AI-BASED MODEL FOR DETECTION AND TRACKING OF DRONES USING CONVOLUTIONAL NEURAL NETWORK

Deepanshu Phartyal[1] Mihir Ghooi[2]

[1]Research Scholar, MCEME, Secunderabad

[2]Faculty, MCEME, Secunderabad

ABSTRACT

The prime objective of this proposed research is to build a model for drones' real-time detection and tracking. In recent years, the security of UAVs emerged as a big concern for many military agencies worldwide. Surveillance of UAVs has been prompted as due to the technical advancement in terror, UAVs are being used widely to harm civilians and military stations. Detection and tracking of such UAVs are one the prominent challenges, and hence the proposed model will help in such security actions with an enhanced measure ability of the suspicious activity by the DRONES. The goal is to develop an algorithm to Optimize the Detection of Drones and check the feasibility of tracking a drone-based on IR technology at night and optimize the algorithm to achieve adequate accuracy (min 70%) in tracking a single target. To adapt the tracking algorithm for tracking selected targets in the presence of clutter and multiple targets at short ranges using Python and OpenCV. The proposed model achieved max accuracy of 100% for drones and 99.4 %, 96.9% and 99.2% for airplanes, Birds and Helicopters, respectively. The model uses 8000 epoch maximum for a total of 3200 images, and the proposed model gives 0.98, 0.99, 1.00, and 0.99 for mAP F1 score, precision and recall, respectively, with an average loss function of 0.1479.

Keywords: Object Detection System, CNN, UAV-Drone, Detection, Tracking, Security and Surveillance.

1. INTRODUCTION

As production technology and drone use progress, more drones are being employed for military, commercial, and safety applications. Due to its success in applications including airport protection, facility protection, and integration with security and surveillance systems, the deployment of various drone types has attracted much attention in recent years. In contrast, drones could seriously threaten these security areas [1]. Hence it is critical to develop a simple method for classifying various drone types in those systems. Such technologies may be used in army and airport security structures to prevent drone invasions or ensure their safety. As a result, it is essential to discover, recognize, and choose them while thinking about public safety and the threats that UAVs present. The process of diagnosing the kind of the target class is known as identification. The method of determining the target class is reputation. Monitoring the target, which can be suspicious and threaten the safety of the surroundings, is the detection process [2].

The present research aims to deploy a model for detecting and tracking Drones using Deep Learning techniques. Four classes are identified in the dataset, which was taken from Github (A publically available dataset source). The dataset has a total of 3200 Tit images shot at night, and four flying objects were captured, namely, Bird, Airplane, Drone and Helicopter. The proposed work would meet the current demand for security; hence the proposed model is built for surveillance purposes only. Detection and tracking of flying objects at night are quite challenging to filter the classes of the object and current position; hence, the posed model will tackle this challenge and can be used for military and civilian use.

1.1 OBJECT DETECTION SYSTEM

Detection and tracking of Drone is part of the object detection system (ODS), which is based on the classification localization of objects. Making a classifier that can categorize photos of an object that has been tightly cropped is one way to start developing an item identification device [4]. This model will now be used to choose drones using a sliding window method. We crop a portion of the picture onto each slide using a sliding window method (similar to the one used in convolutional networks).

The crop's proportions match those of the sliding window. Each cropped picture is then sent into a CNN model, which forecasts the danger that the cropped image is of a drone. After doing so, we enlarge the sliding window and run it over the whole image again. This method is performed multiple times. Given that we trim several photos before running them through the ConvNet, this approach is computationally and computationally time-consuming, making the whole process sluggish [5]. The sliding window's convolutional implementation enables this issue to be resolved. A working model of the ODS system is shown below ***(Figure no-1)***.

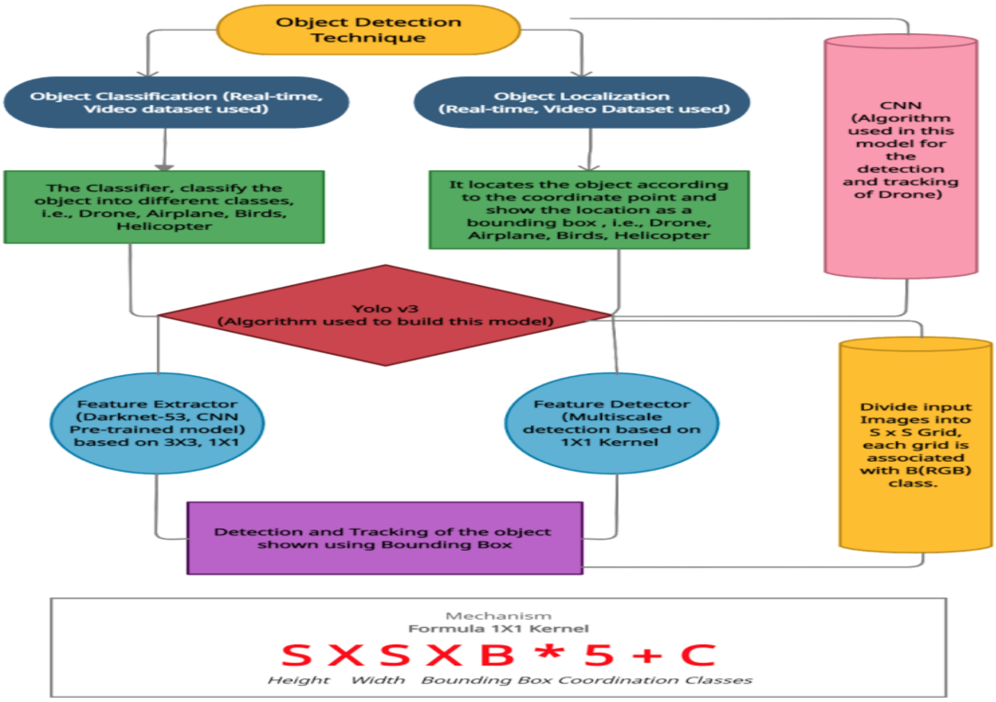


Figure 1. Objective detection system

1.1.1 DETECTION AND TRACKING OF DRONES

The process of object tracking involves identifying a trajectory or direction that an item travels in across several frames. The algorithms CNN and YOLOv3 Darknet 53 are used to find objects in a picture. A bounding box is formed over the devices, and the classification and localization were done using an object class type coordinate system (x, y plane). The observed bounding box is supplied to neural networks as a helpful resource for monitoring. Several frames of multi-object tracking are used to track a bound box.

2. MAJOR CONTRIBUTIONS AND PAST WORKS

2.1 LITERATURE SURVEY

According to Farhad et al. (2022), this study aimed to improve the detection and recognition of drones and birds. In order to attain the goal, the authors used Deep Learning to develop the model in this study. One of the key confusing concerns is the physical forms of birds and drones, which has been addressed in the proposed model that helps distinguish between the two [6]. The data set for this study has 10000 photos, and the proposed model has an accuracy of 83 percent. Xiang Ren et al. (2022) investigated the limitations in the Drone detecting area, and it was previously suggested that alternative variables such as processing speed, storage space, and so on be included. The suggested work represents progress since it has the potential to boost processing performance while reducing Drone storage space. For the UAV TIR video stream dataset, the mask-RCNN technique using MobileNetV3 was employed to overcome the inadequacies. Lizhi Yang et al. (2022) have scholarly handled the dataset problem and devised a technique to eliminate such concerns that require a great deal of attention in this study job. The fusion strategies for RGB-IR hybrid datasets and the issues that come with them were explored. Synthetic IR images were created utilizing simulation techniques such as Sim Simulation Engine and CycleGAN [7]. The developed arrangement was later tested on the NVIDIA Jetson Xavier. In order to reduce the number of false detections, Fredrik Svanstrom et al. (2021) developed a new Drone detection method. With a video stream, the dataset is taken from Sweden's three airports; Machine Learning and Fusion sensor approaches were employed to achieve the goal. Airplanes, Birds, helicopters, and Drone are the four object classes in the dataset.

The dataset employed in this study has audio and video, which aids in the system's optimization. In order to reduce erroneous detection, the authors propose adding two more kinds of objects, such as insects and clouds. In this work, Brian et al. (2021), For both Ir and visible photos, a DNN-based technique was employed to recognize and track the UAV [8]. The dataset employed here has various environmental circumstances, which aids the system in detecting and tracking with greater precision. The best result is 98 percent for visible photos and 82 percent for IR images using the map. This result indicates the optimum architecture that may provide the best performance in the domain of DNNs. Sara Al-Emadi and her colleagues (2021).

Drone identification and tracking were attempted in this study by increasing the acoustic feature of the Drone. Drones were identified and detected using various Deep Learning techniques [9]. The paucity of accessible acoustic datasets is also addressed in the work. Therefore, the authors create a hybrid drone acoustic dataset to fulfill the current need and future research. Deep learning GAN algorithms were used to compose drone audio samples and artificially created noise. The usefulness of several algorithms such as CNN, RNN, and CRNN on the assembled dataset was explored. The discovery also demonstrates the applicability and benefits of the suggested technology over the current one. Tao Hong and colleagues (2021) This research looked at the detection and tracking of drones in real-time. With the help of a Drone tracking dataset that included four different types of drones and photographs from various environments, YOLOv3 with Deep Learning served as the main conduit for implementing the proposed notion [10]. The implementation's concrete finding demonstrates that the maximum degree of accuracy was 94 percent for tracking. The studies also look into how much money the overall arrangement costs. A comparative analysis of the performance of several YOLO technologies was evaluated in this research work by Shuo Wang. (2021). XTDrone UAV simulation software was used to assess and test the Pascal VOC with mAP and FPS. We compared YOLOv3, YOLOv3 tiny, YOLOv3-SPP3, YOLOv4 and YOLOv4-tiny [11].

The test results show that YOLOv4 has a better accuracy percentage than the remainder alternatives, with a score of 87.48. When the alternative models were compared in test time, it was discovered that YOLOv3-tiny provided the greatest results. Wei Xun, Daniel Tan, and others (2021) With the YOLOv3 detector, transfer learning algorithms were used. Compared to the machine learning approach, the findings show that YOLOv3 had an average accuracy of 88.9%. Previously, a test for real-time detection was performed on the NVIDIA Jetson TX2 to assess the correctness of the suggested model [12]. The Deep Learning-based YOLOvs detector was specifically trained for Drone detection using transfer learning. The researchers produced a real captured dataset for the detection of drones by Syed Ali Hassan et al. (2019).

Because one of the most important aspects of this type of implementation is the development of the dataset, the researchers used YOLO v2 and YOLOv3 to meet their goals. The performance of the developed model was assessed in terms of MAP and accuracy. The results show that the YOLOv3 outperforms the competition. According to Behera and Raj (2020), of all the alternatives, the Convolution Neural Network (CNN) helps to extract information from images and detect it with maximum accuracy. Computer Vision is also used in the proposed study and is suggested as the most reliable approach [13]. Due to YOLOv3's large design and lack of class, 150 epochs were employed. The authors urge that future research focus on RF signal detection.

According to the authors of this research report, Yuanyuan Hu et al. (2019), YOLOv3 provides the best performance for detection with maximum accuracy and speed by capturing deep and high-level features. To forecast bounding boxes for anti-UAV, the proposed model uses the last four scale feature maps, with the sizes of all four scales generated from the input data. According to the authors, the k means could be replaced in the future to improve performance [14]. Haipeng Zhao et al. (2020) study describes a model for light-weighted real-time object detection using YOLOv3-LITE.For GPU and mobile devices, the proposed model can be employed. The suggested model's process consists of a residual supplement block, parallel high-to-low resolution, and increasing network depth.

The size of the produced model is 20.5MB, which is 91.70 percent, 38.07 percent, and 74.25 percent smaller than YOLOv3-Tiny, YOLOv3, and SlimYOLOv3spp3, respectively, according to the research result. According to Adarsh and Rathi (2020), the proposed model employs two-stage detection systems with distinct goals. The one-stage detection system encompasses YOLO v1, v2, v3, and SSD, primarily concerned with speed, whereas the two-stage detection system includes RCNN, Fast- RCNN, and Faster RCNN [15].

The data show that the YOLOv3-tiny can help boost speed, but there is no one-size-fits-all solution for accuracy because it differs by necessity. Mingjie Liu et al. (2020) provide a novel technique to model development from the perspective of UAV identification. YOLOv3 Resblock is the foundation for the suggested model. In order to boost performance, a link between two ResNet units was built in Darknet. The performance of YOLOv3 was improved using model training and backbone structure optimization. The model appears to be functioning well in the output.

3. TOOLS AND TECHNIQUES EMPLOYED

3.1 CONVOLUTIONAL NEURAL NETWORK

For problems with computer vision, CNN is a common neural network design. In order to locate relevant features, CNN has the advantage of automatically extracting functions from snapshots ***(Figure no-2)***. The geographic locations of the image are lost during the flattening process. Internal feature representation is learned using small squares of input data to retain the spatial relationship between picture parts.

(a) Convolutional Layer: Filters and feature maps make up the convolutional layer. Consciousness on a certain layer is processed by what is known as filters. There is no way to replace those filters. They begin with the values of the pixels and end up with a distinctive map. The outcome of one clean-out layer is a feature map. Clear out is applied to the whole image, one pixel at a time. A limited number of diverse neurons are triggered to provide a feature map.

(b) Pooling layer: This layer is utilized to reduce dimensionality. Pooling layers are added after one or two convolutional layers to generalize the advancements discovered from prior feature maps. By doing this, the risk of overfitting throughout the academic year is decreased. Three) Fully linked layer: After gathering and combining features from the convolutional layer and pooling them afterward, the fully linked layer is used to allocate the feature to change. Linear activation capabilities or softmax activation characteristics are used in these layers.

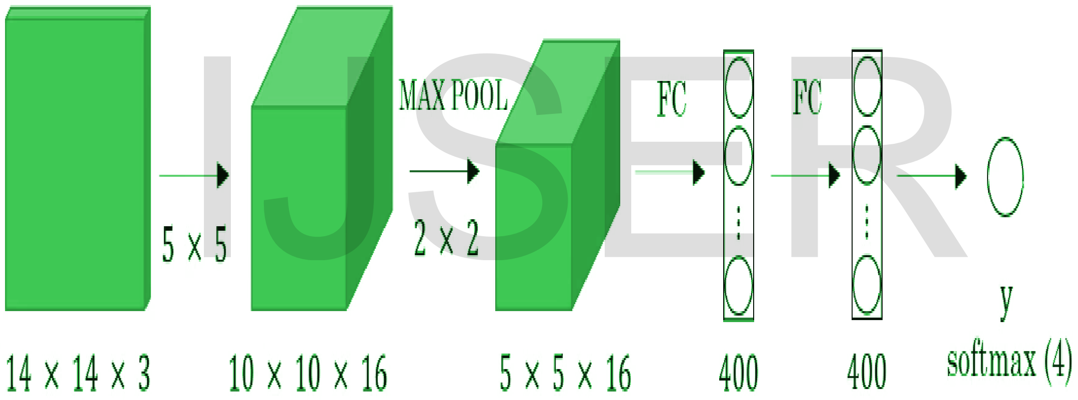


Figure 2 General CNN Architecture

3.1.1 YOLO V3 DARKNET 53

YOLOv3 employs a multi-label classification strategy. An independent logistic classifier calculates the likelihood of entering a positive label. The binary-cross entropy of each label is used to compute loss. The complexity is reduced since the softmax function is not always used.

Bounding field optimization: Regression may be done using the logistic method. YOLO v3 foresees the presence of an item score. A ground fact field defines all gadgets, and if the anchored container overlaps the ground fact container more than any other, the objectness score is 1. There may not be a cost to the anchor field if the overlap rises over the predetermined level. A single anchor field is mapped for each floor reality container. If the anchor box is not chosen and assigned to the bounding field, the self-belief loss is calculated rather than the category and localization losses. The Darknet-fifty three is a 53-layer convolutional neural network community that uses many layers of images. Items in photographs are categorized by a model that has undergone training. The YOLO technology and object detection are enhanced on Darknet-fifty three. After installing 53 more levels, Darknet may have a maximum of 106 layers, 53. This approach has a faster velocity than the previous one. The entry picture is recognized using the animal's facial features.

4. METHODS ADOPTED

4.1 DATASET DESCRIPTION

The dataset was created using four items from the publicly accessible information source Github.com (Airplane, hen, Helicopter and Drone). The data set includes night vision images captured in visible and infrared light ***(Figure no-3)***. The proposed version may demonstrate how well it can recognize and track drones, and the recommended experiment would accurately remember infrared (IR) images.

Fig (a) Fig (b) Fig (c) Fig (d)

Figure 3 IR Images dataset, which contains four classes (a) Drone, (b) Helicopter, (c) Airplane and (d) Bird

4.2 ALGORITHM EMPLOYED FOR THE IMPLEMENTATION OF THE AIM

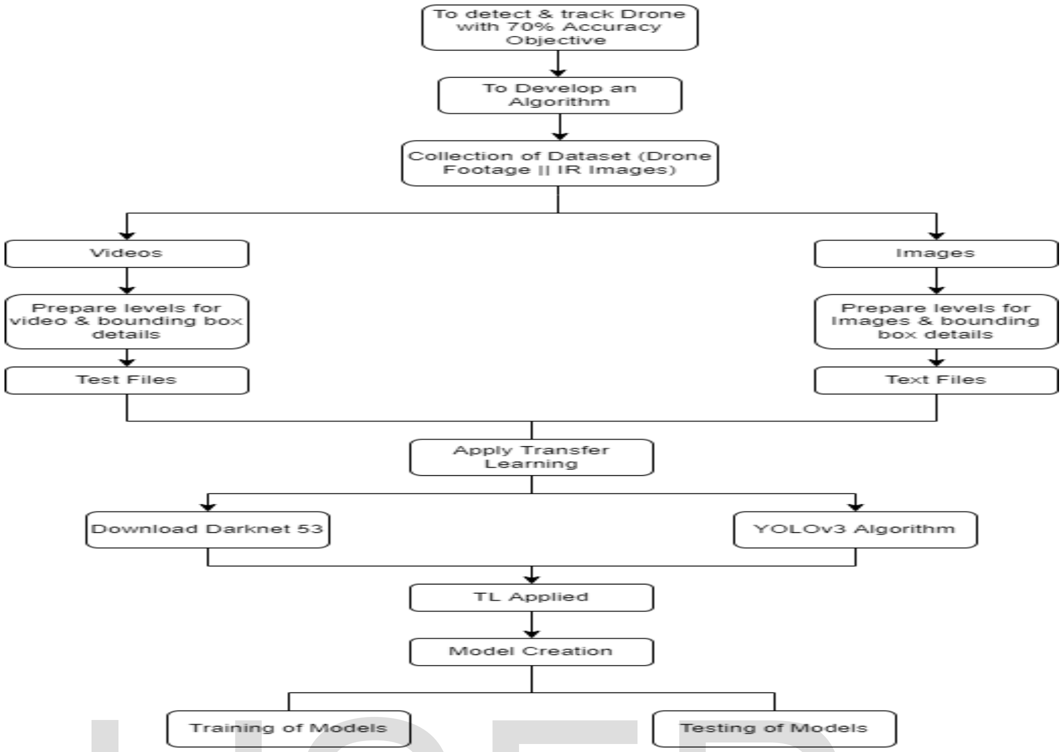


Figure 4 Algorithm Employed for detection and tracking f Drone.

I. Using transfer learning with YOLOv3, we may design a unique item Detection for our unique information. The dataset of IR camera-captured drone footage will be assembled first. After that, such videos will undergo body-through-frame processing and be saved as snapshots. Then, labels for the movies and pertinent bounding box data for the images will be created.

II. The bounding container statistics and magnificence labels are provided in the text papers. The text file that correlates to the picture and contains all the pertinent information is now available. The Darknet53 will then be downloaded and installed to execute transfer learning using the YOLOv3 rules. We will eventually provide training and information verification for the model. The version will then undergo training and testing.

III. The dataset for this project was obtained from Github.com. (Supply of publicly accessible records) The collection includes IR, visual, and acoustic data to train and evaluate drone detection sensors and algorithms. Video of four wonderful object types, including drones, birds, airplanes, and helicopters, is included in the collection. 650 movies are available on it (365 IR and 285 visible).

IV. If all the images from all the movies are extracted, there are 203328 annotated images in the collection overall. The venture's education and inspection records are subsequently created using those videos.

V. The YOLOv3 algorithms of the Darknet53 framework and the computer imaginative and prescient library might be used for this assignment ***(Figure no-4)***. YOLO is a real-time object recognition tool that scans the full image at test time so that the typical picture context prompts its predictions.

VI. Compared to rival algorithms in terms of overall performance, YOLO may make predictions utilizing a single network evaluation in an extraordinarily short time.

VII. The suggested version may be taught using Darknet-fifty three. DarkNet-fifty three is a community of convolutional neural networks with 53 layers. Versions of the DarkNet that have been mastered can classify images into 1,000 different object categories, such as keyboard, mouse, pencil, and a variety of animals.

5. RESULT AND DISCUSSION

5.1 PERFORMANCE MATRIX

It is critical to evaluate how well the deep learning-skilled version performed on the test dataset of unknown statistics. The set of rules analyzed may be impacted by the overall performance metrics. This makes it easier to identify the causes of misclassifications so that corrective action may be taken.

Table No-1 Performance of the model

Class	Accuracy (%)	TP	FP
Drone	100	195	0
Bird	96.9	186	0
Helicopter	99.2	207	1
Airplane	99.4	208	2

The confusion matrix provides prediction data for different devices for the binary class, Accuracy and Loss: The accuracy degree is determined using the formula. Because it provides equivalent costs for all types of mistakes and functions well for a balanced dataset, the accuracy metric is inaccurate as a stand-alone measure. The loss is calculated using loss capabilities utilized for education, and the loss is computed using batch learning, which computes the loss after every batch of schooling. Precision, Recall, and F1-Score: The fraction of exact type effects is known as precision. Keep in mind the proportion of all relevant impacts that the algorithm effectively categorized. The model must be upgraded by maximizing the F-1 score, which considers both accuracy and memory values. The methodology used to calculate such measurements includes (12) The bounding box defines the boundaries of what has been seen. Assessment metrics such as true positive (TP), false positive (FP), as well as precession and imply average precession (mAP) are determined using intersection over union (IoU).

5.2 FINDINGS AND OUTCOME

YOLO v3 Darknet 54 covn74 pre-trained model was used to train the custom dataset, which has a total of 3200 TIR, including four classes and Tensor flow and Sklearn framework. The training of the dataset was done using Google Collab and took approximately 30 hours to train the dataset. The images have 416,416 widths and heights with a 3RGB channel. The batch size of the proposed model was 64, and the subdivision was 16 for four groups. The total no of iterations was 8000 (4*2000) ***(Table No-1)***. The pictures and tables in this section depict the model's overall performance ***(Figure no5-10)***.

Table No -2 Performance of Drone

mAP	Precision	Recall	F1- Score
0.988	1.00	0.99	0.99

Class-I, This class of objects shows the Drone detected and tracked using the YOlov3Darknet 53 and gives maximum accuracy of 100% and 98%, respectively, which is clearly illustrated.

Figure 5(a) Drone with 100% accuracy

Figure 5(b) Drone with 98% accuracy

Class-II, This class of objects shows the Bird detected and tracked using the YOlov3Darknet 53 and gives maximum accuracy of 100% and 99%, respectively, which is clearly illustrated.

Figure 6(a) Bird with 100% accuracy **Figure 6(b)** Bird with 99% accuracy

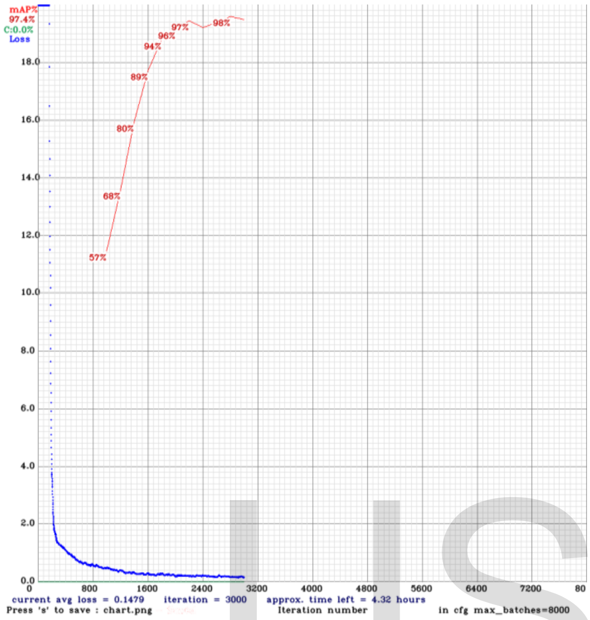
Class-III, This class of objects shows the Helicopter detected and tracked using the YOlov3Darknet 53 and gives maximum accuracy of 100% and 84%, respectively, which is clearly illustrated.

Figure 7(a) Helicopter with 100% accuracy **Figure 7(b)** Helicopter with 84 % accuracy

Class-IV, This class of objects shows the Airplane detected and tracked using the YOlov3Darknet 53 and gives maximum accuracy of 100% and 98%, respectively, which is clearly illustrated.

Figure 8(a) Airplane with 100% accuracy **Figure 8(b)** Airplane with 98 % accuracy

Epoch vs. loss functions for various classes are illustrated here, showing the maximum loss during training and the epoch ***(Table No-2***. Eh, the average loss function found during model training was 0.1479 and mAP 0.988 for a maximum of 8000 iterations for a total of 3200 images ***(Figure no9 -12)***.



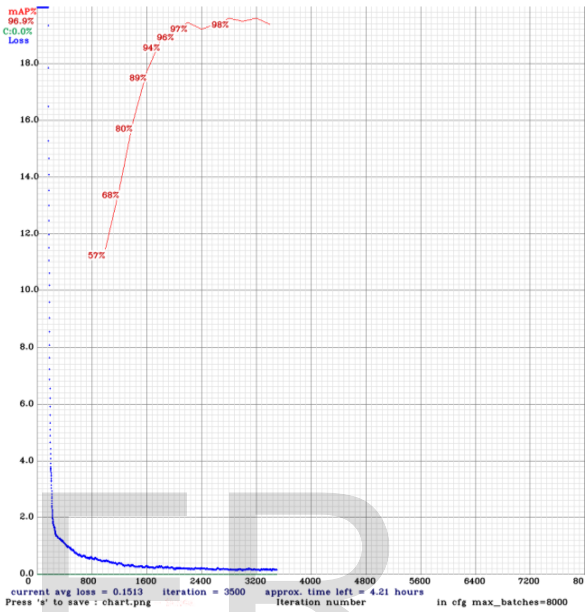
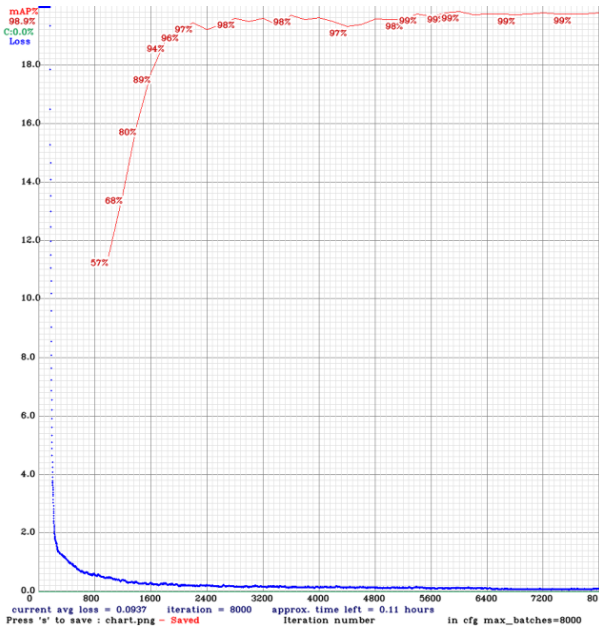


Figure 9 mAp with **97.4%** and LF= 0.1479 **Figure 10** mAp with **96.9 %** and LF= 0.1513



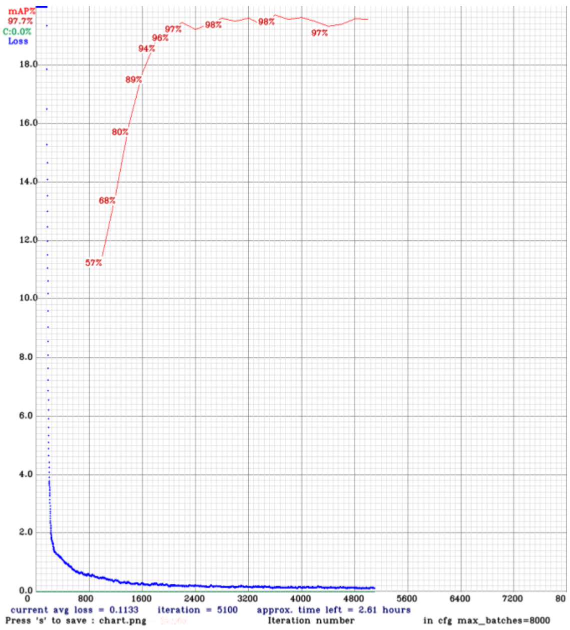


Figure 11 mAp with **98.9 %** and LF= 0.0.937 **Figure 12** mAp with **97.7 %** and LF= 0.1133

5.3 VALIDATIONS AND COMPARISON

Table No-3 Comparison of maximum accuracy achieved by various authors in the

S.NO	AUTHOR	YEAR	NOTEWORTHY CONTRIBUTION
1st	Farhad et al.	2022	83%
2	Xiang Ren et al.	2022	50% (S)
3	Lizhi Yang et al.	2022	90%
4	Fredrik Svanstrom¨et al.	2021	0.9623% (P)
5	Brian et al.	2021	82%
6	Sara Al-Emadi et al.	2021	92%
7	Tao Hong et al.	2021	94%
8	Shuo Wang.	2021	87.48%
9	Daniel Tan Wei Xun et al.	2021	88.9%
10	Syed Ali Hassan et al.	2019	25 % (mAP)
11	Behera & Raj	2020	150 epoch
12	Yuanyuan Hu et al.	2019	89%
13	Haipeng Zhao et al.	2020	48.25 (mAP)
14	Adarsh & Rathi	2020	57% (mAP)
15	Mingjie Liu et al.	2020	72.21% (mAP)
16	Party & Ghooi	2022	98% (mAP)

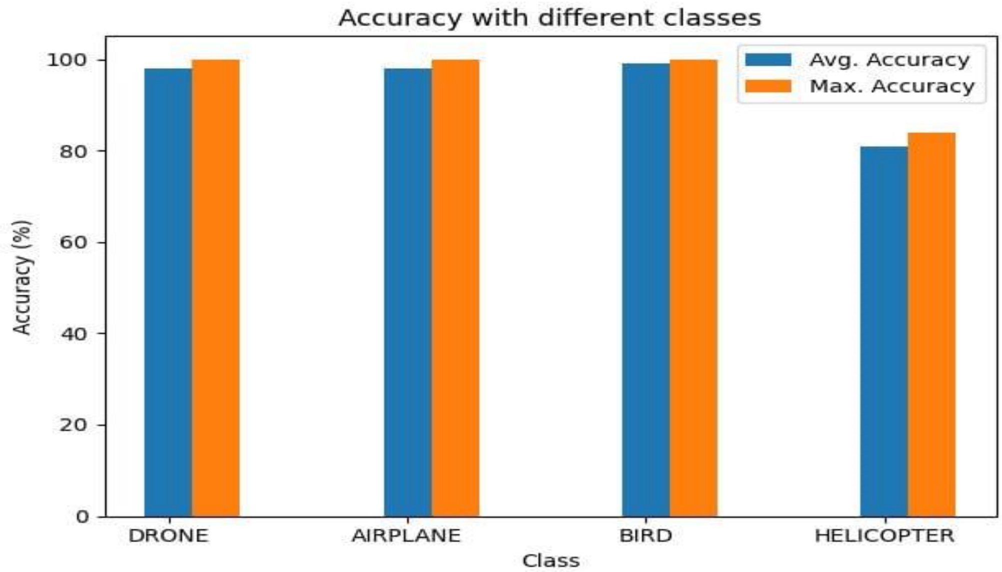


Figure No-12 Comparison of the accuracy of all fur classes

6. CONCLUSION AND FUTURE SCOPE

The model was deployed using Deep Learning using YoloV3 Darknet 53on Google Collab on a maximum of 8000 epochs (4*2000) for 3200 images, and training and testing of the dataset into 70/30 was carried out. The model performed extremely well and achieved maximum accuracy of 100% for drones. The dataset used to train the model has four classes, namely Airplane, Bird, Drone and Helicopter, so the proposed model predicted the accuracy for the rest of the classes too, which was 99.4 %, 96.9% and 99.2% for Airplane, Bird and Helicopter respectively. Henceforth, the proposed model would help detect and track the drones at night with maximum detection and tracking accuracy and help supervise the malicious air. The proposed model is created exclusively for sky surveillance for military applications; however, this could be applicable for civilian uses. Many attempts have been made for detection and tracking, and the proposed model remains the same. Still, there is space for further enhancement. In the future, some other classes can be added to train the model, and the model can also be trained in different environmental conditions, which will help increase the model's performance.

REFERENCES

[1] Farhad Samadzadegan et al.,(2022), "Detection and Recognition of Drones Based on a Deep Convolutional Neural Network Using Visible Imagery," aerospace,9.

[2] Xiang Ren et al.,(2022)," An improved mask-RCNN algorithm for UAV TIR video stream target detection," International Journal of Applied Earth Observations and Geoinformation, v-106.

[3] Lizhi Yang, et al.,(2022), "Drone Object Detection Using RGB/IR Fusion," arxiv.

[4] Fredrik Svanstrom et al.,(2020), "Real-Time Drone Detection and Tracking With Visible, Thermal and Acoustic Sensors," International Conference on Pattern Recognition (ICPR).

[5] Brian K. S. Isaac-Medina et al.,(2021), "Unmanned Aerial Vehicle Visual Detection and Tracking using Deep Neural Networks: A Performance Benchmark," arxiv.

[6] Sara Al-Emadiet al.,(2021), "Audio-Based Drone Detection and Identification Using Deep Learning Techniques with Dataset Enhancement through Generative Adversarial Networks," Sensors, v-21.

[7] Tao Hong et al.,(2021), "Multitarget Real-Time Tracking Algorithm for UAV IoT," Wireless Communications and Mobile Computing,v-2021.

[8] Shuo Wang (2021), "Research Towards Yolo-Series Algorithms: Comparison and Analysis of Object Detection Models for Real-Time UAV Applications," Journal of Physics: Conference Series-2021.

[9] Daniel Tan Wei Xunet al.,(2021), "Drone detection using YOLOv3 with transfer learning on NVIDIA Jetson TX2", Second International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP).

[10] Syed Ali Hassan et al.,(2019), "Real-time UAV Detection based on Deep Learning Network," ICTC, IEEE.

[11] Dinesh Kumar Behera et al.,(2020), "Drone Detection and Classification using Deep Learning," Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020).

[12] Yuanyuan Hu et al.,(2019), "Object Detection of UAV for Anti-UAV Based on Improved YOLO v3", Proceedings of the 38th Chinese Control Conference, China.

[13] Haipeng Zhao et al.,(2020), "Mixed YOLOv3-LITE: A Lightweight Real-Time Object Detection Method", Sensors,v-20.

[14] Pranav Adarsh et al.,(2020), "YOLO v3-Tiny: Object Detection and Recognition using one stage improved model",6th International Conference on Advanced Computing & Communication Systems (ICACCS).

[15] Mingjie Liu et al.,(2020), "UAV-YOLO: Small Object Detection on Unmanned Aerial Vehicle Perspective," Sensors,v-20