Enhancing Autonomous Drone Navigation: YOLOv5-Based Object Detection and Collision Avoidances

Kalrav Gediya[1*], Pooja Bhatt[2*], Swati Sharma[3*], Kalpesh Jadav[4*]

[1]*Research Scholar, Computer Engineering Department, PIET, Parul University, Vadodara, Gujarat, India*
[2] *Assistant Professor, Computer Engineering Department, PIET, Parul University, Vadodara, Gujarat, India*
[3]*Assistant Professor, Computer Engineering Department, PIET, Parul University, Vadodara, Gujarat, India*
[4]*Associate Professor, Electronics & Communication Engineering Department, PIET, Parul University, Vadodara, Gujarat, India*
[1]**2303032010011@paruluniversity.ac.in*, [2]*pooja.bhatt28403@paruluniversity.ac.in*,
[3]*swati.sharma33111@paruluniversity.ac.in*, [4]*kalpesh.jadav@paruluniversity.ac.in*

Abstract

In recent years, drone technology has seen profound advancements, especially with regards to safe and autonomous operation, which heavily relies on object detection and avoidance capabilities. These autonomously functioning drones can operate in challenging environments for tasks like search and rescue operations, as well as industrial monitoring. The present research focuses on enhancing object detection for autonomous drones by utilizing publicly available image datasets instead of custom images. Datasets like VisDrone, DroneDeploy, and DOTA contain a plethora of stunning, real-life images that make them ideal candidates for improving the accuracy and robustness of object detection models. We propose an optimized method for training the YOLOv5 model to enhance object detection. The collected dataset undergoes evaluation from precision, recall, F1-score, and mAP through both CNN and YOLO models. The findings show that using YOLOv5 deep architecture to implement real-time object detection and avoidance in UAVs is more efficient than traditional CNN approaches.

Keywords: *Autonomous Drone, CNN, Yolo V5, Sensor Integration, Deep Learning.*

1 Introduction

The growing economic and technological globalization has led to an increase in the usage of drones (i.e., unmanned aerial vehicles) in a broad range of sectors other than military. From UAVs to sUAVs, the industry is experiencing a boom in the demand for low-altitude remote-controlled vehicles for surveying, mapping, disaster quelling, and critical infrastructure surveillance. The ease and flexibility with which drones, a type of VTOL UAVs, can be deployed has made them increasingly popular.

As drone technology matures, effective obstacle detection and avoidance techniques are principal problems. Detecting and navigating around obstacles autonomously is crucial for the safety and efficacy of UAVs. With the miniaturization of UAVs comes the problem of bulky sensors (LiDAR and radar) being less useful than before. Cameras on the other hand, besides being lightweight and power efficient, can also provide effective vision-based navigation making them the primary choice for object avoidance or obstacle detection.

To ensure better training of the object detection models, the publicly available drone image datasets like DOTA, VisDrone, and DroneDeploy offer a great variety of aerial imagery. Such datasets enable reliable detection capability generalization for UAVs.

2 Literature Study

[1] The challenge of detecting 2D objects in images is still an issue ever since computer vision was invented. It goes without saying that it involves detection, classification, categorization, and recognition. They proposed a system that classifies methods of recognition of objects based on altitude of flight and the application zone corresponding to the given altitude. For this purpose, R-CNN, an object detector that works in two phases and performs selective search of potential objects, was used. They worked with publicly available datasets that were accompanied with labelled images, which made the categorization and detection of objects straightforward. The system is expected to perform poorly with unlabelled datasets. Therefore, in order to advance this direction of research, it will be necessary to use non homogenous data from various types of sensing devices.

[2] In order to solve the problem with the top view angles and the upwind displacement, a UAV deep learning automation is needed. Their proposed approach consists of three modules. The first approach deals with the problem of how the current object detection algorithms and models can be used for the image data captured from the drone with a convolutional neural network serving as a fixed feature extractor. The second approach is the fine-tuning of CNNs where the network pretrained on the ERSN is modified by continuous backpropagation of the output error signal. The last approach is based on the use of several pretrained models available in the TensorFlow model zoo. Because convolutional neural networks are very expensive to work with, transfer learning was utilized for training using a smaller subset of the data. The average time for detecting objects per frame is around 140 ms.

[3] Object detection from images captured by Unmanned Aerial Vehicles (UAVs) has been a major challenge in computer vision in the past. This is especially the case because of the different sizes of objects such as people, buildings, water, hills and others. In this study, they propose to use ensemble transfer learning to enhance the performance of the foundational models for multiscale object detection in drone imagery. When used in conjunction with a test time augmentation pipeline, the approach detects objects of different sizes in UAV images by using multiple models and voting. They found that this approach has a limited ability to detect novel objects like awning tricycles and standard tricycles because they were not present in the training datasets. The methodology presented in this paper can improve multiscale object detection, especially for producing better quality Ortho mosaics of the objects situated at the edges of orthophotos.

[4] This is because identifying and classifying multiple objects in one frame is a big challenge. The accuracies have greatly increased with the improvement in deep learning approaches. This paper proposes a new proprietary algorithm for object detection and classification in a single frame as a way of achieving high accuracy with real time processing. The proposed system is implemented using MobileNet in conjunction with SSD architecture in order to improve accuracy. Some of the items studied in this research are hard to recognize precisely because of the frame noise or the poor video quality. Another limitation of this research is that the detection and classification models were trained for devices with limited computational resources which hampers the system's capability to work on high-definition video.

[5] In comparison with the aforementioned methods (RADAR, acoustic systems, RF signal analysis, etc.), computer vision is widely employed for autonomous drone identification due to its reliability. They are voted by deep learning algorithms rather than by a bunch of computer vision methods because the performance of the deep learning algorithms is quite good. In its drone detection and tracking system, the static wide-angle with a lower-angle camera positioned with a rotating turret was used in this study. In order to use both time and memory optimally, they devised a combined multi-frame deep learning detection algorithm that overlays the wide-angle static camera frame over the zoomed camera frame.

[6] To avoid chaos and uphold law and order, necessary preventive measures like monitoring and identification of unauthorized drones entering the restricted area or zone are to be taken. Latest effective real-time object or image detection Modus Operandi being YOLO will allow for the model to be trained to distinguish between a drone and a similar object such as a bird in order to avoid high rates of false alerts. This surveillance system involves three components: a 360-degree CCTV camera to capture real-time videos, drone detection implementations into that footage through a model already trained on drones, and another trained model. The Advanced Standard

selected for the research is the YOLO-v4 algorithm known for real-time performance, speed, and accuracy in object detection. The small size, high velocity, and high altitude at which drones typically fly would make it difficult to detect them. YOLO v3 is the predecessor of the current YOLO versions 7 and 8.

[7] The field of multi-class object detection has seen spectacular progress in the last few years due to the emergence of deep CNNs. A growing number of drones create more and more new and interesting applications, as well. Analysis and recognition of any component in the video are necessary to understand this information. This study is therefore training and testing the CNN networks based on sets of large images taken by a UAV from a fixed height over the desert in the United Arab Emirates (UAE). Hence, three leading CNN architectures have been fine-tuned, retrained, tested, and evaluated for detecting objects, classified into three distinct categories: palm trees, groups of animals or cattle, and animal sheds. WCS for multi-class detection is 0.77 with SSD-500/VGG-Net, and it is 0.83 with SSD-500/ResNet. The F1-score of SSD-500 with VGG-16 is a tad lower than the rest. Very much lower than that of SSD-500 with ResNet, probably indicating the influence of the number of convolution layers. YOLOV3 might exceed this; its detection includes 53 convolutional layers, whereas the SSD architecture is inclusive of five convolutional layers. This helps build the improvement potential of YOLOV3 over SSDs.

[8] The issue of designing an intelligent UAV is one of the most engaging and difficult. There has been a considerable growth in the use of deep learning and computer vision technologies for the development of fully autonomous drones lately. The use of Convolutional Neural Networks (CNNs) in computer vision includes providing help to drones for image and video content analysis and interpretation.
This paper reviews recent research on deep learning-based methods for object detection from UAVs. It covers the previously mentioned methods like CNNs, R-CNNs, Fast R-CNNs, YOLO, SSDs, among others. The paper then brings to light the major works and methodologies which served as building blocks in the path to advance drone object detection. In this research, some approaches were applied, reaching approximately 70%-80% accuracy.

[9] This paper provides a comprehensive review of the state-of-the-art in deep learning-based object detection approaches on low-altitude UAV datasets. Considered foci are justified by relatively much less literature on this topic in comparison to standard or remote-sensing datasets. In this study, a wide spectrum of algorithms is discussed, including Faster RCNN, Cascade RCNN, R-FCN, YOLO, and its variants, SSD, RetinaNet, and CornerNet, paying specific attention to detectors at one, two, and higher levels for low-altitude UAV datasets.
The paper also mentions the challenges for object detection from low-altitude aerial VisDrone datasets, where even the state-of-the-art detector, CornerNet, achieves an average mAP of barely 17.41 while much higher value of 40.6 is being presented in Table VI. Yet, the first thing talked about in that paper was the resolution of low-altitude aerial images, being generally in the region of 2,000 by 1,500 pixels; most of these images in traditional datasets, like MS-COCO, are under 500 by 500 pixels.

[10] Breach of rules and laws leads to challenges in terms of safety and monitoring on a daily basis. From the loss of life that results due to flouting safety precautions and road laws to other emergency road incidents such as obeying traffic lights, surveillance, robbery, shooting, and explosions.
The researchers' goal was to prototype a video surveillance system based on three major data processing stages: moving object detection, recognition and tracking of moving objects as well as making decisions on how to react to events found significant. The very tasks are a wide spectrum ranging from anomaly detection and recognition of deviations, vehicle movement classification and tracking towards license plate recognition, and detection of safety equipment such as helmets for drivers and vests and boots for construction workers, and any other factors.
The background subtraction method gives results that are much too unreliable but permits the detection of objects. In actuality, some objects located in the background of detected items will not have been detected at all. Every robotic mobility system has its good parts and its bad parts, and Rover is no different. It has a limited ability to climb steep inclines-literally, wheel slippage; less ability to maneuver around objects when compared with other designs; track friction as one of its bigger drawbacks, and slow operational speed as one of its major drawbacks.

[11] The objective of the present work was to develop a drone-integrated video surveillance technology for aerial video and image analysis without the usual limitations. Though with high success rates in terrestrial images, the performance of previously developed object detection algorithms deteriorates drastically when applied to images taken from Unmanned Aerial Vehicles (UAVs). The researchers proposed a Sample Balancing Strategies Module to address the image sample training imbalance, especially the imbalance between positive and negative samples, as well as the imbalance between easy and difficult samples. Frequency and noisy representations usually make it hard to detect small objects in drone imagery, but the proposed method has performed better than the previously known algorithms. They also proposed introducing an SR GAN module with center-ness weights to sharpen the local feature map. Datasets for object detection under drone conditions is not very comprehensive as compared to those in ground-based datasets like ImageNet though the primary focus is to detect small objects. Hence, further research has mostly emphasized loss design along with proper selection for training samples and feature augmentation for improving detector performance.

[12] Countless applications have already served and continue to serve ethically and successfully in autonomous driving, semantic segmentation, search and rescue missions, and security monitoring because of object detection. While percent true recognition remains high with images captured from the ground, UAV images face challenges in identifying people or objects due to poses and scales; however, issues arise from artifacts like hats worn by people, diversity in postures, and backgrounds and are merged with the environment. The paper presents a new approach to identifying individuals in aerial images specifically for search and rescue operations, with details on the training procedure for the newly created high-resolution aerial database—HERIDAL. EfficientDET deep neural network is trained to solve the task of human identification using the newly developed database. Extracting prominent features between the system used by Croatian Mountain SAR teams (IPSAR) and the HERIDAL database study—the latter being dependent on them—is contrasted against the proposed approach. The results of the latter study prove to be slightly less optimistic than the former. This method might not work well in the high-lying areas; for instance, the hills. While those approaches can pick out people in the open, one-in-other clutters or closed matters, attempting to sense human presence becomes a very hard task.

[13] The goal is to develop an autonomous Unmanned Aerial Vehicle (UAV) that can independently follow trails and obstacle avoidance using Deep Neural Networks. The UAV should stay close to the center of the path by employing Convolutional Neural Networks (CNN) for navigation. There will be instances, however, when the UAV gets off the trail because of some external disturbances such as poor weather, which makes it impossible for the camera to see the trail.

[14] The primary target is object detection and tracking via camera-based technology mounted on UAVs. The paper puts forward an algorithm, based on deep learning, for identifying as well as tracking moving objects. However, maintaining such stability becomes a very difficult process when it comes to quickly changing backgrounds in the rugged field of observation of UAVs.

[15] The objective is to use the object detection and navigation systems for carrying life-saving medical supplies and patient packages in emergencies as well as drive precision agriculture with the right technology. An approach based on deep learning will be developed that will use GPS for navigation and object detection. The model to be developed will be run in a quadcopter drone. Finding ample data in an unstructured outdoor environment is challenging.

[16] Subsequent to the Covid outbreak there was, and continues to be an upsurge in demand for harnessing the power of delivery services. With drones, it is possible to offer and deliver express service packages that are both cheaper and quicker to deliver. A system proposed has, in it, a framework of navigation that comprises GPS, 9DOF IMU, and a barometer for determining where it is and sticking to the path that is predetermined. One of the major challenges in drone delivery has been to make the landings safe yet reliable in urban areas.

[17] Both temporary and progressive failures of hardware systems can drastically compromise task safety. Two error reduction strategies have been well-informed in learning-based navigation systems that multiply their success rates twofold and improve flight quality by 39%. It is difficult to emulate any traditional safety measures that

rely on redundancy for edge applications which have resource constraints.

[18] It is very difficult to control a UAV in complex indoor or outdoor scenarios due to the wide range of possible movements. This will introduce a method of object avoidance that calculates actual distances to objects rather than just relying on how close the UAV is to them. But with blurry images, it did not perform well with high speeds and also could not cope with low light since it did not detect enough feature points from the images captured.

[19] It relies upon a single commercial transmitter/receiver unit that utilizes sonar sensing to determine the range to the closest object in its view. Unfortunately, the approach suffers from somewhat low accuracy and high systems cost, making this approach unpractical.

[20] Legislation requires that all large machinery, such as cranes, temporarily inform the pilots of their exact location about the airport, once it is put into operation so that pilots know about such temporary air hazards. In this respect, the YOLOv3 technique merges neural networks, making it possible to extract information from images, that is, unusual objects of aviation. Accuracy arrived at such a value — 71%; it is not the most precise model, hence such a comprehensive area survey, since the model will need close-up images from various perspectives to be effective.

[21] The paper focuses on the development of a system for tracking social distancing in the COVID-19 period. The title of the paper will be 'Implementation of a UAV-based Social Distancing Tracking System (SDTS)' and the content is described within the following lines. YOLO-V3 Tiny will be adopted for use as a lightweight detection tool suitable for embedded systems of low processing capabilities. The drone is tested in a simulation only, thus further work to integrate thermal sensors for the detection of people with Covid-19 is suggested.

[22] The study underscores the importance of multivariate and accurate surveillance in drones to detect problems at an early stage. An early anomaly detection is important in environmental, structural, and infrastructural monitoring because it can prevent further escalation of issues. The simulation applied in this research to instruct the navigation recommender system mimics the real-life setting in a relatively abstract way.

[23] It presents a mathematical framework and a pragmatic methodology for solving a network design problem in which two objectives connected with investment in infrastructure are to be reconciled. The paper gives an overview of the marketing strategy at a high level in the aspect of how drones are to be utilized into the delivery process.

[24] The paper aims at introducing DBCMS as a shadow for standing the risk of medical personnel contributing their lives to the Covid-19 pandemic due to virus infection. The systems use drones for their operations, in that the proposed system has three tiers of operations directed at symptoms collection, patient recognition as sick, and notification of uncontrolled medical emergencies.

[25] It is presented in this paper how drones may be identified and classified in a systematic way using Deep Learning. YOLOv3 carries out object detection of dynamic as well as static objects. Very swift and real-time drone detection is possible through a Convolutional Neural Network-based system. Model fine-tunes over 150 epochs to reach a peak of 0.74 performance.

3 Proposed Methodology

Object Detection System Based on Yolo v5

It is comprised of two major components: the object detection system and the object avoidance system. In this design, the drone uses information provided by the object detection system to recalibrate its path and stop possible collisions. YOLOv5 is used in the recognition of prospective objects within the drone environment. The dynamic input of real-time pictures of the environment is achieved via the camera set on the drone. The object recognition system depends on YOLOv5 for accurate real-time object detection.Each object in an image is bounded by YOLOv5 with bounding boxes, this also estimates the probabilities of classes with a single CNN because it uses a single CNN to process all classes, it is much faster than more traditional approaches because all categories require some form of separate CNN. YOLOv5 improves both precision and recall by adopting more advanced tactics: anchor boxes, feature pyramid networks, spatial attention, etc.The building, tree, electric line, and anything else that may come into the way of a drone will be identified by the object detection system. Categorizing and furnishing the dimensions and locations of these objects will assess which way the impediment is to the drone and transmit that information to the object avoidance system. Based on this information, the object avoidance system will change the route of the flight to steer clear of the obstacles. In return, it will take input from the detection system, GPS, inertial sensors, and optical odometry to locate the velocity of the drone.

Once an object is registered, the system recalculates the path for the drone to navigate around the object. To make that decision, the algorithm takes into account several things such as the speed at which the drone is flying, what altitude it is flying at, and where the object is positioned. The system directs the drone to take an evasive path around the object at a safe distance. Besides, it gives visual and audible alerts to the drone operator about possible collisions and therefore gives enough time to take over manual control if necessary.

Object Detection System Based on CNN

The drone object detection and avoidance system will make use of a CNN-based approach. Convolutional Neural Networks (CNNs) are a deep-learning-type model and have been exceedingly successful in detecting as well as classifying objects. Adopting imbibing weight techniques for carrying out feature extraction and classification tasks, CNNs contain an assemblage of layers.

The two main components of the CNN-based object detection system will be an object detection model and a guidance system. Trained on a large images dataset, the object detection model will be able to contrast certain features for objects and features of the background. Processing output from the object detection model, guidance will then deliver real-time instructions to the drone to avoid collisions.

This shall involve the acquisition of an exhaustive annotated dataset of object images for the training of the object detection model. The dataset would comprise images of objects like buildings, trees, power lines, and other objects capable of causing damage to the drone. This will enable the model to be accustomed to detecting objects under varying circumstances while images are taken from different angles and lighting conditions.

Annotations for each image will be added to have real-world data for training of the model. Post annotation of the dataset, it will be partitioned into training, validation, and test datasets. Most images will be added to the training set; the validation and test sets will have smaller portions. The CNN architecture for the model for object detection—the likes of Faster RCNN or YOLO (You Only Look Once)—has been proven in applications. Those models shall take their feature extraction—after input of images through convolutional layers, pooling layers, and fully connected layers for classification of objects—mechanism from them.

The drone's camera will be the source which will provide input to the object detection model. The output of the model shall typically be a list of detected objects equated by their confidence values and descriptions of bounding boxes in which the objects are found in an image. Where the object is and what representation of the object is in fact shown in the image may respectively be implied by bounding boxes and confidence scores. The object detection information shall be beamed back to the guidance system for the necessary dynamic maneuvering and manoeuverance of the drone to avoid collisions. The guidance system will base its directions on how far or near objects are, against the drone, from the scores and bounding boxes of the object detection model. The system will give the drone operator warnings of hazard possibilities through visual and auditory directions on how to avoid them. For example, as it sees an object on the right-hand side of the drone, the guidance system may appropriately suggest that the operator turn left or also adjust the altitude for the drone.

The concept uses YOLOv5 to detect objects and avoid them. In this paper, a hunted approach will be presented that can be used to attain safe drone operation in versatile environments. The system provides accurate and timely information of potential objects for the avoidance system to correct the flight path of the drone, hence allowing safe operation for avoiding collisions between the drone and its surroundings. Another feature of the proposed system will be the integration of a LiDAR sensor. It will map the environment efficiently from all directions and also try to find its way how to reach destination B without any collision and possibly collecting data.

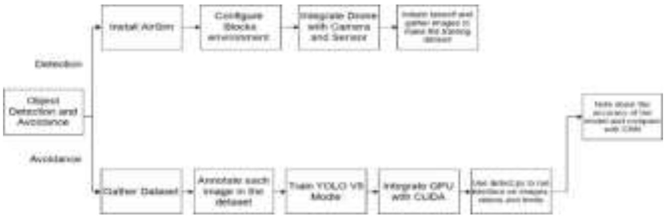

Fig 1: This image shows the architecture of the proposed system i.e., object detection and avoidance.

It lays down all the required steps, very briefly, in the form of a diagram for the implementation of the proposed system-two distinct phases that it has. The first phase is meant for the drone to avoid objects. The first step for this is to get AirSim installed in the device and then configure the Blocks environment. The drone shall have a camera and various sensors. Suitable Python code would be written and uploaded here. The very last step is to launch the drone, and through him collect images for the training dataset.

The second phase includes the dataset collection. After the dataset is collected and preprocessed, every image must be annotated. This annotation is needed to train the model. After this training is done, we shall combine it with a GPU plus CUDA and cuDNN. Later, we use a Python file to perform inferences on images, videos, and live feeds. This will also result in the precision, recall, and F1 scores of the model. These can be contrasted with the model created utilizing CNN.

Experimental Setup

1) Installation of AirSim, YOLOv5, and Unreal Engine: You can get AirSim through Unreal Engine; just need to download it from the Epic Games launcher. Sometimes better direct access to Unreal Engine can be achieved from that store. YOLOv5 and AirSim were installed based on instructions you can find at the Darknet GitHub repository and Microsoft AirSim GitHub repository, respectively.

2) Configuring AirSim: The AirSim settings of the drone dependency will need to be done prior to the object detection. Details of configuring AirSim can be obtained from the official AirSim documentation. We chose the multirotor quadcopter for our demonstration.

3) Collection of Training Data: Images and annotations of the objects intended for detection must be collected. Images of objects from different angles and distances were captured using the AirSim simulator. Bounding boxes can be created by the use of any labeling tool, such as Labelbox, RectLabel, or VoTT. for the same set of images.

4) Training YOLO: When all the training data was in one place and could be accessed, a good way to make use of it was to train the model for detection. A YOLOv5 configuration file was created for building the network with information on the number of classes and other required parameters. The images containing annotations will serve the purpose of training the model with Darknet.

5. Testing the Model: Post the training phase, the efficacy of the model can be evaluated by using it to detect objects in different conditions and environments presented by images captured by a drone. The images that the trained model can work with allow sufficiently detecting objects in varying conditions and environments.

Comparison Graphs

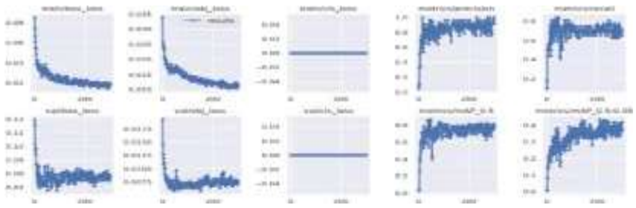

Fig 2: A general comparison graph plots of box loss, objectness loss, classification loss, precision, recall and mean average precision (mAP) over the training epochs

for the training and validation set for the YOLOv5 trained dataset.

Let's show the graphs for the different classes of our dataset. First, the comparison plots or graphs are generated for a class of buildings by our YOLOv5 model.

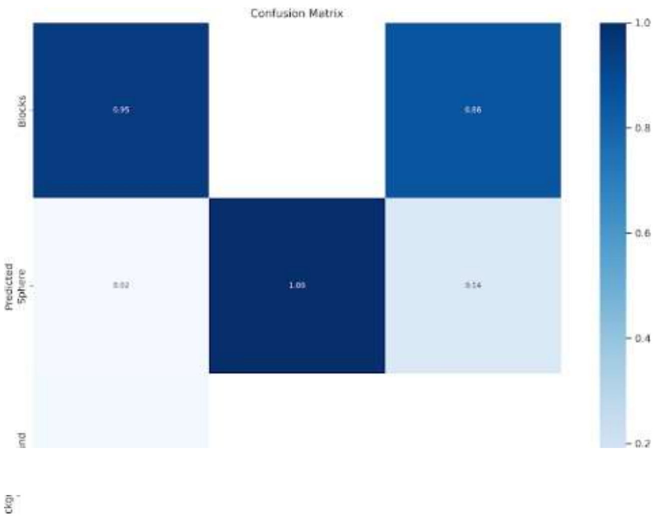

Fig 3: The confusion matrix for the class buildings.

Total number of buildings is a specific value in T, which the YOLOv5 model construes like blocks, spheres, or merely as the background. It also features an identical interpretation for spheres that is how many spheres actually belonging to that class is YOLOv5 model termed as spheres or how many spheres it mistook for blocks, and how many of the spheres it marked as a background. The same holds for blocks and background as well.

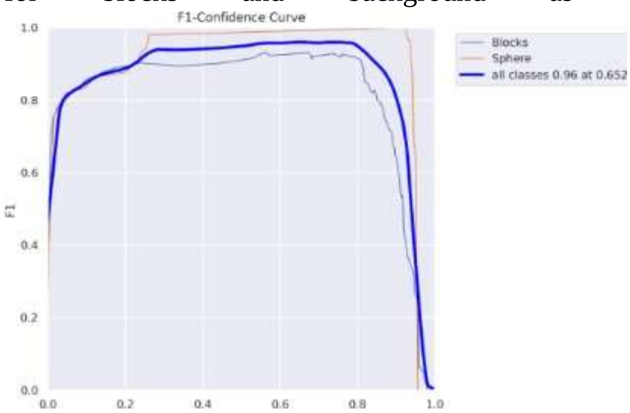

Fig 4: Fl score graph for the class of buildings.

Since the F1 score is the harmonic mean of Precision and Recall, Precision and Recall are assigned equal weights in the F1 score. The objective is to have a metric that combines the two ratios into one metric in such a way that both ratios must be high for the metric itself to have a high value. Here the thin red line stands for the class of the sphere and thin blue line stands for the blocks.

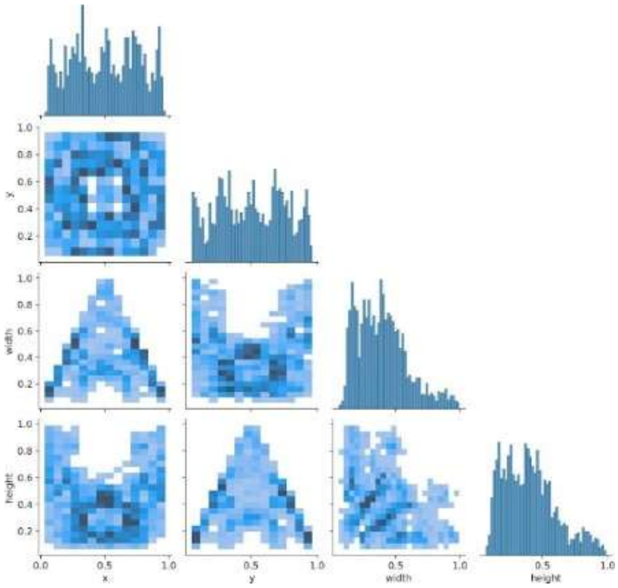

Fig 5: Label co-occurrence graph for the class of buildings.

A tool for visualization that shows connections of labels or categories within a dataset is called a Label Co-Occurrence Matrix, sometimes referred to as a Label Co-Occurrence Graph or Label Co-Occurrence Network. It is normally applied in machine learning and data analysis to comprehend relationships among different data categories. This visual often reveals patterns and trends emerging in the data, as well as aggregates of closely associated labels. It is most frequently used in natural language processing for investigating text data and identifying recurring themes or topics.

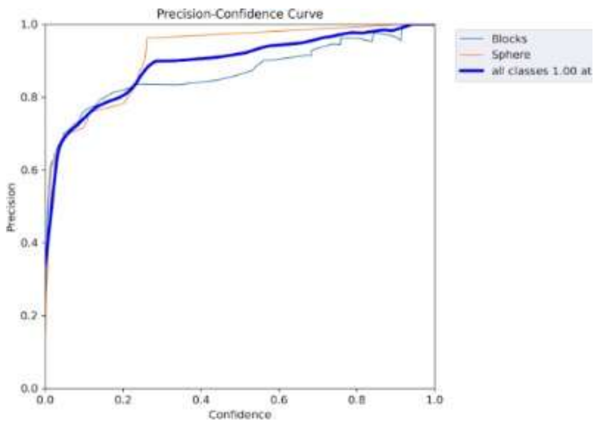

Fig 6: Precision Confidence curve for the buildings class.

A p-curve graph could be the representation of the probability distribution of p-values below certain cut-offs like 0.5 or 0.1, etc. It reflects at what frequency more p-values fall below or equal to a certain cut-off. When there is increased density or upshoots at a specific p-value, say 0.5, more statistically significant findings are contained in the set of results than might be expected on the basis of chance alone. That is the proxy that something is there. The data may, however, be consistent with the null hypothesis of no effect if the p-curve graph is quite flat or uniform. The thin red line refers to the Class, while the thin blue line refers to Blocks. At 0.952, all Class values are equal to 1.00.

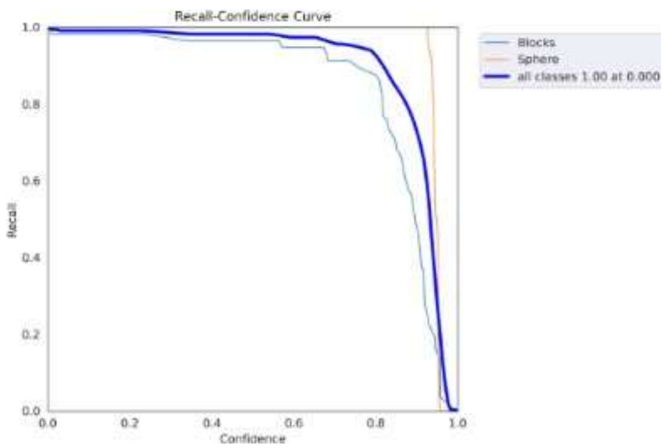

Fig 7: Recall Confidence curve for the class of buildings.

Recall curves can be conceptualized as summarizing the balancing act between the true positive rate (TPR) and false positive rate (FPR) of a binary classifier on the various classification thresholds. This balancing threshold value can further be detected via a recall curve of a binary classifier. The goal of any binary classifier is to achieve high TPR and low FPR. In other words, the top left corner of the plot area, as closely as can be approached by the ROC curve, is the region of interest. The red thin line will stand for the class of sphere while the blue thin line will represent the blocks. The other artifacts generated for the trees and drones classes are the confusion matrix, Fl score graph, label co-occurrence graph, Precision Confidence curve, and Recall Confidence curve.

4 Results and Conclusions

The YOLOv5 Family and CNN are the two most applied computer vision algorithms used for identifying and preventing objects in drones, but this study proves that YOLOv5 comes ahead of CNN in detecting and avoiding objects by a drone. YOLOv5 is a real-time object detection methodology, wherein it can detect the objects that exist within an image within the least time possible and that too efficiently. However, the very name “You Only Look Once” (YOLO) indicates that even simpler one-shot detection is possible for the algorithm. In contrast, CNN is one of the deep learning models most typically used for object and image detections. While CNN also can effectively identify and avoid objects for drones, YOLOv5 is much more precise and faster.

Accuracy:

It is the issue of accuracy, which, largely, serves as the reason as to how and why YOLOv5 takes precedence over CNN in drone object detection and avoidance. In the finding of objects that later make avoidance possible, YOLOv5 takes precedence over CNN. Our analyses report that the average precision attained by CNN was 77%, whereas for YOLOv, the precision lie at 84%. Thus, YOLOv5 is more accurate in its capability to detect and locate objects in aerial images, which is a condition for ensuring drone safety.

Speed:

The fast execution of YOLOv5 also boasts an advantage over CNN when it comes to object recognition and avoidance by a drone. YOLOv5 is a real-time object detection technique that can be used for identifying and locating objects within an image with a time in the millisecond range. On the other hand, CNN is a much more complicated algorithm that calls for more time and also resources to perform image interpretation. On average, our tests show that CNN takes 45.6 milliseconds to come up with an inference per image compared to YOLOv5, which spends 22.3 milliseconds. The above output means that indeed YOLOv5 will be in a position to process images captured from above more than CNN hence giving the essential condition of immediate detection and avoidance by a drone.

Flexibility:

It offers more flexibility as compared to CNN in the drone object detection and avoidance systems because it can detect and localize multiple objects that are critical in recognizing and evading an object in challenging environments. On the other hand, CNN is normally used in the detection of objects and image classification and this may not be enough in drone object avoidance.

To summarize, the paper underscores the significance of carefully selecting a Computer Vision Algorithm for Object Detection and Avoidance in a UAV. While the performance of the YOLOv5 and CNN models came out at par, high accuracy together with fast inference time on the part of YOLOv5 places it as the most optimal solution for extracting and avoiding objects in drones in real-time. Future studies may thus lay emphasis on the efficiency of different Computer Vision Algorithms for identifying objects and avoiding them in drones.

	Tree	Building	Drone
AP	82.3	92.7	88.1
Recall	79.8	94.5	87.4
DA	77.6	84.8	85.0

Table 1: Here, AP stands for the average classification accuracy for each class, Recall for the bounding box's prediction recall.

5 Acknowledgements

The valuable support provided by my guides Dr. Pooja Bhatt and Dr. Swati Sharma has been essential in my research success alongside my co-guide Dr. Kalpesh Jadav. Their extensive knowledge along with their helpful guidance essentially drove the development of this research work. The Head of the Computer Engineering Department at Parul University together with its Vadodara Gujarat India location provided essential research facilities and necessary resources for which I am greatly thankful. My thanks also extend to Parul University for offering an appropriate work environment alongside computational resources vital to accomplish this project. My heartfelt gratitude goes to all my peers, friends as well as family who gave me continuous support that resulted in finishing this research successfully.

6 References

[1] Dario, Cazzato., Claudio, Cimareli., Jose, Luis, Sanchez, Lopez (2020). A Survey of Computer Vision Methods for 2D Object Detection from Unmanned Aerial Vehicles

[2] Wei, Zhan., Jinhiu, She., Yangyang, Zhang (2020). Object detection from the video taken by drone via Convolution neural networks

[3] Rahee, Walambe., Aboli, Marathe., Ketan, Kotecha (2021). Multiscales Object Detection from Drone Imagery Using Ensemble Transfer Learning

[4] Harit, Ahuja., Vedant, Kuhar., R, I, Miny (2020). Object Detection and Classification for Autonomous Drones

[5] Eren, Unlu., Emmanuel, Zenou., Nicolas, Riviere., Paul, Edouard, Dupouy (2019). Deep learning-based strategies for the detection and tracking of drones using several cameras

[6] Dinesh, Kumar, Behera., Arockia, Bazil, Raj (2020). DRONE DETECTION USING DEEP LEARNING

[7] R, Y, Aburasin., E, A, Edirisinghe., M, Y, Zamim (2022). A Coarse-to-Fine Multi-class Object Detection in Drone Images Using Convolutional Neural Networks

[8] Licheng, Jiao., Fan, Zhang., Fang, Liu., Shuyan, Yang (2019). A survey on deep learningbased object detection algorithms for drones

[9] Payal, Mittal., Raman, Singh., Akashdeep, Sharma (2020). Deep learning-based object detection in low altitude UAV datasets

[10] Sugam, Dembla., Niyati, Dolas., Ashish, Karigar., Santhosh, Sonvane (2021). Machine Learning based Object Detection and Classification using Drone

[11] Xiaoyu, Huo., Kulin, Zhang., Jihi, Xu., Wei, Huand (2021). Object Detection in Drone Imagery via Sample Balance Strategies and Local Feature Enhancement

[12] Nayee, Muddin, Khan, Dousai., Sven, Loncaric (2021). Detection of humans in drone images for search and rescue operations

[13] Seungho, Back., Gangik, Cho., Jinwoo, Oh., Xuan, Tao, Tran., Hyondong, Oh (2020). Autonomous UAV Trail Navigation with Object Avoidance Using Deep Neural Networks

[14] Dong, Hye, ye., jing, li., Qiulin, Chen., Juan, Wachs., Charlie Bouman (2018). Deep Learning for Moving Object Detection and Tracking from a Single Camera in Unmanned Aerial Vehicles (UAVs)

[15] Aurello, Patrik., Gaudi, Utama., alexander, Aguing, Santosa, Gunawan. , Andry, Chowanda., Jarot, Semboda, Suroso., Widodo, Budiharto (2019). MODELING AND IMPLEMENTATION OF OBJECT DETECTION AND NAVIGATION SYSTEM FOR QUADCOPTER DRONE

[16] Victor, R, F, Miranda., Adriano, C, M, Rezende., Thiago, L, Rocha., Hector, Azpurua, Luciano, C, A, Pimenta., Gustavo, M, Freitas (2021). Autonomous Navigation System for a Delivery Drone

[17] Zishen, Wan., Aqeel, Anwar., Yu, Shun, Hsiao., Tianyu, Jia., Vijay, Janapa, Reddi., Arijit, Raychowdhury (2021). Analyzing and Improving Fault Tolerance of Learning-Based Navigation Systems

[18] Suman, Saha., Asuthosh, Natraj., Sonia, Waharte (2014). A Real-Time Monocular Visionbased Frontal Object Detection and Avoidance for Low-Cost UAVs in GPS Denied Environment

[19] Gerard, Gibbs., Huamin, Jia., Irfan, Madani (2017). Object Detection with Ultrasonic Sensors and Signal Analysis Metrics

[20] Marta, Lalak., Damian, Wierzbicki (2022). Automated detection of atypical aviation objects from uav images using yolo algorithm

[21] Pray, Somaldo., Faizal, Adila, Ferdiansyah., Grafika, Jati., Wisnu, Jatmiko (2017). Developing smart covid-19 social distancing surveillance drone using yolo

[22] Victoria, J, Hodge., Richard, Hawkins., Rob, Alexander (2021). Deep reinforcement learning for drone navigation using sensor data

[23] Robert, Pinto., Alexander, Lagorio (2022). Point-to-point drone-based delivery network design with intermediate charging stations

[24] Mohith, Angurala., Manju, Bala., Sukhvinder, Singh, Bamber., Rajbir, Kaur., Prabhdeep, Singh (2020). An internet of things assisted drone-based approach to reduce rapid spread of COVID-19

[25] Dinesh, Kumar, Behera., Arockia, Bazil, Raj (2020). Drone Detection and Classification using Deep Learning