# A compact solution for ultra-light drone optical auto-detection and distance estimation using AI

Nguyen Ngoc Xuyen[1], Phan Huy Anh[2], Nguyen Le Cuong[1*]

[1]Electric Power University, Hanoi, Vietnam.
[2]Institute of Electronics, Academy of Military Science and Technology, Hanoi, Vietnam;
[*]Corresponding author: cuongnl@epu.edu.vn

*Received 26 Jul 2022; Revised 15 Sep 2022; Accepted 07 Nov 2022; Published 18 Nov 2022.*
**DOI:** https://doi.org/10.54939/1859-1043.j.mst.83.2022.11-21

## ABSTRACT

*This paper proposes a system for ultra-light drone (ULD) auto–detection using only one non-static optical PTZ camera. The system includes multi-stages of suspect objects detection, clarification, and distance estimation. An AI model for detection and clarification stages is designed based on the YOLOv3 architecture and trained with a practical dataset. In the detection stage, the camera continuously pans, tilts, and zooms to take panoramic images of the detection zone and pass them to the AI model. Once the AI model detects a suspect object, it will switch to the verification stage. In this stage, the camera controlled by the AI model's output focuses on the target to clarify and estimate the distance to ULD. The proposed solution was implemented and tested with popular fly cams. The results show that the system can auto-detect ultra-light drones effectively with high accuracy.*

**Keywords:** Ultra-Light Drones; Black Dot; YOLOv3 Model; Drone detection; Verification.

## 1. INTRODUCTION

The application of ultra-light drones (ULD) [5] has rapidly become popular in the last few years. This type of vehicle is low cost, easy to assemble, and simple to use. Besides providing many valuable utilities for users, ULD also has many negatives. The uncontrolled use of ULDs may bring potential threats of using drones for terrorist attacks and other illegal purposes. So that, solutions for detecting a ULD currently attract great interest. There are many proposed methods of ULD detection and distance estimation, such as radar, lidar, passive RF signal detection; acoustic signal detection; thermal and optical image detection. The above methods all have their own advantages and limitations. The way of using active radar may be limited or confusing due to ULD's small reflective size and echoes from undesired targets [2-5]; passive RF signal detection cannot detect ULDs flying in automatic mode, without communication to the ground control station [3, 5]; acoustic detection or lidar is not effective with small, low flight speed aircraft [1-3]; Thermal image is costly and very close detection distance [2]; the method of using optical images has acceptable detection range and can detect ULD with high accuracy, but it can only be used in suitable light conditions [2, 3].

In recent years, AI in general, and image processing, in particular, have experienced explosive development. The state-of-the-art image processing models are mainly divided into two types: one-stage and two-stage [12]. Some typical one-stage models can be mentioned as You Only Look Once (YOLO), Single Shot Display (SSD), and some typical two-stage models can be listed as Fast Region-based Convolution Neural Networks (Fast R-CNN), Faster R-CNN, Mask R-CNN. The above image processing models are trained based on deep learning (DL) and use Convolution Neural Networks (CNN) for object detection [10-12]. Some models have been applied in drone detection applications, and their performance greatly supports the detection of drones from visible data such as optical images, and thermal images. Studies in [1-4, 6-9, 12] indicated that in drone detection applications, the YOLO model is widely used thanks to its balance between accuracy and speed. The ULD detection systems using the optical image and AI mentioned in [1-4, 6-9] can detect ULD with high accuracy, but there still exists some issues limiting efficiency, such as short range [1, 2, 9]; high quality image requirement [1, 2, 7-9]; inaccurate distance measurement [1]; restricted field of surveillance or complicated system [3, 4]; not real-time detection [7-9].

In order to reduce the system complication as well as improve the efficiency of detection and the precision of distance estimation using the optical images, in this paper, the authors propose a solution that uses only one non-static PTZ optical camera with a YOLO3-based AI model. The algorithm includes multi-stages of suspect objects detection, clarification, and distance estimation. The AI model for detection and clarification stages is designed based on the YOLOv3 architecture and trained with a practical dataset. In the detection stage, the camera continuously pans, tilts, and zooms to provide panoramic images of the zone of interest to the AI model. It is also controlled by the AI model's output to verify suspect objects. Once the AI model detects a suspect object, it will switch to the verification stage. In this stage, the camera focuses on the target to clarify and measure the distance. The proposed solution was implemented and tested with popular fly cams. The results show that the system can detect ultra-light drones effectively with high accuracy.

The above solution is researched and developed based on the theory of optics, image processing, and camera controlling techniques. The rest of the paper is organized as follows: Section 2 is about the methodology; Section 3 shows the experimental setup; Section 4 illustrates results and section 5 concludes the paper.

## 2. METHODOLOGY

### 2.1. System architecture

Figure 1 below shows the architecture of the system to deploy the proposed solution. In the figure, there are three big blocks which present for hardware devices and small blocks present for processing blocks.

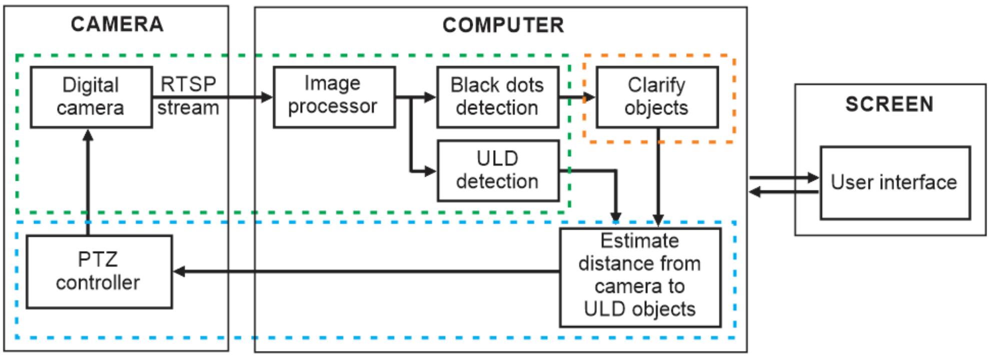

***Figure 1.*** *ULD detection system architecture.*

The system's hardware consists of a pan-tilt-zoom camera, a desktop computer, and a desktop screen. The camera has 2 Megapixel sensor, 48 times optical zooming lens, a pan angle in the range of $0^o$ to $360^o$, a tilt angle from $-90^o$ to $45^o$, and angle controlling accuracy up to $0.1^o$/second. The desktop computer has an Nvidia GTX 2080Ti graphic card, AMD Ryzen 9 CPU, and 16GB of RAM. The Ubuntu 18.04 LTS operation system, OpenCV 4.2.0, Cuda Toolkit 10.2, and CuDNN 7.6.5 library are installed for the application of image processing to detect ULD. The camera is connected to the computer via a Giga-Ethernet link and transmits data via RTSP stream protocol.

### 2.2. Algorithm

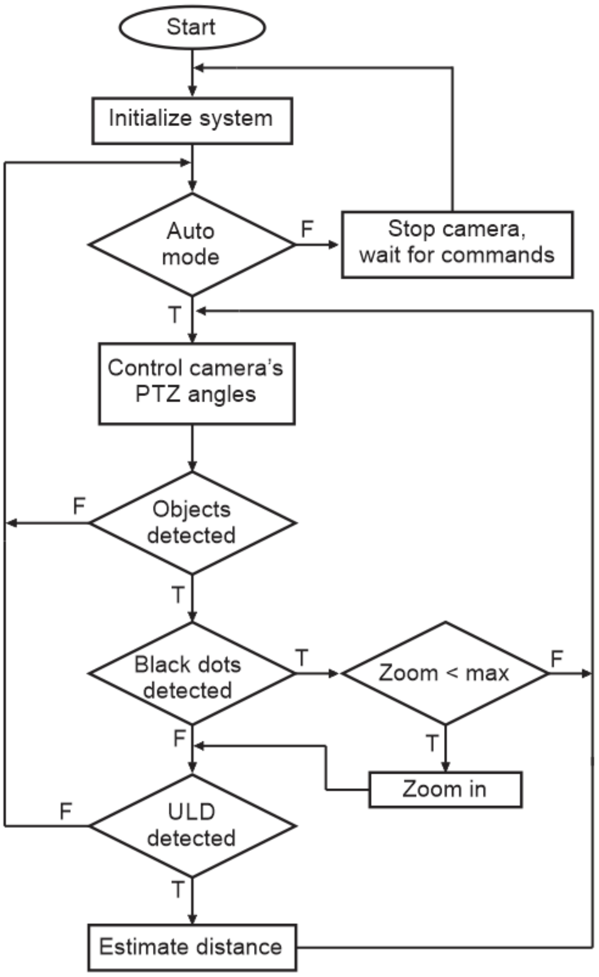

***Figure 2.*** *Software algorithm in detail.*

The ULD detection and estimation system works in a 3-stage process as follows:

- Surveillance stage (the green dash line in figure 1): the PTZ camera turns continuously to scan and look for trained objects. If the detected object is a black dot, go on to stage 2. If the detected object is ULD, skip stage 2, go on to stage 3.
- Verification stage (the orange dash line in figure 1): the PTZ camera zooms and focuses on black dots to verify whether they are ULD or not.
- Distance estimation stage (target locked – the blue dash line in figure 1): the system estimates the distance to ULD and controls the PTZ camera to track the highest confidence object.

The flowchart in figure 2 describes how the software algorithm works in detail.

Upon starting, the system is initialized by 3 parameters: monitoring ground distance, monitoring height, and working mode. When operating, the camera is controlled according to the installed parameters to capture images in the being monitored area. The image processing model detects both ULD and black dots in parallel. At a long distance, out of the effective range of the camera, the ULD may just be a black dot, and this makes the AI model may not detect ULD correctly. Thus, all black dots are labeled as suspected to be ULD objects, the camera will zoom in one by one in order of bigger to smaller bounding box to confirm whether it is ULD or not. When a ULD object is detected, the system will estimate the distance from the camera to the ULD. In case of many ULD objects appear at the same time, the system has the ability to estimate the distance to all of them. Detecting black dots and then clarifying them can help the detection system not miss objects, thereby increasing the system's performance and object detection distance.

### 2.3. Object detection with YOLOv3

YOLO is a one-stage image processing model based on a single CNN, it can predict multiple bounding boxes in a single frame at the same time and calculate probabilities for those boxes [6- 8]. It is extremely faster than two-stage image processing models such as Mask R-CNN, Fast R-CNN, Faster R-CNN because this model skips the stage of determining region proposals, the input image is taken to CNN directly for processing [10-12]. Many versions of YOLO have been launched with improvements in the data processing layers inside the model, processing rate, and accuracy. Among version 1, version 2 and version 3 by Redmon, YOLO version 3 has the highest accuracy, especially with small objects [12]. The architecture of YOLOv3 is shown in figure 3.

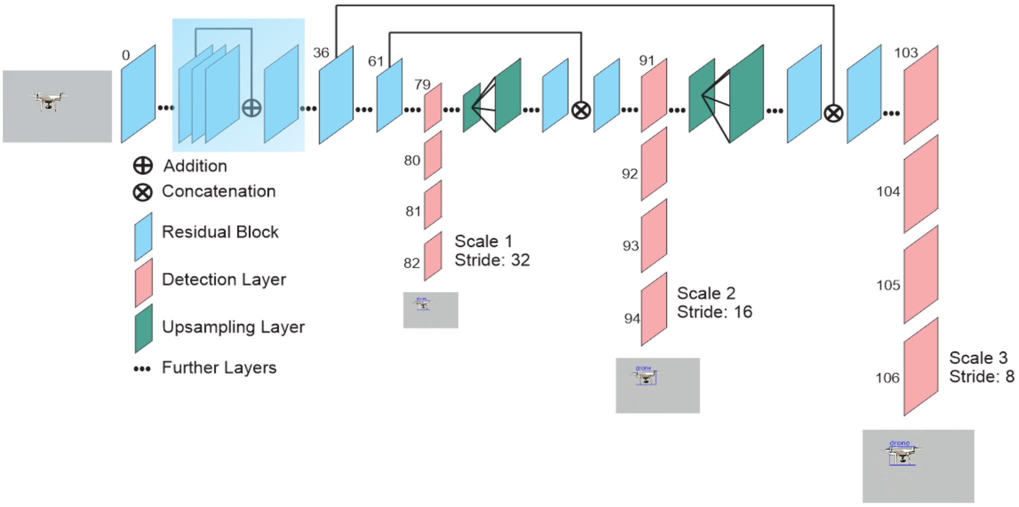

***Figure 3.*** *YOLOv3 network architecture [11, 12].*

YOLOv3 model divides the input images into $S \times S$ square grid cells. Each grid cell predicts the position information of $B$ bounding boxes, and calculates the probability of each learning object, which the bounding box is corresponding to [11, 12]. The weight of YOLOv3 has a total of 106 processing layers [12]. YOLOv3 uses an optimized sum-square error loss function for bounding boxes prediction and binary cross-entropy loss function for class prediction [10, 11]. This model predicts boxes at 3 different scales, with strides of 32, 16, 8 [11, 12]. It means that the resized input images are divided by 32, 16, and 8. The final output of YOLOv3 is a 3D-tensor that contains the coordinates, width, height and object's score of each bounding box in the processed image [11]. Due to the highest accuracy, acceptable processing speed and ability to process large input images, the YOLOv3 is suitable for ULD detection applications.

### 2.4. Distance estimation

A camera lens is made up of one or more converging lenses placed in series. The image obtained from the camera is a real two-dimensional (2D) image. The distance from the camera to the objects in the image can be computed based on the camera's optical parameters. Figure 4 shows how an object's image is created in the camera's sensor.

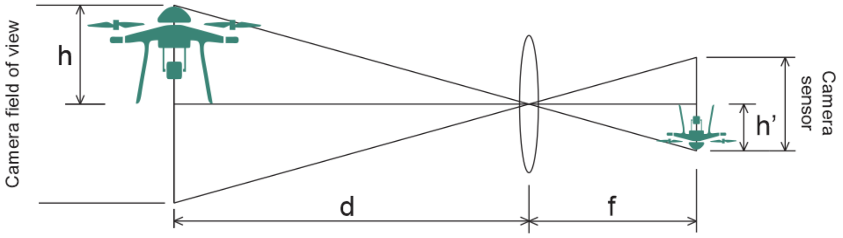

***Figure 4.*** *Distance estimation using optical parameters.*

Distance to the object $d$ can be calculated by the following formula:

$$\mathit{Distance\ to\ object}\ (m) = \frac{\mathit{Real\ object\ size}\ (m) \times \mathit{Focal\ length}\ (mm)}{\mathit{Object's\ size\ on\ sensor}\ (mm)} \tag{1}$$

$$d(m) = \frac{h\ (m) \times f\ (mm)}{h'\ (mm)}$$

whereby:

$f'$: The camera's $\mathit{Focal\ length}$ taken from its specification;

$h'$: The $\mathit{Object's\ size\ on\ sensor}$ can be calculated via the object's size on the image. In this paper, the object's size on an image is the width of YOLOv3's output bounding boxes, which is the number of pixels of the ULD in the image;

$h$: The $\mathit{Real\ object\ size}$ that was taken from the ULD library after clarification.

## 3. EXPERIMENTAL SETUP

### 3.1. Dataset

In this paper, we create our own practical dataset. The dataset includes 53736 images of 2 common types of ULD: DJI Phantom 4 and DJI Mavic 2. Figure 5 shows example images (cropped) of the dataset.

The images' size is 1280 x 720 pixels, all captured by the PTZ camera in many different conditions of background, light, fog, distance to ULD, and camera's focal length. The dataset image quality is at various levels, from very small, and blurred to clear images of ULD. The clear objects are labeled as ***drone***, and the objects which are not clear enough are labeled as ***dot***, all in YOLO format. The dataset includes 10% of background images without objects, 50% of ULD images, and 40% of black dot images.

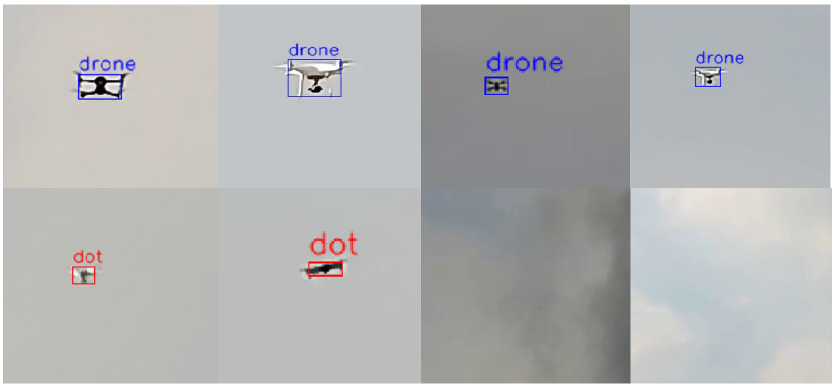

***Figure 5.*** *Dataset example images.*

### 3.2. Training model

When being trained, this dataset is split into two parts with a ratio: 90% is used for training and 10% are used for validation. The YOLOv3 model is trained with the Darknet-53 backbone. The training configurations are set following the Darknet's recommendation for custom object detection.

The best weight file is gotten at the step of 42000. The trained YOLOv3 model on our custom dataset achieved 95.68% of mAP@0.50 (92.46% for black dot and 98.90% for drones), 0.93 precision (thresh = 0.25), 0.96 of recall, 69.00% of IoU, loss value is approximately 0.05 and image processing rate achieved 21.3 fps on the computer mentioned above.

### 3.3. Field trial

The authors tested the detection system in a vacant land area that has straight line vision over 500 meters to evaluate the effectiveness of the ULD detection method using the optical camera and image processing techniques. The layout of the camera in the monitoring area is illustrated in figure 6 and the actual ULD detection system is illustrated in figure 7 below.

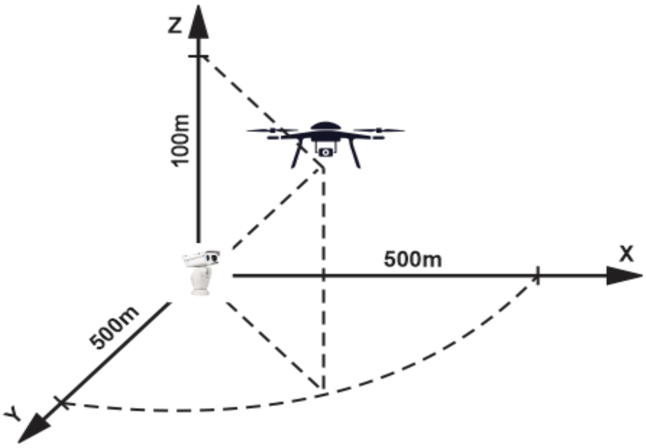

***Figure 6.*** *Camera layout in the monitoring area.*

During the test, the camera's pan angle is limited to the range of $0^o$ to $90^o$; the image resolution is 1280 x 720 pixels; the image rate is 20 fps; the camera's zoom level and tilt angle are tested in real conditions to find the optimal parameters for each distance and altitude.

***Figure 7.*** *Actual ULD detection system.*

The ULDs which are used for testing in this paper are DJI Phantom 4 and DJI Mavic 2. The width dimension (without propellers) of Phantom 4 and Mavic 2 is 350 millimeters and 275 millimeters respectively, their maximum cruise speed is 14 m/s in ideal conditions. The tested altitude is 50 meters and 100 meters. Table 1 below shows the camera's configurations.

***Table 1.*** *System's setting parameters.*

| Monitoring altitude (meters) | Monitoring ground distance (meters) | Tilt angle | Zoom |
|---|---|---|---|
| 50 meters | 100 | $58.96^O$ | 10x |
| | 200 | $16.84^O$ | 15x |
| | 300 | $10.51^O$ | 25x |
| | 400 | $8.00^O$ | 30x |
| | 500 | $6.44^O$ | 35x |
| 100 meters | 100 | $47.65^O$ | 10x |
| | 200 | $29.21^O$ | 15x |
| | 300 | $19.49^O$ | 25x |
| | 400 | $14.91^O$ | 30x |
| | 500 | $12.03^O$ | 35x |

## 4. RESULTS

### 4.1. Detection result

The authors performed 100 detection tests for each pair of altitude/ground distance parameters. The detection performance is evaluated by 2 parameters: $Detection\ precision\ (DP)$ and $Average\ distance\ estimation\ error\ (ADEE)$. The $DP$ is the percentage ratio of ULD true detection times and total tested times. The $ADEE$ is the relative distance error between estimating by the PTZ camera and measuring by GPS. They are calculated as the following formulas:

$$DP = \frac{T_{true}}{T_{total}} \times 100\% \tag{2}$$

$$ADEE = \frac{|D_{GPS} - D_{Camera}|}{D_{GPS}} \times 100\% \tag{3}$$

whereby: $T_{true}$ is true detection times

$T_{total}$ is the total tested times

$D_{GPS}$ is the distance to ULD measured by GPS

$D_{Camera}$ is the distance to ULD estimated by the camera.

Figures 8, 9, 10 below illustrate the detection results of the system.

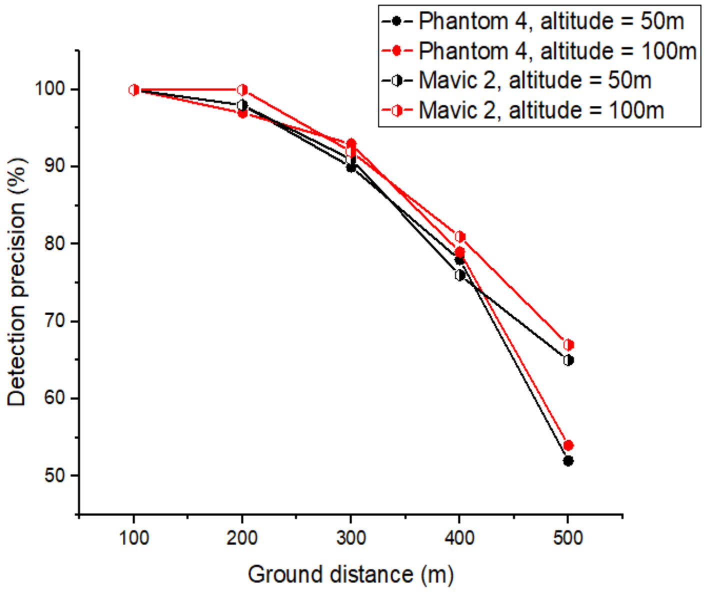

***Figure 8.*** *Detection results of Phantom 4 and Mavic2 at the altitude of 50 meters and 100 meters.*

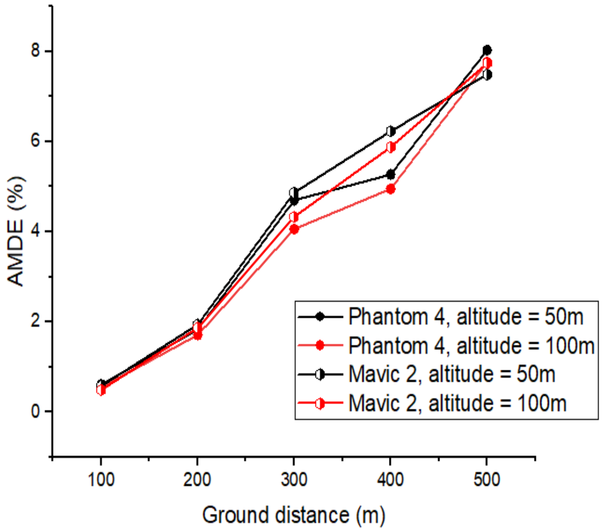

***Figure 9.*** *Average measurement distance error of Phantom 4 and Mavic2 at the altitude of 50 meters and 100 meters.*

***Figure 10.*** *System found a dot (left) and then verify it is a drone (right).*

### 4.2. Discussion

Tested results in figures 8, 9, 10 show that the detection system can detect and clarify ULD objects effectively at a ground distance up to 500 meters, and altitude up to 100 m. The system can detect 100% of drones appearing in the monitoring area at a distance of 100 meters, an average of 98% of drones at a distance of 200 meters, 92% at 300 meters, 79% at 400 meters, and 60% at 500 meters. At a further distance, the $DP$ decrease for all two kinds of drones. The $DP$ of DJI Phantom 4 decreases more quickly than Mavic 2 because of its white color, this can make Phantom 4 easily mix in the white clouds and become very hard to detect. Similar to Phantom 4, the Mavic 2 drones also can be mixed in dark clouds, but it is still easier to be detected due to the black dot detection algorithm. Through the test, the authors recognize that both DJI Mavic 2 and DJI Phantom 4 have detection precision at the altitude of 100 meters is higher than detection precision at the altitude of 50 meters, because at a higher altitude, 4 arms of them are more clear to detect.

Going along with detection precision, the $ADEE$ is also higher at a further distance. At 100 meters, the average $ADEE$ of both kinds of drones is less than 1% (0.56%), and it raises more quickly at a further distance. The average $ADEE$ at 200 meters is 1.84%, at 300 meters is 4.49%, at 400 meters is 5.58% and at 500 meters is 7.76%. Similar to the $DP$, the $ADEE$ of Mavic 2 is better than Phantom 4, and the results at the altitude of 100 meters are better than at the altitude of 50 meters.

The detection system in this paper also has several defects. The first defect is the weakness of the camera when capturing images in bad light conditions. Both too bright light and too dark light make the camera not work effectively although the camera has infrared light to support capturing in night conditions. The second defect is that during operation, the stage of compressing image data into RTSP stream causes a delay, which makes the image being processed slower than the image captured by the camera. Since distance estimation uses YOLOv3 output, a delay may lead to the miscalculation of distance estimation due to the camera's focal length and object size in the image is not time-synchronized. The authors perform a test to measure the delay between the original image (uncompressed into RTSP stream) and the image after being processed by YOLOv3, the result shows that the total delay of image compression and image processing is 0.5 seconds. Another critical factor affecting the system's performance is the camera's vibration, and focusing speed while capturing at high zoom level. This can make lose object's traces due to the camera does not capture images timely, or object's image is not clear enough, as a result, the detection system ignores objects.

## 5. CONCLUSIONS

This paper proposes a system for ultra-light drones (ULD) auto-detection and distance estimation using only one non-static optical PTZ camera. The YOLOv3 model, which is trained with Darknet-53 backbone and custom dataset, achieves 95.68% of mAP@0.50 (92.46% for black dot and 98.90% for drones), 0.93 of precision (thresh = 0.25), 0.96 of recall, and 69.00% of IoU. The tested result shows that the detection system can detect and clarify ULD objects effectively at the ground distance up to 500 m, altitude up to 100m, average detection precision achieves 100% at a distance of 100 m, 98% at a distance of 200 m, and decrease down to 60% at 500 m. The average AMEE achieves 0.56% at 100 meters, 1.84% at 200 meters, and raise to 7.76% at 500 m. The detection precision and the AMEE of DJI Mavic 2 are better than DJI Phantom 4, and the result at the altitude of 100 meters is better than the results at 50 meters. Detecting black dots in an image and then clarifying whether it is ULD or not helps the system increase detection distance and efficiency of ULD detection. To improve the efficiency of object detection and distance estimation, it is possible to upgrade the computer hardware, camera to reduce vibration, image transmitting delay, however, this can increase the cost of hardware.

## REFERENCES

[1]. Y. C. Lai, Z. Y. Huang, *"Detection of a Moving UAV Based on Deep Learning-Based Distance Estimation,"* Remote Sens. (2020). https://doi.org/10.3390/rs12183035.

[2]. F. Svanström, C. Englund and F. Alonso-Fernandez, *"Real-Time Drone Detection and Tracking with Visible, Thermal and Acoustic Sensors,"* 2020 25th International Conference on Pattern Recognition (ICPR), pp. 7265-7272, (2021), doi: 10.1109/ICPR48806.2021.9413241.

[3]. E. Unlu, E. Zenou, N. Riviere, P. E. Dupouy, *"Deep learning-based strategies for the detection and tracking of drones using several cameras,"* IPSJ T Comput Vis Appl 11, 7 (2019). https://doi.org/10.1186/s41074-019-0059-x.

[4]. Igor S. Golyak, Dmitriy R. Anfimov, Iliya S. Golyak, Andrey N. Morozov, Anastasiya S. Tabalina, and Igor L. Fufurin, *"Methods for real-time optical location and tracking of*

*unmanned aerial vehicles using digital neural networks,"* Proc. SPIE 11394, Automatic Target Recognition XXX, 113941B (2020); doi: 10.1117/12.2573209.

[5]. N. H. Hoang, N. L. Cuong, T. V. Kien, *"Measuring the arrival time of signal to determine coordinates of ultra-light drone,"* Journal of Military Science and Technology, FEE (2020), (in Vietnamese).

[6]. Seidaliyeva, Ulzhalgas & Akhmetov, Daryn & Ilipbayeva, Lyazzat & Matson, Eric., *"Real-Time and Accurate Drone Detection in a Video with a Static Background,"* Sensors. 20. 3856. 10.3390/s20143856, (2020).

[7]. Y. Hu, X. Wu, G. Zheng and X. Liu, *"Object Detection of UAV for Anti-UAV Based on Improved YOLO v3,"* 2019 Chinese Control Conference (CCC), 2019, pp. 8386-8390, (2019). doi: 10.23919/ChiCC.2019.8865525.

[8]. D. K. Behera and A. Bazil Raj, *"Drone Detection and Classification using Deep Learning,"* 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1012-1016, (2020). doi: 10.1109/ICICCS48265.2020.9121150.

[9]. Hassan, Syed & Rahim, Tariq & Shin, Soo., *"Real-time UAV Detection based on Deep Learning Network,"* 630-632. 10.1109/ICTC46691.2019.8939564, (2019).

[10]. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, *"You Only Look Once: Unified, Real-Time Object Detection,"* arXiv:1506.02640v5 [cs.CV], (2016).

[11]. Joseph Redmon, Ali Farhadi, *"YOLOv3: An Incremental Improvement,"* arXiv: 1804.02767v1 [cs.CV], (2018).

[12]. S.V. Viraktamath, M. Yavagal, R. Byahatti, *"Object Detection and Classification using YOLOv3"*, International Journal of Engineering Research & Technology, **Vol. 10**, Issue 02, (2021).

## TÓM TẮT

### Giải pháp tinh gọn để tự động phát hiện và ước lượng khoảng cách đến máy bay không người lái siêu nhẹ sử dụng ảnh quang học và trí tuệ nhân tạo

*Bài báo này đề xuất một giải pháp tự động phát hiện máy bay không người lái siêu nhẹ (flycam) sử dụng duy nhất một camera PTZ động. Hệ thống phát hiện flycam theo một quy trình ba bước: phát hiện chấm đen, làm rõ chấm đen có phải flycam không và ước lượng khoảng cách đến flycam. Việc phát hiện và làm rõ chấm đen được thực hiện bởi một mô hình trí tuệ nhân tạo dựa trên kiến trúc của YOLOv3, được huấn luyện với tập dữ liệu về flycam do nhóm tác giả xây dựng. Ở bước phát hiện chấm đen, camera PTZ liên tục quay và chụp lại hình ảnh của khu vực cần giám sát rồi chuyển hình ảnh tới mô hình trí tuệ nhân tạo để xử lý. Khi phát hiện có chấm đen, hệ thống sẽ thực hiện làm rõ chấm đen đó có phải flycam hay không, nếu đúng, hệ thống sẽ bám theo đối tượng, đồng thời ước lượng khoảng cách đến đối tượng. Giải pháp trên được nghiên cứu và thử nghiệm với các loại flycam thông dụng. Kết quả cho thấy, hệ thống có thể tự động phát hiện flycam với độ chính xác cao ở khoảng cách lên đến 500 mét.*

**Từ khoá:** Máy bay không người lái siêu nhẹ; Phát hiện chấm đen; Phát hiện flycam; Mô hình YOLOv3; Làm rõ đối tượng.