Multi-size drone detection using YOLOv5 network

Vo Thanh Hai[1], Le Van Nhu[1*], Doan Quoc Khanh[1],
Nguyen Phuong Nam[2], Doan Van Sang[3]

[1]Le Quy Don Technical University;
[2]Academy of Military Science and Technology;
[3]Naval Academy.
[*]Corresponding author: levannhuktq@gmail.com
Received 30 March 2022; Revised 05 June 2022; Accepted 10 June 2022; Published 28 June 2022.
DOI: https://doi.org/10.54939/1859-1043.j.mst.80.2022.142-148

ABSTRACT

With increasingly modern technology, advanced and flexible functions, compact design, and low cost, drones are recently used in many fields with different effective purposes. Unlike beneficial applications, hostile forces leverage drones to explore the terrain, carry illegal explosives, and so on. Those applications can seriously threaten national security and defense. To prevent illegal drones effectively, we apply deep neural networks to detect the illegal drones in a variety of conditions and different sizes of drones. Accordingly, a computer-based system using modern cameras combined with an algorithmic model can solve the complex drone detection problem. Therefore, an emerging complex neural network approach based on YOLOv5 is proposed in this paper. With the method, we achieve a very expected result (confidence of 0.993 @0.5IOU), which meets the requirements of the drone detection problem.

Keywords: Drone detection; Computer vision; YOLOv5; Complex neural network; IoU.

1. INTRODUCTION

An unmanned aerial vehicle (UAV or drone) is an aircraft with no pilot in the cockpit. The system consists of a drone, a remote controller on the ground, and a communication system. Drones can operate with different modes of autonomy: under remote control by the operator or automatically by a computer-based automation system. Drones are usually equipped with cameras for filming or taking pictures. Detection of illegal drones can help warn, prevent and track their activities.

Nowadays, the problem of drone detection is studied with many approaches such as radar, acoustic, radiofrequency (RF) sensors, and computer vision (Optics). Each method has its own strengths but also several certain limitations. For instance, a system named ND01, which develop by US experts to detect and intercept drones of small to large sizes are used in border security and protection, important infrastructure protection, prison protection, and anti-terrorism. The system integrates two modules that work independently of each other. The first is a detection module, including Radar, Camera Thermal imager, and PTZ (pan, tilt, and zoom) camera, and the second is an interceptor module (Jammer). Radar, a crucial solution for drone detection, is an electromagnetic (EM) system used to detect the presence and measure the position and distance of an object in the observing range. It works by radiating EM energy into space and tracking the signals reflected from the object. Radar-based detection method has many advantages such as being fast, highly sensitive, and accurate. However, installation and operation are labor-intensive and consume a lot of energy. Besides using the Radar system, RF sensors are also applied to detect drones, but they cannot detect fully autonomous drones [1]. On the other hand, sound waves are also used for object detection, but the acoustic sensor will be less effective for long-range and noisy environments. Finally, optical-based drone detection is often considered the primary and most effective method due to its robustness, accuracy, range, and interpretability. In addition to the visible cameras, thermal or infrared cameras are also used to improve the drone detection performance in dark or low light conditions.

The optic-based approach has an outstanding advantage in computer vision applications using recent emerging deep learning algorithms. For example, Schumann *et al.* [2] propose a model that detects a region covering an object using background subtraction. The detected area is then classified into a drone or distractor, such as a bird, by applying a complex neural network (CNN). Another study also uses CNN to conduct tests with different CNN-based network architectures, such as Zeiler and Fergus (ZF), Visual Geometry Group (VGG16), etc [3]. The test results show that VGG16 with Faster R-CNN performs better than other architectures on the same training dataset. Recent emerging network architecture is YOLO, which stands for You Only Look Once. This network has many versions released such as YOLO [4], YOLOv2 [5], YOLOv3 [6], YOLOv4 [7], and the latest is YOLOv5 [8]. Accordingly, research work in 2017 [9] used YOLOv2 to detect objects belonging to two classes (drones vs birds), and obtained a good performance with precision and recall values of approximately 0.9. However, a limitation of the work is that the ability to distinguish between drones and birds in the same frame is low if they are close to each other. In May 2021, Pham [10] proposed a very good method that is a combination of R-CNN and YOLOv2 for drone detection applications. The results of Pham's study show that combining two models can significantly increase the detection probability because they can alter each other to detect drones. However, the training dataset is limited and does not show the diversity of the objects. Moreover, using one network to replace the other alternately raises computational complexity.

Despite having outstanding achievements in the drone detection problem, the above-mentioned works do not comprehensively consider the influence of the drone size in the image on its detection accuracy. In addition, the YOLOv5 network has been widely used, achieving very high results in object detection. Therefore, we propose to use the YOLOv5 network to train and evaluate the drone detection performance on a diverse dataset of different drone contexts, sizes, and environmental conditions. As a result, our trained YOLOv5 has overcome previous methods' limitations and clarified the drone size's dependency on detection accuracy.

The rest of the paper is arranged as follows: section 2 presents the proposed method and dataset. Then, the test process, results, and analysis of the influence of drone size on detection accuracy are described in section 3. Finally, in section 4, we conclude the research results and define the future works.

2. METHODS AND DATASETS

2.1. YOLOv5-based Drone Detection

For the drone detection problem, we decide to use YOLOv5, which was released on May 18, 2020 [13]. The basic operating principle of YOLO is to threat detection as a regression problem. It divides the image into grid cells, detects objects in each grid, and generates a confident bounding box for each object at classifier output. This technique has high performance and efficiency for real-time video data. YOLOv5 inherits the above technique and is developed to a higher level. Compared with YOLOv4, which is considered to have the most outstanding performance currently, YOLOv5 has the same accuracy but much higher speed.

As shown in figure 1, YOLOv5 consists of three blocks, including backbone (CSPDarknet), Neck (PANet), and Head (Yolo Layer) [14]. Specifically, YOLOv5 incorporated CSPNet into the Darknet, which plays the role of backbone. CSPNet solves the problem of repetitive gradient information in backbones at a large scale and updates gradient changes into feature maps. Thereby, it can reduce parameters and FLOPS. The drone detection process requires a high speed and accuracy; therefore, the lightweight model allows inference efficiency on poor-resource devices. In the Neck block, the Path Aggregation Network (PANet) [11] aims to push information flow in the fragment template. It enhances the use of correctly localized signals in the underlying layers, improving the position accuracy of the object. Finally, the head of

YOLOv5 generates three different sizes {16 × 16, 32 × 32, and 64 × 64} of the object map to achieve multi-scale prediction capability for small, medium, and large objects. The drone is always changing its position, so its size in the frame is always changing. The multi-scale of YOLOv5 detection ensures tracking of the size change of the drone.

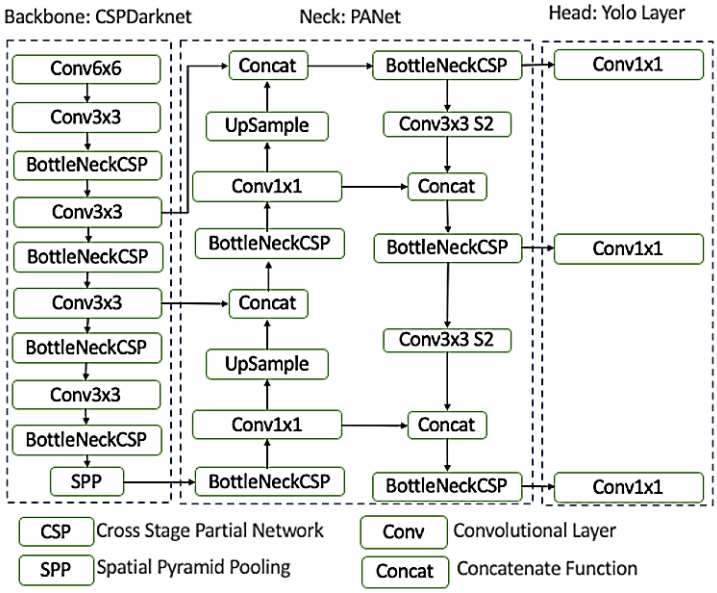
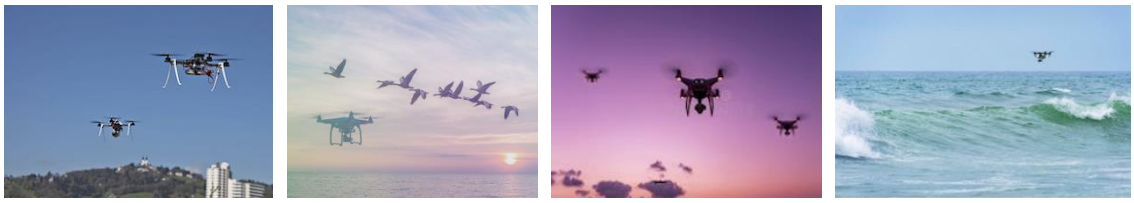


Figure 1. *Yolov5's network architecture [15]. It consists of three blocks: (1) Backbone: CSPDarknet, (2) Neck: PANet, and (3) Head: Yolo Class.*

2.2. Dataset

To ensure the YOLOv5 model can handle different types of drones of various sizes, we have collected images from multiple drone-type datasets, such as from Google, Dreamtime, etc. After manual filtering, we build a composite dataset containing 3600 images of different drone types and sizes.



a) *b)* *c)* *d)*

Figure 2. *Representative drone images in the dataset.*

All images in the dataset have the same size of 512×512, whereas the drones have various sizes from 12×10 to 509×509. In this work, we use 3000 images to train the YOLOv5 model, 200 images for validation, and 400 images for testing. Representative image samples in our dataset are shown in figure 2. In figure 2a, there are two drones appearing in a city where drones can hide behind buildings or other obstructive objects. The second picture (figure 2b) shows the presence of birds, which might cause a mis-detection when they fly near the drone. Figure 2c shows the drone in low light conditions, which limits the visibility of the optical equipment. The last image (figure 2b) is a drone operating above the sea, where the sunlight or strong light intensity can increase the contrast resulting in the subject being blurred, which greatly affects the drone detection. These are just a few representative images that we would like to focus on. Additionally, there are still many other contexts such as drones in schools, hospitals, parks, airports, and so on in our dataset.

3. RESULTS AND DISCUSSION

The YOLOv5, the latest version of YOLO networks, is developed in the Pytorch framework. We conduct training YOLOv5 on the Linux system, CUDA version 10.1, Pytorch version 1.6.0, Python version 3.8, and NVIDIA Tesla K80 by Google Colab.

In the first experiment, the Stochastic Gradient Descent (SGD) and ADAM optimizers are taken into account for training the YOLOv5 network with the same dataset and other training parameters. The training metrics after 300 epochs are demonstrated in figure 3, where we can observe that the SGD optimizer provides higher performance in terms of recall and precision than the ADAM optimizer. Specifically, at the epoch 300, the model trained by the SGD optimizer obtains a recall of 0.96 and a precision of 0.98; meanwhile, the model trained by the ADAM optimizer gets only a recall of 0.77 and a precision of 0.79, which are significantly lower than that of SGD optimizer.

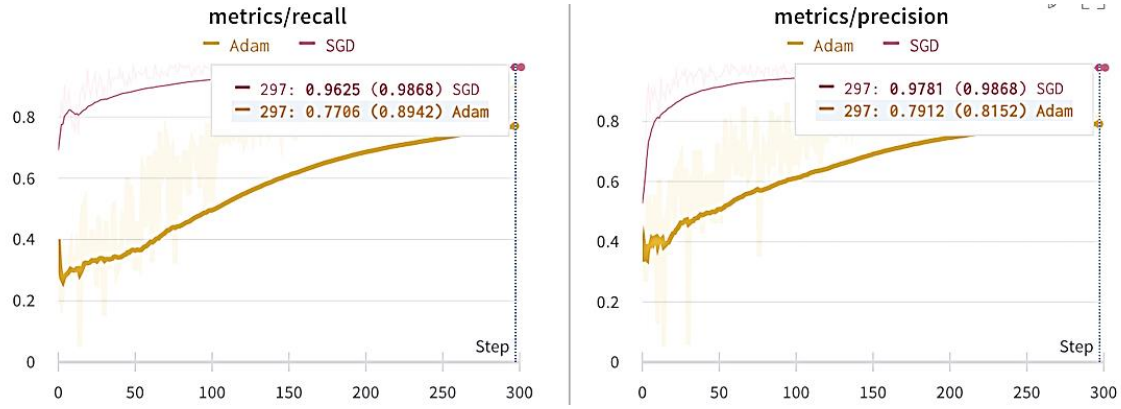


Figure 3. *The comparison results of Stochastic Gradient Descent (SGD) and Adam optimizers.*

Based on the result of the first experiment, the SGD optimizer is selected to train the YOLOv5 for the next experiment. In this stage, the YOLOv5 network is trained in 500 epochs using the SGD optimizer with other training parameters as reported in table 1.

Table 1. *Training parameters of YOLOv5.*

Parameter	Learning rate	Epochs	Batch size	Image size	Optimizer	GPU type	Training time
Value	0.01	500	32	512	SGD	Tesla K80	1d 3h

In this study, precision and recall are used as metrics for evaluating YOLOv5 performance. The precision is calculated as the ratio of the number of positive samples correctly predicted to the number of samples predicted to be positive. The recall is calculated as the ratio of the number of correctly predicted positive samples to the number of actually positive samples. They are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

where, TP is the true positive, which is the number of points of the positive class that are correctly classified as positive; FP is the false positive, which is the number of points of the negative class that are mistakenly classified as positive; FN is the false negative, which is the number of points of the positive class was misclassified as negative. When the drone's prediction

is correct, and the intersection on the conjugate (IoU) [12] measures the intersection of the prediction box with the truth on the ground greater than the threshold (0.5 in our tests), the detection is considered true.

In order to improve the performance of the trained YOLOv5 model, data enhancement techniques, including 90-degree rotation, and grayscale manipulation, are applied to the dataset. After training, the model performance is evaluated using Precision, Recall, and mAP (Mean Accuracy Precision) when the IoU is at 0.5 (50%) and 0.95 (95%) to measure the scale model's inference in real-time. The specific Precision, Recall, and mAP score values are reported in table 2. This result confirms the effectiveness of our approach in accurately predicting drones.

Table 2. *Training results.*

Metrics	**Precision**
Recall	0.982
Precision	0.978
mAP @0.5IOU	0.993
mAP @0.95IOU	0.725

In the next experiment, 400 images of size 512×512, which contain drones of different sizes, are used to test the drone recognition performance relying on the drone size. The results in Table 3 indicate that the larger the drone size is, the more accurate the YOLOv5 model recognizes. Specifically, the drone size of 60×30 allows the model to predict the drone with a precision of 0.85 to 0.9. When the drone size is 90×60, the precision increases greater than 0.9.

Table 3. *The precision with the pixel numbers of the drone.*

Precision	**Drone Size**
0.85 – 0.9	60 × 30
> 0.9	90 × 60

Some representative images, as shown in figure 4, with drone size, detection accuracy, and detection time are presented in table 4. In figure 4, the size of the drones in the standard 512×512 frame decreases, which degrades the detection accuracy. This result demonstrates that the larger the size of the drone, the greater the detection accuracy.

a) *b)* *c)* *d)*

Figure 4. *Representative results.*

Table 4. *Drone size and detection accuracy.*

Figure	Drone size	Accuracy
3a	95×65	0.92
3b	90×55	0.91
3c	50×30	0.90
3d	40×30	0.88

Finally, the YOLOv5 model is taken into account to compete with two other well-known models, including Faster-RCNN [16] and Single Shot Detector (SSD) Mobilenet [17], in the same task of drone detection. In order to ensure a fairly comparison, we conduct training for the three mentioned models with the same condition of training parameters and hardware source. Afterward, all three models are tested with 400 drone images of the test set. The average accuracy and speed of drone detection are considered in this comparison. The comparison results reported in table 5 indicate that YOLOv5 achieves the best accuracy of 0.89 and runs fastest with a detection time of 17.6 ms. The Faster-RCNN model yields the second-best accuracy of 0.82; however, it runs very slow, with a detection time of approximately 1 second. The SDD Mobilenet obtains the worst accuracy and runs faster than Faster-RCNN but still much slower than YOLOv5.

Table 5. *Comparison results.*

Metrics	YOLOv5	Faster-RCNN	SSD Mobilenet
Accuracy	0.89	0.82	0.75
Inference time	17.6 ms	1010 ms	742 ms

In figure 5, we demonstrate some results of drone detection when using the three models of Faster-RCNN, SSD Mobilenet, and YOLOv5. We can observe that the Faster-RCNN model can detect a drone in the image; however, it has false detected birds as drones (figure 5a). Besides false detecting the birds as drones, the SSD Mobilenet model has even ignored the small drone (figure 5b). Due to cleverly designed with the three parts of CSPDarknet Backbone, PANet Neck, and Yolo layer Head, the YOLOv5 model has correctly detected the drone with 87% confidence without false detection of birds (figure 5c).

a) Faster-RCNN. *b) SSD Mobilenet.* *c) Yolov5.*

Figure 5. *Some representative drone detection results using three different methods.*

4. CONCLUSIONS

This paper used the YOLOv5 network to build a drone detection model. A dataset with various drone sizes in a standard 512×512 image frame was built for training. The training results gained 0.993 for mAP@0.5IOU and 0.725 for mAP@0.95IOU. Our model has provided fast object detection and good accuracy. In addition, we have investigated the dimensions of the drone in an image frame with outstanding detection accuracy. In comparison, the YOLOv5 model has remarkably outperformed two other well-known models of Faster-RCNN and SSD Mobilenet in terms of average detection accuracy and execution time. In the orientation of research work, we intend to develop and implement the YOLOv5 network into practical applications in the future.

REFERENCES

[1]. Allahham, Mhd Saria et al. "*Deep Learning for RF-Based Drone Detection and Identification: A Multi-Channel 1-D Convolutional Neural Networks Approach*." 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), pp. 112-117, (2020).

[2]. A. Schumann, L. Sommer, J. Klatte, T. Schuchert, and J. Beyerer, "*Deep cross-domain flying object classification for robust UAV detection*," 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1-6, DOI: 10.1109/AVSS.2017.8078558. (2017).
[3]. Saqib, Muhammad et al. "*A study on detecting drones using deep convolutional neural networks*." 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1-5, (2017).
[4]. Chaudhari, Sujata et al. "*Yolo Real Time Object Detection*." International Journal of Computer Trends and Technology 68, pp. 70-76, (2020).
[5]. Huang, Xin et al. "*PP-YOLOv2: A Practical Object Detector*." ArXiv abs/2104.10419, (2021).
[6]. Redmon, Joseph, Ali Farhadi. "*YOLOv3: An Incremental Improvemen*t." ArXiv abs/1804.02767, (2018).
[7]. Bochkovskiy, Alexey et al. "*YOLOv4: Optimal Speed and Accuracy of Object Detection*." ArXiv abs/2004.10934, (2020).
[8]. Chen, Yuwen et al. "*Ship detection in optical sensing images based on YOLOv5*." International Conference on Graphic and Image Processing, (2021).
[9]. Aker, Cemal and Sinan Kalkan. "*Using deep networks for drone detection"*. 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1-6, (2017).
[10]. Viet, Pham Van. "*A Combination Of Faster R-Cnn And Yolov2 for Drone Detection in Images*." TNU Journal of Science and Technology, (2021).
[11]. Wang, Chien-Yao et al. "*CSPNet: A New Backbone that can Enhance Learning Capability of CNN*." 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1571-1580, (2020).
[12]. Gao, Yan et al. "*Decoupled IoU Regression for Object Detection*." Proceedings of the 29th ACM International Conference on Multimedia, (2021).
[13]. Jocher, Glenn R. et al. "*Ultralytics/yolov5: v3.0*.", (2020).
[14]. *"A Forest Fire Detection System Based on Ensemble Learning"* - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/The-network-architecture-of-Yolov5-It-consists-of-three-parts-1-Backbone-CSPDarknet_fig1_349299852.
[15]. Xu, Renjie & Lin, Haifeng & Lu, Kangjie & Cao, Lin & Liu, Yunfei. *"A Forest Fire Detection System Based on Ensemble Learning. Forests",* Electronics, Close-Range Sensors and Artificial Intelligence in Forestry, (2021).
[16]. X. Farhodov, O. -H. Kwon, K. W. Kang, S. -H. Lee and K. -R. Kwon, "*Faster RCNN Detection Based OpenCV CSRT Tracker Using Drone Data*", International Conference on Information Science and Communications Technologies (ICISCT), pp. 1-3, (2019).
[17]. W. Budiharto, A. A. S. Gunawan, J. S. Suroso, A. Chowanda, A. Patrik and G. Utama, "*Fast Object Detection for Quadcopter Drone Using Deep Learning*", 3rd International Conference on Computer and Communication Systems (ICCCS), pp. 192-195, (2018).

TÓM TẮT

Phát hiện drone nhiều kích thước sử dụng mạng YOLOv5

Ngày nay, máy bay không người lái được sử dụng rộng rãi với nhiều mục đích khác nhau. Với công nghệ ngày càng hiện đại, được trang bị nhiều chức năng cao cấp, linh hoạt với thiết kế nhỏ gọn mà giá thành lại không quá đắt. Drone được sử dụng trong nhiều lĩnh vực với nhiều mục đích khác nhau, đặc biệt là trong lĩnh vực quân sự, các thế lực thù địch sử dụng nó để thăm dò địa hình, mang vật liệu nổ trái phép, có thể đe dọa đến an ninh. Thị giác máy tính có thể được áp dụng để phát hiện một cách hiệu quả máy bay không người lái bất hợp pháp trong nhiều điều kiện khác nhau và các kích thước đa dạng của máy bay không người lái. Một hệ thống dựa trên máy tính sử dụng camera hiện đại kết hợp với một mô hình thuật toán có thể giải quyết tốt bài toán phức tạp trong phát hiện máy bay không người lái. Bài báo này đề xuất một phương pháp tiếp cận mạng nơ-ron phức tạp mới nổi đó là Yolov5. Với phương pháp này, chúng tôi đã được một kết quả hết sức mong đợi (0,993 cho @0,5IOU), đáp ứng được yêu cầu trong bài toán phát hiện đối tượng.

Từ khoá: Phát hiện máy bay không người lái; Thị giác máy tính; Yolov5; Mạng nơron phức tạp; IoU.