DOI: 10.6977/IJoSI.202409_8(3).0008
M. Kumar, U. Pilania /Int. J. Systematic Innovation, 8(3), 99-114 (2024)

A framework for detection of drone using yolov5x for security surveillance system

Manoj Kumar[1], Urmila Pilania[2]*
[1,2]*Department of Computer Science, Manav Rachna University, Faridabad, India
urmilapilania@gmail.com
(Received 27 March 2024; Final version received 29 June 2024; Accepted 29 June 2024)

Abstract

In response to the increasing use of drones, there is an emerging need for dependable security surveillance systems capable of detecting them. This article presents a theoretical framework for drone detection using the YOLOv5x (You Only Look Once) deep learning algorithm, which aims to improve security surveillance systems. The framework is composed of various hardware and software components, including drones, cameras, and computer systems, and utilizes YOLOv5x to accurately detect drones. To carry out the research, the Roboflow Drone 1 dataset is used, which includes a variety of images captured under various experimental conditions. The YOLOv5x algorithm is selected due to its high accuracy and low computational cost, which makes it an ideal solution for security surveillance systems. To measure the performance of the system, metrics such as F1-Score, Recall, Precision, Area under the Curve (AUC), and Mean Average Precision (mAP) are employed. The mAP is calculated at 0.95 at a learning rate of 0.5, precision is 0.901, recall is 0.97, AUC is 95.3, and F-1 score is 0.94. Overall, this framework provides a dependable and efficient approach for detecting drones, resulting in improved security of critical infrastructure, public events, and other sensitive locations. Furthermore, the proposed research is compared with existing state-of-the-artwork, and the experimental results verified that the proposed research outperformed the state-of-the-art.

Keywords: Drone Detection, YOLOv5x, Security, Surveillance, Unmanned Aerial Vehicle.

1. Introduction

A drone, named an Unmanned Aerial Vehicle (UAV), is a jet which is generally operated remotely by a human as per Kumar, M, et al. (2024). Drones have an extensive variety of applications, containing military operations, videography, surveying, inspection, and aerial photography. Modern drones are prepared having advanced features such as GPS, cameras, and autonomous flight capabilities. Drones can be categorized based on their cost, size, weight, capabilities, and intended use according to Taha, B., & Shoufan, A. (2019). This information is used to determine the appropriate level of regulations, safety measures, and limitations for the use of drones in various contexts. Some common categories of drones include as represented in **Table 1**:

Table 1. Types of Drones Zhu, X., et al. (2021)

Characteristic	Consumer Drone	Commercial Drone	Military Drone	Agriculture Drone	Emergency Drone	Education Drone
Purpose	These types of devices are lightweight, small in size, and designed for leisure use, typically weighing less than 55 pounds	Larger, more capable drones designed for commercial use like midair photography, survey, package distribution, inspection of pipelines and power lines.	These types of drones are developed for military use like investigation, reconnaissance, and beset killings.	Drones specifically designed for agricultural applications, such as crop monitoring, mapping, and precision agriculture.	Drones are used for emergency response and disaster relief operations, such as search, rescue, and firefighting.	These are used for educational uses like teaching learners about drone machinery and flight.
Size	Small	Varies with application	Medium	Large depending upon application	Varies	Design as per use

Range	Medium range	Short to long range	Long range	Medium depends on application	Varies depending upon application	Range varies with use
Sensor	Cameras and gyroscopes	Sensors for data collection	Cameras, radar, infrared	Multispectral and hyper-spectral sensors	Advanced sensors for detecting heat and survivors.	Basic Sensors
Cost	Affordable	Cost is high due to more features	Very costly due to military grade material	Very costly due to specific instrument and sensors	Depending on features cost varies.	More affordable

Yolov5x-based drone classification: It is a version of the YOLO object recognition method which could be utilized for drone classification. It is a deep learning technique which could automatically identify the type and size of a drone. Its intended use and capabilities can be analyzed using aerial images captured by the drone. The YOLOv5x works by separating an image into a network of cells by using a deep neural network to foresee the presence and class of objects within each cell. The model is skilled on a large dataset of pictures of different kinds of drones, so it can learn to identify specific features and patterns associated with each class of drone Pilania, U., et al. (2023).
After training the model, it could be applied to classify new images in real-time. By using YOLOv5x for drone classification, it is possible to quickly and accurately determine the type along with the abilities of a drone. It could be valuable for numerous applications, like security, surveillance, law enforcement, emergency response, and border protection. The proposed research works by training the model utilizing the roboflow drone 1 dataset of images to accurately detect drones. The YOLOv5x algorithm is used because of its good accuracy, very lower computational cost, allowing for efficient and effective surveillance. The method could be combined into existing surveillance systems to provide real-time monitoring, and alerts, enabling security personnel to take appropriate action when unauthorized drones are detected. The basic difference between YOLOv5 and YOLOv5x is given as in **Table 2**:

Table 2. Features of YOLOv5x

Feature	YOLOv5	YOLOv5x
Architecture	YOLOv5 comes in varying number of layers as per the requirement of the application.	It is designed with deeper and wider architecture.
Networks	It use CSPDarknet53 (Cross Stage Partial) network for extraction of features.	It uses ResNet with Cross Stage Partial connections Network (CSPNet) as backbone for extraction of more complex features.
Size	It uses small size variant resulting into low accuracy.	It uses large model size resulting into better accuracy.
Hardware Requirement	It can work with CPU and uses GPU for better performance.	It may use additional hardware for real-time applications because of high computational demands.
Resolution	It may work with varying resolution depending upon application.	It operates on high resolution resulting into good feature extraction and high accuracy.
Speed	Its speed is high due to due to smaller size of model.	Its speed is low due to large model size and complex structure.

The main objective of the proposed work is to design a framework for drone detection using the YOLOv5x, precisely for security surveillance systems. The objective is to improve security and safety in domains like airports, public events, critical infrastructure, and private properties, by effectively identifying and monitoring the presence of drones in real-time.

Research Contributions:

- The framework focuses on achieving real-time drone detection, which is crucial for security surveillance applications.
- To enhance the practicality and reliability of the system, the work contributes to reducing false alarms.

- It is applicable to different real-time security systems due to scalability and versatile nature.
- Proposed approach could be adapted to various types of cameras, sensors, and deployment scenarios.
- The research contributes to the growth of a user-friendly interface for security personnel. This includes a dashboard or software that provides a clear and intuitive display of detected drones and relevant information.
- Comprehensive performance evaluation and validation are conducted to measure the accuracy, precision, recall, and real-world applicability of the developed framework. This includes testing the model under numerous environmental conditions and drone types.

In summary, the research "A Framework for Detection of Drones using YOLOv5x for Security Surveillance System" aims to provide a robust, real-time solution for drone detection in security surveillance applications, with a focus on accuracy, speed, and practicality. The contributions lie in the customization of the YOLOv5x model, its integration, and the reduction of false alarms, all while addressing the security needs of various sectors.

2. Literature Review

The drones have been used gradually in recent years, as they have been established to be useful in numerous fields such as agriculture, construction, and aerial photography. However, the rise in unauthorized drone usage has led to significant security concerns. This literature review focuses on the framework for detecting drones using YOLOv5x for security surveillance systems.

These days' drones have been used widely due to their affordability and commercial availability. However, the use of drones also has increased the risk of illegal activities such as drug smuggling and terrorism. To prevent such actions, it is crucial to monitor and detect drones in restricted or special zones. One of the biggest challenges in drone detection is the resemblance between drones and birds, particularly when viewed against multifarious backgrounds in surveillance videos. To address this, a new image-based drone-detection model has been designed using YOLOv5 in Al-Qubaydhi, et al. (2022). Due to the limited availability of information, transfer learning was used to enhance the performance of the system. The results were outstanding, with an average precision rate of 94.7%. This system could be useful for safeguarding restricted areas and preventing illegal drone interventions.

The increasing quantity of drones has been used for both marketable and recreational purposes raised concerns about their potential misuse, including privacy violations and drug smuggling. However, spotting drones could be challenging due to other items in the sky, like aircraft, birds, and computerized systems requiring a huge amount of information and highly configured devices for real-time detection. To overcome the mentioned issues, the study Kumar, M., et al. (2024) proposed the use of a one-shot sensor called YOLOv5, which can be trained with pre-existing weights and data augmentation. The system was assessed using mAP, and recall, and attained a mAP of 90.40%, which is a 21.57% enhancement over the earlier model which applied YOLOv4 and was tried on the same dataset. This technology could suggest a solution to the problem of drone abuse and help protect people's privacy.

The advancements in drone technology have led to the emergence of object identification technologies which could be useful to various scenarios, including detecting illegal immigrants, natural disasters, missing people or objects, and industrial accidents. The authors De Galiza Barbosa, et al. (2023) aimed to discover ways to improve object recognition recitals under challenging conditions. Experimental data were gathered through photography in a confusing environment with varying environmental conditions. The F11 4K PRO drone and VisDrone dataset were used for the experiment. The study proposed a better form of the original YOLOv5 model and compared its performance with the original model. Key performance indicators, including precision, recall, F-1 score, and mAP (0.5), were calculated for both models. The improved YOLOv5_Ours model validated enhanced performance as mAP (0.5) and function loss than the original YOLOv5 model. Constructed on the statistical analysis, the study concludes by identifying a good object identification model for various challenging conditions. Detecting objects from drone-captured scenarios is a popular and recent task, but it comes with challenges. Drones navigate at different altitudes, resulting in varying object scales, which makes optimizing networks difficult. Additionally, motion blur from high-speed, low-altitude flights causes objects to become indistinguishable, adding to the challenge. To solve these challenges, the authors in Zhu, X., et al. (2021) proposed TPH-YOLOv5, a modified version of YOLOv5 that includes further prediction head to detect objects of diverse scales. TPH-YOLOv5 improved by about 7% compared to the baseline YOLOv5 model, which was promising and competitive.

Human action detection from drones has developed a significant challenge in current years, with potential applications in environmental monitoring, search, and rescue operations. Though, this challenge was complicated by the variability of human subjects' scales, orientations, and occlusions in drone-captured images. Authors Ahmad, T., et al. (2022) proposed low-resource machine learning approaches for action detection using the "Okutama-Action" dataset, which includes images with controlled image acquisition parameters. The proposed approach combines object recognition with a gradient-boosting classifier to identify actions in single images. The authors integrated YoloV5 with the classifier to achieve both scalability and efficiency in the object recognition system while accommodating the variable difficulty of samples. The proposed method beat erstwhile architectures used on the same dataset, which we attribute to the performance of Yolov5 and the adequacy of our pipeline to the Okutama dataset's specificities in terms of bias-variance tradeoff. The goal of drone detection is to locate the drone(s) within a video by identifying the smallest rectangle that encloses it. To resolve this issue, authors Aker, C., & Kalkan, S. (2017) presented an end-to-end object recognition method using Convolutional Neural Networks (CNN). To overcome the lack of information for training the network, the author developed a system that combines background-subtracted real images to create an extensive artificial dataset. This approach allowed us to achieve high precision and recall values simultaneously. After the literature survey, it has been concluded that there are several challenges associated with drone detection as represented by the **Table 3**:

Table 3. Mapping of issues with proposed work

Challenge	Existing Work	Proposed Work
Size and Speed	Drones can vary in size and speed, making them difficult to detect. Small drones can be particularly challenging to detect, as they may not produce enough heat or sound to be detected by traditional sensors as per Singha, S., & Aydin, B. (2021)	Diverse training of the model on different datasets helps in handling varying size and shape of the drone. The proposed model is able to detect drones of varying size and shape.
Flight Patterns	Drones can fly in unpredictable patterns, making it difficult to predict where they will be at any given time. They can also fly at low altitudes, making them difficult to detect with radar as per paper Akyon, F. C., et al. (2021)	ResNet with CSPNet efficiently work on features such as shape, texture, color, and motion patterns. The model efficiently detects drones at very low height as well as at very high altitudes.
Background Noise	In noisy environments, it can be challenging to detect the sound of a drone, particularly if it is small and has a low acoustic signature as in paper Anwar, M. Z., et al. (2019)	As in feature extractions ResNet is used which effectively filter out the background noise.
Signal Interference	Drones can interfere with GPS and other communication signals, which can make it difficult to track their location Wu, M., et al. (2018)	Pre-processing and data augmentation techniques are used in YOLOv5x for signal inference. These techniques help in tracking the location of the signal accurately.

3. Proposed Work

Traditional methods of detecting drone activity such as human surveillance, and radar systems, can be time-inefficient, resource-demanding, and could not always be effective. So, there is a requirement of more efficient and reliable method of detecting drones. To acknowledge the above challenges, the work proposes a framework for the identification of drones using the YOLOv5x deep learning algorithm. The framework provides a comprehensive solution for detecting drones in real-time, enhancing the security of critical infrastructure, public events, and other sensitive locations. The proposed work aims to develop an efficient and accurate drone identification system using modified YOLOv5x. For modification of YOLOv5x the backbone network, CSPDarknet53 is replaced with Residual Network (ResNet) and head network is updated to include extra layers for capturing the features more accurately. Additionally, the system will utilize advanced data augmentation techniques to improve its performance and robustness. The proposed model is trained using a large dataset of drone images under many environmental and lighting conditions. The proposed system's performance will be evaluated based on many constraints such as detection accuracy, recall, precision, and F1-score.

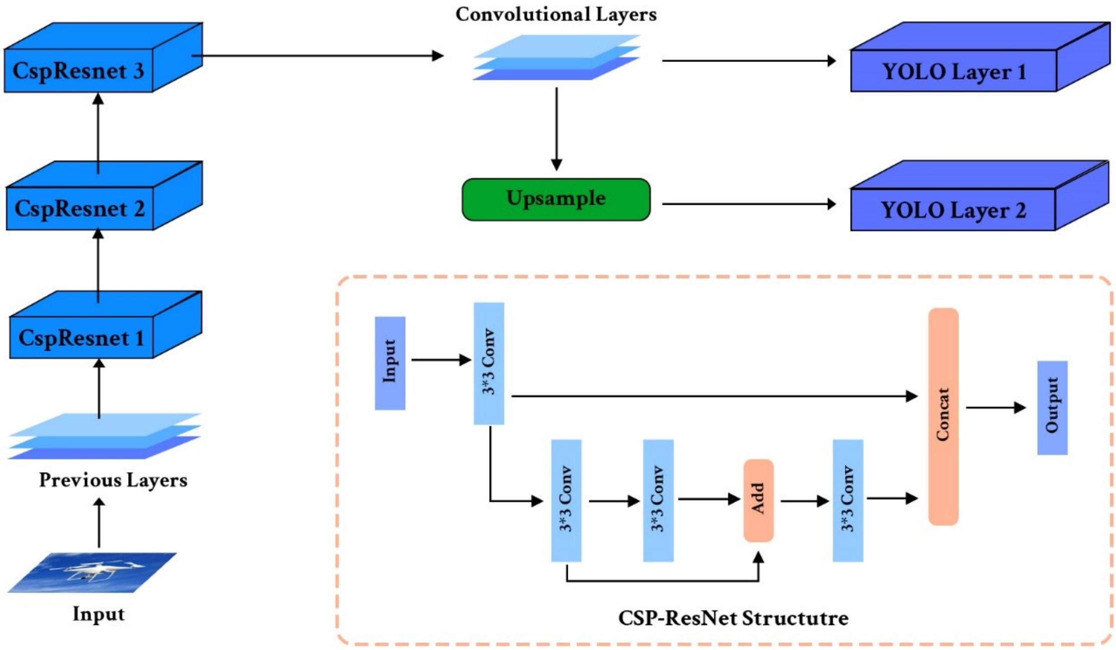


Fig 1. Layered Architecture of YOLOv5x

In the proposed work, Roboflow Drone 1 dataset is used which is a publicly available dataset containing images and annotations for drone detection. The dataset contains images of drones in various orientations, sizes, and backgrounds. The annotations include the bounding boxes and labels for the drones in the images. The dataset contains over 7,500 images with a resolution of 1080 x 720 pixels. The dataset includes a diversity of objects, such as buildings, trees, vehicles, birds and to simulate real-world scenarios. The Roboflow Drone 1 dataset is used for developing and testing deep-learning models for drone detection. It could be used to train and evaluate models based on deep learning algorithms, like YOLO, Faster R-CNN, and SSD. The dataset is available for download on the Roboflow website and can be accessed for free.

YOLOv5x is a larger and more accurate version of YOLOv5, designed to run on high-performance GPUs. It has a larger number of layers and parameters, allowing it to perform more complex computations and achieve higher accuracy in object detection tasks. In summary, YOLOv5s is optimized for speed and efficiency, while YOLOv5x is optimized for accuracy. The choice of a particular model depends on the specific use case and the available computational resources.

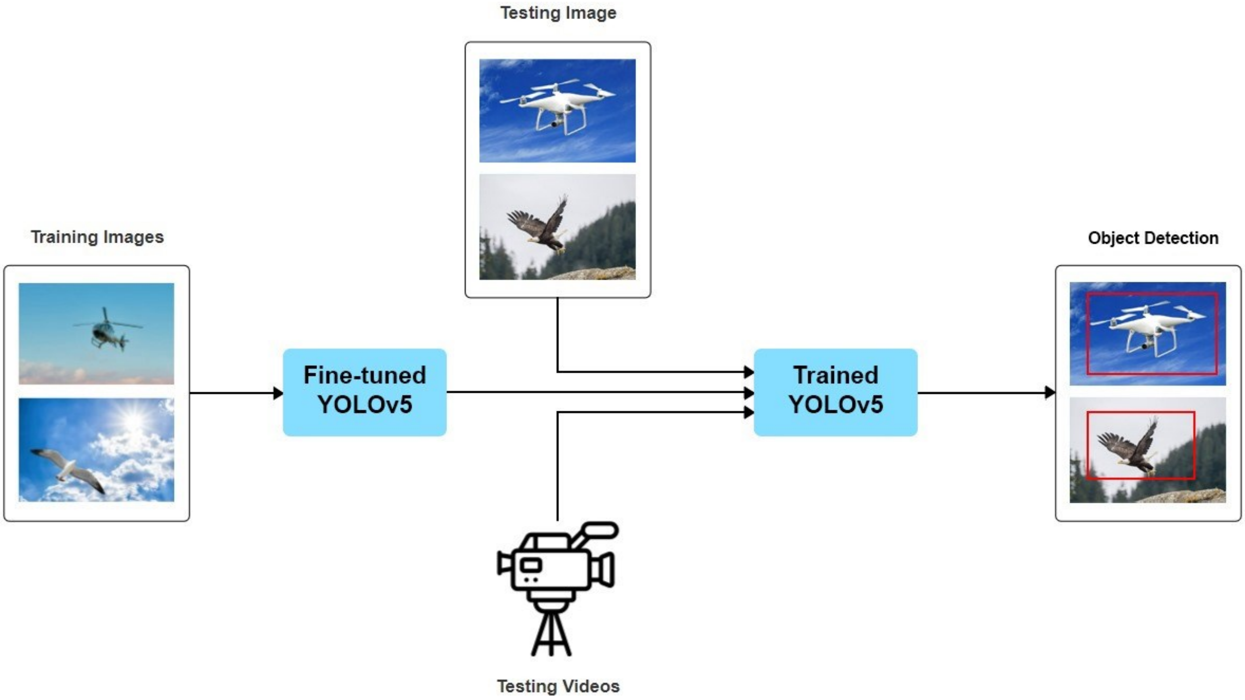


Fig 2. Proposed Flowchart

YOLOv5x is a standard object recognition model which uses a layered architecture to identify and categorize objects in images as depicted in **Figure 1**. **Figure 2** represents the proposed flowchart for the research work. **Figure 3** represents the scattering of the sizes and positions of drones in the dataset. The architecture of YOLOv5x is divided into several key components.

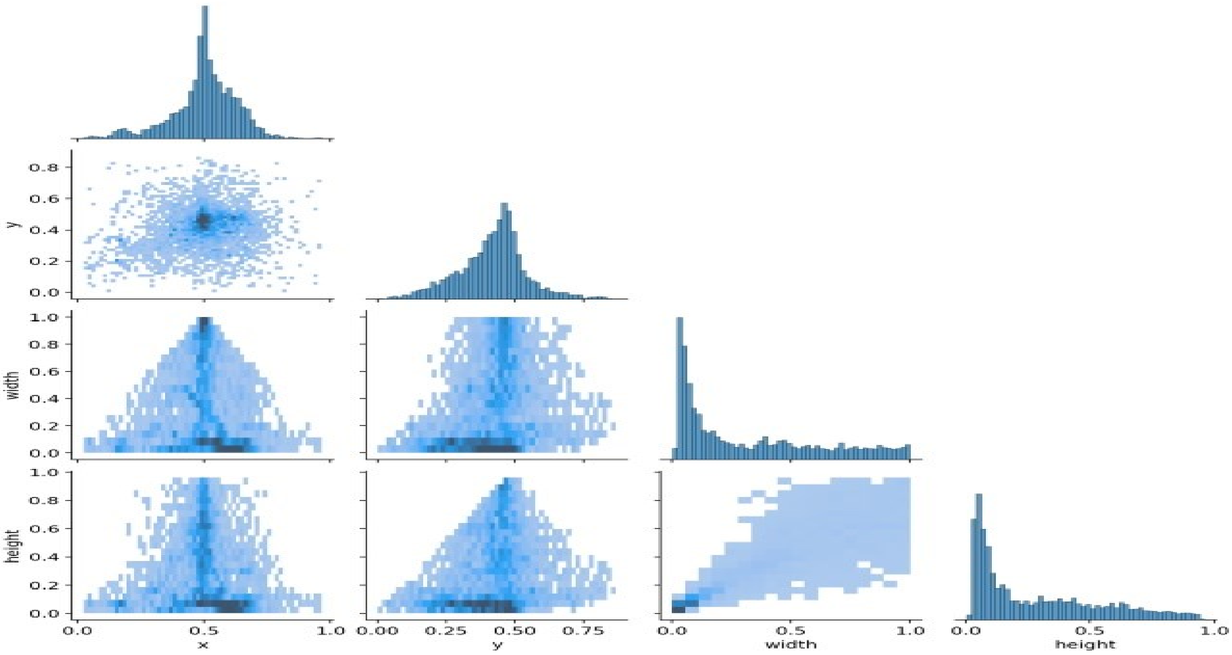


Fig 3. Distribution of the Sizes and Positions of drone in dataset

The backbone network of YOLOv5x consists of a deep CNN which retrieves features from the input image. It is built on the ResNet design, which improves the efficiency of feature extraction. CSPDarknet53 in the basic architecture of YOLOv5x has been replaced by the ResNet. ResNet has the ability to extract robust features from the input images. The required changes have been done to adjust ResNet with the neck and head network of the YOLOv5x. The feature map generated by the ResNet is compatible with different layers of YOLOv5x. ResNet has a sufficient number of resources for training the dataset and interfaces. The integrity of the ResNet is monitored using performance metrics. The purpose of changing the CSPDarknet53 to ResNet help in down-sampling operations, and skip connections which help in feature extraction at different spatial resolutions. ResNet uses CSP connections at different levels in the network which enable reuse and reduced computational efficiency. The CSP connections enhance the way of representation of the features and retrieve contextual data from the images. This network can handle images with different backgrounds, shapes, poses, and sizes. It makes a foundation for the detection of drones in the input images.

It is responsible for aggregating and refining the features retrieved by the backbone network. To achieve this, YOLOv5x employs a Spatial Pyramid Pooling (SPP) module in its neck network, which enables the model to capture features at various scales in an efficient manner. It incorporates many stages after the backbone network for the refinement of features extracted by the backbone network. All these stages focus on enhancing the presentation of extracted features, context capturing, and information aggregation. All these steps lead to more accurate detection of drones in minimum time.

The head network of YOLOv5x is in charge of forecasting bounding boxes and class likelihoods for objects in the image. Contrasting classical object recognition approaches that use anchor boxes, YOLOv5x uses an anchor-free detection approach that eliminates the need for anchor boxes, resulting in more precise object detection. It is trained by the backbone network and intermediate stages to detect drones in input images. Due to the process of training, it is able to put the boundary box on the detected drones.

The last stage is post-processing. The predictions made by the head network of YOLOv5x are post-processed to enhance the final results by eliminating redundant detections. This is accomplished using a Non-Maximum Suppression (NMS) algorithm, which removes overlapping detections and retains the most confident detections for each object. YOLOv5x employs a layered architecture that empowers it to achieve top-notch performance in object detection tasks, exhibiting high precision and efficiency.

4. Result Analysis

The evaluation of performance metrics for drone detection may differ based on the technology employed and the intended use case. YOLOv5x algorithm's drone detection performance may be influenced by several factors, like drone size, speed, lighting conditions, and weather. Hence, it's necessary to assess the performance metrics in various scenarios to ensure that the algorithm can function efficiently in diverse environments. YOLOv5x is a more advanced version of the Yolov5 object detection algorithm, which employs a larger and more intricate model architecture. The training and validation procedures for Yolov5x are similar to those for the standard YOLOv5 algorithm. To train a Yolov5x model, the dataset is first split into training and validation sets using the --split flag when running the train.py script. During the training phase, YOLOv5x generates several metrics like training loss, validation loss, average precision, and recall. These metrics are used to access the model's performance during training and make necessary adjustments to hyperparameters.

DOI: 10.6977/IJoSI.202409_8(3).0008
M. Kumar, U. Pilania /Int. J. Systematic Innovation, 8(3), 99-114 (2024)

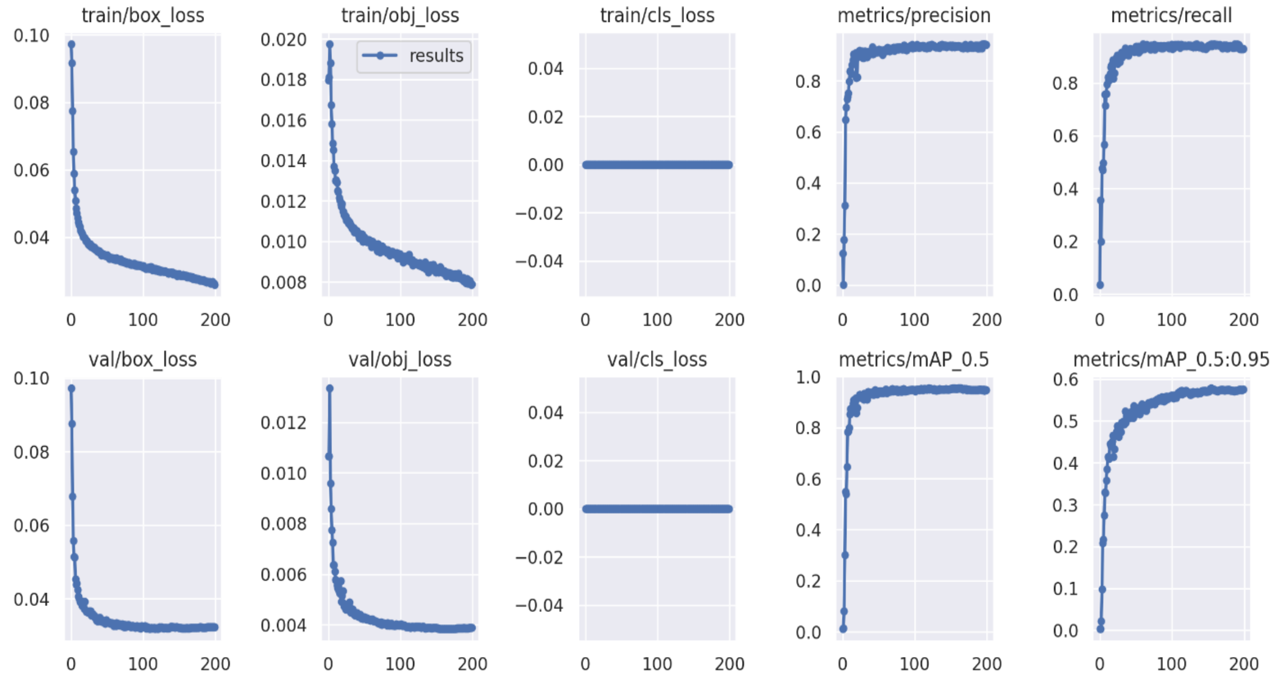


Fig 4. Training and Validation Loss

Training a YOLOv5x model can be computationally demanding, requiring a high-performance GPU and sufficient memory to effectively train the model. The x-axis in **Figure 4** represents the number of epochs, which is set to 200 for this work. The required number of epochs to train YOLOv5x for drone detection may vary depending on factors such as the dataset size, the complexity of the detection task, and the hardware configuration.

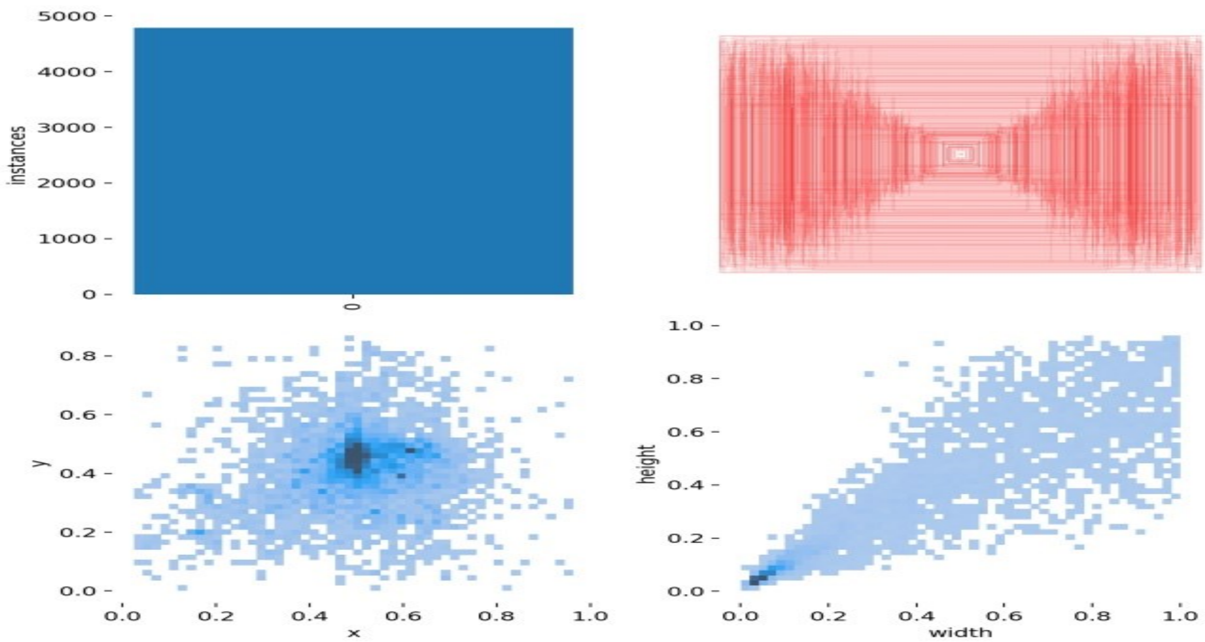


Fig 5. Distribution and Histogram of Drone in Training Set

Box loss, as represented in **Figure 4**, measures the change between the expected and true bounding box coordinates of an object. The Fig. 5 represents the distribution and histogram of the drone training set. Distribution means how data is stored in the dataset and histogram is used to represent this data graphically. Distribution helps in gaining knowledge about the features of data in the dataset. A histogram is also used to check the patterns in the dataset. The dataset is properly annotated and formatted as per the requirement of YOLOv5x. During training the number of input images, average number of drones per image, and maximum number of drones per image have been calculated. The occurrence of drones is retrieved and counted to check the patterns in the dataset. By using these patterns, it creates the histogram for the input. It is typically calculated using a regression loss function such as MSE or plane L1 loss. Eq. (1) and (2) represent the bounding box and object box mathematically.

$$L_{box} = \lambda_{coordinate} \sum_{n=0}^{k^2} \times \sum_{m=0}^{l} [\, 1_{n,m}^{obj} [(o_n - \hat{o}_n)^2 + (h_n - \hat{h}_n)^2]] + \lambda_{coordinate} \sum_{n=0}^{k^2} \times \sum_{m=0}^{l} [\, 1_{n,m}^{obj} [(\sqrt{c_n} - \sqrt{\hat{c}_n})^2 + (\sqrt{e_n} - \sqrt{\hat{e}_n})^2]] \quad (1)$$

where k is the numeral of grid cells in the input, l is the numeral of bounding boxes expected by each grid cell, ô, ĥ, ĉ, ê are the predicted coordinates of the center, width, and height of the box in cell n, and o_n, h_n, c_n, e_n are the ground-truth coordinates of the center, width, and height of the box in cell n. The term $1_{n,m}^{obj}$ is an indicator function that equals 1 if the m bounding box in cell n is responsible for detecting an object, and 0 otherwise. The parameter $\lambda_{coordinate}$ controls the weight of the box localization loss.

$$L_{obj} = \sum_{n=0}^{k^2} \times \sum_{m=0}^{l} 1_{n,m}^{obj} \left[\log\left(\sigma(\hat{\imath}_{n,m})\right) + (1 - 1_{n,m}^{obj}) \log\left(1 - \sigma(\hat{\imath}_{n,m})\right) \right] \quad 1 \quad (2)$$

where $\hat{\imath}_{n,m}$ is the predicted objectness score for the m^{th} bounding box in cell n, and σ is the sigmoid function. In the term $1 - 1_{n,m}^{obj}$ is an indicator function that equals 1 if the m^{th} bounding box in cell n is not responsible for detecting an object and 0 otherwise.

$$loss_{YOLOv5} = loss_{bounding\ box} + loss_{classification} + loss_{confidence} \quad (3)$$

$$\mathrm{loss_{bounding\ box}} = \lambda_{if} \sum_{a=0}^{b^2} \times \sum_{c=0}^{d} E_{a,c}^{g} h_g (2 - K_a X n_a) \left[(x_a - x_a^{\prime c})^2 + (y_a - y_a^{\prime c})^2 + (w_a - w_a^{\prime c})^2 + (h_a - h_a^{\prime c})^2 \right] \quad (4)$$

$$loss_{classification} = \lambda_{classification} \sum_{a=0}^{b^2} \times \sum_{c=0}^{d} E_{a,c}^{g} \times \sum_{C \in c_l} L_a(c) \log\left(LL_a(c)\right) \quad (5)$$

$$loss_{confidence} = \lambda_{confidence} \sum_{a=0}^{b^2} \times \sum_{c=0}^{d} E_{a,c}^{confidence} (c_i - c_l)^2 + \lambda_g \sum_{a=0}^{b^2} \times \sum_{c=0}^{d} E_{a,c}^{g} (c_i - c_l)^2 \quad (6)$$

Loss function in YOLOv5x plays significant role in enhancing the outcomes. It is the combination of boundary box, classification and the confidence as represented by Eq. (3). Eq. (4) shows how bounding box can be calculated. Where h′ and w′ shows the height and width of target object in the input image. x_a and y_a represent the coordinates of the input image. λ_{if} indicated whether the target object exists in the image or not. $\lambda_{confidence}$ in Eq. (5) and (6) represent loss coefficient. $\lambda_{classification}$ is classification loss coefficient, c_l is class, and c is the confidence score.

DOI: 10.6977/IJoSI.202409_8(3).0008
M. Kumar, U. Pilania /Int. J. Systematic Innovation, 8(3), 99-114 (2024)

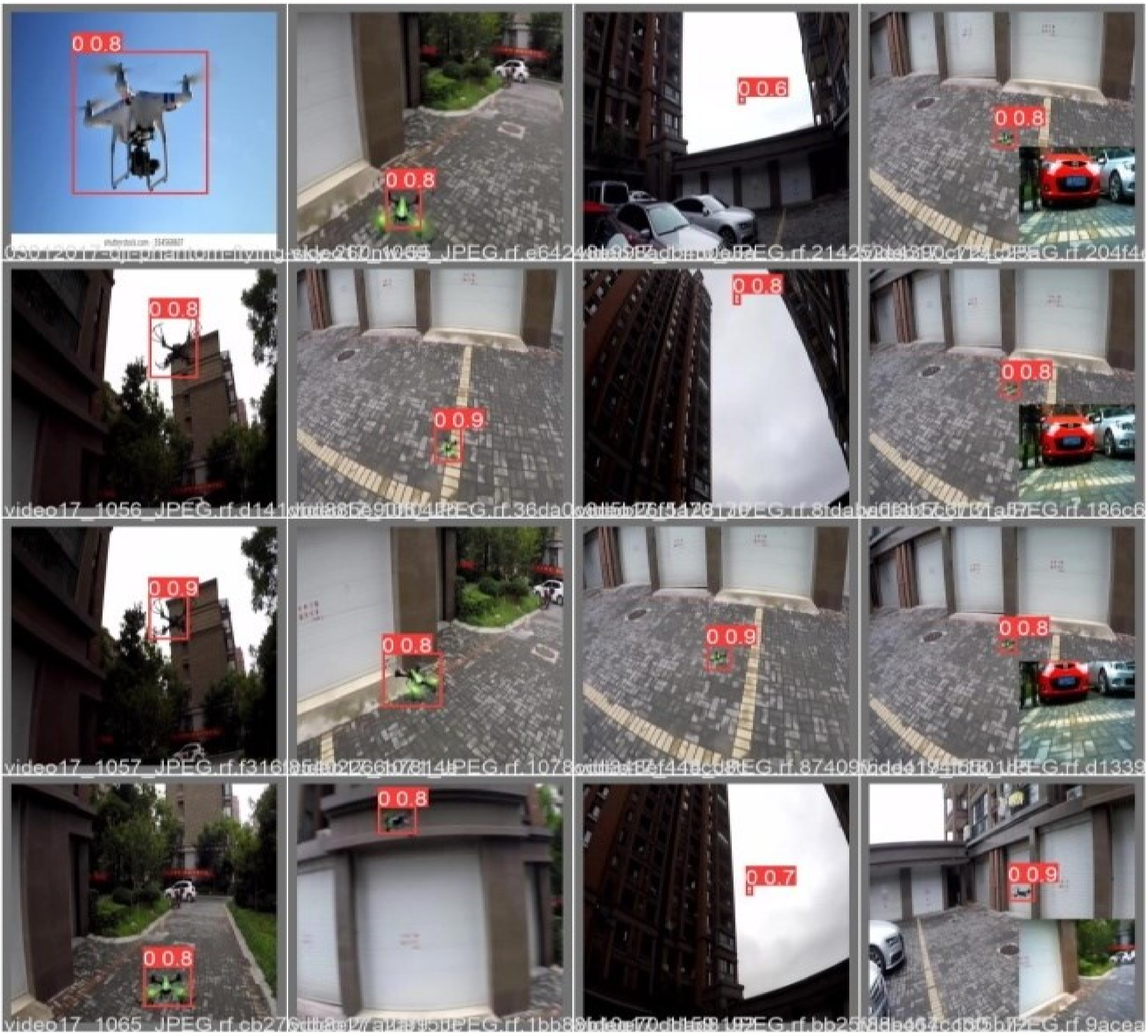


Fig 6. Drone Detection

Object loss is the loss function which is used to decide if an object exists or not in input image. It is calculated using a sigmoid activation function and binary cross-entropy loss Liu, X., et al. (2023) The purpose of object loss is to adjust the model's parameters so that it can more accurately detect objects in an image. Class loss is used for multi-class classification responsibilities. It measures the difference between the predicted probabilities and the true probabilities of an object belonging to different classes. The class loss is computed using a softmax activation function and categorical cross-entropy loss. The purpose of class loss is to enhance the model's parameters so that it can better classify objects into their respective categories. For the reduction of object loss in the proposed work, we have used high quality dataset and a regress training process.

DOI: 10.6977/IJoSI.202409_8(3).0008
M. Kumar, U. Pilania /Int. J. Systematic Innovation, 8(3), 99-114 (2024)

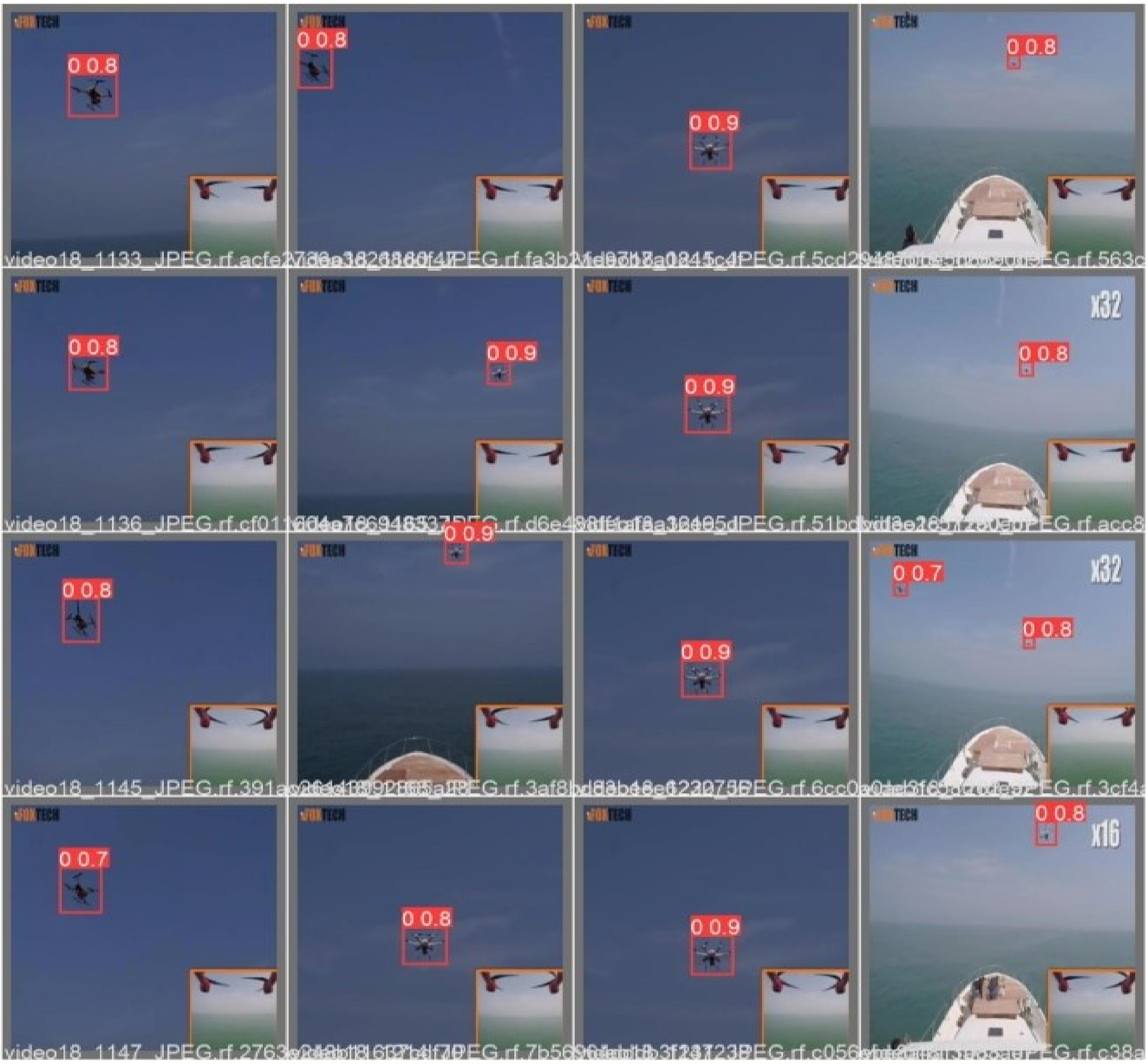


Fig 7. Drone Detection

False detection could be defined in terms of false positives which means the detection of a drone where it does not exist, false negative means the actual drone is present but the model is not able to detect it, and sometimes the model may label incorrectly to the detected object. As shown in Figure 7 the drones are detected but the accuracy of detection is not 100% which means the model is not 100% sure whether the detected object is a drone or some other object. As per the results model has a 4.7% false detection rate.

Confusion matrix is a commonly used table for assessing the enactment of a classification approach. It matches the expected labels of the approach with the actual labels in the validation or test set. In the context of object detection, a confusion matrix can be utilized to assess the model's ability to detect objects of various classes Abdallah, S. M. (2024). As shown in Figures 6 and 7, the results of YOLOv5x detection demonstrate that the deep learning method could detect drones with high accuracy.

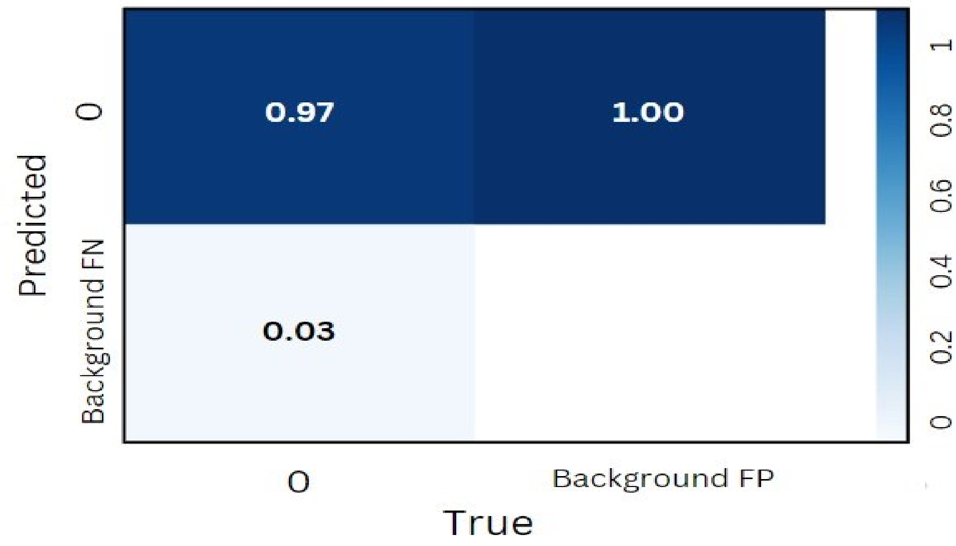
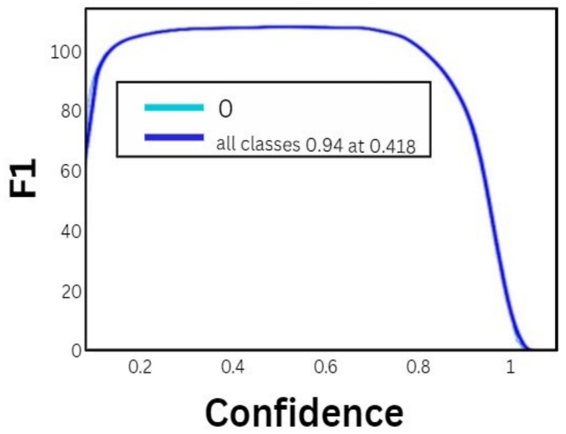


Fig 8. Confusion Matrix

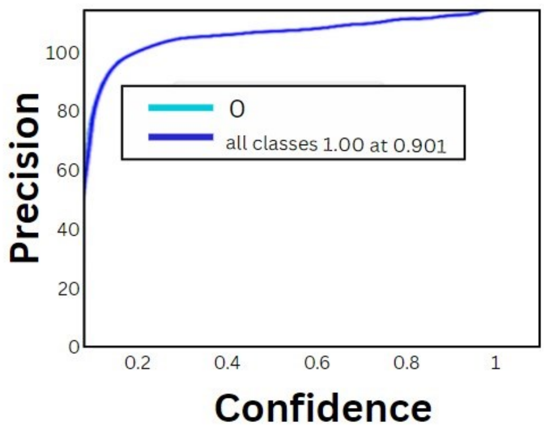
The validation or test set produces a confusion matrix that represents the numeral of True Positive (TP), false positive, false negative, and true negative expectations for each class, as illustrated in **Figure 8**. This information can be analyzed to evaluate the method's performance in identifying objects of diverse classes and identify opportunities for improvement Cheng, S., Zhu, Y., & Wu, S. (2023).

i) Reliability and Accuracy Analysis

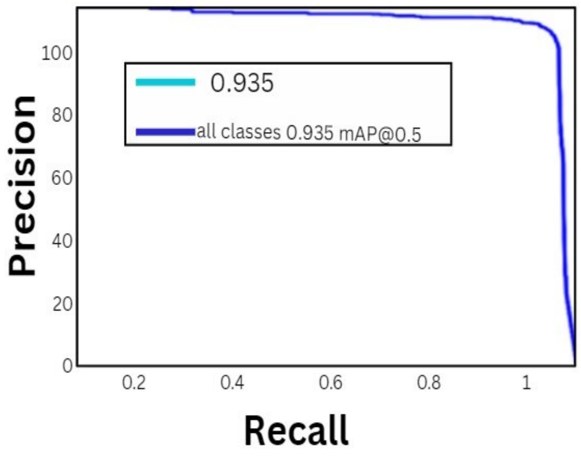
The reliability and accuracy analysis of drone detection using YOLOv5x assesses the model's ability to consistently perform over time and under different conditions. This analysis includes measuring the precision, F1-Score, AUC, and recall rates over multiple test runs, and checking for false positives and false negatives. Also, evaluating the model's performance under variable environmental circumstances such as changes in light, weather, drone size, and drone distance.



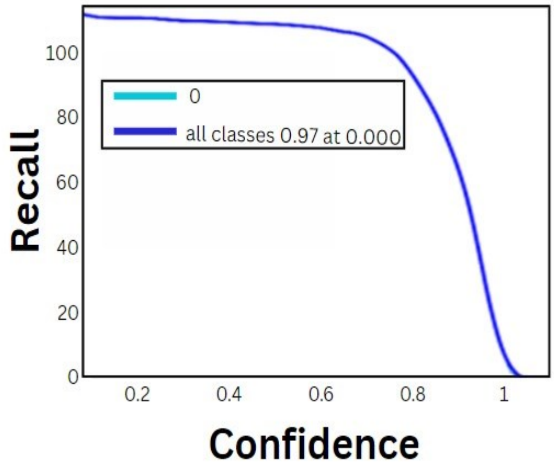
a) F1-Score



b) Precision



c) AUC



d) Recall

Fig 9. Performance Metrics

The proposed work is found to be reliable and accurate based on values of performance metrics obtained from experimental results.

F1-Score: It is a performance metric which uses precision and recall in measuring a model's accuracy. It is the harmonic mean of precision and recall as per authors Li, S., et al (2023) in Eq. (7), and the F1-Score value was determined to be approximately 0.94 for the research work.

$$F1 = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \quad (7)$$

Precision: It measures the accuracy of the positive predictions made by the system, as represented in Eq. (8). In YOLOv5x, precision is utilized to assess the model's capability to accurately identify instances of a particular object class according to Alsanad, H. R.,. et al. (2023) The precision was calculated to be around 0.90 from the experimental work.

$$Precision = \frac{True\ Positives}{(True\ Positives + False\ Positives)} \quad (8)$$

TP refers to a numeral of positive detections accurately made by the system, while False Positives refer to the numeral of positive findings incorrectly made by the system.

AUC: The AUC is a widely used metric to assess the Performance of machine learning systems in diverse applications, including drone detection. In drone detection, the AUC is commonly used to find the machine learning model's ability to accurately differentiate between the drone and non-drone signals as per authors Kumar, S., et al. (2024). At a learning rate of 0.5, the mAP or AUC was determined to be 0.95.

Recall: Recall measures the accuracy of the positive cases detected by the system. In the context of YOLOv5x, recall can be utilized to evaluate the model's ability to detect all instances of a particular object class in an image according to Kumar, S., et al (2023). It can be calculated using Eq. (9). The recall for the work was calculated to be equal to 0.97.

$$Recall = \frac{True\ Positives}{(True\ Positives + False\ Negatives)} \quad (9)$$

The TP denote to the numeral of accurate positive detections made by the system, while the False Negatives correspond to the numeral of positive detections missed by the model.

The computational outcomes of the work are illustrated in **Figure 9**. The F1-Score value was determined to be approximately 0.94, while the precision was calculated to be around 0.90. At a learning rate of 0.5, the mAP or AUC was determined to be 0.95. Furthermore, the recall for the work was calculated to be equal to 0.97. From the simulation, it has been determined that the proposed work is found to be reliable and accurate.

ii) Complexity Analysis

The complexity analysis of drone detection using YOLOv5x entails assessing the computational resources necessary to perform object detection on a given dataset. This includes measuring the computational workload per image, the memory consumption during processing, and the detection and tracking time for drones. Moreover, the analysis involves identifying any processing bottlenecks and optimizing the model to reduce computational complexity. Figure 10 depicts the detection time required for drones, which was found to be shorter than that of previous works in the field. For storing the data and instruction cloud google collab was used. Furthermore, YOLOv5x's image processing does not require significant memory compared to video databases.

iii) Security Analysis

YOLOv5x serves as a powerful tool for real-time object detection, extending its capability to both images and videos. To bolster the security of the proposed work, we meticulously curated a diverse dataset encompassing various critical variables, including backgrounds, lighting conditions, drone types, and other essential factors. Subsequently, we applied data augmentation techniques, introducing features like rotation, transformation, scaling, and lighting variations.

Our integration strategy extends to encompass security infrastructure elements, including CCTV cameras, alarm systems, and autonomous response mechanisms. Moreover, we diligently ensured that our proposed system adhered to all relevant regulations pertaining to privacy and surveillance.

iv) Comparison of Proposed Work

The proposed research has yielded superior results compared to the work in the paper Aydin, B., & Singha, S. (2023). In this research work, a mAP of 0.95 at a learning rate of 0.5 has been achieved. YOLOv5x, which employs a lightweight design and ResNet architecture, outperformed YOLOv5 in drone detection speed. YOLOv5x's PyTorch framework and fine-tuning the last layers of the architecture helped to optimize for the customized dataset. The values of the momentum, learning rate, and batch size are also customized. The proposed work underwent 200 iterations of training on top of reassigned weights for the roboflow drone 1 dataset.

The mAP is calculated 0.95 at a learning rate of 0.5 which is found to be better compared to the work by authors Aydin, B., & Singha, S. (2023). The precision is calculated 0.901 for the proposed work which is 0.01 smaller than the state of art work. The metric recall is calculated 0.97 which is much better than the work performed by authors Aydin, B., & Singha, S. (2023). F-1 score is calculated as 0.94 and for the state of art work, it is found to be 0.89.

Figure 10 represents a comparison of the performance of the proposed system and the previous system in terms of 4 assessment metrics: precision, recall, F1 score, and mAP@0.5. We also calculated the AUC which is found to be 95.3% but in comparison work, it has not been calculated so we are not able to compare the AUC metric.

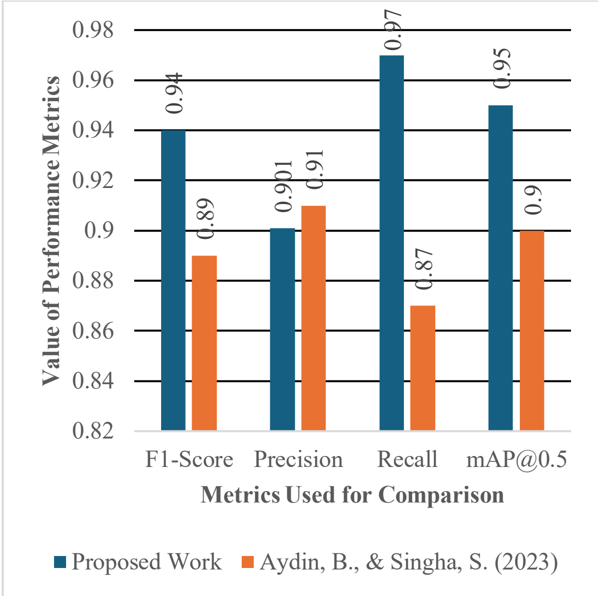


Fig 10. Comparison Graph

5. Conclusion

In conclusion, the proposed framework for the detection of drones using the YOLOv5x deep learning algorithm provides an effective solution for enhancing security surveillance systems. The framework leverages a combination of hardware and software components, including drones, cameras, and computer systems, to accurately detect unauthorized drone activity. The framework incorporates real-time monitoring capabilities, allowing for quick and effective responses to security threats. The YOLOv5x algorithm is chosen due to its high accuracy and low computational cost, making it an ideal solution for security surveillance systems. The mAP is calculated 0.95 at a learning rate of 0.5, precision is 0.901, recall is 0.97, AUC is 95.3, and F-1 score is 0.94.

The proposed framework is designed to be scalable, and customizable, enabling easy integration into existing security systems. It provides a reliable and effective method for detecting drone activity and enhancing the security of critical infrastructure, public events, and other sensitive locations.

Overall, the proposed framework is a valuable contribution to the field of security surveillance, providing a comprehensive solution for detecting unauthorized drone activity. Further development and customization of the framework can be done to meet specific security requirements.

Reference

Abdallah, S. M. (2024). Real-Time Vehicles Detection Using a Tello Drone with YOLOv5 Algorithm. Cihan University-Erbil Scientific Journal, 8(1), 1-7.

Ahmad, T., Cavazza, M., Matsuo, Y., & Prendinger, H. (2022). Detecting human actions in drone images using YOLOv5 and stochastic gradient boosting. Sensors, 22(18), 7020.

Aker, C., & Kalkan, S. (2017). Using deep networks for drone detection. In 2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS) (pp. 1-6). IEEE.

Al-Qubaydhi, Nader, Abdulrahman Alenezi, Turki Alanazi, Abdulrahman Senyor, Naif Alanezi, Bandar Alotaibi, Munif Alotaibi, Abdul Razaque, Abdelaziz A. Abdelhamid, and Aziz Alotaibi. (2022). Detection of unauthorized unmanned aerial vehicles using YOLOv5 and transfer learning, Electronics 11, no. 17 (2022): 2669.

Akyon, F. C., Eryuksel, O., Ozfuttu, K. A., & Altinuc, S. O. (2021). Track boosting and synthetic data aided drone detection. In 2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 1-5). IEEE.

Anwar, M. Z., Kaleem, Z., & Jamalipour, A. (2019). Machine learning inspired sound-based amateur drone detection for public safety applications. IEEE Transactions on Vehicular Technology, 68(3), 2526-2534.

Aydin, B., & Singha, S. (2023). Drone detection using yolov5. Eng, 4(1), 416-433.

Cheng, S., Zhu, Y., & Wu, S. (2023). Deep learning based efficient ship detection from drone-captured images for maritime surveillance. Ocean Engineering, 285, 115440.

De Galiza Barbosa, Felipe, Samuel J. Galgano, Ariel L. Botwin, Aline Bobato Lara Gongora, Giovanna Sawaya, Ronaldo H. Baroni, and Marcelo. (2023). A. Queiroz. "Genitourinary imaging." In Clinical PET/MRI, pp. 289-312. Academic Press, 2023.

Kumar, M., Pilania, U., Bhayana, T., & Thakur, S. (2024). Utilizing YOLOv5x for the Detection and Classification of Brain Tumors. In 2024 2nd International Conference on Disruptive Technologies (ICDT) (pp. 1343-1348). IEEE.

Kumar, M., Pilania, U., Thakur, S., & Bhayana, T. (2024). YOLOv5x-based Brain Tumor Detection for Healthcare Applications. Procedia Computer Science, 233, 950-959.

Kumar, S., Pilania, U., & Bala, R. (2024). A novel hybrid deep belief Google network framework for brain tumor classification. International Journal of Systematic Innovation, 8(2), 95-112.

Li, S., Yang, X., Lin, X., Zhang, Y., & Wu, J. (2023). Real-time vehicle detection from UAV aerial images based on improved YOLOv5. Sensors, 23(12), 5634.

Pilania, U., Kumar, M., Singh, S., & Adhalkha, P. (2023). A Deep Ensemble Learning Network-based Approach to Detect Neurodegenerative Diseases. In 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), 2023, pp. 1033-1038. IEEE.

Taha, B., & Shoufan, A. (2019). Machine learning-based drone detection and classification: State-of-the-art in research. IEEE access, 2019, 7, 138669-138682.

Wu, M., Xie, W., Shi, X., Shao, P., & Shi, Z. (2018). Real-time drone detection using deep learning approach. In Machine Learning and Intelligent Communications: Third International Conference, MLICOM 2018, Hangzhou, China, July 6-8, 2018, Proceedings 3 (pp. 22-32). Springer International Publishing.

Zhu, X., Lyu, S., Wang, X., & Zhao, Q. (2021). TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 2778-2788).

AUTHOR BIOGRAPHIES

Dr. Urmila Pilania is currently working as Associate Professor in the Department of Computer Science & Technology at Manav Rachna University, Faridabad. She is having 14 years of experience in academics and research. Dr. Urmila Pilania received the Engineering degree (B. TECH. in Computer Science & Engineering) in 2007 from Guru Jambheshwar University of Science and Technology, Hisar - Haryana (India). Dr. Urmila has earned master's degree (M.TECH. in Computer Science & Technology) in 2009 from Vaish College of Engineering, Rohtak-Haryana. Dr. Pilania has been awarded with the doctoral degree (Ph.D. in Computer Science & Technology) degree in 2021 from Manav Rachna University, Faridabad, India. Her primary research interests are in Information Security, Image Processing, Machine Learning and Artificial Intelligence. Dr. Urmila has published many research papers related to Information Security, object detection and classification. She is author of more than 34 research publications with journals/Conferences of repute and some papers are under review or in press for publication with leading publisher.

Dr. Manoj Kumar is presently working as Associate Professor in the Department of Computer Science & Technology, at Manav Rachna University, Faridabad. He has over 14 years of experience including academics, industry & research. He has earned his Ph.D. in the year 2022, in the area of Computer Vision. The title of the thesis work is "Automatic Detection of Moving Object in a Video for smart Surveillance system". His area of expertise is Computer Vision, Artificial Intelligence, Machine Learning, Image Processing, IOT and Blockchain. Dr. Kumar has published many research papers related to moving object detection and tracking in videos and also extracts features moving object/vehicles from video in real-time video surveillance system. Dr. Kumar has published more than 10 papers in SCIE, SCOPUS, WEB OF SCIENCE indexed journals. Dr. Kumar has presented more than 20 papers in various Scopus Indexed International Conferences. The proceedings of the conferences are Scopus indexed and few of them published as a Book chapters. He has published 4 National Patents. He has participated in various workshops & Faculty development programs.