Aerial Supervision of Drones and Other Flying Objects Using Convolutional Neural Networks

Vivian Ukamaka Ihekoronye, Simeon Okechukwu Ajakwe, Dong-Seong Kim, Jae Min Lee
Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea
(ihekoronyevivian, simeonajlove)@gmail.com,(dskim, ljmpaul)@kumoh.ac.kr

***Abstract*—Accurate detection of distant drones in clustered environment amidst other flying objects such as birds is of critical importance in anti-drone system design. This study proposed a novel object detection model that efficiently detect and differentiate drones from other flying objects under different weather conditions. The custom dataset consists of manually generated drone images and bird samples under sunny, cloudy and evening conditions. The simulation result shows that KITYOLO outperformed YOLOv5 both in precision (sunny 96.2% vs 85%; cloudy 73.7% vs 26.3%; evening 58.5% vs 26.1%) and recall (evening 42.4% vs 15%) in all aspects with an overall F1-score of 98% as against 91.9% while maintaining timeliness and memory usage.**

***Index Terms*—Aerial supervision, CNN, Drone, Detection, YOLO.**

I. Introduction

Drone companies are currently experiencing tremendous increase in sales due to the varying application of drones in different sectors. Gone are the days when drones are mostly deployed in the military sector for supervision and most times as a means of penetrating adverse terrain. Currently, drones are widely used by individuals and in the entertainment industries as hobbyist drones for aerial photography, agriculture application, object detection and logistics as seen by the Amazon groups. The high applications of drones is majorly due to the reduced cost and the miniature size of drones, experienced in its very swift maneuverability. The increase in the mis-usage of drones internationally is devastating, leading to economic loss of lives and properties. In early 2019, different airports in the UAE, USA and UK experienced mishaps as a result of drone operations with the recent violations in Saudi's Oilfield Aramco and Abha airport, where drones were illegally flown [1]. These circumstances require austere security measures because most drones have cameras mounted on them, making them capable of spying and retrieving confidential information in restricted areas. Also, transportation of explosives can easily be achieved with drones, making them very dangerous when used by attackers or terrorists. Thus, detecting and preventing such malicious practice implemented through the deployment of drones is crucial for the security of the society.

Drone detection, also known as anti-drone technology, is an act of detecting and/or tracking unwanted drones in any given restricted area or territory [2]. However, the similarity of drones and other flying objects in aerial view is the major challenge of detecting and restricting drones. Different techniques have been adopted for the detection of drones, ranging from acoustic [3]method that uses sensors to determine the sound emitted by the drone; radar approach [4] that implements radio waves to determine the distance, angle and velocity of the target object, infrared sensor; which uses the heat signature of the drone for detection [5], to the most paving technique which is computer vision (CV) technology, a field of Artificial Intelligence (AI) that enables computer to retrieve information from digital images, videos and other visual inputs, while reporting to the ground control stations.

Currently, computer vision is being used in solving object detection problems which is a peculiar AI problem, by deploying deep learning algorithms. Innovations in the different deep learning models have displayed consequential usage for object detection in ground based applications [6]. The extraction of meaningful information from images and videos can be achieved through detecting the image and also classifying it. To achieve the detection and classification of objects, Convolution Neural Networks (CNN), a deep learning algorithm is used for this purpose. CNN is responsible for the deep extraction of image features at different layers [7]. This paper seeks to address the challenge of aerial supervision of drones as target whilst accurately recognizing and predicting it from other objects such as birds in any weather condition. A state of the art CNN model was designed to solve this challenge, and also the proposed model was further compared with a very fast object detector based on computational complexity, accuracy and timeliness while achieving the following specific objectives:

1) Deploying computer vision for image capturing and processing of targets(drones and birds);
2) Gathering and labelling different datasets of 2 different drones and birds on flight;
3) Designing a state-of-the-art model that can optimally detect, predict and classify drones as well as birds based on computer vision and CNN;
4) Evaluation of the feasibility of the proposed model with the state-of-the-art model based on accuracy, sensitivity and computational complexity.

The remaining sections of this paper are categorized as: Section II, captures Related Works; the System Design is extensively discussed in Section III, while the Result Discussion, Evaluation Performance and Conclusion are captured in Sections IV and V respectively.

II. Related works

Computer vision is one of the most essential domain of AI, having different sectors such as object detection, image recognition and surveillance [8], with object detection being the most blooming sector as a result of its enormous applications. Object detection is the ability of computer (i.e anti-drone system in this research work) and software systems (i.e the proposed system) to locate objects in an image/ video and accurately distinguish each object. The rapid adoption of deep learning in computer vision has brought breakthrough to highly accurate object detection algorithms such as You Only Look Once (YOLO) [9], Single Shot Detectors(SSD), Fast-Regional Convolution Network (FRCNN) and Faster RCNN [9]. Nowadays, object detection is being deployed for surveillance operations, face recognition and security systems. Also, UAVs application is on the rise due to their high mobility, suitable incorporation in object detector models, easy deployment and their capacity to capture images at any altitude in respect to views, angles and scalar differences [10]. This in turn has posed challenges such as densely distributions of target objects, scale variance of aerial objects, and differentiation of UAVs and other flying objects in different weather conditions in airborne.

To contribute towards solving these recurring challenges, researchers have resorted to CNN for optimal solutions [11]. CNN also known as a deep learning algorithm, receives images as inputs, delegates learnable weights and biases to the objects in the image, then carry out prediction tasks on the objects based on their various classes. Several layers exist and are interconnected in the architecture of the CNN which is analogous to the connective patterns of neurons of the human brain, making it to be trained on any particular tasks based on the given parameters. With a focus on visual capturing and detection, most anti-drone detecting systems [5]are equipped with cameras that aids in the panning, tilting and zooming of target objects. The automatic techniques employed by CNN in image processing has also resulted to the wide interest of researchers, owing to the fact that CNN displays excellent performance in object detection and classification .

Classification of birds, drones and backgrounds were carried out by [5], evaluating several CNN models such as Resnet-50, Resnet-18, VGG16, Gogglenet, AlexNet and SqueezeNet to ascertain the best classifier. Although, these classifiers have already been validated in the ImageNet Large Scale Visual Recognition Competition (ILSVRC) with 1000 labels classification, their experiment however displays that for the classification of small number of labels, a simplified CNN model results to better performance. As Alexnet, Restnet18 and Squeezenet performed optimally than the other models when classifying just 3 labels, which was contrary to the result gotten from the 1000 label classification of ILSVRC. While, author [12] compared two variants object detectors, that is YOLO versions 2 and 3, for the detection and classification of drones from no drones with a total of 149 images for the training and validation of the model, having a higher accuracy of 95.20% for YoloV3.

To solve the problem of scale variations and densely distributions of objects, researchers [13] designed SPB-YOLO model, an end-to-end detector that has the strip bottleneck (SPB) module which used the attention mechanism approach to solve the dependency of scalar variations of UAV images. Also, by the upsampling of the detection head of YOLOv5 in the addition of a detection head based on Path Aggregation Network, the challenge of dense object distribution was mitigated. However, the disparity experienced in the detection and classification of aerial targets in different weather conditions is still a research gap. For anti-drones to be efficacious in drone detection and prediction of similar targets even during weather obscurity, an efficient state-of-the-art model needs to be embedded in it for optimal accuracy and speed.

YOLO architecture is a plausible innovation of Artificial intelligence for computer vision. YOLO is a single-shot object detector that is extremely fast when compared with its counterpart; multiple-shot detectors such as Fast-RCNN and Faster-RCNN. This is as a result of the YOLO technique, that uses the features extracted from the entire captured image while predicting the classes of the images simultaneously from bounding boxes. The architecture and mode of detecting and predicting drone images from other images by the YOLO model which is incorporated in the anti-drone system will be explained in the subsequent sections.

III. System Design

The operational processes of the proposed model is captured in Fig. 1. This model adopts YOLO architecture for the detection and classification of drones and other objects in aerial perspective. During surveillance, the anti-drone system captures all aerial images within its peripheral and central vision; drones and birds alike. The input to the proposed model are images extracted from the anti-drone system, which are subjected to further processing deploying the model's architecture.

The main functionality of this model is to detect and distinguish drones from birds, without being deterred by obscure weather conditions nor altitude of the object as it was trained under a sunny, cloudy and gloomy (evening) scenarios and at various heights, so as check the robustness of the model to accurately detect and predict the movement of drones in restricted places. Therefore, the inputs to the system is either drones or birds, relying on the architectural framework of the system; it processes the input and generate outputs based on the classifications of the object.

A. Custom Drone Detection Strategy

The standard YOLO architecture detector is designed on three distinct modules. The Backbone Module that adopts the Cross Stage Partial Network (CSPNet) [14] responsible for drone/bird feature extractions. Next, is the Neck built on the Feature Pyramid Network, for features aggregation. Lastly, the Head Module that aids the model to handle varying sizes of objects and capable of generating multi-scale predictions.

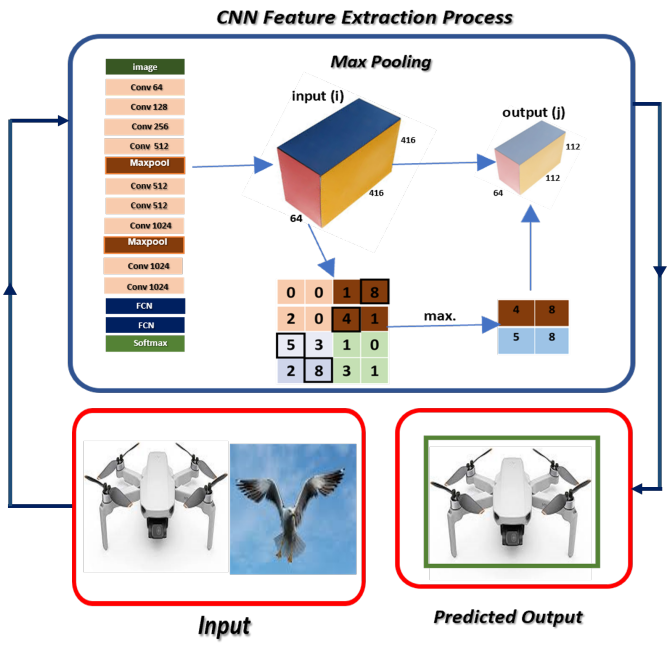

Fig. 1. Overview of System Design

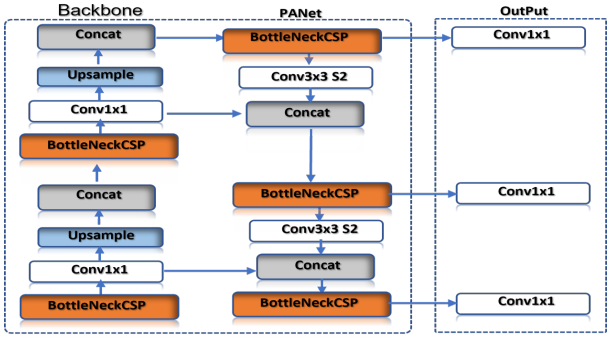

Fig. 2. Custom KITYOLO Drone Detector Model

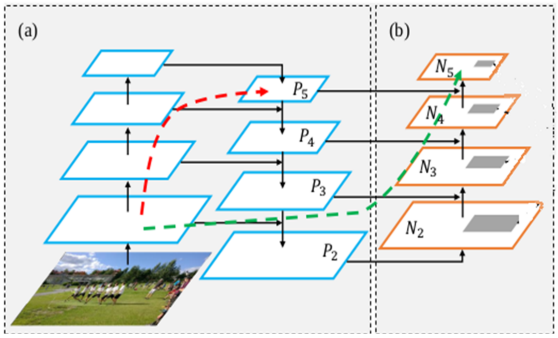

Fig. 3. Custom backbone PANet (a) FPN backbone used in YOLOv5 (b) Bottom-up Path Supplement

Due to the dynamic and flexible nature of the architecture of YOLOv5, the model can be augmented considering the relative challenge to be solved, hence, the addition of Path Aggregation Network (PANet) to the proposed model which is an integral part,so as to mitigate the sparsely and densely distribution of the nature of the target and accurately classify it from other objects in airborne.

As earlier stated, the proposed model, known as KITYOLO, is designed deploying the framework of YOLOv5 ; being the latest version of the YOLO series (v1, v2, v3 and v4). Though YOLOv5 is a very fast object detector that is capable of extracting 140 frames per second (fps) in real time, its major challenge is extracting features from densely distributed objects with optimal accuracy. Therefore, KITYOLO is designed to solve this inherent problem peculiar to YOLOv5, which is a vital issue in detecting and preventing drones in restricted areas. The disparity, similarity, and tininess of the targets (drones and birds) created the need for instance segmentation. That is, the need to explicitly detect, classify and localize various object instances in an image. Hence, the addition of Path Aggregation Network (PANet) to KITYOLO (which is missing in the standard YOLOv5), so as to enhance the propagation of low-level features, captured in Fig. 2; depicting an improved and better architecture.

As feature extraction takes place in the network, from high level to low level layers, the complexity of consecutive layers increases, leading to a corresponding decrease in spatial resolution. Fig. 3 explains the PANet adopted in KITYOLO and Feature Pyramid Network (FPN) used in the architectural design of YOLOv5. The FPN deployed in YOLOv5 follows a top-down path (Fig. 3(a)) integrating rich features from high level layers with accurate localization from lower level layers, by upsampling the layers. This approach follows a longer path which increases the network complexity as well as the latency, thereby reducing the model's accuracy when detecting very tiny objects. Unlike FPN, PANet deployed in the proposed model follows a bottom-up path approach (Fig. 3(b)), which reduces the number of paths as well as the network's complexity, having a resultant positive effect in the accurate detection of very tiny objects.The Neck compartment generally is responsible for feature aggregation and to improve the accurate localization of features in lower layers, leading to the overall object location accuracy.

The difference between drones and birds seems to fade off once the detection distance reaches or supersedes 100meters as in the case of this research, making both objects appear similar during detection when at flight. To detect and classify drones from birds, KITYOLO captures the entire image during run time using a single convolution network, making it capable of predicting objects of different classes based on confidence at a faster rate. The input image in the YOLO architecture is splitted into S X S number of grids, with each grid having B bounding boxes along side their confidence scores as displayed in Fig.4. Also, each bounding box is made up of 5 predictions (x,y,w,h and c); where x, y depicts the coordinates representing the center of the box of the grid cell, w, h representing the width and height of the grid and c the confidence prediction, representing the Intersection Of Union (IOU) between the

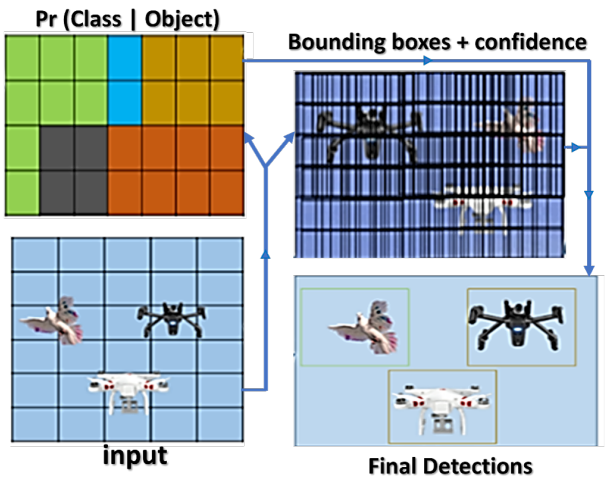

Fig. 4. Processing of Capturing Image value using bounding boxes

predicted box and ground truth box as shown in equation (1):

$$Box_{(cs)} = Pr_{(object)} \times IOU_{(b,\,object)}, \tag{1}$$

where $Box_{(cs)}$ is the box confidence score, $Pr_{(object)}$ is the probability of an object in the grid, and $IOU_{(b,\,object)}$ is Intersection of Union express as area of union of two boxes. For non-linearity in the network, YOLOv5 uses Softmax activation function to classify its multi-classes output. Softmax function returns the probability of each class using the given equation (2):

$$\sigma(Z^{\rightarrow})_i = \frac{e^{Z_i}}{\sum_{j=i}^{k} e^{Z_j}} \tag{2}$$

where σ is softmax, $(Z^{\rightarrow})_i$ is input vector, e^{Z_i} is the standard exponential function for input vector, k is the number of classes in the multi-class classifier, e^{Z_j} is the standard exponential function for output vector.

B. Dataset Capturing and Description

The dataset used for this research comprises drones and birds images. Two different drones; Mavic-Air and Mavic-Enterprise were separately flown in three different scenarios of *sunny, evening(gloomy)* and *cloudy* weather conditions. The videos of flown drones were captured at different time of the day to reflect their distinct scenario characteristics. Image frames were extracted from the video sequence of 1190 data frames from both drones, and labelled using Makesense software to generate ground truth values from the initial background values viz bounding boxes. The bird datasets of 950 images were gotten from Kaggle, a free online resource for datasets, but the images were manually labelled.

C. Simulation and Experimental Setup

The datasets of birds and drones were combined and distributed in a ratio of, 70% for training of the model. 20% for model testing and 10% for the validation of the model, so as to avoid overfitting and to achieve optimal performance. In addition, the simulation was done in a Python environment on a system configuration of Intel(R) Core(TM) i5-7400 CPU @ 3.00GHz, 8GB RAM, GPU Tesla K80. Several hyper-parameters such as best weights for weights initiation, input size of 416, batch size of 16 and learning rate with an epoch of 100 were used.

IV. Result Discussion and Performance Evaluation

To test the model's effectiveness in detecting and classifying drones from birds under different climatic conditions, metrics such as F1-score, precision, recall, number of frame per second (fps), and memory usage (GLOPS) were used to compare the proposed model with YOLOV5 model, as both models were trained and tested with the same dataset .

The result in Table I highlights the detection performance of KITYOLO and YOLOv5 models under different weather conditions and heights using a uniform dataset to prevent every form of bias.

TABLE I
Detection Results of Drone-Bird

Drone-Bird Detection				
Scenario	**KITYOLO (%)**		**YOLOv5 (%)**	
	Precision	**Recall**	**Precision**	**Recall**
Mavic_Enter_Cloudy	69.9	73.7	27.4	26.3
Mavic_Enter_Evening	57.5	60.0	47.0	90.0
Mavic_Enter_Sunny	92.9	90.0	90.5	95.5
Mavic_Air_Cloudy	96.1	1.00	95.6	1.00
Mavic_Air_Evening	58.5	42.4	26.1	15.0
Mavic_Air_Sunny	96.2	1.00	85.1	1.00
Bird	79.8	91.6	66.8	94.7

For drone-bird detection, the result from Table I indicates that KITYOLO has a superior precision value of 79.8% than YOLOv5 which is 66.8%. Across weather conditions, KITYOLO had a higher recall of 73.7% as against 26.3% of YOLOv5 in a cloudy weather. Also for evening, KITYOLO had a better precision and recall values of 58.5% and 42.4% than 26.1% and 15.0% of YOLOv5. Lastly, in sunny condition, a 96.2% precision value by KITYOLO affirms its detection capability than the 85.1% value of YOLOv5.

A. Performance Evaluation

The result in Table II highlights the comparison of KITYOLO with YOLOv5 in terms of speed, rationality of detection (F1-score), and memory usage (GFLOPS). F1-score is a test

TABLE II
Models Performance Evaluation

Performance of Models for Weapon Detection			
Models	**F1Score (%)**	**Time (FPS)**	**GFLOPS**
KITYOLO	**98.0**	**0.022s**	**16.4**
YOLOv5s	91.9	0.022s	16.4

of the behaviour of model with changes in its precision and recall expressed as:

$$\hookrightarrow F1score = \frac{2(Precision \times Recall)}{Precision + Recall}, \tag{3}$$

From Table II and Fig. 5, it can be clearly seen that KITYOLO achieved a superior detection performance of 98% than YOLOv5 of 91.9%; which is a significant 6.1% increase in detection rationality despite that the two models had the same time of detecting each object per second (0.022s) and 16.4 GFLOPS.

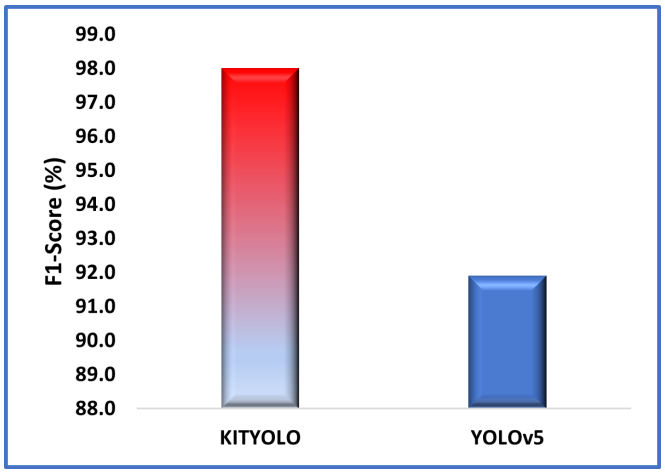

Fig. 5. F1-score Performance Comparison of KITYOLO and YOLOv5s

The confusion matrix in Fig. 6 indicates that KITYOLO can not only accurately detect different types of drones under different weather conditions, but also differentiate it from all kinds of birds in a timely manner and with less computational complexity and minimal false alarm rate. The images on Fig. 7 are samples of drones and birds detection by KITYOLO under different weather conditions and heights. The displayed results are detection and classification tasks carried out concurrently by the proposed model showing degree of accuracy and sensitivity.

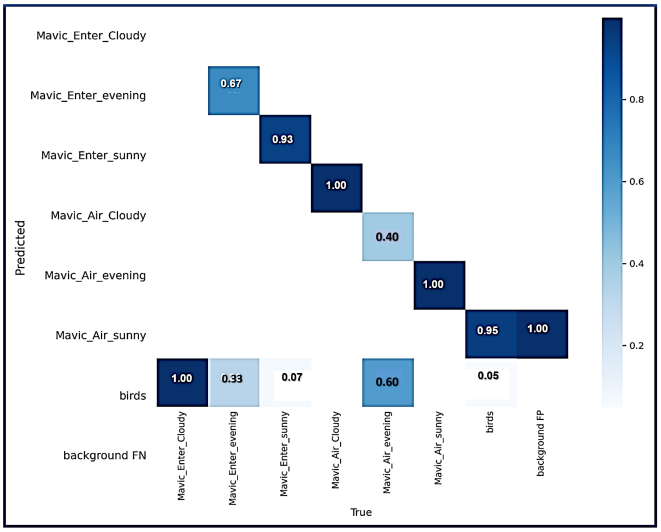

Fig. 6. Confusion Matrix of KITYOLO

These results show a high detection improvement by the proposed model in comparison with YOLOv5 in detecting tiny objects under different weather conditions in a timely manner and less resource usage. However, a closer look at the results in Table I indicates a drop in the detection performance during evening/gloomy conditions which is an ongoing research challenge in computer vision.

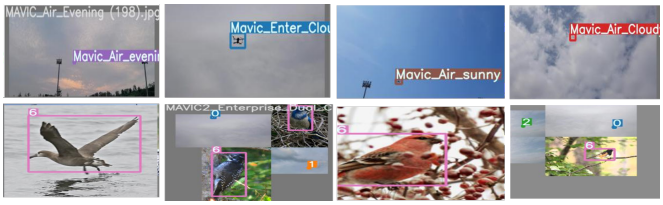

Fig. 7. Samples of drone-bird detection by KITYOLO

V. Conclusion

This work presents a novel drone detection model; KITYOLO that improved the accuracy and precision of tiny objects in different weather condition while maintaining timeliness and computational complexity. In the future, we hope to increase our dataset and improve the model for robust performance.

Acknowledgment

This research work was supported by Priority Research Centers Program through NRF funded by MEST (2018R1A6A1A03024003) and the Grand Information Technology Research Center support program (IITP-2021-2020-0-01612) supervised by the IITP by MSIT, Korea.

References

[1] S. Al-Emadi and F. Al-Senaid, "Drone Detection Approach Based on Radio-Frequency Using Convolutional Neural Network," in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, 2020, pp. 29–34.

[2] S. Ajakwe, R. Arkter, D. Kim, and J.-M. Lee, "Lightweight cnn model for detection of unauthorized uav in military reconnaissance operations," in *2021 Korean Institute of Communication and Sciences Fall Conference*. (KICS), 11 2021.

[3] M. Z. Anwar, Z. Kaleem, and A. Jamalipour, "Machine Learning Inspired Sound-Based Amateur Drone Detection for Public Safety Applications," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2526–2534, 2019.

[4] R. H. Geschke, A. Shoykhetbrod, R. Brauns, C. Schwäbig, S. Wickmann, S. Leuchs, C. Krebs, A. Küter, and D. Nüssler, "Post-integration Antenna Characterisation for a V-band Drone-detection Radar," in *2021 15th European Conference on Antennas and Propagation (EuCAP)*, 2021, pp. 1–4.

[5] H. M. Oh, H. Lee, and M. Y. Kim, "Comparing Convolutional Neural Network(CNN) models for machine learning-based drone and bird classification of anti-drone system," in *2019 19th International Conference on Control, Automation and Systems (ICCAS)*, 2019, pp. 87–90.

[6] P. Gajalakshmi, J. V. Satyanarayana, G. V. Reddy, and S. Dhavale, "Detection of Strategic Targets of Interest in Satellite Images using YOLO," in *2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP)*, 2020, pp. 1–5.

[7] 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," in *2018 3rd International Conference on Computer and Communication Systems (ICCCS)*, 2018, pp. 192–195.

[8] J. Harikrishnan, A. Sudarsan, A. Sadashiv, and R. A. Ajai, "Vision-face recognition attendance monitoring system for surveillance using deep learning technology and computer vision," in *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*, 2019, pp. 1–5.

[9] D. K. Behera and A. Bazil Raj, "Drone Detection and Classification using Deep Learning," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020, pp. 1012–1016.

[10] S. Ajakwe, R. Arkter, D. Kim, G. Mohatsin, D. Kim, and J.-M. Lee, "Anti-drone systems design: Safeguarding airspace through real-time trustworthy ai paradigm," in *The 2nd Korea Artificial Intelligence Conference*. (KAIC), 09 2021.

[11] D. T. Wei Xun, Y. L. Lim, and S. Srigrarom, "Drone detection using YOLOv3 with transfer learning on NVIDIA Jetson TX2," in *2021 Second International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP)*, 2021, pp. 1–6.

[12] S. A. Hassan, T. Rahim, and S. Y. Shin, "Real-time UAV Detection based on Deep Learning Network," in *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, 2019, pp. 630–632.

[13] X. Wang, W. Li, W. Guo, and K. Cao, "SPB-YOLO: An Efficient Real-Time Detector For Unmanned Aerial Vehicle Images," in *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2021, pp. 099–104.

[14] K. Luo, R. Luo, and Y. Zhou, "Uav detection based on rainy environment," in *2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, vol. 4, 2021, pp. 1207–1210.