



ORIGINAL RESEARCH ARTICLE

MODELLING OF A REAL-TIME AERIAL SURVEILLANCE SYSTEM FOR QUADCOPTER APPLICATION USING MACHINE VISION

K. Okhiria¹, J.D. Jiya², I.S. Sintali³

¹Department of Mechanical Engineering, Nigerian Army University Biu, Borno State

²Department of Mechatronics and Systems Engineering, ATBU, Bauchi State

³Department of Mechanical Engineering, ATBU, Bauchi State

Corresponding author's email: okhiria.kenneth@naub.edu.ng

ARTICLE INFORMATION

Received: 25th March 2025

Revised: 2nd May 2025

Accepted: 2nd May 2025

Keywords:

Convolutional neural network

YOLO

Quadcopter

Dataset

Surveillance

ABSTRACT

Unmanned Aerial Vehicles (UAVs) have gained significant traction for real-time surveillance applications such as object tracking and search-and-rescue missions. A key enabler of these capabilities is the integration of a real-time, highly accurate object detection system. While available two-stage detectors like R-CNN, Fast R-CNN, Faster R-CNN, and Mask R-CNN achieve high detection accuracy and precise localization by dividing the image into regions and classifying each region, they are often slow, complex, and resource-intensive, making them unsuitable for real-time applications. To address this limitation, this research utilizes a Darknet-based YOLOv3 (You Only Look Once, Version 3), a state-of-the-art algorithm for real-time object detection in videos, live feeds, and images. Leveraging a deep Convolutional Neural Network (CNN), YOLOv3 efficiently learns features and predicts object locations and class probabilities in a single pass, ensuring high-speed and reliable detection. A real-time aerial surveillance system for quadcopter application using machine vision is proposed. The training dataset, obtained from the internet and self-taken images from a camera, was manually annotated into three critical categories: suspect, bandit, and weapon. The dataset was subsequently divided into training and testing subsets. Experimental results demonstrate that the proposed system achieves outstanding detection accuracy, with an overall mean average precision (mAP@0.5) of 93.82%, precision of 94%, and recall of 83%. Compared to a ResNet-50-based Faster R-CNN model, the YOLOv3-based approach outperformed, achieving success rates of 0.81 for the Bandit class and 1.00 for both Suspect and Weapon classes. This research improves drone-based AI for real-time object detection in security systems, enhancing efficiency and adaptability. It also creates specialized datasets, including a tailored bandit dataset, to support future UAV security research.

© 2025 Faculty of Engineering, University of Maiduguri, Nigeria. All rights reserved.

1.0 Introduction

In past years, the evolution of unmanned aerial vehicles has experienced substantial growth. Recently, these UAVs have proven to be a promising solution for a diverse range of applications, including firefighting, mapping, reconnaissance, disaster relief, and search and rescue operations (Rendón & Martins, 2017). Other civilian applications include Inspection and search activities, aerial photography of landscapes, precise agricultural practices, traffic and crowd control, observing meteorological conditions, and providing emergency healthcare and parcel delivery services. These are feasible due to the current design of UAVs. UAVs are capable of autonomous navigation, operating beyond the line of sight, particularly in challenging terrains, and even in adverse weather conditions (Rendón & Martins, 2017;Janarathanan et al., 2019;Dilshad et al., 2020). The most famous UAV, that is being used for social and security applications, is the quadrotor, compared to other rotorcraft, The geometric symmetry of quadcopters simplifies modeling and controller design. Additionally, their unique maneuverability and capability for vertical take-off and landing (VTOL) render them an edge over other UAVs, Utilizing a quadcopter allows for easy initiation of surveillance in locations where traditional security cameras cannot be installed, (Pashaei et al., 2020);(Iqbal et al., 2021). Face recognition is a method

utilized to verify or identify an individual's identity by examining and correlating patterns derived from their facial characteristics. Image recognition and image localization in environmental monitoring have emerged as popular and highly studied topics in the field of computer vision. The Convolutional Neural Network (CNN) model stands out as the predominant image processing technique utilized for these purposes. CNN does three tasks: feature extraction by convolution layer, feature selection by pooling layer and classification by fully connected layer. A convolutional neural network, a prominent deep learning technology, is commonly applied in image recognition tasks. An extension of object classification is object detection, aimed at recognizing the object in the image. Object detection detects all occurrences of pre-conceived classes and draws bounding boxes to provide localization of the image. This study employs a quadcopter for image acquisition and utilizes the YOLOv3 algorithm for object classification, localization, and detection aiming to facilitate environmental monitoring. Typically, investigations into image recognition and localization technology involve identifying possible targets and drawing insightful conclusions from the acquired image data (Pashaei *et al.*, 2020; Zhao *et al.*, 2018; Lin *et al.*, 2021; Zaidi *et al.*, 2022).

YOLO ("You Only Look Once") is a single-stage detector that uses a fully convolutional neural network to predict anchor boxes and class probabilities in a single pass, without extracting region proposals like F-RCNN (Ammar *et al.*, 2020). It divides the image into grid cells to predict object locations and classifications. YOLOv3, featuring the Darknet-53 backbone, improved upon its predecessors. It is a real-time object detection algorithm that employs a single convolutional layer that spans the entire image. It divides the image into grid cells, and each grid cell predicts a certain quantity of bounding or anchor boxes around objects, assigning high scores to those associated with predefined classes. Every bounding box is assigned a confidence score representing the estimated accuracy of that particular prediction, and each box is designed to detect only one object. YOLOv3 demonstrates speed and accuracy in relation to mean average precision (mAP) and intersection over union (IOU) values. Moreover, it undergoes pre-training on the COCO dataset. (Chen *et al.*, 2021; Bao *et al.*, 2021; Ammar *et al.*, 2020).

Recently, drones have found increased acceptability for real-time surveillance applications. An object detection device is an important component of a UAV for air-to-ground surveillance, for it to be employed for tasks such as object tracking, search, and rescue, it requires real-time data processing and must have a high detection accuracy. Furthermore, available two-staged detectors like R-CNN, fast-RCNN, faster-RCNN, and Mask R-CNN divide the image into regions, classify each region then combine them. They have the characteristics of having high detection accuracy, precise localization, and improved feature extraction ability. But they are often slow, complex, and resource intensive thus, not suited for real-time object detection (Li *et al.*, 2018). Single-staged detectors are faster, more accurate, and simpler object detection methods predicting the object in an input image directly in one pass through the network. When compared to other detection algorithms, YOLOv3, a single-stage object detector provides high performance in terms of accuracy and speed when employed for object detection (Ammar *et al.*, 2020); (Bao *et al.*, 2021). To overcome these issues, this research aims to model a Real-Time Aerial Surveillance System for Quadcopter Applications Using Machine Vision which involves employing deep learning techniques like YOLOv3 to automatically retrieve information from an image. This extracted information can range from a basic signal to a detailed set of data, encompassing aspects like the identity, position, and orientation of objects within the image. The goal is to create a quadcopter-based solution for surveillance missions, capturing real-time data and using YOLOv3 to classify, localize, and detect relevant objects in the image.

Convolutional neural networks have long been used for target tracking using UAVs, (Sun *et al.*, 2022) proposed a transformer-based aerial tracking framework combining spatial and temporal dimensions to solve challenging problems in aerial tracking like target loss, misclassification, and imprecise boundary boxes. Security surveillance using quadcopters has been shown by (Iqbal *et al.*, 2021) to be effective with deep learning techniques, they proposed a deep learning-based security and surveillance method and real-time transmission of data to the control room by a Comparative analysis of the built model with four pretrained CNNs; squeeZNet, GoogleNet, ResNet-50, and ResNet-18. It consists of four phases, collection of the dataset, preprocessing phase to reduce image features, training phase for model building, and query image entity prediction. ResNet-50 interfaced with Faster-RCNN model performed better than models built from other feature extractors. Deep learning have been also been employed for drone detection and classification, (Behera & Raj, 2020) used YOLOv3 as an object detector, CNN for getting features from images and binary cross-entropy loss function for class prediction. Drone technology using wireless sensor networks (WSN) and the Internet of Things (IoT) have been employed in border security, (Sharma *et al.*, 2021) used TensorFlow and a face recognition model to classify persons and animals. The intrusion coordinates are sent to a nearby patrol team and the drone surveillance system is activated and sent to the intrusion-detected area to track the intruder using the GPS coordinates provided by the intrusion-detection system. to reduce the response time to threat, Quadcopters in combination with Artificial Intelligence has recently been used for facial

recognition and surveillance, (Morbale et al., 2022) used Raspberry Pi module and Pi camera mounted on the UAV to capture image of an intruder in real time. The use of quadcopter with CNN based detection model for real time surveillance is recently gaining ground in the engineering field, (Abdou & Aziz, 2021) leveraged on the use of AI and UAV technology for a real-time human and cars detection and tracking system using a CNN object detection algorithm YOLOv4 and a multi-tracking algorithm DeepSORT. The system consists of four (4) stages; object detection, object tracking, object localization and drone control. (Kim & Cho, 2021) employed Region based convolutional neural networks for object detection. The system includes a pre-processing filter (PF) to create a newly combined color image, a Faster R-CNN for detection, and a histogram network (HN) to estimate the distance between the UAV and the object. Drones have also been employed for search and rescue in time of natural disaster, (Mishra et al., 2020) proposed a novel dataset of aerial recognition for humans in need of help during disaster situations and search and rescue operation for disaster management application, and a new identification classification model. Pre-processing of dataset involved frame extraction, frame selection and action annotation of image frame. The detailed review of various security and surveillance methods highlights the potential of this field to achieve significant advancements. Modern deep learning techniques can improve surveillance mechanisms. Therefore, this study introduces a technique for detecting bandits, weapons, and suspects using advanced deep learning model-Yolov3.

2. Materials and Method

This section deals with the methodology used in the development and implementation process. The schematic representation outlining the methodology is shown in figure 1. The system's framework includes the subsequent key stages: First, there's the gathering and collection of a dataset containing instances of the entities intended for detection. Subsequently, the preprocessing stage converts the images into condensed features to enable efficient processing. The Training Phase involves training and the construction of the model. The fourth step entails predicting entities in a query image, where the image is captured by the camera, processed, and fed into the trained model to anticipate potential matches.

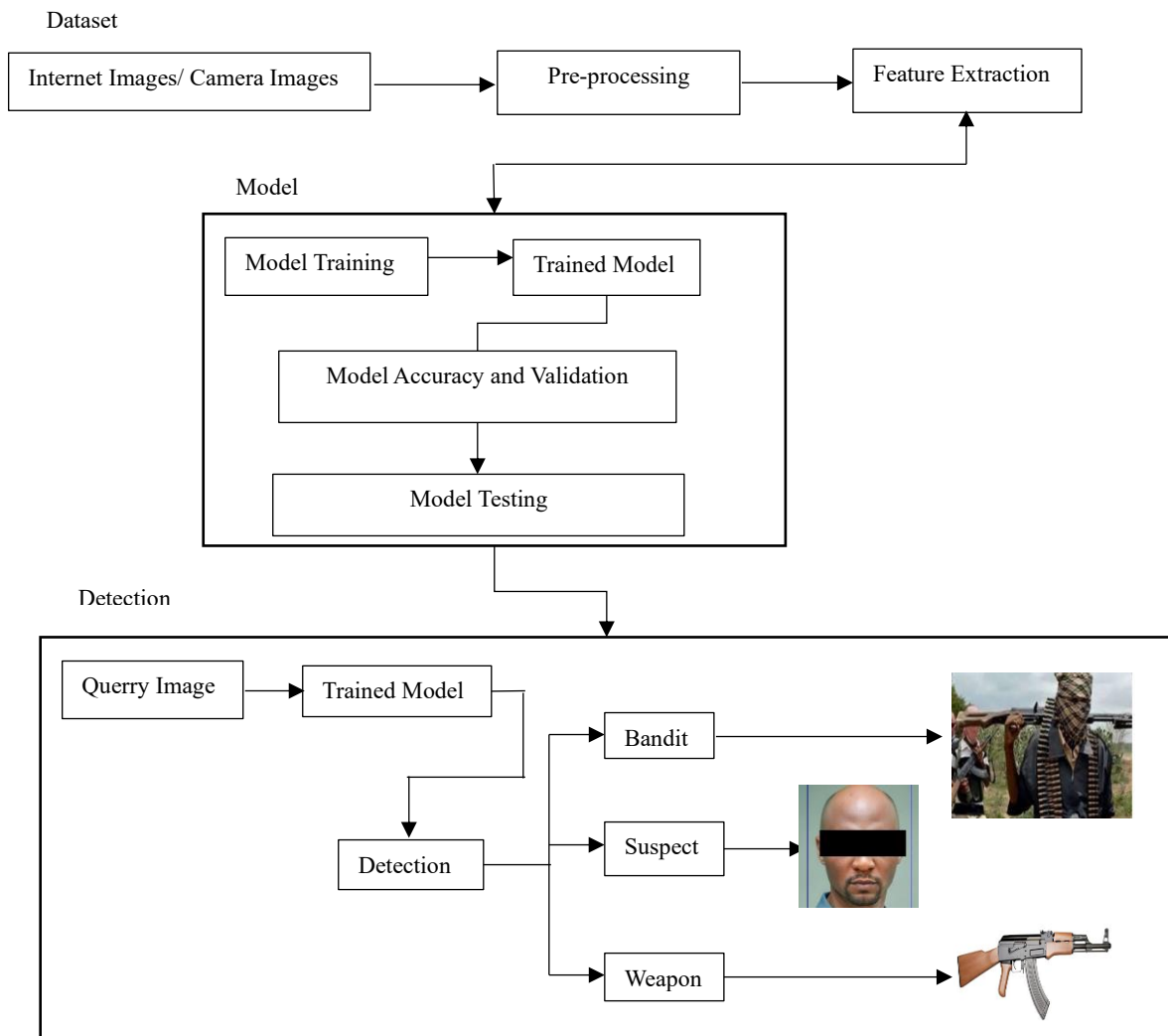


Figure 1: Block Diagram of Proposed Methodology

2.1 Data Acquisition

During data acquisition, Bandit images were collected online, weapon data came from Kaggle dataset from the internet, and suspect images were taken with a camera. Images in the dataset were classified as bandit, suspect, and weapon. A yolov3 model, aimed at detection was trained using images sourced from the created dataset.

2.2 Preprocessing of Images in the Dataset

Image processing is the steps taken to format images before they are used by the model for training and inference. The key steps include resizing the data set image frames, extracting image features, and data augmentation.

2.2.1 Resizing of Image Frames and Data Set Construction

Utilizing an image with a high pixel count is computationally demanding for modeling and can potentially slow the learning process. Consequently, all extracted images were resized automatically by the yolo network to dimensions of $416 * 416$ pixels. YOLOv3 relies on Darknet-53 as its feature extraction backbone. This network features 53 convolutional layers combined with batch normalization and Leaky ReLU activation functions. It also incorporates residual connections to enhance information flow and support the training of deeper networks. The presence of a dataset is a crucial step in assessing algorithm quality. Therefore, a dataset of 2000 images was created, comprising 500 images for WWeapon, 1000 for Bandit, and 500 for Suspect object classification.

2.2.2 Data Augmentation

The data augmentation techniques include *Grey Scale*; RGB images are computationally expensive and dimensionally large hence the images used for feature extraction were converted to grey scale. *Rotation*; Some of the images captured by the camera were rotated horizontally by 90° , some rotated vertically (upside down) while others were flipped. This gives variations in the dataset and further improve the training process. *Contrast adjustment*: the contrast was adjusted for the images in the training data set, this improves the appearance of the images for visual interpretation and increase the performance of the model for subsequent task.

2.2.3 Feature Extraction from Image Frames

For feature extraction, a transfer learning approach was employed. Transfer learning enables the incorporation of prior knowledge through a pre-trained model, which can subsequently be adjusted for a particular task with a smaller dataset. This approach leads to quicker convergence, enhanced accuracy, and diminished overfitting. In this research, the proficient Darknet-53 model, pre-trained on the ImageNet dataset, was employed for precise feature extraction. Leveraging this pre-trained model as a foundation streamlined the training process and enhance accuracy, saving valuable time (Shah et al., 2023). Darknet-53 architecture is shown in figure 2 (Ma et al., 2020).

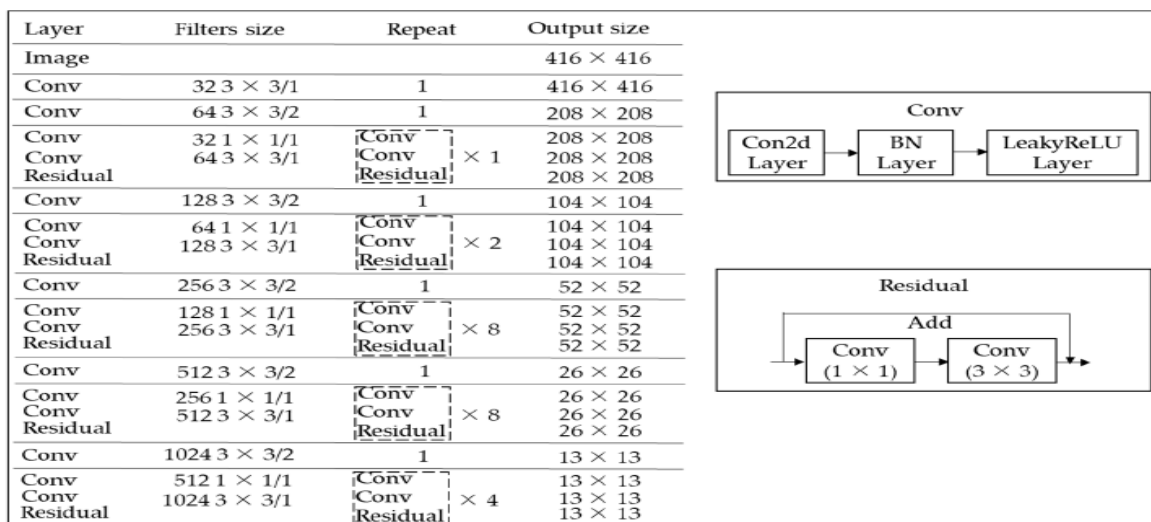


Figure 2: Darknet-53 Architecture (Ma et al., 2020)

2.3 YOLOv3

Features from the image frames, their related object’s location and class identification are needed to be learned by some CNN-based machine learning algorithm. In this research, YOLOv3 object detection algorithm known for its speed and accuracy and well suited for real-time applications was employed for object classification, localization and detection. YOLOv3 is trained on labeled datasets to learn how visual features correspond to specific classes. It predicts bounding boxes by outputting the object’s center coordinates, width, and height, allowing precise localization. The model combines object classification and localization, providing both the object class and a confidence score for each detection. The basic working of YOLOv3 is shown in figure 3 as obtained from (Aswini *et al.*, 2022).

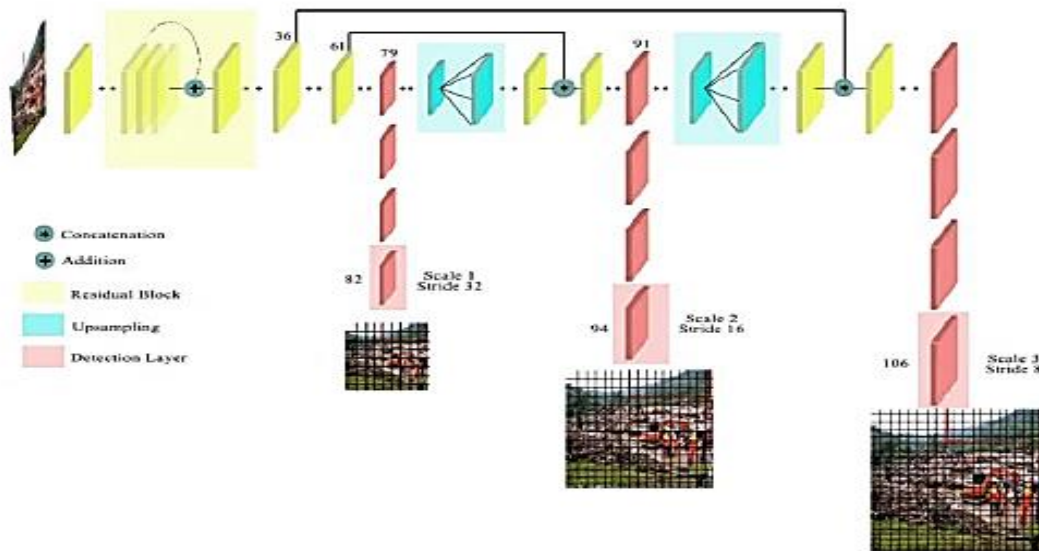


Figure 3: YOLOv3 Structure (Aswini *et al.*, 2022)

The algorithm employs a single convolutional layer across the image, dividing it into segments to predict bounding boxes and probabilities. It features 106 convolutional layers, with 53 pre-trained for feature extraction and 53 for detection and training. Detection occurs at three layers: the 82nd (stride 32, 13x13 feature map), 94th (stride 16, 26x26 feature map), and 106th (stride 8, 52x52 feature map). Residual layers enable skip connections, and the image is down sampled to the 82nd layer and up sampled to the 106th. YOLOv3 uses 1x1 kernels at these layers to ensure consistent feature map dimensions.

2.4 Training

A total of 2000 images were prepared as experimental samples for the data set. These was divided into a training set, and test set in the ratio 80:20 which results to 1600 training, and 400 testing images respectively as shown in Table 1. The training data was used for the actual training, and the test set was used to test the accuracy of the model after the training most have been completed.

Table 1: Data Set Distribution

Classification	Training Set	Testing Set
Weapon	400	100
Bandit	800	200
Suspect	400	100

YOLOv3 was used and trained on Google Colab that offers free GPU. The model was trained for 192 epochs, with a batch size of 64, and a learning rate of 1×10^{-3} the parameter configuration is shown in Table 2. For detection of the image, query videos obtained by using camera mounted on a quadcopter and trained models were given for processing. The system detects the images and show the information and classification of that image.

To train the model the images in the dataset were labelled to specify their coordinates and boundaries for the machine learning model to recognize and predict. MAKESENSE.AI a user-friendly annotation tool for computer vision tasks was used to draw bounding box on all the images on the dataset pointing to the region of interest within a given image. The dataset used for the training was divided into two groups 80% training set and 20% test set and Google Collaboratory used for the training. Darknet an open-source neural network was downloaded to create a darknet network for training the detector. Within the darknet folder, changes were made to the make file to enable GPU and OpenCV and build the darknet. The YOLO configuration file, which includes details about the network, convolutional layers, three YOLO detection layers, and additional layers along with their properties was utilized. In the darknet configuration (cfg) folder exist the YOLO configuration files containing details about the model’s architecture, changes made to the cfg file include: Maximum batch size = classes × 2000 which determines the maximum number of iterations for which the network will be trained, Filter = (classes + 5) × 3 = 24 for 3 classes, and Steps = 80% and 90% of the maximum batches.

Table 2: Training Parameters

<i>Related Parameter</i>	<i>Value</i>	<i>Meaning</i>
Batch size	64	Number of pictures per training
Learning rate	1×10^{-3}	Initial learning rate
Epoch	192	Cycle through training dataset
Number of iterations	6000	iterations

3. Results and Discussion

Pretrained weight was downloaded for darknet model and the model was trained for 192 epochs, tested and built. The average precision for each class (Bandit, Suspect and Weapon) as well as the overall performance metrics for the model was obtained and the result obtained after the training is as recorded in Tables 3 and 4. The model loss graph for the object detector or the graph of loss function for detection task is shown in Figure 4. The dataset was further trained for 50 and 150 epochs on the Roboflow platform. Roboflow is a computer vision platform designed to help users create computer vision models more quickly and accurately by offering enhanced tools for data collection, preprocessing, and model training. It was used to adjust orientations of the images, resize images, alter image contrast, and apply data augmentation. Additionally, it was employed for training of the model for 50 and 150 epochs respectively. The performance metric at 50 and 150 epochs are as shown in Table 5.

Table 3: Detailed Results of Training

<i>Object detector</i>	<i>Class</i>	<i>No of training images</i>	<i>Average precision</i>	<i>No of test images</i>	<i>Epoch</i>
Darknet-based YOLOv3	Bandit	800	0.81	200	192
	Suspect	400	1.00	100	192
	Weapon	400	1.00	100	192

Table 4: Performance Metrics

<i>Object detector</i>	<i>Precision</i>	<i>recall</i>	<i>F1-score</i>	<i>Average IoU</i>	<i>mAP@0.50</i>	<i>Detection time</i>	<i>Epoch</i>
Darknet-based YOLOv3	0.94	0.83	0.87	74.69%	93.82%	09 sec	192

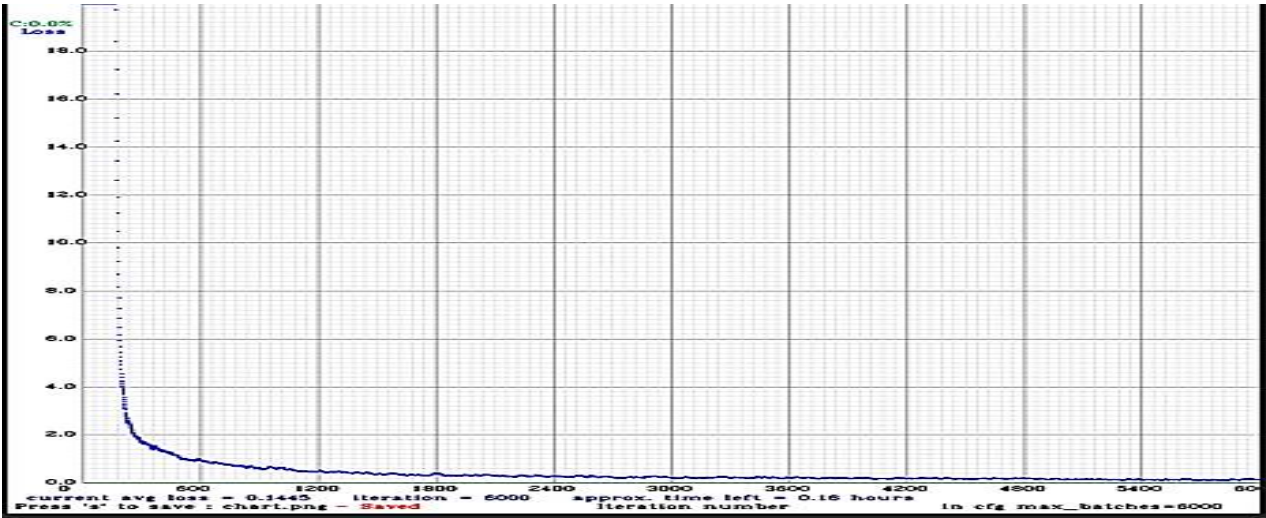


Figure 4: Model Loss Graph

Table 5: Performance Metrics @ 50 And 150 Epochs

Object detector	Epochs	Precision	Recall	mAP@0.5
Darknet-based YOLOv3	50	0.85	0.88	89.2%
	150	0.91	0.86	91.2%

The object detection inference process handles tasks such as image classification and localization. Thus, to test the model for accurate prediction of the classes, inference was carried out by given images which are not part of the dataset to the model for detection and classification, the model accurately classified the images based on their respective classes. The results are shown in Plates 1, 2 and 3.



Plate 1: Bandit Detection



Plate 2: Weapon Detection



Plate 3: Suspect Detection

Images which are not part of the dataset were also given to the model for validation, the model accurately classified the images based on their respective classes. The result is shown in plates 4, 5, and 6.



Plate 4: Suspect Detection



Plate 5: Weapon Detection



Plate 6: Bandit Detection

Comparison based on average precision for each class was made with the result of the study by (Iqbal et al., 2021). Table 6 shows the result for the comparison.

Table 6: Comparison of Result At 150 Epochs

Object detector	Class	No of Training images	No of Test images	AP@150 epoch
<i>restNet-50 based faster-RCNN Iqbal et al., (2021)</i>	Face	1805	695	0.98
	Weapon	1700	550	0.97
	intruder	1550	500	0.85
<i>Darknet-based YOLOv3 (proposed method)</i>	Suspect	400	100	1.00
	Weapon	400	100	1.00
	bandit	800	200	0.70

In this study, a deep-learning-based artificial intelligence model which can detect Bandit, Suspect and Weapon and also distinguish which of the 3 different classes a detected object belongs to was developed for use in autonomous security control. After training at 192 epochs, the average precision obtained for each class was 0.81 for Bandit, 1.00 for Suspect and Weapon respectively. An average precision of 1.00 and 0.81 indicates the model is performing detection accurately. The result obtained for performance metrics for the model indicate a high IoU of 74.69% which indicates the predicted bounding box coordinates closely resembles the ground truth box coordinates. A mean average precision ($mAP@0.5$) of 93.82% (0.9382) and a precision of 0.94 shows a very high precision and prediction ability by the model. With a recall of 0.83, the model is able to find true positive out of all the prediction with high rate of satisfaction. The model loss graph or the graph of the loss function for detection tasks shown in figure 4 is the sum of the loss functions for bounding box, objectiveness score and class prediction. It is seen that the model converges hence the loss or error reaches

Corresponding author's email address: okhiria.kenneth@naub.edu.ng

a minimum level of acceptable error reaching a point of stability for the model to make accurate predictions. Yolo algorithm possesses a fast detection rate which is seen in the detection time of 9 seconds for the model.

(Iqbal et al., 2021) studied a real-time surveillance using ResNet-50-based Faster-RCNN to detect face, weapons and intruder. Their results showed an average precision (AP) of 0.98, 0.97, and 0.85 for face, weapon, and intruder respectively at 150 epochs. The proposed model had an AP of 1.00 for both suspect (face) and weapon, and 0.70 for bandit (intruder) at 150 epochs. The model proposed in this research used a Darknet-based Yolov3 algorithm to detect suspect (face), weapon and bandit (intruder). From the result obtained, the proposed model has a higher AP for suspect (face), and weapon but showed a slightly lower value for class Bandit, this occurs because a bandit can be masked or bare faced, and carries different kinds of weapons which was tricky for the learning algorithm to detect. The proposed method used a lower epoch and less resources to achieve a better performance. The AP (93.82%) recall (83%), and precision (94%) of the proposed model has an excellent success rate. To further compare the result obtained by (Iqbal et al., 2021) training their model for 150 epochs the proposed model was trained for same epochs and result shown in table 6 indicates the propose model has an excellent success rate and higher precision as compared to the study in the literature. In this study, the importance of the proposed model is increased due to the fact that it is trained and tested on computers that are less expensive and do not have sufficient infrastructure. Unlike the study by Iqbal et al., (2021), which utilized 6,800 images, the proposed model achieved strong results using only 2,000 images. This efficiency is made possible by YOLOv3's backbone, Darknet-53, which is pretrained on the extensive ImageNet dataset, enabling effective feature extraction with fewer training samples.

3.1 Test

The object detection model was effectively deployed, integrated into a Python script, and implemented on a laptop to process video frames captured by its camera in real time. The model accurately detected and classified objects within the live video stream from the laptop's camera. The test results along with their confidence scores are presented in Plates 7, 8, and 9.

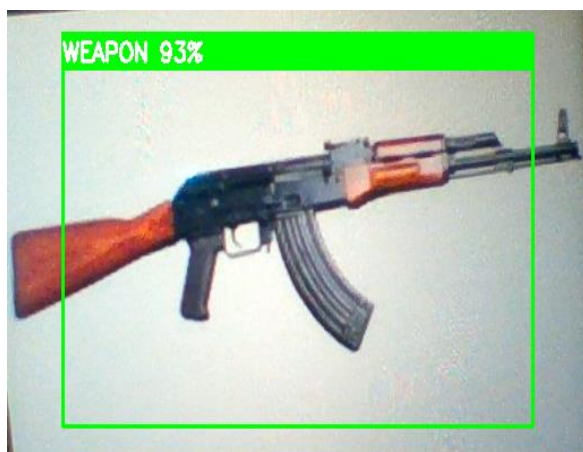


Plate 7: Weapon Detection



Plate 8: Bandit Detection



Plate 9: Suspect Detection

A confidence score of 78% for bandit, 93% for weapon and 80% for suspect shows high detection accuracy of the proposed model.

3.2 Limitation

The algorithm is designed to detect only one specific type of firearm. Given the complexity and performance demands of developing an algorithm capable of identifying multiple firearm types, such as rifles, grenade launchers, pistols, and handguns, this research focuses on implementing an algorithm specifically for detecting AK47 firearms. The algorithm can only detect firearms that are in plain sight and not hidden by the person carrying them. The algorithm sometimes fails to detect weapon accurately, this is because the dataset had less variation for weapon and often bandits carry same AK47 rifle, which in this case seems tricky for the learning algorithm.

To address any bias that may occur in the labelling of images as suspect and bandit, A suspect is defined as someone involved in a crime but not yet proven guilty, often under investigation based on evidence or suspicion. Hence it is assumed the security forces have the image (facial image) of the suspect in their database thus, the study focused on only facial recognition for a suspect. A bandit is someone who commits crimes like robbery, often violently or in an organized manner, including acts like killing and destruction of properties. The study focused on the nature of banditry in Nigeria and their visual features like clothing, accessories, mask and gun possessed, thus, the model might face challenges and perform poorly when deployed to detect similar activities in other parts of the world.

4. Conclusions

This study demonstrates the effectiveness of a deep learning-based object detection model for real-time autonomous security surveillance using the YOLOv3 algorithm. Trained on a custom dataset, the model successfully classified and localized three critical object classes Bandit, Suspect, and Weapon with high average precision values of 0.81, 1.00, and 1.00 respectively. Achieving a mean average precision of 93.82%, precision of 94%, and recall of 83%, the system proved not only accurate but also efficient, with fast inference time and minimal computational overhead. Compared to a ResNet-50-based Faster R-CNN model, the proposed system maintained competitive or superior detection performance, particularly for the Suspect and Weapon classes, even with fewer training epochs and limited hardware resources. These results affirm the model's practical applicability in real-world UAV surveillance scenarios, especially in low-resource environments.

Future research will aim to enhance the model by enabling detection of a wider variety of threats, including multiple types of weapons, and by improving its effectiveness in challenging conditions such as occlusion and complex and dynamic environments. Efforts will also explore upgrading to more advanced architectures like YOLOv5 or YOLOv8 and integrating thermal or infrared imaging to improve the detection of hidden weapons. Detection Accuracy for AK47 Firearms can be improved by Increasing the dataset's variation to include different AK47 variants to enhance dataset diversity and improve detection accuracy.

This research advances automated security and surveillance systems by providing a foundation for drone-based AI solutions in real-time object detection. It enhances the efficiency of security operations, addresses critical security gaps, and offers adaptability to various scenarios. Additionally, the study contributes to the creation of specialized aerial surveillance datasets, such as a tailored bandit dataset, serving as a valuable resource for future research and development in UAV-based security solutions.

References

- Abdou, AAA. and Aziz, NAA. 2021. CNN-based Off-board Computation for Real-time Object Detection and Tracking Using a Drone. *Journal of Tomography System and Sensor*, 4(2): 11–20.
- Ammar, A., Koubaa, A., Ahmed, M. and Saad, A. 2020. *Aerial Images Processing for Car Detection using Convolutional Neural Networks: Comparison between Faster R-CNN and YoloV3*. 2012.
- Aswini, N., Uma, SV. and Akhilesh, V. 2022. Drone to Obstacle Distance Estimation Using YOLO V3 Network and Mathematical Principles. *Journal of Physics: Conference Series*, 2161(1).
- Bao, W., Huang, M., Zhang, Y., Xu, Y., Liu, X. and Xiang, X. 2021. Boosting ship detection in SAR images with complementary pretraining techniques. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, PP (8), 1.

- Behera, DK. and Raj, AB. 2020. Drone Detection and Classification using Deep Learning. *Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020, Iccics*, 1012–1016.
- Chen, W., Huang, H., Peng, S., Zhou, C. and Zhang, C. 2021. YOLO-face: a real-time face detector. *Visual Computer*, 37(4): 805–813. <https://doi.org/10.1007/s00371-020-01831-7>
- Dilshad, N., Hwang, JY., Song, JS. and Sung, NM. 2020. Applications and Challenges in Video Surveillance via Drone: A Brief Survey. *International Conference on ICT Convergence*, 728–732.
- Iqbal, MJ., Iqbal, MM., Ahmad, I., Alassafi, MO., Alfakeeh, AS. and Alhomoud, A. 2021. Real-Time Surveillance Using Deep Learning. *Security and Communication Networks*, 2021.
- Janarthanan, A., Ho, HW., Gopal, L. and Wong, VSWK. 2019. *An Unmanned Aerial Vehicle Framework Design for Autonomous Flight Path*. 0–4.
- Kim, J. and Cho, J. 2021. Rgdinet: Efficient onboard object detection with faster r-cnn for air-to-ground surveillance. *Sensors*, 21(5): 1–16.
- Li, S., Zhang, W., Li, G., Su, L. and Huang, Q. 2018. Vehicle Detection in UAV Traffic Video Based on Convolution Neural Network. *Proceedings - IEEE 1st Conference on Multimedia Information Processing and Retrieval, MIPR 2018*, 1–6.
- Lin, C., Zhang, W. and Shi, J. 2021. *Tracking Strategy of Unmanned Aerial Vehicle for Tracking Moving Target*. 19(61573286), 1–12.
- Ma, H., Liu, Y., Ren, Y. and Yu, J. 2020. Detection of collapsed buildings in post-earthquake remote sensing images based on the improved YOLOv3. *Remote Sensing*, 12(1).
- Mishra, B., Garg, D., Narang, P. and Mishra, V. 2020. Drone-surveillance for search and rescue in natural disaster. *Computer Communications*, 156: 1–10.
- Morbale, J., Vidyapeeth, B., Rohan, P. and Vidyapeeth, NB. 2022. Quadcopter Drone with Face Recognition. *International Journal of Scientific Research in Engineering and Management (IJSREM)*, 6(6): 1–6.
- Pashaei, A., Ghatee, M. and Sajedi, H. 2020. Convolution neural network joint with mixture of extreme learning machines for feature extraction and classification of accident images. *Journal of Real-Time Image Processing*, 17(4): 1051–1066.
- Rendón, MA. and Martins, F. F. 2017. Path Following Control Tuning for an Autonomous Unmanned Quadrotor Using Particle Swarm Optimization. *IFAC-PapersOnLine*, 50(1): 325–330.
- Shah, SA., Lakho, GM., Keerio, HA., Sattar, MN., Hussain, G., Mehdi, M., Vistro, RB., Mahmoud, EA. and Elansary, HO. 2023. Application of Drone Surveillance for Advance Agriculture Monitoring by Android Application Using Convolution Neural Network. *Agronomy*, 13(7): 1764.
- Sharma, M.K., Singal, G., Gupta, SK., Chandraneil, B., Agarwal, S., Garg, D. and Mukhopadhyay, D. 2021. INTERVENOR: Intelligent Border Surveillance using Sensors and Drones. *2021 6th International Conference for Convergence in Technology, I2CT 2021*: 1–7.
- Sun, N., Zhao, J., Wang, G., Liu, C., Liu, P., Tang, X. and Han, J. 2022. Transformer-based moving target tracking method for Unmanned Aerial Vehicle. *Engineering Applications of Artificial Intelligence*, 116(September): 105483.
- Zaidi, SSA., Ansari, MS., Aslam, A., Kanwal, N., Asghar, M., and Lee, B. 2022. A survey of modern deep learning-based object detection models. *Digital Signal Processing: A Review Journal*, 126: 1–18.
- Zhao, K., He, T., Wu, S., Wang, S., Dai, B., Yang, Q. and Lei, Y. 2018. Application research of image recognition technology based on CNN in image location of environmental monitoring UAV. *Eurasip Journal on Image and Video Processing*, 2018(1).