Multidisciplinary Journal of Engineering and Technology
ISSN: 3007-6900
DOI: 10.61784/mjet3005

IDENTIFICATION AND TRACKING OF AERIAL UAVS BASED ON DEEP LEARNING VISUAL ALGORITHMS

JianJun Song
Shanghai Technical Institute of Electronics Information, Shanghai 201411, China.
Corresponding Author: JianJun Song, Email: songjianjun151@163.com

Abstract: When conducting long-range wireless charging and monitoring of Unmanned Aerial Vehicles (UAVs) in the air, remote identification and tracking of the drones in the air are required. To address this issue, a deep learning-based algorithm for aerial UAVs identification and tracking is proposed. The YOLO algorithm is utilized for UAVs identification in the air, and the Deep Sort algorithm is used for tracking the identified UAVs. A model and training dataset are constructed, and the deep learning model is trained. The trained model is then invoked to verify the identification and tracking effectiveness of the UAVs in the air.

Keywords: Deep learning visual algorithms; YOLO; Deep SORT; UAVs identification and tracking

1 INTRODUCTION

Long-distance wireless power transmission, applied to Unmanned Aerial Vehicles (UAVs), has the potential to be a breakthrough technology for efficiently charging remote-flying UAVs [1]. Currently, the use of microwave and laser for long-distance wireless charging of UAVs has become feasible [2]. The rapid advancement of technology will gradually enhance the safety and reliability of long-distance wireless charging for UAVs. To achieve long-range wireless charging for UAVs, real-time alignment and tracking of the UAVs must be performed. Therefore, utilizing visual methods for UAVs identification and tracking is an economical and feasible approach.

In recent years, the rapid development of UAVs has raised many security issues. Various Anti-Unmanned Aerial Vehicle Defense Systems (AUDS) have been developed worldwide, including radars, high-energy laser guns, and other anti-drone defense systems. However, these systems are expensive and difficult to be widely accepted in civilian scenarios [3]. The use of visual methods for UAV identification and tracking is evidently more economical and easier to apply.

To address the security issues caused by UAVs, researchers have begun to use deep learning-based image algorithms to identify UAVs, achieving good results [4]. Some researchers have improved the YOLO algorithm to enhance the identification of UAVs [5, 6]. In addition to using deep learning algorithms for aerial UAV identification, researchers have also used deep learning algorithms to predict the trajectories of UAVs [7].

Moreover, the hierarchical image-based target tracking of UAVs is also an important task. Previous researchers have conducted a significant amount of research and application on common multi-target tracking algorithms in various fields, such as droplet identification and tracking in microfluidic control and object recognition and tracking in logistics warehouses, achieving certain results [8-11]. Kumar S et al. utilized a deep learning algorithm for vehicle detection and tracking [12]. Razzok M et al. based on deep learning image algorithms for pedestrian detection and tracking [13]. Some researchers have attempted to use deep learning-based methods for the identification and tracking of UAVs [14], but overall, research and application in the identification and tracking of aerial UAVs are relatively limited.

This study proposes a deep learning-based method for the identification and tracking of aerial UAVs. The YOLO algorithm will be used to identify aerial UAVs, and the Deep SORT algorithm will be used to track the identified UAVs. A recognition and tracking model will be constructed, and a training dataset will be established to train the deep learning model. The trained model will be invoked to validate the effectiveness of aerial UAV identification and tracking.

2 METHODOLOGY

2.1 Method Overview

A method for the identification and tracking of UAVs is proposed based on the combination of YOLO algorithm and Deep SORT algorithm, as illustrated in Figure 1. The method mainly includes model construction, training deployment and invocation, image acquisition, image preprocessing, image distortion correction, recognition, and tracking.

Model Construction: A model combining the YOLO algorithm and Deep SORT algorithm is built for object detection and tracking of UAVs.

Training, Deployment, and Invocation: The constructed algorithms are trained and subsequently deployed for the recognition and tracking of aerial UAVs.

Image Acquisition: UAV images are captured using a camera for further processing.

Image Preprocessing: Preprocessing of the captured images includes tasks such as scaling, smoothing, and conversion to grayscale.

Image Distortion Correction: Image distortion correction aims to eliminate distortions that may occur during image capture, enhancing accuracy in presenting the original scene.
UAV Recognition and Tracking: Utilizing the constructed and trained deep learning model, images or videos are processed to identify UAVs, followed by visual tracking and analysis using deep learning tracking algorithms.

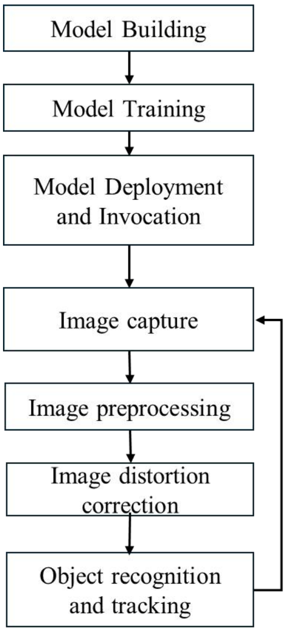
Model Building
Model Training
Model Deployment and Invocation
Image capture
Image preprocessing
Image distortion correction
Object recognition and tracking

Figure 1 Method Overview

2.1.1 Experimental platform

An experimental setup, as illustrated in Figure 2, is constructed for the purpose of capturing images of aerial UAVs. This experimental platform primarily consists of a host computer, a camera, and a support frame. The system is designed to perform real-time, multi-angle imaging and processing of UAV images. The camera is responsible for capturing images of the UAVs, while the host computer runs the deep learning visual model and handles the real-time processing of the images collected by the camera.

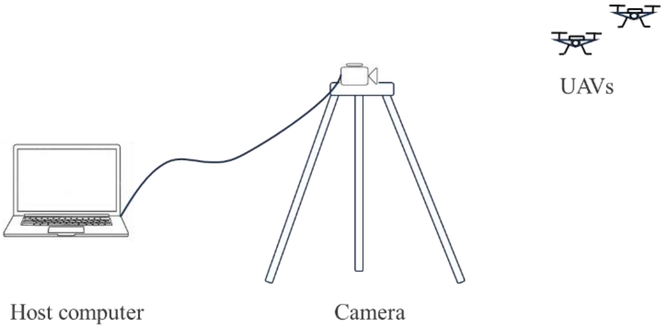

Figure 2 Experimental Setup

2.1.2 Principle of YOLO + deep SORT algorithm

Combining the YOLO and Deep SORT algorithms enables the accomplishment of multi-object detection and tracking tasks. As shown in Figure 3, initially, the YOLO algorithm performs object detection, identifying the positions and class information of target objects in each frame. Subsequently, the Deep SORT algorithm utilizes these detection results for data association and trajectory updating, thereby achieving multi-object tracking.

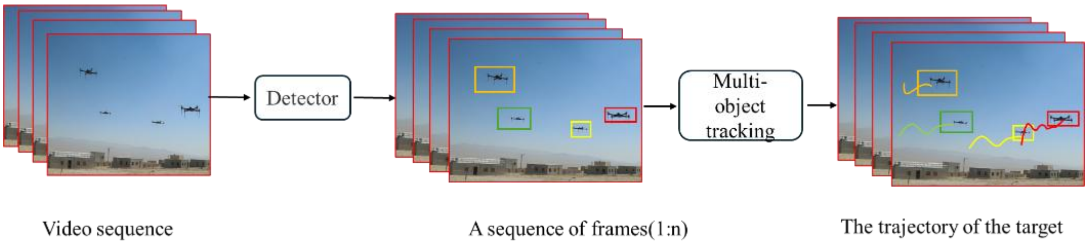

Figure 3 The Schematic Diagram of the Algorithm Principle

The tracking process of the SORT algorithm is as follows:
(1) Utilize the YOLO algorithm for target detection in the video sequence to obtain the position and size information of target objects in each frame.
(2) Extract features from each target object in every frame to obtain its deep feature vector.
(3) Initialize a tracking list to store information about currently tracked target objects.
(4) For each target object in every frame, compute its Mahalanobis distance and Hungarian matching cost with target objects in the tracking list to find the best match.
(5) If a target object successfully matches with a target object in the tracking list, update its position and velocity information.
(6) If a target object fails to match with any target object in the tracking list, add it to the tracking list and initialize its position and velocity information.
(7) Repeat the above steps until the entire video sequence is processed.
The process of the Deep SORT algorithm is more complex compared to the SORT algorithm, incorporating appearance feature extraction and cascade matching.

2.1.3 Method process

Just like Figure 4, integration of YOLO and Deep SORT algorithms for UAV recognition and tracking applications:
(1) Network Model Construction: Develop a deep learning model combining YOLO and Deep SORT.
(2) Dataset Construction: Build a UAV dataset for training the YOLO algorithm for target recognition. Construct a REID dataset for training the Deep SORT tracking algorithm to track UAVs.
(3) YOLO Model Training: Train the YOLO model for UAV recognition.
(4) YOLO Model Invocation and Evaluation: Invoke the trained YOLO model for practical use, evaluation, and optimization.
(5) Deep SORT Model Training: Train the Deep SORT algorithm model for multi-object tracking of UAVs.
(6) Deep SORT Model Invocation and Evaluation: Invoke the trained Deep SORT model for practical use, evaluation, and optimization.
(7) Overall Model Invocation: Invoke the models for UAV recognition and multi-object tracking in a comprehensive manner.

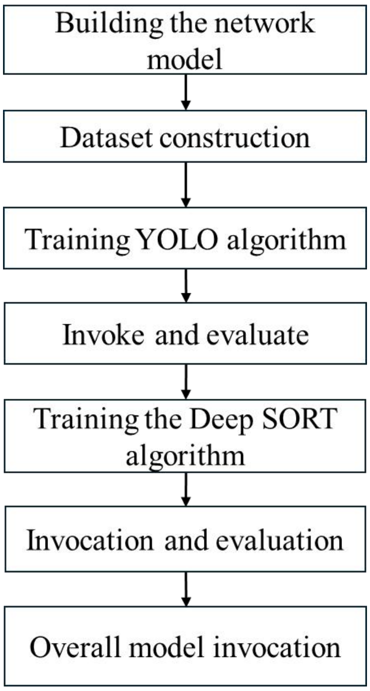

Figure 4 Method Process

2.2 UAV Recognition Based on YOLO Algorithm

2.2.1 Object recognition and performance metrics

Object detection refers to the process of using computer vision techniques to identify and localize specific objects or targets in images or videos. The following are performance metrics for object detection algorithms:
Precision: Precision evaluates how accurate the predictions are in terms of correctness:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{1}$$

Recall Rate, which evaluates how comprehensive the search is in terms of completeness:

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{2}$$

Intersection over Union (IOU), as shown in Figure 5, is a metric used to evaluate the overlap between predicted and ground truth bounding boxes. If the IOU is greater than a threshold, the prediction is considered positive. The calculation method is as follows:

$$IOU = \frac{Area\ of\ Overlap}{Area\ of\ Uion} \tag{3}$$

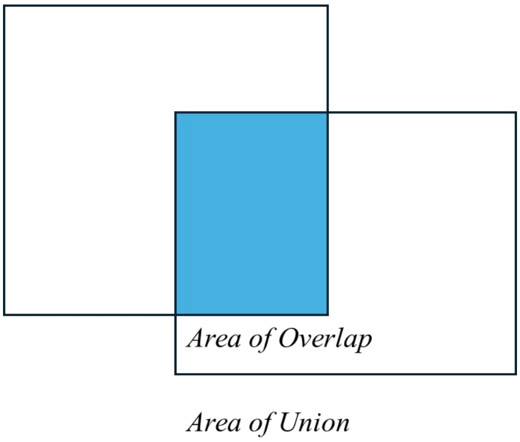

Figure 5 The Schematic Diagram of the IOU

Average Precision (AP) is a measure of how well a learned model performs on each category.

Mean Average Precision (MAP) measures the overall performance of the learned model across all categories, calculated by taking the average of AP values for all categories.

Frames Per Second (FPS) refers to the number of image frames processed per second, used to evaluate the speed and efficiency of computer vision systems in handling images or videos.

2.2.2 YOLO algorithm principle

YOLO stands as a paradigmatic example of a single-stage object detection algorithm. Figure 6 provides a schematic depiction of the YOLO algorithm's core mechanics. The process can be broadly demarcated into three principal stages:

(1) Adjustment of the input image to a standardized size.

(2) Acquisition of feature maps through the utilization of a convolutional neural network.

(3) Application of Non-Maximum Suppression (NMS) to the entire set of predicted results, followed by the outputting of the final outcomes.

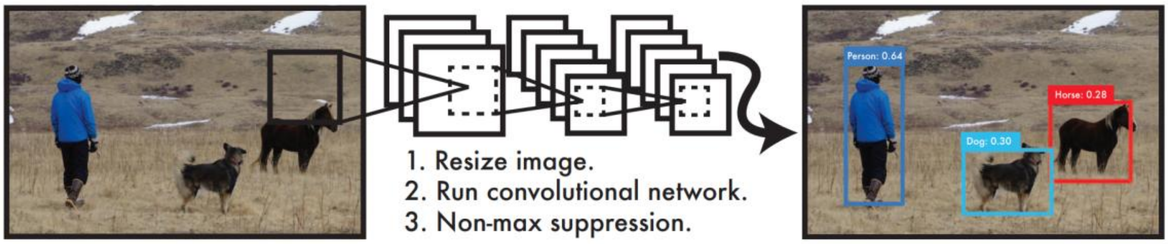

Figure 6 Schematic Diagram of the YOLO Algorithm [15]

Using the YOLOv5 algorithm as an example, the network structure is explained, which mainly consists of three parts: Backbone, Neck, and Head.

Backbone: A convolutional neural network that aggregates and forms image features at different levels of granularity. YOLOv5 uses CSPDarknet53 as its Backbone, which, while maintaining high accuracy, improves model inference speed through enhanced network structure.

Neck: A series of network layers that mix and combine image features and pass them to the prediction layer. YOLOv5 uses Path Aggregation Network (PANet) as its Neck, achieving the fusion and utilization of multi-scale features through top-down path enhancement and bottom-up feature fusion.

Head: Predicts image features, generates bounding boxes, and predicts categories. YOLOv5 adopts the Anchor-Free approach from YOLOv3, optimizing network parameters by calculating the loss between predicted boxes and ground truth bounding-boxes to achieve accurate object detection.

Figure 7 depicts the procedure of the YOLO algorithm, which consists of the following steps:

(1) The image is divided into an S*S grid.

(2) Each grid cell predicts M bounding boxes, with the prediction information for each bounding box including its position, confidence level, the probability of an object's presence, and the probability of belonging to a specific class.

(3) Non-Maxima Suppression (NMS) is conducted, utilizing the predicted probabilities of objects and IOU values to filter out the necessary bounding boxes for output.

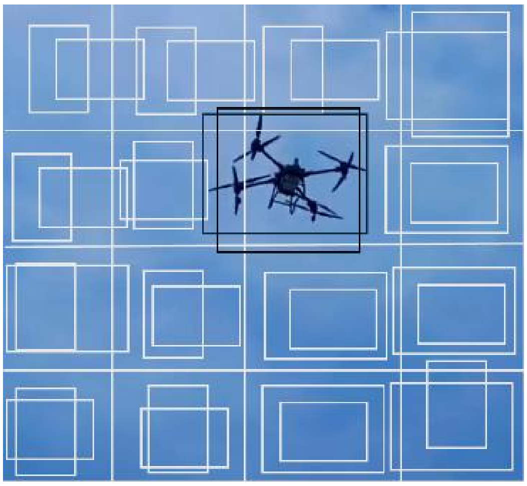

Figure 7 Schematic Diagram for UAVs Forecasting with YOLO Algorithm

2.3 Multi-Target Tracking of UAVs Based on the Deep SORT Algorithm

2.3.1 Multiple object tracking and evaluation

The Multiple Object Tracking (MOT) refers to the process of locating each target in every frame of an image and tracing their trajectories. An input video sequence yields the trajectories of targets and a unique ID for each target, with each target represented by a bounding box.

The evaluation criteria for MOT are primarily categorized into the following four types:

(1) Accuracy: This metric measures the accuracy of the target tracking performed by the algorithm. The ID Switches (IDSW) metric quantifies the number of times the MOT algorithm switches between different objects. The Multiple Object Tracking Accuracy (MOTA) metric calculates the false positive rate, false negative rate, and unmatched detection rate, combining these into a numerical value that relatively fairly reflects the overall tracking performance. Despite its limitations, this is the most widely accepted MOT evaluation metric to date.

(2) Precision: Metrics such as the Multiple Object Tracking Precision (MOTP), Tracking Distance Error (TDE), and Optimal Sub-Pattern Assignment (OSPA) are widely used as critical evaluation criteria. These metrics consider both the overlap of bounding boxes and distance measurements, describing the degree to which objects are accurately tracked. Additionally, they consider false alarms and label errors, providing a comprehensive perspective on the performance evaluation of multi-target tracking systems.

(3) Completeness: Completeness metrics indicate the extent to which the ground truth trajectories are tracked. This includes measures such as Mostly Tracked (MT), Partially Tracked (PT), Mostly Lost (ML), and Fragmentation (FM), which belong to this category.

(4) Robustness: To assess the MOT algorithm's capability to recover from occlusions, metrics referred to as Recovery from Short-Term Occlusions (RS) and Recovery from Long-Term Occlusions (RL) have been introduced.

2.3.2 Principles of deep SORT

The Deep SORT algorithm employs a target detector to detect targets, as shown in Figure 8, and associates the detected target with the Kalman filter-predicted trajectory using association metric. The so-called association metric considers appearance features and Mahalanobis distance, conducts feature extraction and similarity estimation, constructs a cost matrix using Mahalanobis distance and appearance features, and performs matching using the Hungarian algorithm.

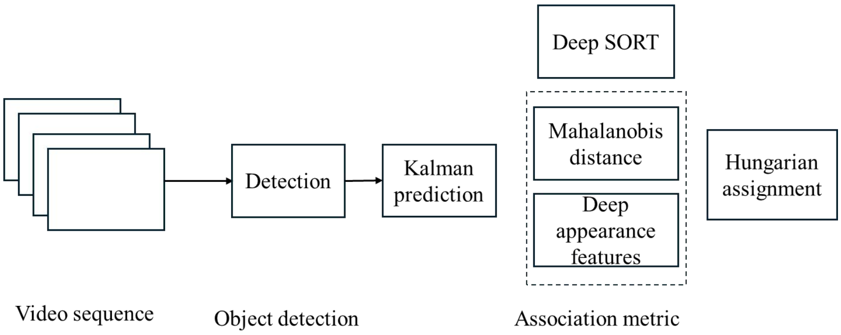

Figure 8 Schematic Diagram of the Deep SORT Algorithm Principle

The Kalman filter utilizes a linear system state equation to perform optimal estimation on the system state through observations of the system's inputs and outputs. The recursive optimal estimation theory of the Kalman filter, adopting the state-space representation method, is capable of handling multi-dimensional and non-stationary stochastic processes. The Kalman filter predicts and updates the state from one frame to the next based on the previous frame's state and uses the measurements from the current frame to update this prediction. The Kalman filter can be used to estimate the motion trajectory of targets, which can enhance the accuracy of tracking, especially in scenarios where the target moves rapidly or is obscured.

The Mahalanobis distance is a method for measuring the distance between multidimensional data points, considering the covariance of the data. Unlike Euclidean distance, Mahalanobis distance effectively handles variables with different scales and correlations. The square of the Mahalanobis distance should conform to a chi-square distribution with p degrees of freedom.

The assignment problem deals with assigning machines to tasks, workers to jobs, and so on. The goal is to determine the optimum assignment that, for example, minimizes the total cost. The Hungarian algorithm is a combinatorial optimization algorithm that solves the assignment problem in polynomial time.

3 CASE STUDY

In the context of long-range, wireless aerial charging for UAVs, accurate target and tracking of airborne UAVs are essential. This study employs a combined method using YOLO and Deep SORT for the detection and tracking of aerial multi-rotor drones. A network model was constructed, a dataset was prepared, and the model was trained and deployed. Figure 9 illustrates the scenario of target detection and tracking for UAVs.

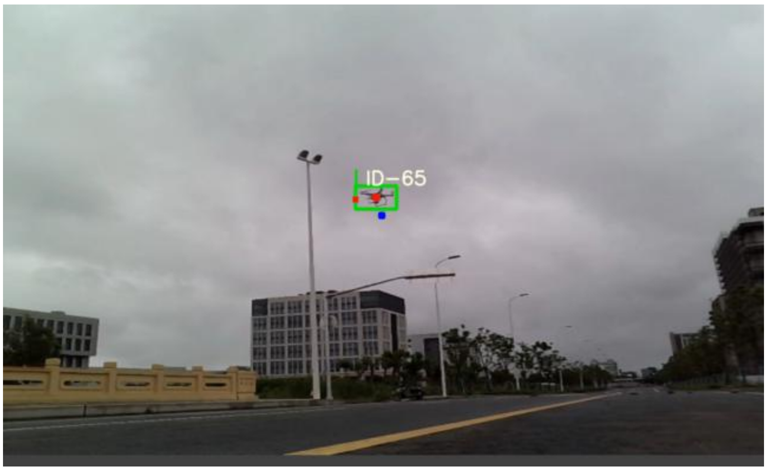

Figure 9 Identification and Tracking of UAVs

3.1 Dataset Creation

3.1.1 Algorithm dataset identification

10,000 images of UAVs were collected to establish the dataset, as shown in Figure 10. These images were annotated to mark the positions of UAVs and label their categories. Figure 11 illustrates the use of the Label-image tool for annotating UAV images.

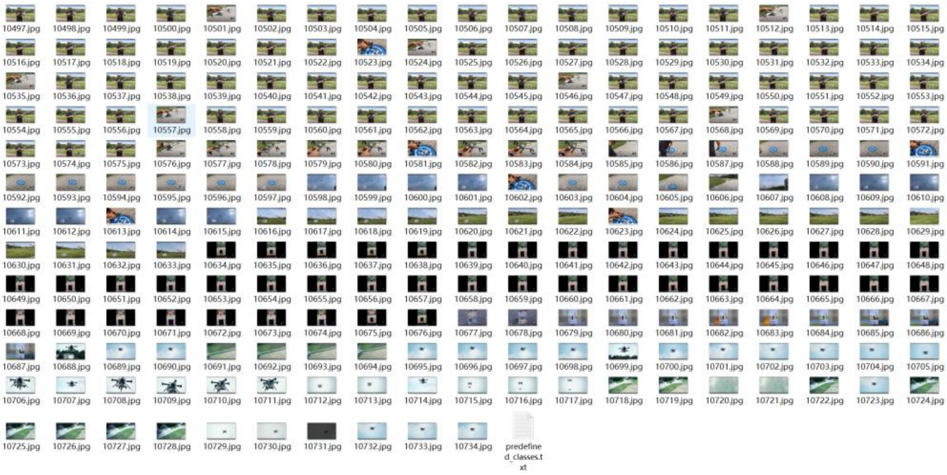
10497.jpg 10498.jpg 10499.jpg 10500.jpg 10501.jpg 10502.jpg 10503.jpg 10504.jpg 10505.jpg 10506.jpg 10507.jpg 10508.jpg 10509.jpg 10510.jpg 10511.jpg 10512.jpg 10513.jpg 10514.jpg 10515.jpg
10516.jpg 10517.jpg 10518.jpg 10519.jpg 10520.jpg 10521.jpg 10522.jpg 10523.jpg 10524.jpg 10525.jpg 10526.jpg 10527.jpg 10528.jpg 10529.jpg 10530.jpg 10531.jpg 10532.jpg 10533.jpg 10534.jpg
10535.jpg 10536.jpg 10537.jpg 10538.jpg 10539.jpg 10540.jpg 10541.jpg 10542.jpg 10543.jpg 10544.jpg 10545.jpg 10546.jpg 10547.jpg 10548.jpg 10549.jpg 10550.jpg 10551.jpg 10552.jpg 10553.jpg
10554.jpg 10555.jpg 10556.jpg 10557.jpg 10558.jpg 10559.jpg 10560.jpg 10561.jpg 10562.jpg 10563.jpg 10564.jpg 10565.jpg 10566.jpg 10567.jpg 10568.jpg 10569.jpg 10570.jpg 10571.jpg 10572.jpg
10573.jpg 10574.jpg 10575.jpg 10576.jpg 10577.jpg 10578.jpg 10579.jpg 10580.jpg 10581.jpg 10582.jpg 10583.jpg 10584.jpg 10585.jpg 10586.jpg 10587.jpg 10588.jpg 10589.jpg 10590.jpg 10591.jpg
10592.jpg 10593.jpg 10594.jpg 10595.jpg 10596.jpg 10597.jpg 10598.jpg 10599.jpg 10600.jpg 10601.jpg 10602.jpg 10603.jpg 10604.jpg 10605.jpg 10606.jpg 10607.jpg 10608.jpg 10609.jpg 10610.jpg
10611.jpg 10612.jpg 10613.jpg 10614.jpg 10615.jpg 10616.jpg 10617.jpg 10618.jpg 10619.jpg 10620.jpg 10621.jpg 10622.jpg 10623.jpg 10624.jpg 10625.jpg 10626.jpg 10627.jpg 10628.jpg 10629.jpg
10630.jpg 10631.jpg 10632.jpg 10633.jpg 10634.jpg 10635.jpg 10636.jpg 10637.jpg 10638.jpg 10639.jpg 10640.jpg 10641.jpg 10642.jpg 10643.jpg 10644.jpg 10645.jpg 10646.jpg 10647.jpg 10648.jpg
10649.jpg 10650.jpg 10651.jpg 10652.jpg 10653.jpg 10654.jpg 10655.jpg 10656.jpg 10657.jpg 10658.jpg 10659.jpg 10660.jpg 10661.jpg 10662.jpg 10663.jpg 10664.jpg 10665.jpg 10666.jpg 10667.jpg
10668.jpg 10669.jpg 10670.jpg 10671.jpg 10672.jpg 10673.jpg 10674.jpg 10675.jpg 10676.jpg 10677.jpg 10678.jpg 10679.jpg 10680.jpg 10681.jpg 10682.jpg 10683.jpg 10684.jpg 10685.jpg 10686.jpg
10687.jpg 10688.jpg 10689.jpg 10690.jpg 10691.jpg 10692.jpg 10693.jpg 10694.jpg 10695.jpg 10696.jpg 10697.jpg 10698.jpg 10699.jpg 10700.jpg 10701.jpg 10702.jpg 10703.jpg 10704.jpg 10705.jpg
10706.jpg 10707.jpg 10708.jpg 10709.jpg 10710.jpg 10711.jpg 10712.jpg 10713.jpg 10714.jpg 10715.jpg 10716.jpg 10717.jpg 10718.jpg 10719.jpg 10720.jpg 10721.jpg 10722.jpg 10723.jpg 10724.jpg
10725.jpg 10726.jpg 10727.jpg 10728.jpg 10729.jpg 10730.jpg 10731.jpg 10732.jpg 10733.jpg 10734.jpg predefined_classes.txt

Figure 10 UAVs Dataset

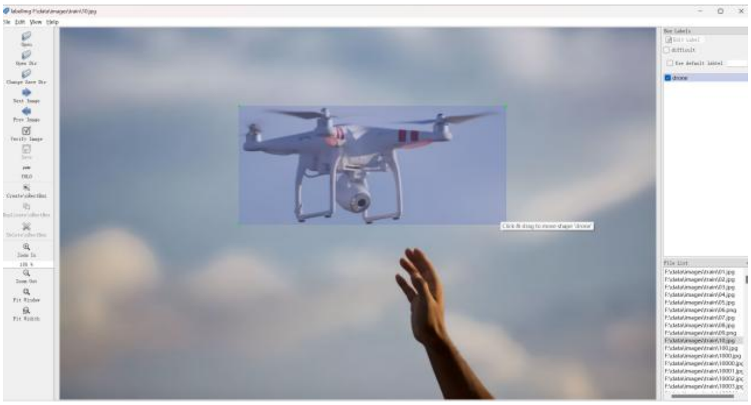

Figure 11 Labeling UAVs

3.1.2 REID dataset

Figure 12 presents the UAV REID dataset, using 0009_c001_00016401_0_c01_t4.jpg as an example to elucidate the image naming convention:
0009 represents the unique identifier for each UAV, ranging from 0001 to 0020, indicating a total of 20 UAVs.
c001 denotes the first camera (where "c" stands for Camera), with a total of N cameras.
00016401 indicates the 16401st frame captured by camera c1, with a video frame rate (FPS) of 25.

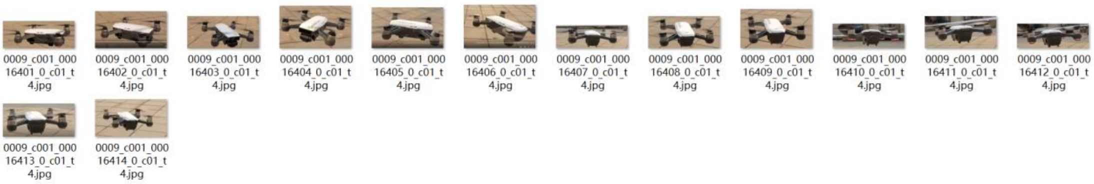
0009_c001_00016401_0_c01_t4.jpg
0009_c001_00016402_0_c01_t4.jpg
0009_c001_00016403_0_c01_t4.jpg
0009_c001_00016404_0_c01_t4.jpg
0009_c001_00016405_0_c01_t4.jpg
0009_c001_00016406_0_c01_t4.jpg
0009_c001_00016407_0_c01_t4.jpg
0009_c001_00016408_0_c01_t4.jpg
0009_c001_00016409_0_c01_t4.jpg
0009_c001_00016410_0_c01_t4.jpg
0009_c001_00016411_0_c01_t4.jpg
0009_c001_00016412_0_c01_t4.jpg
0009_c001_00016413_0_c01_t4.jpg
0009_c001_00016414_0_c01_t4.jpg

Figure 12 UAVs REID Dataset

3.2 Training and Analysis of YOLO Algorithm

3.2.1 Confusion matrix

The confusion matrix is an n*n matrix, where n represents the number of classes. Each row represents the actual class, and each column represents the predicted class by the model. The size of the confusion matrix depends on the number of classes involved, making it applicable to various classification tasks, whether binary or multiclass.

Figure 13 depicts the confusion matrix for drone detection. From the figure, it can be observed that when the prediction is for a UAVs, the actual probability of being a drone is 0.96; when predicted as background, the actual probability of being UAV is 0.04. This indicates that the model exhibits high accuracy in UAVs detection.

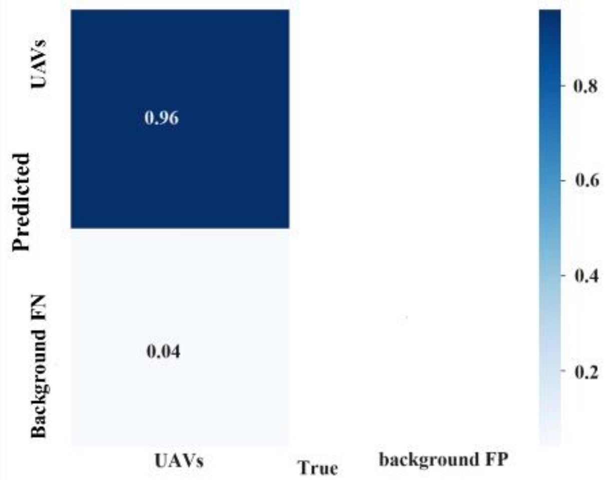

Figure 13 Confusion Matrix for UAVs Detection

3.2.2 F1-Score

The F1 score is the harmonic mean of Precision and Recall, simultaneously considering the accuracy and coverage of the model's predictions. The F1 score ranges from 0 to 1, where 1 indicates that the model's Precision and Recall are both perfect, while 0 suggests that either the Precision or Recall is extremely poor.

From Figure 14, it can be observed that when the confidence is 0.6, the F1-score for drones is 0.84. This indicates that the model performs well in predicting the “UAV” class, demonstrating high precision and recall.

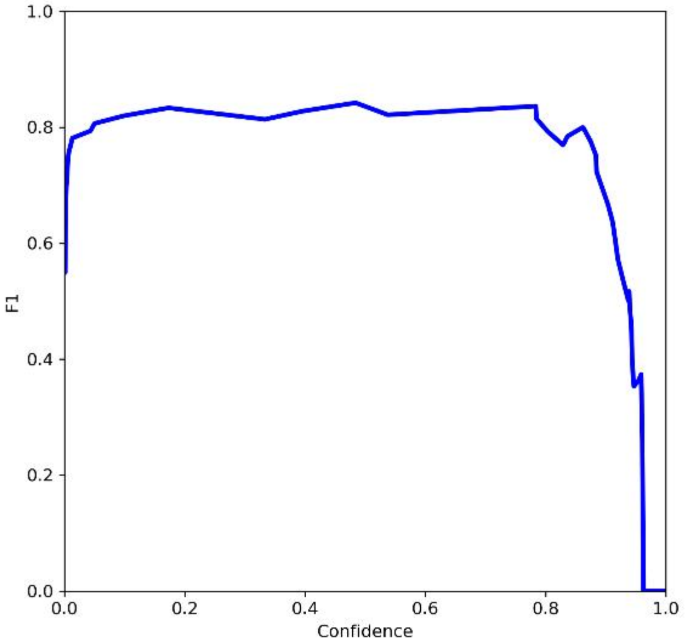

Figure 14 The F1 Score for UAV Detection

3.2.3 Precision

The blue line in the figure 15 represents the “UAV” category. The overall trend shows an increase in precision as confidence increases. Precision reaches 1 at a confidence level of 0.95.

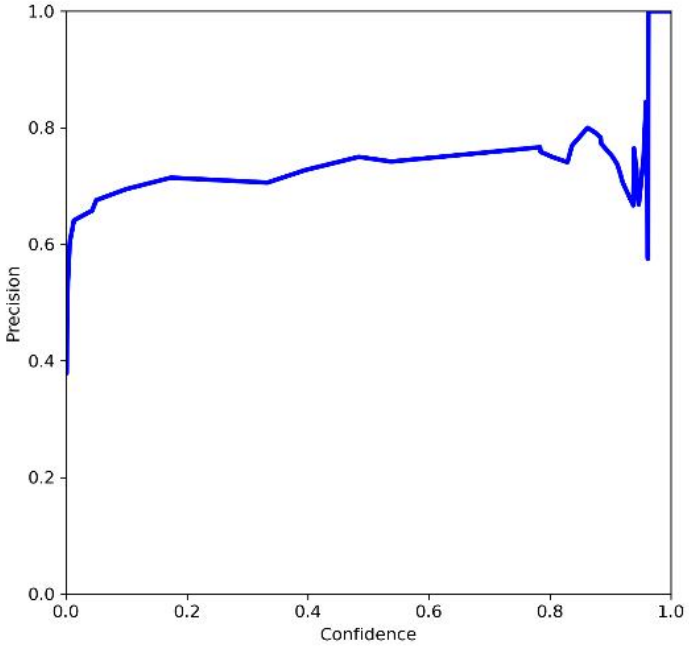

Figure 15 Precision Curve

3.2.4 Recall curve

Figure 16 displays the Recall-Confidence Curve (R curve), illustrating recall at various confidence levels. Specifically, the horizontal axis represents confidence ranging from 0.0 to 1.0, while the vertical axis represents recall, also ranging from 0.0 to 1.0.

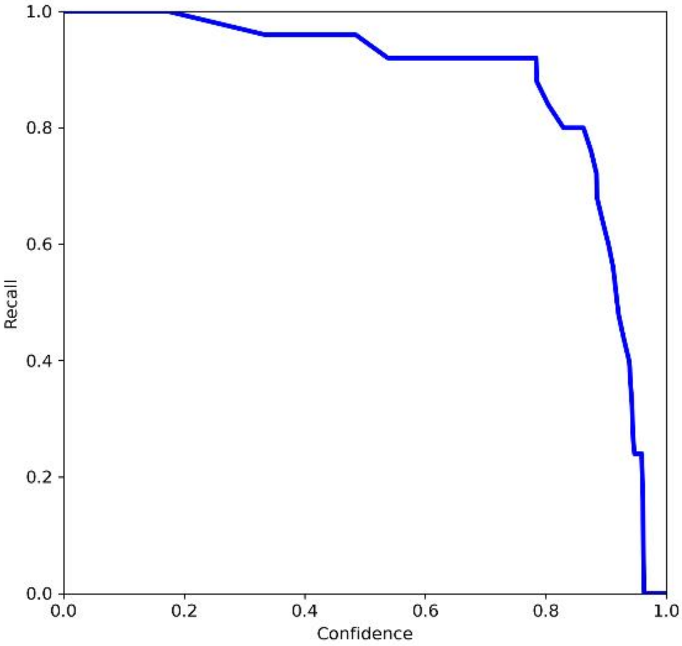

Figure 16 Recall-Confidence Curve

This curve indicates that the recall for the “UAV” category initially remains stable with, followed by a rapid decline after a certain point.

3.2.5 Precision-recall curves

Figure 17 presents the Precision-Recall Curve, the trade-off between precision and recall achieved by the at different threshold settings, the x-axis Recall (ranging from 0.0 to 1.0), while the y-axis Precision (ranging from 0.0 to 1.0). The Precision-Recall Curve for the “UAV” category exhibits a stable precision around 0.8.

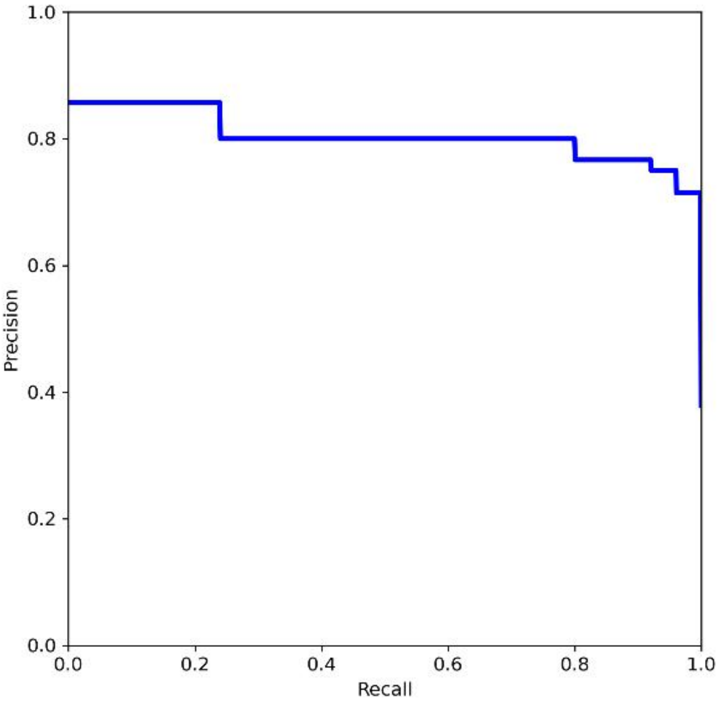

Figure 17 Precision-Recall Curve

3.2.6 Loss function

Figure 18 illustrates the loss during the training process of the YOLOv5 algorithm. The trends of different loss categories are shown in the figure:

(1) Localization loss (Box Loss): YOLOv5 utilizes generalized intersection over union (GIOU) loss as the loss function for bounding boxes. The box loss is inferred as the mean GIOU loss of the predicted boxes, and a smaller box loss value indicates more accurate box positioning.

(2) Confidence loss (Obj Loss): It is presumed to be the mean loss for object detection, where a smaller obj loss value indicates more accurate object detection.
(3) Validation set localization loss (Val Box Loss): This represents the bounding Box Loss on the validation set.
(4) Validation set confidence loss (Val Obj Loss): This represents the mean Object loss on the validation set.

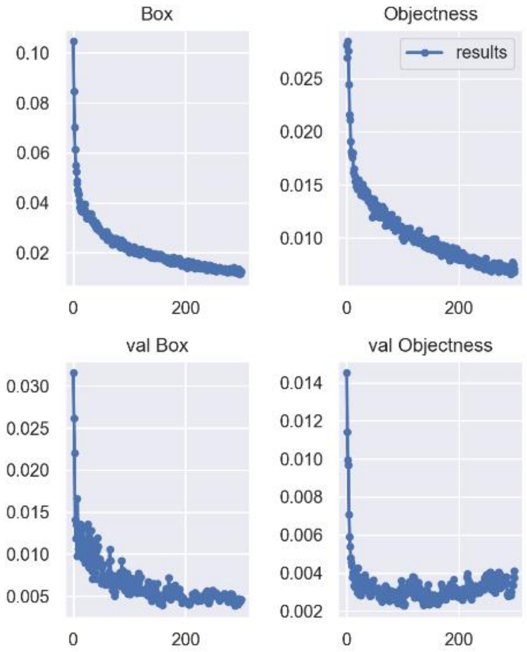

Figure 18 The Loss during the Training Process

3.2.7 Performance on the test set

Figure 19 illustrates the recognition performance of UAVs on the test dataset. The results indicate that the trained algorithm demonstrates good recognition capability, effectively identifying UAVs of various types, sizes, and within different background environments.

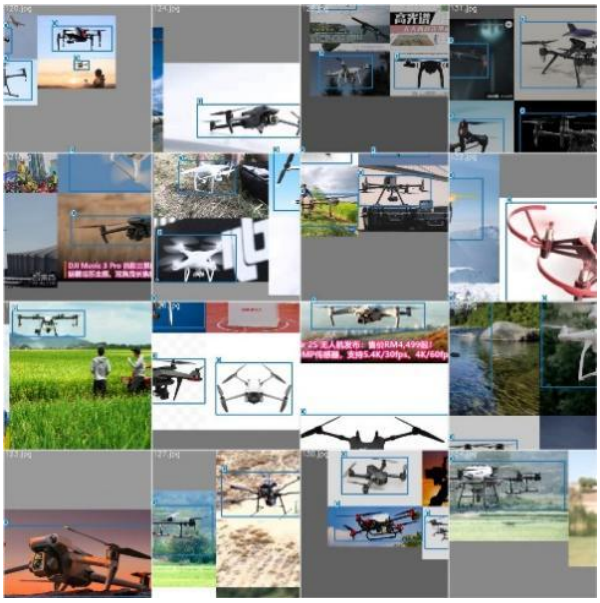

Figure 19 Performance of the Recognition Algorithm

3.3 Tracking Algorithm Training and Analysis

The appearance recognition network of Deep SORT was trained for 300 epochs. The value of loss function gradually decreased over the training period, as depicted in Figure 20. As training progressed, its accuracy steadily improved.

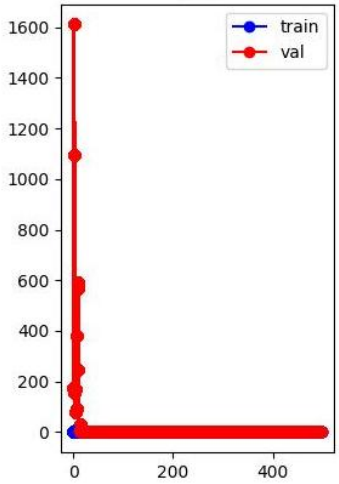

Figure 20 Loss of the Appearance Recognition Network

3.4 Experimental Results

Recognition and tracking were conducted on test videos, successfully identifying UAVs within the video sequences. The model effectively tracked the UAVs visually throughout the video. Figure 21 presents a diagram of the recognition and tracking results.

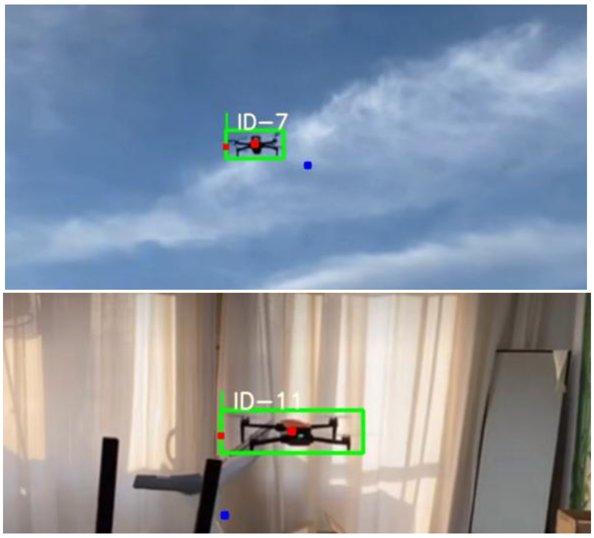

Figure 21 UAV Recognition and Tracking Results

3.5 Summary

A UAV recognition and tracking method based on the combination of the YOLO algorithm and Deep SORT algorithm is proposed. A UAV recognition and tracking model integrating the YOLO algorithm and Deep SORT algorithm is constructed. Then, a dataset specifically for training the drone recognition and tracking algorithms is curated. The constructed model undergoes training, and the training results are analyzed. Subsequently, the trained model is deployed, effectively accomplishing recognition and tracking tasks with good real-time processing performance.

4 CONCLUSION

This study proposes a deep learning-based method for aerial UAVs recognition and tracking. The YOLO algorithm is employed for UAVs detection in the air, while the Deep SORT algorithm is used to track the identified UAVs. The model is constructed, the training dataset is prepared, and the constructed deep learning model is subsequently trained. The trained models are then utilized to validate the effectiveness of UAVs recognition and tracking in aerial scenarios. The trained models successfully accomplish the tasks of recognition and tracking, demonstrating excellent real-time processing capabilities.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

The author received support from Association of Fundamental Computing Education in Chinese Universities for the research. The project is titled "Research on Wildlife Conservation Applications Based on Deep Neural Network Models", with the project number 2024-AFCEC-103.

REFERENCES

[1] Naing KM, Zakeri A, Iliev O. Wireless energy transfer to long distance flying intelligent Unmanned Aerial Vehicles (UAVs) using reactive power transfer techniques. Technol, 2020, 7(9), 2458-9403.

[2] Chittoor PK, Chokkalingam B, Mihet-Popa L. A review on UAV wireless charging: Fundamentals, applications, charging techniques and standards. IEEE access, 2021, 9, 69235-69266.

[3] Sheu BH, Chiu CC, Lu WT, Lien CC, Liu TK, Chen WP. Dual-axis rotary platform with UAV image recognition and tracking. Microelectronics Reliability, 2019, 95, 8-17.

[4] Samadzadegan F, Dadrass Javan F, Ashtari Mahini F, Gholamshahi M. Detection and recognition of drones based on a deep convolutional neural network using visible imagery. Aerospace, 2022, 9(1), 31.

[5] Dadrass Javan F, Samadzadegan F, Gholamshahi M, Ashatari Mahini F. A modified YOLOv4 Deep Learning Network for vision-based UAV recognition. *Drones*, 2022, *6*(7), 160.

[6] Sun H, Wang X, Cao Z, Bai F, Wang X, Hao Y, Wang J. Deep learning-based counter-UAV early warning neural network. Science Technology and Engineering, 2021, 21(22), 9461-9469.

[7] Shi Y, Zhu J, Ling Z. Research on UAV detection algorithm based on feature-enhanced YOLOv4. Journal of Electronic Measurement and Instrumentation, 2022, 36(7), 16-23.

[8] Durve M, Orsini S, Tiribocchi A, Montessori A, Tucny JM, Lauricella M, ... Succi S. Benchmarking YOLOv5 and YOLOv7 models with DeepSORT for droplet tracking applications. The European Physical Journal E, 2023, 46(5), 32.

[9] Xie T, Yao X. Smart logistics warehouse moving-object tracking based on yolov5 and deepsort. Applied Sciences, 2023, 13(17), 9895.

[10] Bai T. Multiple Object Tracking Based on YOLOv5 and Optimized DeepSORT Algorithm. In Journal of Physics: Conference Series. IOP Publishing, 2023, 2547,012022.

[11] Huang Y, Xiao D, Liu J, Tan Z, Liu K, Chen M. An improved pig counting algorithm based on YOLOv5 and DeepSORT model. Sensors, 2023, 23(14), 6309.

[12] Kumar S, Singh SK, Varshney S, Singh S, Kumar P, Kim BG, Ra IH. Fusion of deep sort and Yolov5 for effective vehicle detection and tracking scheme in real-time traffic management sustainable system. Sustainability, 2023, 15(24), 16869.

[13] Razzok M, Badri A, El Mourabit I, Ruichek Y, Sahel A. Pedestrian detection and tracking system based on Deep-SORT, YOLOv5, and new data association metrics. Information, 2023, 14(4), 218.

[14] Ma L, Meng D, Zhao S, An B. Visual localization with a monocular camera for unmanned aerial vehicle based on landmark detection and tracking using YOLOv5 and DeepSORT. International Journal of Advanced Robotic Systems, 2023, 20(3), 17298806231164831.

[15] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, 779-788.