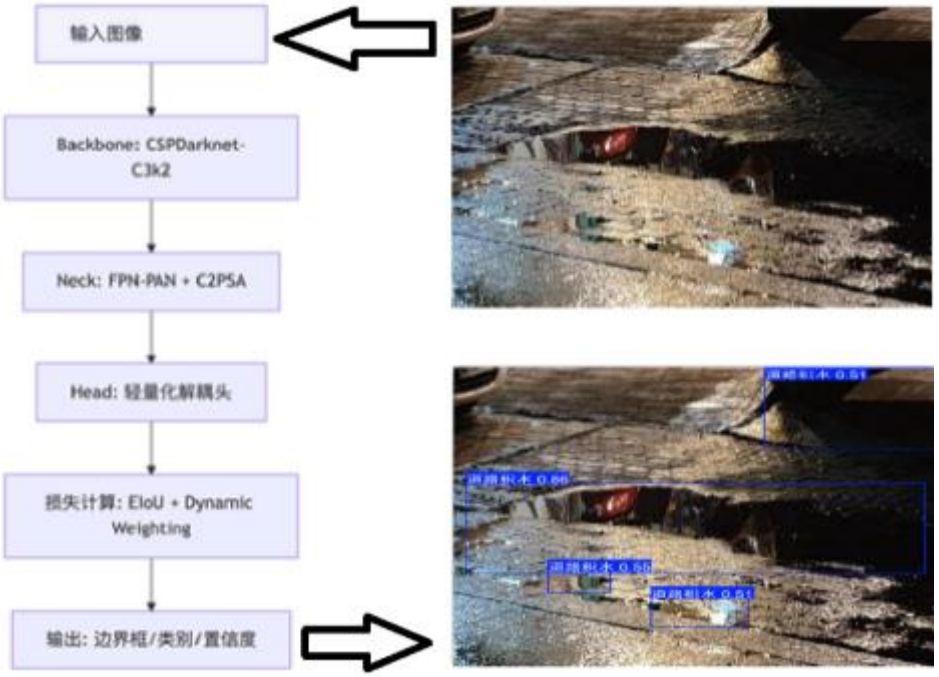

Figure 2. Framework of the YOLOv11 Model

The backbone is responsible for efficiently extracting multi-scale features from the input images, thereby reducing computational redundancy and providing high-quality representations for subsequent processing. The neck integrates the multi-scale features extracted by the backbone through top-down and bottom-up fusion using a Feature Pyramid Network (FPN) combined with a Path Aggregation Network (PANet). The head, built upon the fused feature pyramid provided by the neck, employs independent branches to perform parallel predictions of classification probabilities and bounding box coordinates/confidence scores across multiple scales. Its primary function is to achieve task

decoupling so as to minimize conflicts and enhance both classification and localization precision. In the final output stage, suspected targets are enclosed within bounding boxes and assigned confidence scores, with only those exceeding the predefined threshold being displayed as valid detection results.

4. Dataset Description

The dataset used in this experiment was constructed by extracting 1,100 high-definition images from a total of 20 minutes of video footage recorded by FPV racing drones. The captured UAV models include Air3s, Mavic2, and Mavic3. Data collection was carried out at three different time periods—midday (11:00-12:00), afternoon (14:00-15:00), and evening (17:00-18:00)—thus covering both strong and weak illumination conditions. The dataset also incorporates frequent variations in viewpoints and backgrounds, including complex environmental settings. During the shooting process, the distance between the target and the camera ranged from 30 cm to 20 m, while the flight altitude ranged from 5 m to 40 m, providing rich variations in both distance and altitude. The UAV images are annotated with between 1 and 4 target instances.

5. Experimental Results and Analysis

5.1 Experimental Hardware and Software

To verify the effectiveness of the UAV aerial target detection method, a corresponding experimental hardware and software platform was established. The hardware part was mainly used for data acquisition and model operation, including the aircraft, imaging equipment, and computing resources; while the software part covered the operating system, programming language, and a deep learning framework. The specific experimental hardware and software configurations are shown in Table 1 and 2.

Table 1. FPV Racing Drone Parameters

Component	Parameter
Frame size	5-inch
Camera CMOS size	1/1.7-inch
Max video quality	4k&60fps
Video transmission	DJI O3 Air Unit
Max speed	150-200 km/h
Endurance time	5 minutes

Table 2. Operating Environment

Condition	Environment
Operating system	Ubuntu 20.04
Programming language	Python3.8
GPU	RTX2080.Ti
Deep learning frameworks	Pytorch_2.4.1
Libraries	OpenCV, math, torch, numpy

5.2 Experimental Results

In the experiments, we used mAP50 (Mean Average Precision) and mAP50-95 to measure detection precision. mAP is a commonly used evaluation metric in object detection models. It refers to the mean of average precision (AP) values across all classes and all images, and is widely applied to evaluate algorithm performance in object detection tasks. mAP50 represents the mean average precision when the IoU (Intersection over Union) threshold is set to 0.5. mAP50-95 is calculated as the mean average precision over IoU thresholds ranging from 0.5 to 0.95. Compared with

mAP50, which considers only an IoU threshold of 0.5, mAP50-95 provides a more comprehensive evaluation and can more precisely reflect the model's performance under different levels of overlap. The experimental results are shown in Table 3.

Table 3. Detection Targets and Experimental Results

Target type	Number of detection images	Total detections	Precision	Recall	mAP50	mAP50-95
All types	194	514	0.592	0.24	0.295	0.103
Mavic3	192	322	0.519	0.339	0.34	0.129
Mavic2	172	172	0.532	0.25	0.267	0.108
Air3	20	20	0.72	0.132	0.279	0.0718

6. Conclusion and Outlook

In this experiment, we used FPV drones to capture images of Mavic2, Mavic3, and Air3 UAVs, and manually constructed a dataset containing over one thousand images with annotated UAV target positions. On this basis, the YOLOv11 object detection algorithm was employed for training. The experimental results show that YOLOv11 can precisely identify and locate UAVs in images and distinguish their categories, demonstrating overall excellent detection performance. However, some limitations persist with respect to detection precision. For instance, when the FPV drone is positioned at a greater distance from the UAV under observation, the target appears at a very small scale in the image, which hinders precise localization by the algorithm. Moreover, in complex background scenarios (such as scenes with dense trees or multicolored buildings rather than a single-color sky or grassland), UAV targets tend to blend into the environment, thereby reducing detection reliability. In summary, this study reveals the limitations of existing algorithms in complex environments. In the future, it is urgent to develop target detection methods with higher precision and stronger robustness to effectively meet the needs of UAV detection in diverse scenarios and to provide technical support for preventing potential UAV-related security risks.

References

[1] YANG Y, JIANG W, GAO Z. Real-time target detection algorithm for low altitude UAVs. Acta Aeronautica et Astronautica Sinica [Internet]. [cited 2025 Aug 19];1-13. Available from: (Note: The journal's official citation style may differ; this follows Vancouver conventions for an online first article with a known citation date but no volume/issue/pagination).

[2] Lin F, Peng K, Dong X, Zhao S, Chen BM. Vision-based formation for UAVs. In: Proceedings of the IEEE International Conference on Control and Automation (ICCA). 2014. p. 1375-80.

[3] Gökçe F, Üçoluk G, Şahin E, Kalkan S. Vision-based detection and distance estimation of micro unmanned aerial vehicles. Sensors [Internet]. 2015 [cited 2025 Oct 11];15(9):23805-46. Available from: https://doi.org/10.3390/s150923805

[4] James J, Ford JJ, Molloy TL. Learning to detect aircraft for long-range vision-based sense-and-avoid systems. IEEE Robot Autom Lett. 2018 Oct;3(4):4383-90.

[5] Rozantsev A, Lepetit V, Fua P. Flying objects detection from a single moving camera. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2015. p. 4128-36.

[6] Chen Y, Aggarwal P, Choi J, Kuo CCJ. A deep learning approach to drone monitoring. In: Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). 2017. p. 686-91.

[7] Peng X, Sun B, Ali K, Saenko K. Learning deep object detectors from 3D models. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2015. p. 1278-86.

[8] Walter V, Vrba M, Saska M. On training datasets for machine learning-based visual relative localization of micro-scale UAVs. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). 2020. p. 10674-80.