I.J. Image, Graphics and Signal Processing, **2024, 2, 83-95**
Published Online on April 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2024.02.07

Drone Detection from Video Streams Using Image Processing Techniques and YOLOv7

Muhammad K. Kabir
Department of Computer Science and Engineering, Brac University, Dhaka 1212, Bangladesh
E-mail: muhammad.khubayeeb.kabir@g.bracu.ac.bd
ORCID iD: https://orcid.org/0000-0001-6868-8882

Anika N. Binte Kabir
Department of Computer Science and Engineering, Brac University, Dhaka 1212, Bangladesh
E-mail: anika.nahian.binte.kabir@g.bracu.ac.bd
ORCID iD: https://orcid.org/0000-0002-8712-5046

Jahid H. Rony
Department of Computer Science and Engineering, Dhaka University of Engineering and Technology, Gazipur, Bangladesh
E-mail: 174067@student.duet.ac.bd
ORCID iD: https://orcid.org/0000-0002-6492-745X

Jia Uddin*
Artificial Intelligence and Big Data Department, Woosong University, Daejeon, South Korea
E-mail: jia.uddin@wsu.ac.kr
ORCID iD: https://orcid.org/0000-0002-3403-4095
*Corresponding author

Received: 14 April, 2023; Revised: 06 July, 2023; Accepted: 23 October, 2023; Published: 08 April, 2024

Abstract: For ensuring the safety issues, a country should establish a secure monitoring system around the most important places. Due to the huge development in unmanned aerial vehicles (UAV), drone detection is a vital part of the safety monitoring system for reducing threats from neighboring countries or terrorist groups. This paper presents a deep learning-based drone detection method. A You Only Look Once (YOLO) v7 architecture is used to train on the dataset. The training dataset consists of drone images in various environments. The trained model was tested on multiple videos of drones from YouTube. Experimental results demonstrate that the model exhibited a recall of 0.9656 and a precision of 0.9509. In addition, the performance of the model compares with the state-of-art models with YOLOv8, YOLO-NAS, Faster-RCNN architectures and it outperforms the other models by maintaining a more stable precision and recall curve.

Index Terms: Unmanned aerial vehicles (UAV), automatic drone detection, image annotation, You Only Look Once (YOLO)

1. Introduction

A drone, also known as an unmanned aerial vehicle (UAV) refers to a flying vehicle that can be controlled remotely. Depending on the size, different kinds of sensors can be equipped with it to provide various services and features. Depending on its robustness and multifunctionality demands for drones are sky-rising for various purposes, such as mapping, search and rescue, surveillance, film production as well military defense works [1]. Drones can reach places that would be difficult for humans as a result it is one the most reliable and widely used in mapping, surveying, security surveillance, and videography [2-5]. To cope with the market demand, rapid advancement in drone technology is going on, and the growth and advancement are expected to continue for quite a long term. However, along with the positive side, drone technology can be seriously harmful to mankind if it falls in the wrong hands. Drones were reported to be used for illegal activity such as drug smuggling, illegal surveillance of sensitive areas, conducting espionage, or using them as weapons by and against terrorists [6]. Such activities cause an alarming situation for national security and public safety issues [7]. In recent days, the use of drones for illegal surveillance and spying on important people is at an alarming rate [8, 9]. It is crucial for people, governments, and companies to think about the possible privacy implications of drone technology as it develops and to take actions to lessen any potential harm.

A few years back, a drone attack at London's Heathrow Airport affected numerous flights and loss of important work hours [10]. This incident has alarmed us about the possibility and effect of a drone attack. Before that, drones have been tried to use for assassination [11]. The report says that during a speech by Venezuelan President Nicolas Maduro, several drones carrying explosives flew toward the stage where he was speaking. The explosives detonated mid-air, causing panic among the crowd and leading to the speech being abruptly cut short [12]. Later, an attack on Saudi Arabia's Abqaiq oil facility, allegedly carried out by drones led to a 5% drop in global oil production and caused geopolitical tensions around the world [13]. Even in live sports events drones have raised issues for security. A man in the UK was fined for flying a drone over Trent Bridge cricket ground during a match, breaching air navigation laws, and flying close to a police helicopter [14]. The incident raised concerns about the potential risks posed by drones. Therefore, the development of drone detection technologies is crucial to ensure the safety and security of individuals and communities, and our society.

But the fact is that it is also important to use drones for ensuring the proper utilization of technological advancement in our life. From that perspective, drone detection and a classification system are introduced. In this process, first of all, a drone is detected using various methods such as radar, audio, or video processing. Then it can be classified and find out whether it is authorized or not. The process is represented in Fig. 1. Here in this study, we are focusing on the detection stage of the system.

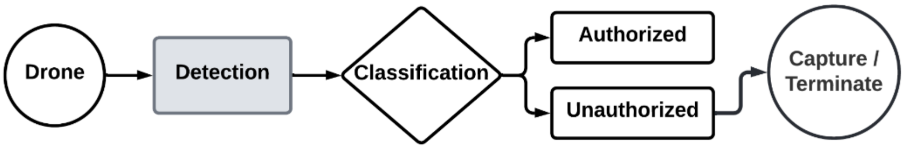

Fig. 1. Drone detection and classification system.

Now, for detection or identification of a drone can be described into the following groups:

One of the state-of-the-art procedures of aerial vehicle detection is using Radar. Usually, radar follows the Doppler effect and reflection time of emitted waves to determine an aerial object's position. Different methods and bands of frequency can be utilized for this purpose. Such small drones can be detected using distributed frequency modulation continuous wave (FMCW) connected with a fiber optic link instead of radio frequency within a 500-meter range [15]. Then the Multiple Input Multiple Output Orthogonal Frequency Division Multiplexing (MIMO OFDM) radars which have multiple transmitters and receiver antenna, can be used for identifying drones by Benjamin Nuss *et al.* [16]. Besides that, different bands such as X-Band (8.75GHz) and Ku-Band (12-18GHz) are also effective for detecting aerial vehicles, especially drones [17, 18]. However, because of having a smaller size, higher speed, and a very lower altitude of fly, it becomes very tough to detect all kinds of drones with the radar system.

Another way is to use the power of high-precision acoustic sensing. Drones generate a sound signature from its rotor. By detecting and analyzing the pattern of this sound drones can be detected easily. This method can be very useful in cases of lower altitudes where the radar system faces difficulties. As a result, various methods and processes can be followed to use the acoustic signature for drone detection [19–26]. CNN method can be followed which in some cases outperformed other popular methods [19], However, CRNN is faster and more stable. Support Vector Machine (SVM), medium Gaussian SVM, SVM along with AlexNet, and Mel Frequency Cepstral Coefficients (MFCC) can also perform well [20, 21, 23]. Noise interruptions and different modifications in the case of drone detection via audio are also tested [24]. However different motorized equipment could have similar types of motors that are used in drones, in that case, it becomes complex to make a difference between those acoustic signatures. Besides that, the ambient noise of the environment also interferes with the acoustic signature. Even if the system can detect the drone successfully, it is extremely complex to classify them.

Among the other ways of drone detection, visual analysis is one of the major and dependable processes. This method depends on the use of a camera or sensors to capture the image or video of the surrounding environment and then analyze the data to detect the position, size, shape, speed, and other important features of the drone. This procedure is very effective not only for detection but also for classification which is the main goal to ensure security. From those understandings, we have focused on a method using visual analysis procedures and enhancing drone detection accuracy.

The contribution summary of this paper is as follows:

A novel approach is presented for drone detection using YOLOv7 [27, 28] and is an attempt to discern the possible limitations of the dataset.

We compare the results of this YOLOv7 model with YOLOv8 [29], YOLO-NAS, and Faster RCNN to analyze the results.

A dataset comprising 57,311 images was accumulated of which, 40,878 images consisted of a minimum of one instance of UAV each whereas the remaining 16,433 images were instances of birds flying in the open sky which were used as background images to mitigate the false positive rates in our approach. Images and labels were collected from a wide range of sources [30-32].

The paper follows a specific organization that includes a literature review in Chapter 2, which provides a background on previous research in the field. Chapter 3 presents the proposed methodology, outlining the steps taken to achieve the research objectives. Chapter 4 offers a detailed analysis of the experimental results, providing insights into the effectiveness of the proposed methodology, and finally, Chapter 5 concludes the paper.

2. Literature Review

In past years, various convolutional neural network (CNN) based approaches have been applied in this area [32–35]. Where Saqib *et al.* evaluate deep learning-based object detection techniques for detecting drones [33]. Different CNN-based network architectures- ZF, and VGG16 were experimented with using pre-trained models with transfer learning due to sparse training data. Among the architectures, VGG16 with Faster R-CNN performed the best on the training dataset, according to the results. The study suggests that considering birds as a separate class could improve results. Then in [32], propose an end-to-end object detection model based on CNN for drone detection and identifies the smallest rectangle that encloses the drone. The study proposes an algorithm to overcome the issue of limited training data by generating a large synthetic dataset merging real images with their background removed. After that, the challenge of limited training data for detecting UAVs using deep neural networks is presented in a study [34]. To overcome this challenge, the authors propose using the PBRT rendering software to generate a large number of photorealistic UAV images with high variation. They trained the Faster R-CNN network using their rendered images and achieved an AP of 80.69% on the test set, which is much higher than the network trained only on the COCO and PASCAL VOC datasets. The rendered image dataset also contains information on important parts of UAVs and pixels covered by UAVs, which can be used for more complex applications.

Besides that, Ulzhalgas Seidaliyeva *et al.* presents a novel approach for image-based detection and classification of UAVs using a deep learning-based framework [36]. The proposed method utilizes a combination of two pre-trained models: a Faster R-CNN model for detecting UAVs in images and a VGG-16 model for classifying the detected UAVs into different types. The proposed method got an overall precision of 96.7%. Then most recently Dong-Hyun Lee presented a deep learning-based approach for detecting drones using a single camera mounted on a moving vehicle [35] where a two-stage detection framework was used with an RPN and CNN achieving an AP of 0.873. However, the CNN methods require large datasets and have limited generalizability and interpretability.

Another way to address drone detection is to use various sensing techniques like RF signals, video, sounds, and thermal imaging to detect drones. In [37], a machine learning RF-based drone detection and identification (DDI) system uses low band RF signals from drone-to-flight controller communication and employs three XGBoost models to detect the presence of a drone, identify its type, and operational mode. The system achieved high accuracy rates of 99.96%, 90.73%, and 70.09% for these tasks, respectively. In [38], the OpenCV-based method is used for drone detection and it can operate on drones equipped with cameras. The system analyzes the camera images to determine the location of the image and the vendor model of the drone using machine classifiers. The system achieved an accuracy of about 89%. However, those models are in some cases, computationally expensive and require a significant amount of memory and processing power. In [39], a machine-learning framework is proposed for detecting and classifying amateur drones using sound in noisy environments. MFCC and linear predictive cepstral coefficients (LPCC) are used for feature extraction, and SVM with different kernels are used for classification. The proposed scheme achieves around 97% detection accuracy for amateur drones using SVM with cubic kernel and MFCC features, outperforming correlation-based drone sound detection schemes. They plan to conduct larger-scale experiments with bigger datasets to improve detection accuracy.

You Only Live Once (YOLO) has gained significant attention for object detection in real-time video streams. Different versions of YOLO have been used in various studies for drone detection. The YOLOv3 object detector is used in combination with a convolutional neural network to detect and classify drones with high accuracy in real-time [40]. The YOLOv3 model is trained with a limited number of epochs due to its large architecture and may have difficulty detecting certain types of drones. To improve accuracy, the paper suggests adding counter mechanisms such as, RF signal detection and acoustic systems. The suggested methods include X band and micro-doppler RADAR for detecting drones based on their RF signals, and an acoustic system to detect the type of drone based on its sound. Another study presents a modified version of the YOLO-V3 algorithm for real-time detection of drones in images with enhanced performance using a CNN with fewer layers and densely connected modules [41]. The modified algorithm achieved an accuracy of 95.6% and an average precision of 96% on a newly designed drone dataset. The study suggests the potential of the modified YOLO-V3 algorithm for drone detection and tracking, but it does not discuss its limitations or potential shortcomings.

After that, an automated drone detection system was developed using the YOLOv4 deep learning-based object detection algorithm, trained on drone and bird datasets [42]. They evaluated its performance on various metrics and found it performed better than previous similar studies, achieving a mAP of 74.36% and an FPS of 20.5 on the DJI

Phantom III. The study was limited to YOLO implementation, but the researchers plan to use more diverse image datasets and compare their results with other object detection algorithms in future work. Another study proposes a drone detection system using YOLOv5 and compares its performance to YOLOv4 [43]. The system is trained with pre-trained weights, data augmentation, and transfer learning, achieving a 90.40% mAP, a 21.57% improvement over the previous model. The system was tested on two drone models at different altitudes, reaching maximum FPS values of 23.9 and 31.2 using an NVIDIA Tesla T4 GPU. Future plans include exploring different YOLO versions, larger datasets, and additional object detection algorithms to achieve better results. Then most recently few researchers trained a model on a dataset of various drone sizes in a standard image frame and achieved high accuracy with fast object detection [44]. The performance of the YOLOv5 model is compared to two other models and found to be superior in terms of detection accuracy and execution time. The model achieves a confidence of 0.993 @0.5IOU, providing fast object detection and accuracy. In future work, the authors plan to develop and implement the YOLOv5 network into practical applications.

Besides that, studies related to modifying and enhancing the YOLOv4 were conducted by Cheng *et al.* where light weight MobileViT was used with YOLOv4 to address the problems of large model parameters and false and missing detections of multi-scale drone targets [45]. The performance of the YOLOv5 model is compared to two other models and found to be superior in terms of detection accuracy and execution time. The model achieves a confidence of 0.993 @0.5IOU, providing fast object detection and accuracy. In future work, the authors plan to develop and implement the YOLOv5 network into practical applications. Similar way, YOLOv5 has been used in few other recent studies such as, to classify drones into different categories Valaboju *et al.* [46] used the YOLOv5 model. YOLOv5 and synchronized multi-camera data used to accurately detect drones in 3D using an asymmetric cross approach and specialized performance metrics in another study [47].

Now it can be stated that different algorithms, such as CNN, R-CNN, SVM, XGBoost, and YOLO have been used in recent studies to detect drones efficiently. However, this paper proposes the use of YOLOv7, an advanced and new version of YOLO for drone detection, and compares its performance with other popular object detection algorithms. The study aims to explore the potential of YOLOv7 in drone detection, where its capabilities and limitations can be evaluated to improve the accuracy and efficiency of drone detection systems.

3. Proposed Methodology

Fig. 2 illustrates details of the workflow of this study.

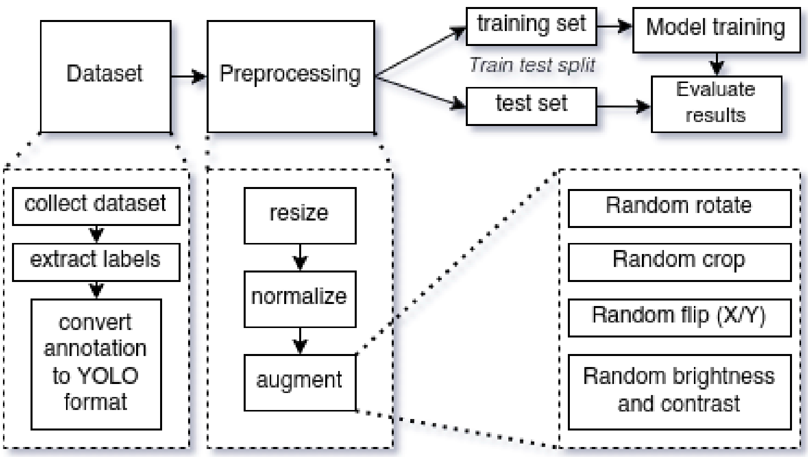

Fig. 2. A block diagram of our workflow.

3.1. Dataset and Preprocessing

The dataset was curated by collecting samples of drone and bird images from labeled and unlabeled datasets from various sources throughout the internet. Table 1 summarizes the details of the dataset, where 35,496 samples from multimodal drone dataset, 2194 samples from drone dataset, and 16034 samples from drone vs bird dataset.

Table 1. A summary of the datasets acquired.

Dataset	Number of samples
(Multimodal) Drone Detection Dataset	35,496
Drone Dataset (UAV)	2,194
Drone vs Bird	16,034
Note: 16,324 background images and 37,400 images of drones	

Fig. 3 shows some training samples from our acquired datasets. The labels were transformed from MSCOCO format to YOLO format. Fig. 4 shows some drone samples and their corresponding annotated bounding boxes.

Fig. 3. Some sample drones from the training dataset.

Fig. 4. Some drone samples and their corresponding annotated bounding boxes from the dataset.

Pre-processing techniques such as resizing, normalization, and data augmentation were applied to the dataset to ensure consistency, quality, and variability. The pre-trained YOLOv7 model was then utilized to train the drone detection system, which is evaluated using the mean average precision (mAP) metric.

The images were resized to a consistent size of 640 by 640 pixels to avoid inconsistencies in the input size. Normalization techniques are then applied to adjust the pixel values of the images to a common scale, eliminating variations in lighting and contrast between images. Additionally, data augmentation techniques are used to increase the dataset's size and variability, which enables the detection model to generalize better to new data. Rotation, flipping, and blurring are some augmentation techniques that were used in this study.

3.2. Model Architecture

The YOLOv7 model improves upon its predecessors by introducing new model reparameterization strategies and a noble label reassignment method. In addition, numerous optimization techniques were incorporated into the training modules and the use of noble extend and compound scaling methods significantly reduced its parameters and complexity. Fig. 5 visualizes the YOLOv7 architecture, the feature maps are extracted by the backbone while the bounding boxes and class probabilities are generated using the lead head and auxiliary head networks. The neck is responsible for enhancing feature maps by performing feature aggregation.

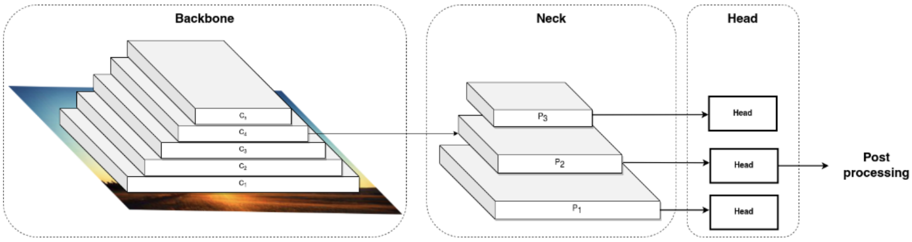

Fig. 5. The YOLO object detection framework.

EELAN computational block: With the model being significantly more lightweight and faster compared to its predecessors, its backbone feature extraction network is based on Extended Efficient Layer Aggregation Network (E-ELAN) computational block as detailed in Fig. 6. The E-ELAN network is used to preserve its original gradient path while enhancing its feature cardinality.

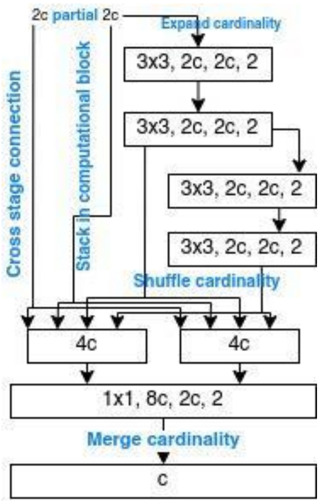

Fig. 6. Extended Efficient Layer Aggregation Network (E-ELAN)

Coarse for auxiliary and fine for lead loss: The detection head is composed of two components: 1) the lead head, and 2) the auxiliary head. The lead head is responsible for producing the final outputs. In contrast, the auxiliary head is responsible for producing hierarchical labels or soft labels to help facilitate the inference at the lead head. The lead and auxiliary heads utilize soft labels for training the model, which is generated based on the correlation and distribution between the source and target data. Soft labels created by the lead head have higher learning abilities than those generated by other methods. The second step involves a "coarse-to-fine lead head-guided label assigner," which produces two sets of soft labels. The first set is a coarse label, which selects more positive targets among grids, and the second set is a fine label that is similar to the soft label created by the lead head. Both soft labels are created by utilizing predicted outcomes from the ground truth and lead head. The method ensures that the training model can learn more efficiently and accurately from the available data by providing different types of soft labels. This approach can enhance the overall performance of the model in the long run. This approach allows for dynamically adjusting the importance of coarse and fine labels during training.

Re-parameterized Convolution: The network includes model-level and module-level reparameterizations to improve the model's performance without compromising the training cost. Reparametrized convolution (RepConv) is a type of convolutional layer that utilizes a combination of 3x3 convolution, 1x1 convolution, and identity connection. The identity connection in RepConv disrupts the residual aspect in ResNet and the concatenation feature in DenseNet. This disruption can be beneficial as it provides a greater variety of gradients for different feature maps.

Compound Model Scaling: Model scaling is a technique used to modify a model's key features to suit specific application requirements, including its resolution, depth, and width. However, in concatenation-based architectures like ResNet, scaling parameters are interdependent and cannot be examined independently. Increasing the model depth can impact the transition layer's input and output channel ratio, leading to decreased hardware utilization. To address this issue, YOLOv7 uses a compound scaling approach in concatenation-based models. This allows the model to maintain its original properties while achieving optimal design.

As a result, the model is computationally less expensive than its predecessors, and less expensive hardware will be able to run the detection algorithm faster and with improved accuracy.

Intersection over Union and Non-Max Suppression: The Intersection over Union (IoU) is a metric that measures the degree of overlap between two bounding boxes in object detection tasks. It is calculated by taking the intersection area of two boxes and dividing it by their union. The IoU score ranges from 0 to 1, where a score of 1 indicates a perfect match between the two boxes, and a score of 0 indicates no overlap. IoU is commonly used to determine if a predicted bounding box is a true positive detection by comparing it to the ground truth bounding box. If the IoU score is above a certain threshold, the predicted box is considered a true positive. IoU is an important metric in object detection since it can help evaluate the accuracy of a model's predictions. By using IoU, we can determine the reliability of the model's detections and adjust the threshold accordingly. The predicted box is considered a false positive detection if the score is below the threshold. This is mathematically represented as follows:

$$IoU = \frac{area\ of\ intersection}{area\ of\ union} \tag{1}$$

Non-Maximum Suppression (NMS) is a post-processing technique used to eliminate redundant detections in object detection tasks. When an object is detected, the detection algorithm may output multiple overlapping bounding boxes around the object. NMS is used to remove these redundant detections and keep only the most accurate ones. The NMS algorithm works by first sorting the detections based on their confidence scores. It then selects the detection with the highest confidence score and removes any other detections that have a high IoU score. This process is repeated for the remaining detections until all redundant detections have been eliminated.

4. Experimental Setup and Result Analysis

The pre-trained version of the model (on the MSCOCO dataset) was utilized in our experiments. A train-test split of 80-20 was performed. The model was trained and evaluated on an RTX Titan GPU with 24 GB VRAM. The training was performed with an IOU threshold of 0.4 and a batch size of 16 for 100 epochs for the YOLOv7 model whereas the YOLOv8 model was trained for up to 150 epochs with the same configurations. Table 2 presents detailed hyperparameters used in our experiments.

Table 2. Hyperparameter values for training the pre-trained YOLOv7 and YOLOv8 models.

Parameters	Values
Optimizer	Adam
Learning rate	0.001
Momentum	0.937
Weight decay	0.0005

Mean average precision (mAP), F1 score, precision, and recall are metrics that are commonly used to evaluate the performance of classification and object detection models. Precision is a measure that calculates the proportion of true positive predictions in relation to all positive predictions. Recall calculates the proportion of true positive predictions compared to all actual positives in the dataset. These metrics are essential for assessing the effectiveness of models that deal with class imbalance. It is important to understand and use these metrics to ensure that models are accurately evaluated and compared.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{2}$$

$$Recall = \frac{True\ Positive}{True\ Positive + True\ Negative} \tag{3}$$

F1 is a harmonic mean of precision and recall and provides a single scalar value that summarizes the model's accuracy in predicting positive and negative instances. When dealing with imbalanced datasets, the F1 score is valuable because it considers precision and recall, both of which are crucial in such situations. A higher F1 score indicates better model performance, with a perfect score of 1.0 indicating perfect precision and recall.

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{4}$$

Mean Average Precision (mAP), on the other hand, is a widely used performance metric for evaluating object detection models in computer vision. The mAP score measures the average precision of a model's predictions across all object categories and detection thresholds, providing a comprehensive summary of its detection accuracy. Higher mAP value indicates better performance. To calculate mAP, a model's predicted detections are compared against the ground truth annotations by computing the precision and recall values. A precision-recall curve is then plotted using which the average precision (AP) is computed. This value is averaged across all classes to obtain the final mAP score. The YOLOv8 (large variant) model had 43.7 million parameters compared to the base YOLOv7 model's 36.9 million.

Fig. 7 and Fig. 8 demonstrate the training curves for our experiments, respectively.

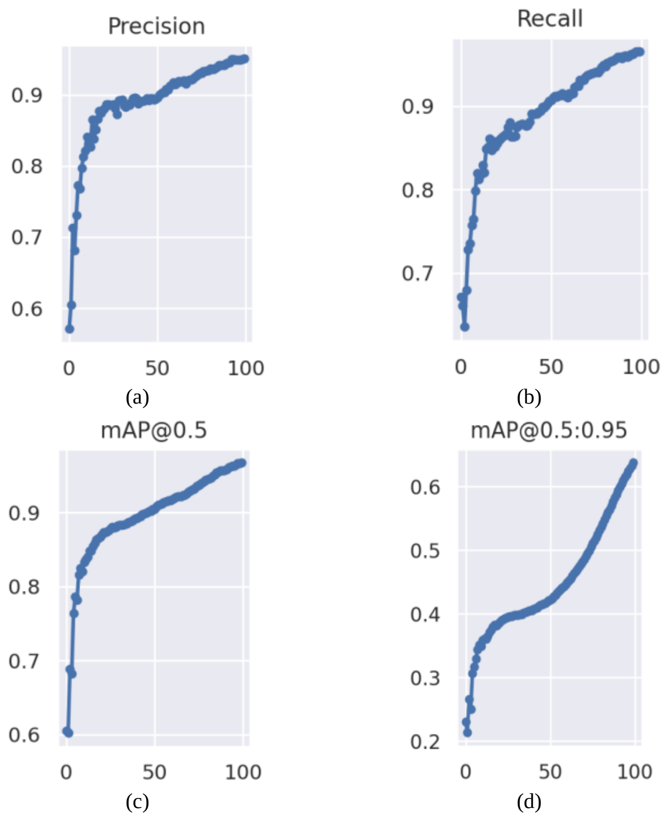

Fig. 7. Training curves of the YOLOv7 model. (a) Precision curve, (b) Recall curve, (c) mAP@0.5 curve, (d) mAP@0.5-0.95 curve.

The precision and recall curves of the YOLOv7 model appeared to be much smoother than that of the YOLOv8 model. Similarly, by looking at the mAP curves, the YOLOv7 model outperforms the YOLOv8 model by 55 percent in our combined drone dataset. Overall, the training curves of the YOLOv8 model appeared to be more jagged and, hence, more unstable when compared to that of the YOLOv7 model.

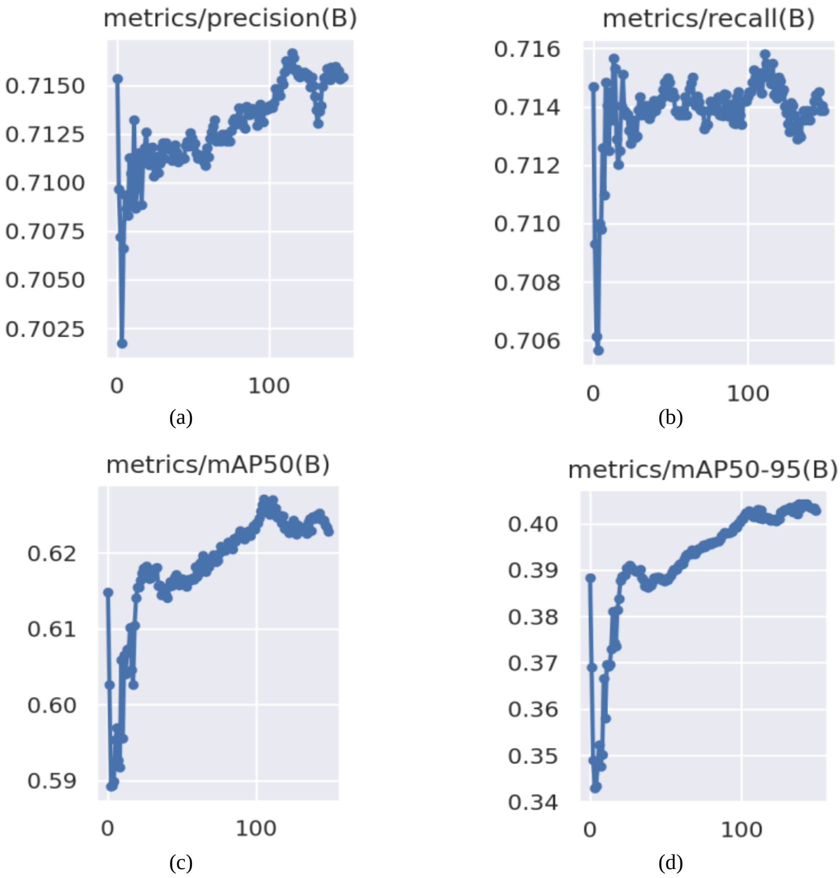

Fig. 8. Training curves of the YOLOv8 model. (a) Precision curve, (b) Recall curve, (c) mAP@0.5 curve, (d) mAP@0.5-0.95 curve.

We assessed the models using mAP, precision, recall, and F1 metric and tested their performance using video footage of drones from YouTube. We tested its capability by comparing it to the base YOLOv8 model. Table 3 summarizes the results of our experiments.

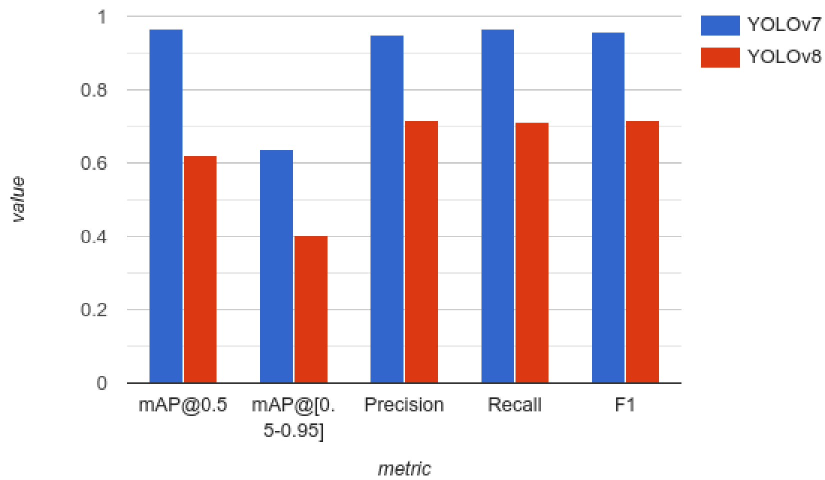

Fig. 9. A side-by-side comparison of the YOLOv7 and YOLOv8 models.

The YOLOv7 model achieved an mAP@0.5 of 0.9673. Furthermore, the model precision and recall were recorded as 0.9509 and 0.9656, respectively. In contrast, YOLOv8 achieved a mAP@0.5 of 0.6229 with precision and recall of 0.7154 and 0.7138, respectively. Table 4 summarizes the training results of the YOLOv8 model. Fig. 9 provides a side-by-side comparison of the performance between YOLOv7 and YOLOv8. The YOLOv7 model was observed to have 55.29 % more mAP@0.5, 58.29% more mAP@[0.5-0.95] and 34.10% increased F1 score.

To assess the generalization ability of both models, we conducted a comparative analysis using a YouTube video. Fig. 10 presents the results of this analysis, displaying a side-by-side comparison of the inference outcomes for some sample frames extracted from the video output.

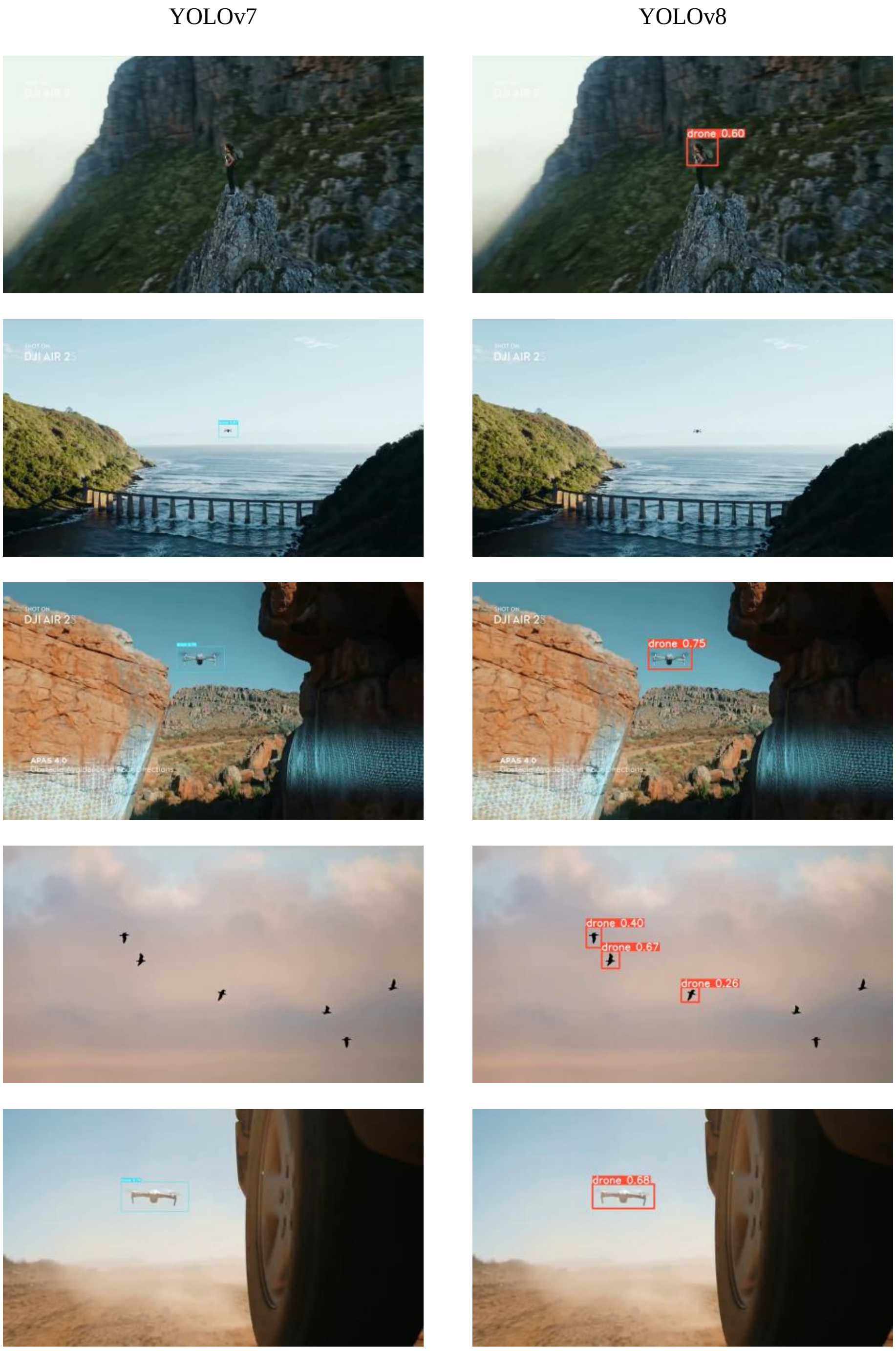

Fig. 10. A side-by-side comparison of both models' inference capabilities on unseen data.

The performance of the proposed model was compared with the state-of-art models, as depicted in Table 3, where Kim *et al.* utilized YOLOv8 from done detection [48], for our dataset, the YOLOv8 shows comparatively better performance than the other models by exhibiting F1 score of 0.715 though in different scenario another study found 96% using YOLOv8 for obstacle detection [49]. On the other hand, YOLO-NAS was used for drone detection by Munir *et al.* [50]. Although the model demonstrated improved recall and mAP values compare to YOLOv8, due to it very lower precision, the model shows F1 score of 0.308 for the dataset. Among the models, the Faster RCNN used in [51] has exhibited a very poor F1 score of 0.148 for the dataset. It happens due to the small dataset, the Faster R-CNN requires a substantial amount of labeled training data to perform well, where the dataset preparation and annotation is time-consuming and expensive.

Table 3. Comparison of different state-of-art models for drone detection

	mAP@0.5	mAP@[0.5-0.95]	Precision	Recall	F1-Score
YOLOv8 [48]	0.6229	0.4028	0.7154	0.7138	0.7146
YOLO-NAS [50]	0.8903	0.5816	0.1825	0.9855	0.3080
Faster-RCNN [51]	0.29	0.19	0.10	0.29	0.148
YOLOv7	0.9673	0.6376	0.9509	0.9656	0.9582

5. Conclusions

In this research, advanced image processing techniques, and deep learning algorithms were employed to detect and recognize drones from various sources. To achieve this, YOLOv7, YOLOv8, YOLO-NAS, Faster-RCNN models were trained on a combined UAV dataset. The experimental results showed that the YOLOv8 model was able to detect drones, but the YOLOv7 model outperformed the YOLOv8 model in terms of accuracy and reliability by 55.29% on mAP@0.5 and 34.10% on the F1 score. In addition, although the YOLO-NAS exhibited improved recall and mAP values, due to its lower precision, it shows F1 score of 0.308 only. Among the models, Faster-RCNN shows very poor performance due to the small size of our dataset. The deep learning-based approach using YOLOv7 has proven to be a highly precise and accurate method for drone recognition. In future work, the focus will be on testing this approach in real-time environments and dynamic scenarios to improve its robustness and practicality.

Acknowledgment

This research is funded by Woosong University Academic Research 2023.

References

[1] M. Ayamga, S. Akaba, and A. A. Nyaaba, “Multifaceted applicability of drones: A review,” Technological Forecasting and Social Change, vol. 167, p. 120677, 2021. https://doi.org/10.1016/j.techfore.2021.120677

[2] B. Mishra, D. Garg, P. Narang, and V. Mishra, “Drone-surveillance for search and rescue in natural disaster,” Computer Communications, vol. 156, pp. 1–10, 2020. https://doi.org/10.1016/j.comcom.2020.03.012

[3] G. Ding, Q. Wu, L. Zhang, Y. Lin, T. A. Tsiftsis, and Y.-D. Yao, “An amateur drone surveillance system based on the cognitive internet of things,” IEEE Communications Magazine, vol. 56, no. 1, pp. 29–35, 2018. https://doi.org/10.1109/MCOM.2017.1700452

[4] A. Gynnild, “The robot eye witness,” Digital Journalism, vol. 2, no. 3, pp. 334–343, 2014. https://doi.org/10.1080/21670811.2014.883184

[5] A. Singh, D. Patil and S. N. Omkar, "Eye in the Sky: Real-Time Drone Surveillance System (DSS) for Violent Individuals Identification Using ScatterNet Hybrid Deep Learning Network," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, USA, 2018, pp. 1710-17108, doi: 10.1109/CVPRW.2018.00214.

[6] M. Nalamati, A. Kapoor, M. Saqib, N. Sharma and M. Blumenstein, "Drone Detection in Long-Range Surveillance Videos," *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Taipei, Taiwan, 2019, pp. 1-6, doi: 10.1109/AVSS.2019.8909830.

[7] A. N. O. U. K. S. RIGTERINK, “The Wane of Command: Evidence on Drone Strikes and Control within Terrorist Organizations,” *American Political Science Review*, vol. 115, no. 1, pp. 31–50, 2021. https://doi.org/10.1017/S0003055420000908.

[8] R. Luppicini and A. So, “A technoethical review of commercial drone use in the context of governance, ethics, and privacy,” Technology in Society, vol. 46, pp. 109–119, 2016. https://doi.org/10.1016/j.techsoc.2016.03.003.

[9] Y. Yao, H. Xia, Y. Huang, and Y. Wang, “Free to fly in public spaces: Drone controllers’ privacy perceptions and practices,” in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI ’17, (New York, NY, USA), p. 6789–6793, Association for Computing Machinery, 2017. https://doi.org/10.1145/3025453.3026049

[10] B. Kotkova, "Airport defense systems against drones attacks," *2022 26th International Conference on Circuits, Systems, Communications and Computers (CSCC)*, Crete, Greece, 2022, pp. 85-90, doi: 10.1109/CSCC55931.2022.00025.

[11] N. P. Walsh, N. Gallón, E. Perez, D. Castrillon, B. Arvanitidis, C. Hu, “Inside the August plot to kill Maduro with drones,” *CNN*, Mar. 14, 2019. https://edition.cnn.com/2019/03/14/americas/venezuela-drone-maduro-intl/index.html (accessed Feb. 21, 2023).

[12] J. P. Daniels, “Venezuela’s Nicolás Maduro survives apparent assassination attempt,” *The Guardian*, Aug. 05, 2018. Available: https://www.theguardian.com/world/2018/aug/04/nicolas-maduros-speech-cut-short-while-soldiers-scatter (accessed Feb. 21, 2023).

[13] K. Taneja, “Saudi Aramco attack shakes up fragile global ecosystem,” *Policycommons.net*, Sep. 18, 2019. https://policycommons.net/artifacts/1352659/saudi-aramco-attack-shakes-up-fragile-global-ecosystem/1964818/ (accessed Feb. 21, 2023).

[14] “Man fined after flying drones over Premier League stadiums,” *BBC News*, Sep. 15, 2015. Accessed: Apr. 12, 2023. [Online]. Available: https://www.bbc.com/news/uk-england-nottinghamshire-34256680 (accessed Feb. 21, 2023).

[15] D.-H. Shin, D.-H. Jung, D.-C. Kim, J.-W. Ham, and S.-O. Park, “A Distributed FMCW Radar System Based on Fiber-Optic Links for Small Drone Detection,” *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 2, pp. 340–347, Feb. 2017, doi: https://doi.org/10.1109/tim.2016.2626038.

[16] B. Nuss, L. Sit, M. Fennel, J. Mayer, T. Mahler and T. Zwick, "MIMO OFDM radar system for drone detection," *2017 18th International Radar Symposium (IRS)*, Prague, Czech Republic, 2017, pp. 1-9, doi: 10.23919/IRS.2017.8008141.
[17] Á. D. de Quevedo, F. I. Urzaiz, J. G. Menoyo and A. A. López, "Drone Detection With X-Band Ubiquitous Radar," *2018 19th International Radar Symposium (IRS)*, Bonn, Germany, 2018, pp. 1-10, doi: 10.23919/IRS.2018.8447942.
[18] J. Ochodnický, Z. Matousek, M. Babjak and J. Kurty, "Drone detection by Ku-band battlefield radar," *2017 International Conference on Military Technologies (ICMT)*, Brno, Czech Republic, 2017, pp. 613-616, doi: 10.1109/MILTECHS.2017.7988830.
[19] S. Al-Emadi, A. Al-Ali, A. Mohammad and A. Al-Ali, "Audio Based Drone Detection and Identification using Deep Learning," *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, Tangier, Morocco, 2019, pp. 459-464, doi: 10.1109/IWCMC.2019.8766732.
[20] S. Jamil *et al.*, "Malicious UAV Detection Using Integrated Audio and Visual Features for Public Safety Applications," *Sensors*, vol. 20, no. 14, p. 3923, Jul. 2020, doi: 10.3390/s20143923. [Online]. Available: http://dx.doi.org/10.3390/s20143923.
[21] S. Mandal, L. Chen, V. Alaparthy, and M. L. Cummings, "Acoustic Detection of Drones through Real-time Audio Attribute Prediction," *AIAA Scitech 2020 Forum*, Jan. 2020, doi: https://doi.org/10.2514/6.2020-0491.
[22] F. Svanstrom, C. Englund and F. Alonso-Fernandez, "Real-Time Drone Detection and Tracking with Visible, Thermal and Acoustic Sensors," in 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 2021 pp. 7265-7272. doi: 10.1109/ICPR48806.2021.9413241
[23] S. Salman, J. Mir, M. T. Farooq, A. N. Malik and R. Haleemdeen, "Machine Learning Inspired Efficient Audio Drone Detection using Acoustic Features," *2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)*, Islamabad, Pakistan, 2021, pp. 335-339, doi: 10.1109/IBCAST51254.2021.9393232.
[24] H. Kolamunna et al., "DronePrint," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 5, no. 1, pp. 1–31, Mar. 2021, doi: 10.1145/3448115.
[25] S. Al-Emadi, A. Al-Ali, and A. Al-Ali, "Audio-Based Drone Detection and Identification Using Deep Learning Techniques with Dataset Enhancement through Generative Adversarial Networks," *Sensors*, vol. 21, no. 15, p. 4953, Jul. 2021, doi: 10.3390/s21154953. [Online]. Available: http://dx.doi.org/10.3390/s21154953
[26] G. Lacava, F. Mercaldo, F. Martinelli, A. Santone, and M. Pizzi, "Drone Audio recognition based on Machine Learning Techniques," *Procedia Computer Science*, vol. 207, pp. 848–857, 2022, doi: 10.1016/j.procs.2022.09.140.
[27] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arXiv:2207.02696 [cs]*, Jul. 2022, Available: https://arxiv.org/abs/2207.02696
[28] C.-Y. Wang, H.-Y. M. Liao, and I.-H. Yeh, "Designing Network Design Strategies Through Gradient Path Analysis," *arXiv:2211.04800 [cs]*, Nov. 2022, Accessed: Apr. 12, 2023. [Online]. Available: https://arxiv.org/abs/2211.04800
[29] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics," *GitHub*, Jan. 01, 2023. https://github.com/ultralytics/ultralytics (accessed Feb. 21, 2023)
[30] Mehdi (2023) drone-dataset. In: GitHub. https://github.com/dasmehdix/drone-dataset. (accessed Feb. 21, 2023)
[31] Birds vs Drone Dataset. https://www.kaggle.com/datasets/harshwalia/birds-vs-drone-dataset. (accessed Feb. 21, 2023)
[32] C. Aker and S. Kalkan, "Using deep networks for drone detection," *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Lecce, Italy, 2017, pp. 1-6, doi: 10.1109/AVSS.2017.8078539.
[33] M. Saqib, S. Daud Khan, N. Sharma and M. Blumenstein, "A study on detecting drones using deep convolutional neural networks," *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Lecce, Italy, 2017, pp. 1-5, doi: 10.1109/AVSS.2017.8078541.
[34] J. Peng *et al*, "Using Images Rendered by PBRT to Train Faster R-CNN for UAV Detection," Václav Skala-UNION Agency: 2018.
[35] D.-H. Lee, "CNN-based single object detection and tracking in videos and its application to drone detection," *Multimedia Tools and Applications*, Oct. 2020, doi: 10.1007/s11042-020-09924-0.
[36] U. Seidaliyeva, D. Akhmetov, L. Ilipbayeva, and E. T. Matson, "Real-Time and Accurate Drone Detection in a Video with a Static Background," *Sensors*, vol. 20, no. 14, p. 3856, Jul. 2020, doi: 10.3390/s20143856. [Online]. Available: http://dx.doi.org/10.3390/s20143856
[37] O. O. Medaiyese, A. Syed and A. P. Lauf, "Machine Learning Framework for RF-Based Drone Detection and Identification System," *2021 2nd International Conference on Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS)*, Tangerang, Indonesia, 2021, pp. 58-64, doi: 10.1109/ICON-SONICS53103.2021.9617168.
[38] D. Lee, W. Gyu La and H. Kim, "Drone Detection and Identification System using Artificial Intelligence," *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju, Korea (South), 2018, pp. 1131-1133, doi: 10.1109/ICTC.2018.8539442.
[39] M. Z. Anwar, Z. Kaleem and A. Jamalipour, "Machine Learning Inspired Sound-Based Amateur Drone Detection for Public Safety Applications," in *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2526-2534, March 2019, doi: 10.1109/TVT.2019.2893615.
[40] 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)*, Madurai, India, 2020, pp. 1012-1016, doi: 10.1109/ICICCS48265.2020.9121150.
[41] H. R. Alsanad, A. Z. Sadik, O. N. Ucan, M. Ilyas, and O. Bayat, "YOLO-V3 based real-time drone detection algorithm," *Multimedia Tools and Applications*, Mar. 2022, doi: 10.1007/s11042-022-12939-4.
[42] S. Singha, B. Aydin, "Automated Drone Detection Using YOLOv4," Drones, vol. 5, no. 95, 2021, pp. 1-20, https://doi.org/10.3390/drones5030095
[43] B. Aydin, S. Singha, "Drone Detection Using YOLOv5," Eng, vol. 4, pp. 416–433, 2023. https://doi.org/10.3390/eng4010025.
[44] V. T. Hai, V. N. Le, D. Q. Khanh, N. P. Nam, and D. V. Sang, "Multi-size drone detection using YOLOv5 network," *Journal of Military Science and Technology*, no. 80, pp. 142–148, Jun. 2022, doi: 10.54939/1859-1043.j.mst.80.2022.142-148.
[45] Q Cheng, Li X, Zhu B, *et al.* "Drone Detection Method Based on MobileViT and CA-PANet," *Electronics*, vol. 12, no. 223, pp. 1-16, 2023. https://doi.org/10.3390/electronics12010223

[46] Valaboju R, Vaishnavi, Harshitha C, *et al.* “Drone Detection and Classification using Computer Vision,” *7th International Conference on Trends in Electronics and Informatics (ICOEI)*, 2023, pp 1320–1328

[47] Lindenheim-Locher W, Świtoński A, Krzeszowski T, *et al.*, “YOLOv5 Drone Detection Using Multimodal Data Registered by the Vicon System,” *Sensors*, vol. 23, no. 6396, pp. 1-16, 2023, https://doi.org/10.3390/s23146396

[48] J. -H. Kim, N. Kim and C. S. Won, "High-Speed Drone Detection Based on Yolo-V8," *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, 2023, pp. 1-2, doi: 10.1109/ICASSP49357.2023.10095516.

[49] S. Rahman, J. H. Rony, J. Uddin, and M. A. Samad, “Real-Time Obstacle Detection with YOLOv8 in a WSN Using UAV Aerial Photography,” J. Imaging, vol. 9, no. 10, Art. no. 10, Oct. 2023, doi: 10.3390/jimaging9100216

[50] A. Munir, A. J. Siddiqui, “Vision-based UAV Detection in Complex Backgrounds and Rainy Conditions,” 2023, arXiv preprint arXiv:2305.16450.

[51] B. Benjdira, T. Khursheed, A. Koubaa, A. Ammar, K. Ouni, “Car detection using unmanned aerial vehicles: Comparison between faster r-cnn and yolov3,” *1st International Conference on Unmanned Vehicle Systems-Oman (UVS)*, 2019, pp. 1-6.

Authors’ Profiles

Muhammad K. Kabir has completed Bachelor degree in Computer Science and Engineering from Brac University, Dhaka, Bangladesh. Currently he is working as a team member in the Multimedia Signal & Image Processing Research Group (MSiP) under the supervision of Dr. Jia Uddin, faculty member of AI and Big Data Department, Endicott College, Woosong University, Daejeon, South Korea. His research area includes-machines learning, deep learning for smart system design, anomaly detection.

Anika N. Binte Kabir has completed Bachelor degree in Computer Science and Engineering from Brac University, Dhaka, Bangladesh. Currently she is working as a team member in the Multimedia Signal & Image Processing Research Group (MSiP) under the supervision of Dr. Jia Uddin, faculty member of AI and Big Data Department, Endicott College, Woosong University, Daejeon, South Korea. Her research area includes-machines learning, deep learning for smart system design, anomaly detection.

Jahid H. Rony is currently pursuing his B.Sc. in Computer Science and Engineering degree from Dhaka University of Engineering & Technology (DUET), Gazipur. Jahid has been actively involved in various extracurricular activities during his studies. He is the President of DUET Robotics Club, where he leads a team of students in developing autonomous systems for various applications for the year of 2023. He is also the Newsletter Editor of IEEE DUET Student Branch, where he is responsible for managing and publishing newsletters to keep the members updated about the latest developments in the field of electrical and electronics engineering. During his undergraduate studies, he worked on developing machine learning models for automated audio classification, image processing and so on. His research interests include IoT, Robotics, Automation, Machine learning, and AI.

Jia Uddin is as an Assistant Professor, Department of AI and Big Data, Endicott College, Woosong University, Daejeon, South Korea. He received Ph.D. in Computer Engineering from University of Ulsan, South Korea, M.Sc. in Electrical Engineering (Specialization: Telecommunications), Blekinge Institute of Technology, Sweden, and B.Sc. in Computer and Communication Engineering, International Islamic University Chittagong, Bangladesh. He was a visiting faculty at School of computing, Staffordshire University, United Kingdom, Telkom University, Indonesia, and University of Foggia, Italy. He is an Associate Professor (now on leave), Department of Computer Science and Engineering, Brac University, Dhaka, Bangladesh. His research interests are Industrial Fault Diagnosis, Machine Learning/Deep Learning based prediction and detection using multimedia signals.

How to cite this paper: Muhammad K. Kabir, Anika N. Binte Kabir, Jahid H. Rony, Jia Uddin, "Drone Detection from Video Streams Using Image Processing Techniques and YOLOv7", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.16, No.2, pp. 83-95, 2024. DOI:10.5815/ijigsp.2024.02.07