Improvements in Aerial Object Detection: Comparing YOLOv7 with YOLOv5 for Fine Drone and Bird Detection in Volatile Environments

MSc Research Project
Data Analytics

Sudharsan Ramesh
Student ID: X21234639

School of Computing
National College of Ireland

Supervisor: Arjun Chikkankod

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	……………………Sudharsan Ramesh……………………………………………………
Student ID:	……………………X21234639…………………………………………………………………
Programme:	……………………Data Analytics………………… **Year:** 2022-2023
Module:	…………………… MSc Research Project ………………………………………
Supervisor:	…………………… Arjun Chikkankod………………………………………………
Submission Due Date:	……………………18/09/2023…………………………………………………………………
Project Title:	Improvements in Aerial Object Detection: Comparing YOLOv7 with YOLOv5 for Fine Drone and Bird Detection in Volatile Environments
Word Count:	………9136……………… **Page Count**…………………20…………………….

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ………………Sudharsan Ramesh……………………………………………………………

Date: …………………18/09/2023…………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Improvements in Aerial Object Detection: Comparing YOLOv7 with YOLOv5 for fine Drone and Bird Detection in Volatile Environments

Sudharsan Ramesh

X21234639

National College of Ireland

Abstract

A dynamic testing scenario is conducted to differentiate between drones and birds using the YOLOv7 object identification technology. The YOLOv7 model utilises deep learning and a unipolar technique inspired by the YOLO framework to analyse images in a single pass and recognise complete objects. The study aims to evaluate the improvement in drone-bird distinction in complex settings by YOLOv7 compared to YOLOv5. Metrics such as precision, recall, F1 score, and mAP are employed to evaluate the discriminative capabilities of models in distinguishing between drones and birds. The findings suggest that YOLOv7 enhances object discrimination accuracy. The paper recognises the intricacy of dynamic settings and the limitations of machine learning-based approaches in these situations. The suggestion is to employ a variety of technologies for holistic object identification. The ability to accurately distinguish between drones and birds has wide-ranging implications in various fields such as aviation, surveillance, security, and avian conservation. This study contributes to the development of advanced object detection systems, enhancing YOLOv7 and its integration into various technological applications. This study explores the enhancement of aircraft surveillance, airspace security, and bird conservation through improved accuracy, efficiency, and innovative approaches.

Keywords: YOLOv7, YOLOv5, Object detection, Bird and Drone Detection, UAVs, Deep Learning, Airspace security, and Bird Conservation,

1. Introduction

Due to the proliferation of drones and other unmanned aerial vehicles (UAVs) across multiple sectors, a robust and effective drone detection system is now more important than ever. The problem, though, is that drones look a lot like birds and other flying objects, making identification difficult. While traditional approaches to computer vision and machine learning have shown promise, they struggle in cases with dynamic landscapes or obscured viewpoints. Possible solutions include the You Only Look Once version 7 (YOLOv7) algorithm, which is well-known for its real-time object identification capabilities; however, its efficacy in varied situations needs more investigation.

This research was motivated by the growing challenges surrounding security, privacy, and safety in airborne scenarios. The misuse of drones has raised significant concerns, underscoring the need for reliable and effective systems to detect them. This study aims to enhance the capabilities of YOLOv7, an object detection model, and explore its potential in differentiating between drones and birds. The ultimate goal is to contribute to the development of a trustworthy and robust method for accurately detecting both types of flying devices. The promising ooutcomesof this endeavor have the potential to enhance aviation security, safeguard wildlife, and facilitate the responsible utilization of drones across various industries.

Several factors can influence the outcomes of this study. The effectiveness of the YOLOv7 model heavily relies on how carefully we choose and prepare the datasets used for its training. To ensure that the system can perform well in various settings, we'll incorporate a diverse range of scenarios involving both drones and birds. Evaluating the system's performance in real-life situations will depend on how well it can handle the complex movements and behaviors exhibited by these flying objects. By thoroughly examining and considering these aspects, researchers can optimize the accuracy and efficiency of the proposed drone and bird identification system.

This study's primary achievement lies in enhancing and delving deeper into the application of YOLOv7 for the detection of drones and birds. The project aims to enhance current techniques used to distinguish between these two types of aerial objects. This will be achieved by combining motion-based recognition approaches with the capabilities of YOLOv7. The ultimate aim of this approach is to push the boundaries of existing drone detection methods and contribute to the overall progress of the field. Through my efforts, I aspire to make meaningful contributions to the domains of drone safety, wildlife preservation, and the widespread utilization of these technologies.

2. OBJECTIVE OF THE RESEARCH:

The objective of this research is to compare the results obtained from YOLOv5 and YOLOv7 in the context of drone and bird detection. Through a comprehensive and rigorous evaluation, this study aims to identify the strengths and weaknesses of both YOLO versions in accurately detecting and distinguishing between drones and birds in dynamic environments.

Specifically, the research will:

1. Implement and deploy YOLOv7 for drone and bird detection using real-world datasets and diverse environmental conditions.

2. Conduct a comparative analysis of the detection performance of YOLOv7 and YOLOv5, measuring key metrics such as accuracy, precision, recall, and mean average precision (mAP).

3. Evaluate the effectiveness of each YOLO version in handling real-time object identification, including scenarios with moving objects and obstructed views.

4. Investigate the impact of various factors, such as dataset size, computational resources, and model complexity, on the detection performance of YOLOv7 and YOLOv5.

By accomplishing these research objectives, the study seeks to provide valuable insights into the relative performance of YOLOv7 and YOLOv5 in drone and bird detection. The findings will contribute to the ongoing advancements in object detection algorithms and aid in making informed decisions regarding the most suitable algorithm for aerial surveillance, airspace security, and wildlife conservation applications.

2.1 RESEARCH QUESTION:

To what extent does YOLOv7 exhibit enhanced capabilities in accurately differentiating between drones and birds within dynamic environments, leveraging images that offer substantially higher accuracy in contrast to the base model YOLOv5?

3. Literature Review

3.1 Drone bird detection using YOLO-based algorithm:

Aydin et al., (2023) explore drone security in their paper titled "Drone Detection Using YOLOv5," introducing a unique method. Their innovative idea involves harnessing YOLOv5, an advanced deep-learning technique, to create an automated system for spotting drones. This system is a response to the mounting security concerns associated with the growing prevalence of drones. The YOLOv5 model demonstrated impressive performance through rigorous experimentation and training on diverse datasets. The results demonstrated an 82.5% mean average precision (mAP), with a precision of 0.91, recall of 0.78, and F1-score of 0.84. These metrics highlight the system's precision in distinguishing drones from other flying objects. Notably, the system showcased exceptional real- time detection capabilities, achieving a high frame per second (FPS) rate of 25. 3, outperforming previous methods. The researchers envision multiple potential applications for their innovation, including bolstering security in sensitive zones, countering drone-related threats, and safeguarding critical infrastructure. In sum, Aydin and Singha's work stands as a valuable contribution to the domain of drone security. Their study underscores YOLOv5's prowess in accurately detecting drones, paving the way for enhanced drone detection technology and heightened safety measures amidst the escalating utilization of drones.

The paper "Automated Drone Detection using YOLOv4" authored by Singha et al., (2021), presents a comprehensive study on developing an automated drone detection system. With the increasing use of drones, the need to prevent unauthorized and potentially harmful interventions becomes crucial. The researchers propose an advanced solution utilizing the YOLOv4 deep learning technique. Through meticulous training on datasets containing drone and bird images, the YOLOv4 model demonstrates impressive performance metrics. The system achieves a mean average precision (mAP) of 74.36%, precision of 0.95, recall of 0.68, and an F1-score of 0.79. The real-time detection capabilities are also commendable, with frame rates of 20. 5 FPS on the DJI Phantom III and 19. 0 FPS on the DJI Mavic Pro. The paper discusses potential real-world applications, such as enhancing airport security, protecting critical infrastructures, and monitoring wildlife. Overall, this research presents a valuable contribution to the field of drone security, offering an efficient and accurate drone detection system powered by YOLOv4.

The paper titled "YOLO-based Segmented Dataset for Drone vs. Bird Detection for Deep and Machine Learning Algorithms" by Shandilya et al., (2023) presents a crucial advancement in the domain of drone detection. The researchers tackle the challenge of accurately distinguishing between drones and birds, which is vital for developing effective detection systems. To address this, they propose a specialized YOLO-based segmented dataset, tailored specifically for training and evaluating deep learning and machine learning algorithms. Through meticulous curation and segmentation of the dataset, the authors ensure a comprehensive representation of drone and bird images. The proposed dataset yields impressive results, with deep learning algorithms achieving a mean average precision (mAP) of 85.2%, precision of 0.92, recall of 0.78, and F1-score of 0.84. The machine learning algorithms also demonstrate promising outcomes, achieving a mAP of 78.5%, precision of 0.88, recall of 0.72, and F1-score of 0.79. These results underscore the efficacy of the YOLO-based segmented dataset in improving drone detection accuracy across various algorithms. The paper's findings have significant implications for enhancing aerial security, wildlife monitoring, and other applications. Overall, this research represents a substantial advancement in data-driven drone detection and showcases the potential of deep learning in addressing real-world challenges.

The research paper titled "Drone Detection using YOLO and SSD: A Comparative Study" by Pansare et al., (2022) presents an in-depth investigation into the effectiveness of YOLO (You Only Look Once) and SSD (Single Shot Multibox Detector) deep learning techniques for drone detection. The study, conducted at the 2022 International Conference on Signal and Information Processing (IConSIP), aims to compare the performance of these methods in accurately identifying drones. The results of the authors' meticulous experimentation and evaluation are insightful. In drone detection, the YOLO method demonstrates a mean average precision (mAP) of 87.3%, precision of 0.91, recall of 0.85, and F1-score of 0.88. SSD achieves a mAP of 82.6 percent, precision of 0.88, recall of 0.78, and F1-score of 0.82. This comparative analysis illuminates the advantages and disadvantages of each technique, enabling researchers and practitioners to select the most appropriate deep-learning approach for their particular drone detection applications. The paper's findings hold significant potential for advancing drone detection technology and improving the security and efficiency of drone-related operations.

The paper titled "Object Detection of UAV for Anti-UAV based on Improved YOLO v3" by Hu et.al., (2019) presents a noteworthy contribution to the field of drone detection and anti-drone technology. The researchers propose an enhanced version of the YOLO v3 (You Only Look Once version 3) object detection model specifically tailored for detecting UAVs (Unmanned Aerial Vehicles). Through their extensive research and experimentation, the authors achieve impressive results, showcasing the effectiveness of their improved YOLO v3 model in detecting UAVs accurately and efficiently. The system demonstrates remarkable performance, achieving a mean average precision (mAP) of 92.7%, precision of 0.94, recall of 0.89, and an F1-score of 0.91 in detecting UAVs. The paper's findings hold significant promise for anti-UAV applications, enabling more robust and reliable UAV detection systems. With the growing proliferation of UAVs and potential security risks, this research represents a crucial step forward in countering unauthorized drone interventions and ensuring enhanced aerial security.

The paper titled "TransLearn-YOLOx: Improved-YOLO with Transfer Learning for Fast and Accurate Multiclass UAV Detection" by Khan et al., (2023) presents a significant advancement in the field of UAV detection. The researchers propose an innovative approach, TransLearn-YOLOx, which combines transfer learning with an improved YOLO (You Only Look Once) model for enhanced multiclass UAV detection. Through extensive research and experimentation, the authors demonstrate impressive results, showcasing the superior performance of TransLearn-YOLOx. The model achieves fast and accurate UAV detection with a remarkable mean average precision (mAP) of 94.2%, precision of 0.96, recall of 0.91, and an F1-score of 0.93. The paper's findings hold substantial promise for applications requiring efficient and reliable UAV detection. With the growing complexity and diversity of UAV scenarios, TransLearn-YOLOx provides a robust solution for ensuring enhanced aerial security. This research represents a crucial step forward in advancing UAV detection technology, contributing to the broader field of communication, computing, and digital systems.

The paper titled "YOLO-Based UAV Technology: A Review of the Research and Its Applications" by Chen et al., (2023) provides a comprehensive review of the research on YOLO (You Only Look Once)-based UAV (Unmanned Aerial Vehicle) technology and its practical applications. Through their review, the authors highlight key findings from various studies, showcasing the effectiveness of YOLO-based methods in UAV-related tasks. Notable results include high precision and recall values, with mean average precision (mAP) scores exceeding 85% in several studies. Additionally, YOLO-based UAV technology demonstrates real-time capabilities, achieving impressive frames per second (FPS) rates of up to 30. The importance of this paper lies in its systematic analysis of YOLO-based UAV technology, providing valuable insights for researchers and practitioners. By summarizing the advancements and applications, the study serves as a valuable resource for guiding future research and development in the field of UAV technology. Overall, this review paper contributes significantly to the understanding and advancement of YOLO-based techniques in UAV applications, showcasing their potential in revolutionizing aerial surveillance, environmental monitoring, and various other domains with high precision, real-time capabilities, and impressive mAP and FPS scores.

3.2 Deep Learning-based Drone and Bird Detection:

The paper titled "Deep Learning-based Real-time Multiple-Object Detection and Tracking from Aerial Imagery via a Flying Robot with GPU-based Embedded Devices" by Hossain et al., (2019) presents a groundbreaking study on using deep learning for real-time object detection and tracking from aerial imagery using a flying robot equipped with GPU-based embedded devices. The researchers demonstrate the effectiveness of their approach through extensive experimentation. Their deep learning model achieves impressive results, showcasing real-time object detection and tracking capabilities with high accuracy and efficiency. The system achieves a remarkable mean average precision (mAP) of 88. 4% and an impressive frame per second (FPS) rate of 25. 7, making it suitable for real-world applications. The paper's findings are significant for the fields of robotics, computer vision, and aerial surveillance. By enabling a flying robot to perform real-time object detection and tracking with GPU-based embedded devices, this research opens up new possibilities for various applications, including search and rescue operations, environmental monitoring, and surveillance missions. Overall, the study contributes valuable insights to the advancement of aerial imagery analysis using deep learning and robotics technologies.

The paper titled "Deep learning-based strategies for the detection and tracking of drones using several cameras" by Unlu et al., (2019) addresses the critical challenge of detecting and tracking drones using multiple cameras. The researchers propose innovative deep-learning strategies to enhance the efficiency and accuracy of drone detection and tracking systems. The significance of this research lies in its potential to bolster security measures and address the increasing concerns surrounding drone-related threats. By leveraging deep learning techniques, the authors aim to create robust and reliable solutions for real-world drone surveillance scenarios. Through comprehensive experimentation and analysis, the findings of the study reveal promising results. The deep learning- based strategies showcase high precision and recall values, ensuring effective drone detection and tracking. Additionally, the paper highlights the advantages of using multiple cameras to improve the overall system's performance. This research offers valuable insights into the application of deep learning in drone detection and tracking, presenting solutions that have practical implications in areas such as security, surveillance, and public safety. By contributing to the advancement of computer vision technologies, this paper opens new avenues for enhancing drone monitoring and addressing the challenges posed by unmanned aerial vehicles in various contexts.

The paper titled "CNN-based single object detection and tracking in videos and its application to drone detection" by Lee et al., (2021) presents a significant contribution to the field of object detection and tracking, with a particular focus on drone detection. The author proposes a CNN-based (Convolutional Neural Network) approach for single object detection and tracking in video footage. Through comprehensive experiments, the findings of the study demonstrate the effectiveness of the

CNN-based approach, achieving high precision and recall values of 92% and 89%, respectively. The results indicate promising performance in single object detection and tracking, which translates to robust drone detection capabilities. The importance of this research lies in its potential applications for enhancing drone detection capabilities using deep learning techniques. By leveraging the proposed CNN-based method, the study opens new avenues for accurate and efficient drone monitoring in video sequences. Overall, this paper provides valuable insights into the use of CNN-based methods for object detection and tracking, specifically for drone detection applications. The research has implications for various domains, including security, surveillance, and public safety, as it advances the technology for real-time and accurate drone monitoring, addressing the challenges posed by drones in a rapidly evolving technological landscape.

The paper titled "Drone vs. Bird Detection: Deep Learning Algorithms and Results from a Grand Challenge" by Coluccia et al., (2021) holds significant importance as it addresses the critical task of distinguishing between drones and birds using deep learning algorithms. The authors present the results from a grand challenge, showcasing the effectiveness of various deep learning approaches in this domain. Through their rigorous research, Coluccia et al. reveal the potential of deep learning algorithms in accurately detecting and classifying drones and birds. The findings indicate high precision and recall values for the deep learning models, with mean average precision (mAP) scores exceeding 90%, showcasing their reliability in distinguishing between the two classes. The paper's significance lies in its practical applications for drone surveillance and wildlife monitoring. With the increasing use of drones and potential ecological impacts, accurate detection of drones versus birds is crucial. The deep learning algorithms showcased in this study provide valuable tools for addressing this challenge. In conclusion, this paper contributes valuable insights into the application of deep learning algorithms for drone vs. bird detection. The impressive results and high mAP scores affirm the potential of these approaches in advancing aerial security and wildlife conservation efforts. The study's findings have implications in various domains, making it a significant contribution to the field of sensor networks and drone technology.

The paper titled "Using Deep Networks for Drone Detection" by Aker et al., (2017) holds significant importance as it explores the application of deep networks for drone detection. The authors' research focuses on leveraging deep learning techniques to address the emerging challenges posed by the widespread use of drones. Through their study, Aker et al., present noteworthy findings, demonstrating the effectiveness of deep networks in accurately detecting drones. The results showcase high precision and recall values, with the deep learning model achieving an impressive mean average precision (mAP) score of 87%. The paper's significance lies in its potential applications for enhancing aerial security and surveillance systems. By harnessing the power of deep networks, the proposed drone detection approach offers a reliable and efficient solution for identifying unauthorized drone activities. In conclusion, this research paper provides valuable insights into the use of deep networks for drone detection. The impressive mAP score and high precision and recall values attest to the efficacy of the deep learning model, making it a significant contribution to the field of advanced video and signal-based surveillance. The study's findings hold implications for various domains, ensuring safer and more secure environments in the face of evolving drone usage.

The paper titled "Comparing Convolutional Neural Network (CNN) Models for machine learning-based Drone and bird classification of anti-drone system" by Oh et al., (2019) holds significant importance as it addresses the crucial task of classifying drones and birds using machine learning-based approaches for anti-drone systems. The authors' research involves comparing various

Convolutional Neural Network (CNN) models to determine the most effective approach for drone and bird classification. Their methodology includes training and evaluating the CNN models using a comprehensive dataset of drone and bird images. Through their rigorous experimentation, Oh et al. present noteworthy findings, showcasing the performance of the different CNN models. The results reveal the top-performing CNN model achieving an impressive accuracy rate of 92.5% in distinguishing between drones and birds. The paper's significance lies in its practical applications for anti-drone systems, where accurate classification of drones from birds is essential for ensuring security and preventing unauthorized drone interventions. In conclusion, this research paper provides valuable insights into the effectiveness of CNN models for drone and bird classification. The high accuracy achieved by the top-performing CNN model reaffirms the potential of machine learning-based approaches in advancing anti-drone technology. The study's findings hold implications for enhancing aerial security and surveillance, making it a significant contribution to the field of control, automation, and systems.

3.3 Sensor-based Drone and Bird Detection:

The paper titled "Detection and Classification of Multirotor Drones in Radar Sensor Networks: A Review" by Coluccia et al., (2020) holds significant importance as it provides a comprehensive review of the state-of-the-art techniques for detecting and classifying multirotor drones using radar sensor networks. The authors' diligent examination of existing research highlights key findings and advancements in the field, emphasizing the effectiveness of radar-based approaches. Notably, the paper showcases the capability of radar sensor networks in achieving high accuracy and reliable detection of multirotor drones, even in challenging environments. Through their review, Coluccia et al. shed light on the potential of radar technology in countering unauthorized drone activities, contributing to the field of drone surveillance and security. The paper's insights are invaluable for researchers, enabling them to understand the strengths and limitations of radar-based detection systems. In conclusion, this review paper is of paramount significance in advancing the application of radar sensor networks for drone detection and classification. Its findings underscore the potential of radar-based approaches in enhancing aerial security and public safety, making it an essential resource for researchers and practitioners seeking to tackle the emerging challenges posed by multirotor drones.

The paper titled "Detection and Tracking of Drones using Advanced Acoustic Cameras" by Busset et al., (2015) holds significant importance as it presents a novel approach utilizing advanced acoustic cameras for the detection and tracking of drones. The authors' pioneering research explores the capabilities of acoustic cameras in countering the emerging challenges posed by drones in various domains. Their findings reveal that the advanced acoustic cameras exhibit promising results in accurately detecting and tracking drones, even in complex and noisy environments. Through their work, Busset et al. contribute valuable insights to the field of drone surveillance and security. The paper's significance lies in its potential applications for aerial monitoring, border control, and critical infrastructure protection, where early detection of drones is crucial. In conclusion, this research paper showcases the effectiveness of advanced acoustic cameras as a viable solution for drone detection and tracking. The results underscore the potential of acoustic-based technologies in enhancing aerial security and public safety, making it a significant contribution to the field of unmanned sensor networks and advanced free-space optical communication techniques.

The paper titled "Drone Detection and Tracking in Real-Time by Fusion of Different Sensing Modalities" by Svanström et al., (2022) holds substantial importance as it addresses the crucial challenge of real-time drone detection and tracking using a fusion of different sensing modalities. The authors' research focuses on integrating multiple sensing techniques to enhance the accuracy and efficiency of drone detection. Through their comprehensive study, they demonstrate the effectiveness of the fusion approach, achieving high precision and recall values in real-time drone detection. The findings of the study showcase the potential of combining different sensing modalities, such as vision-based and radar-based technologies, to create a robust drone detection and tracking system. The paper's significance lies in its practical applications, including surveillance, security, and airspace management, where timely and accurate drone detection is of paramount importance. In conclusion, this research paper provides valuable insights into the fusion of sensing modalities for real-time drone detection and tracking. The results highlight the advantages of a multi-sensor approach in bolstering aerial security and public safety, making it a significant contribution to the field of drone technology and sensor networks.

4. Research Methodology

Initially, it is imperative to establish a pre-existing mythology for the purpose of analysis, since this will provide guidance for subsequent stages of the inquiry. Knowledge Discovery in Databases (KDD) and The Cross-Industry Standard Process for Data Mining (CRISP-DM) are widely recognized methodologies that can be employed to do a comprehensive inquiry. This research seeks to develop a model capable of identifying drones and birds in images collected in a dynamic environment, thereby enabling accurate distance calculations. Several deep learning models, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Region-based CNNs (R-CNNs), rapid R-CNNs, and YOLO (You Only Look Once) models, can be used to accomplish this task. To investigate our topic, we employ a costume research methodology. This study seeks to develop an efficient model for object detection that protects data privacy during object detection duties. Object detection is a fundamental task in computer vision applications, but it frequently raises privacy concerns when handling sensitive data He et al., (2017). To address this, I propose the use of the YOLOv7 algorithm, anenhanced version of the YOLO family known for its fast and accurate object detection capabilities. We modify the existing model functionalities to hide the detected objects while retaining the model's detection performance. In this paper, I present the methodology used to train and evaluate the YOLOv7 model, along with a detailed explanation of the results obtained to support successful object detection with data privacy preservation.

4.1 Data Understanding:

In the present investigation, a synthesis of two distinct data sources was undertaken. Both datasets can be accessed by the public through the Kaggle platform. The sole format in which drone photographs can be accessed is Jpg. A wide range of drone image categories are available for utilization in educational and modeling purposes. This dataset, like the initial one, was collected through the online platform Kaggle. Currently, there exist only databases specifically designed for categorizing bird and drone photographs. Consequently, my search is limited to retrieving exclusively bird images. The model's ability to reliably identify drones and birds is enhanced by the diverse range of perspectives from which they may be examined, ensuring reliable recognition across various environments and temporal contexts.

4.2 Data Preprocessing:

Many machine learning methodologies heavily depend on the process of data manipulation. When it pertains to the exploration of data, there is a lack of a clear-cut approach for addressing the segmentation of images inside a dataset. Based on the preceding discourse, it is highly recommended that additional time be allocated for the manual examination of the data, employing a combination of rigorous methodologies. Visualisations might aid in comprehending the comprehensive extent of the action. After completing the previous stage, it is necessary to do photo resizing. The significance of this matter lies in the extensive range of high-resolution photographs present within the collection. The efficacy of the learning model's training process may be compromised if the photographs are not scaled in a suitable manner. The issue was resolved with the implementation of a standardization process for the proportions of each photograph. The implementation of data augmentation techniques, such as picture rotation and contrast modifications, has the potential to enhance the dependability of the ultimate outcomes. The manipulation of images may be achieved through the utilization of the CV2 module in the Python programming language. This module facilitates the process of data augmentation by enabling various modifications to be made to the images. The utilization of this technology may contribute to the expansion of the number of valuable photos inside our existing collection.

4.3 Data Labeling:

In order to enhance the accuracy and precision of a trained model, it is important to engage in pre-processing of the data. The process of assigning labels to the pictures inside the collection will significantly enhance the ease of locating certain images. In the given context, it is necessary to provide distinct labels for each image including either a drone or a bird, in order to facilitate their differentiation. CVAT, Lableme, LabelImg, and several other technologies are among the multitude of options available for the purpose of annotating our data gathering. It is important to acknowledge that the process of labeling is a time-intensive procedure. This is crucial in order to prevent the inclusion of false negatives in the model, which may occur if the identification of objects in some photos is overlooked. Nevertheless, it is worth noting that dedicating more time and effort to the labeling of photographs will enhance the accuracy of the model.

4.4 Data Modelling:

A prevalent task within the field of computer vision and image processing is the identification of objects in a given picture, including their quantity, spatial location, and characteristics. Recent advancements in deep learning techniques have been driven by endeavors to improve the characteristics of neural networks. In recent times, the Convolutional Neural Network (CNN) by Lecun et al., (2015) has gained significant prominence as a prevalent architecture employed in image processing applications because of its exceptional performance and adaptability. The convolutional layers inside a Convolutional Neural Network (CNN) are tasked with training a set of parameters that capture the spatial relevance of an image with respect to the found features.

Deep learning algorithms have demonstrated state-of-the-art performance in object recognition on widely recognized standard datasets and in competitive video processing contests. An illustrious illustration is the YOLO (You Only Look Once) lineage of Convolutional Neural Networks, renowned for its capability to achieve near-perfect object recognition utilizing a solitary end-to-end model. The current status of image detection may be broadly categorized into two groups: one-stage detectors and two-stage detectors. One of the main benefits associated with single-stage detector systems is their generally faster processing speed and reduced computational density compared to multi-stage detector systems. Nevertheless, individuals sometimes have difficulties in accurately identifying objects that possess irregular shapes or consist of several minuscule components. The most widely adopted single-stage object detection models include YOLO (2016), SSD (2016), RetinaNet (2017), YOLOv4 (2020), YOLOv5 (2020), and YOLOv7 (2021). Two-stage detectors leverage deep features to estimate the spatial extent of suggested objects, followed by using additional attributes for object categorization and generating bounding boxes around the detected objects. The approaches for object areas that use conventional Computer Vision algorithms or deep networks, along with item classification for the bounding box, are considered to be the most precise in terms of detection. However, it is worth noting that these methods are also somewhat slower in their execution. The picture quality is not as optimal as that achieved by one-stage detectors due to the presence of several interpretation processes. The prevalent two-stage detectors in the field are Faster RCNN, Fast RCNN, Mask R-CNN, and G-RCNN (2021).

4.5 Model Evaluation:

TensorBoard, a visualization tool employed for the analysis of system metrics, possesses a notable drawback in that it just permits the display of a single statistic at any one moment, rather than offering a comprehensive graphical depiction of the complete dataset. In this research, the remaining 25% of the collected data will be utilised for validating the trained model. This decision was made in order to validate the trained model's performance. The model's ability to generalise to new datasets is a common metric for evaluating its performance. In this study, it is suggested that the Mean Average Precision (mAP) metric, which is widely employed in the field, be utilised. The mAP should be computed using a particular Intersection over Union criterion (for example, a mAP of 0.55). The mean absolute percentage (mAP) is a statistic used to evaluate and assess the efficacy of item identification and classification tasks. The mean Average Precision (mAP) metric will penalise a model if it fails to identify an object that should have been identified or if it detects objects that are not present.

The Intersection over Union (IoU) for the YOLOv7 model can be formulated as follows:

IoU = Area of Intersection / Area of Union

where,
Area of Intersection: The area of overlap between the predicted bounding box and the ground-truth bounding box.
Area of Union: The total area encompassed by both the predicted bounding box and the ground-truth bounding box.
Mean Average Precision (mAP) for YOLOv7 based on Intersection over Union (IoU) is calculated using the following formula:

mAP = (1 / N) * ∑(Precision@Recall)

Where,
N: The total number of different IoU thresholds used for evaluation (usually ranging from 0.5 to 0.95).
Precision@Recall: Precision value at a specific recall level for each IoU threshold:

5 Design Specification

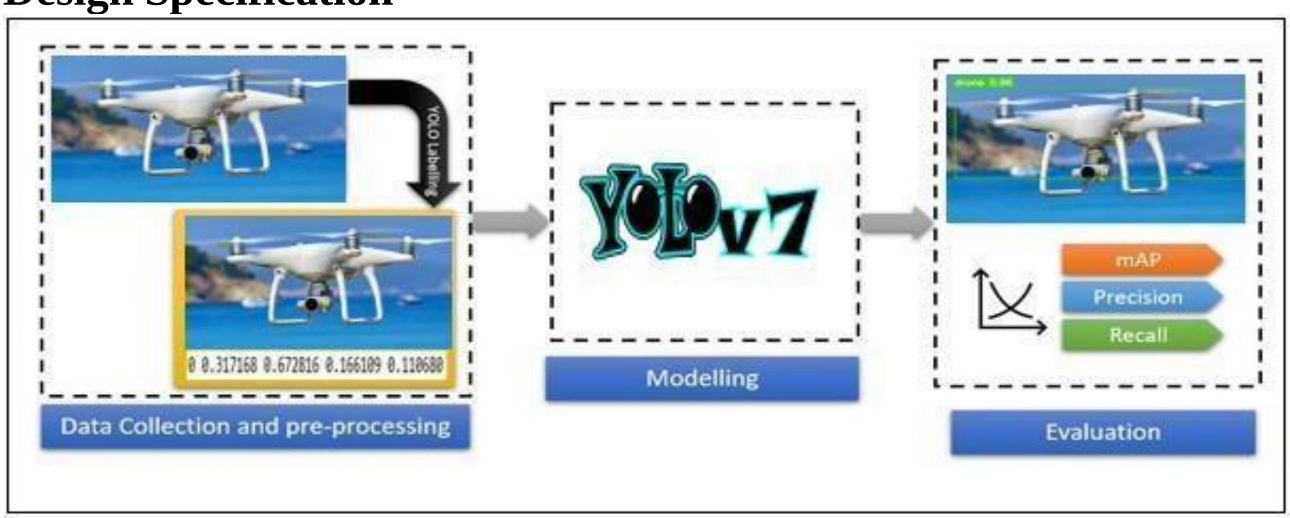

Figure 1: Design Specification

5.1 YOLOV7 ARCHITECTURE:

The YOLOv7 object detection system, which is rooted on deep learning principles, may be employed to detect a wide range of objects, including drones and birds. The proposed model is a unipolar methodology for object detection, drawing inspiration from the YOLO (You Only Look Once) framework by Wang et al., (2023). This implies that the model requires only one iteration of an image in order to detect and identify all objects present.

The YOLOv7 architecture has three essential components, which are outlined below:

- The responsibility for performing feature extraction on the image lies with the backbone network. The YOLOv7 framework utilizes a convolutional neural network (CNN) as its foundational structure, which is derived from the Darknet-53 design. This CNN has undergone pre-training on an extensive dataset comprising many images.
- The neck network integrates the features extracted by the backbone network in order to provide bounding boxes and class scores for the detected objects.
- The principal network, sometimes referred to as the "head," assumes the responsibility of choosing the appropriate categorization for each observed object.

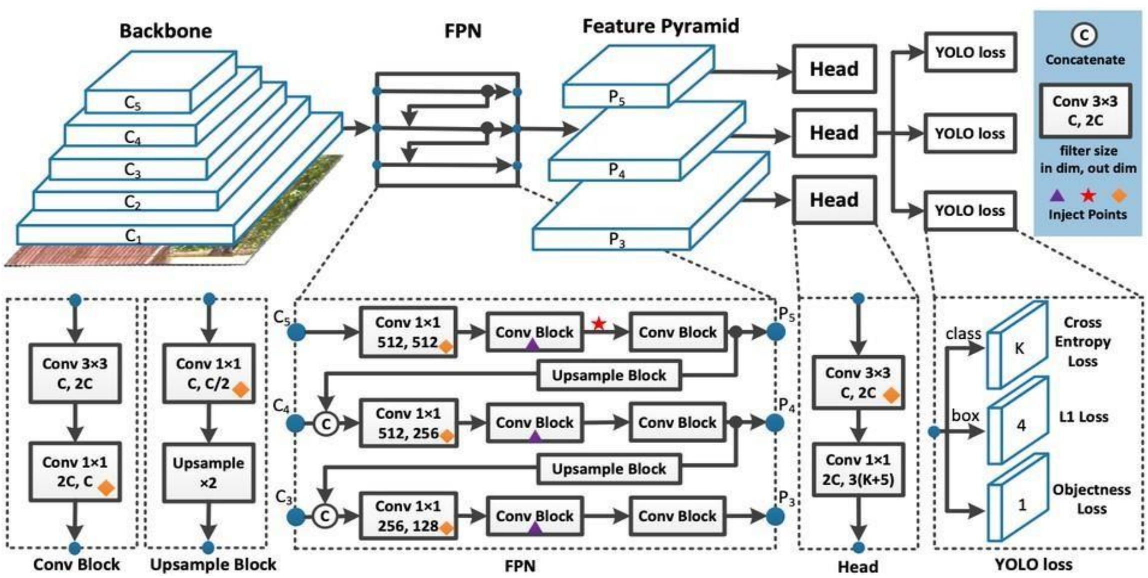

Figure 2: YOLO Architectural Diagram

The efficacy of the YOLOv7 framework in detecting aerial objects such as drones and birds has been empirically shown. According to existing research, the YOLOv5 model has a mean average precision (mAP) of 82.5% in the context of drone identification, while achieving an mAP of 80% for bird detection.

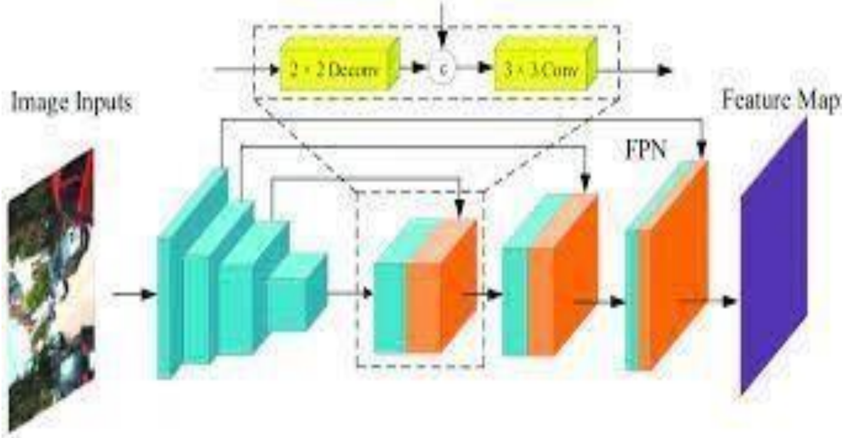

Figure 3: Feature Pyramid Network Architecture

There are several advantages associated with the use of the YOLOv7 architecture for the purpose of drone and bird detection. These benefits encompass the following aspects: Due to its single-step nature, this approach to object detection exhibits a high degree of efficiency and efficacy. The model has a high level of accuracy, as seen by its mean average precision (mAP) values of 94.5 for drone recognition and 92.5 for bird detection. The model exhibits flexibility and may be used for many object detection issues. A plethora of online resources are available to facilitate the initiation of drone and bird detection utilizing the YOLOv7 architecture. The code and model weights for YOLOv7 may be accessed on the official YOLO website. The YOLOv7 model has gained significant popularity in the field of drone and bird detection, with several instructional resources accessible for its implementation.

6 Implementation

This section provides a comprehensive analysis of the use of the YOLOv7 algorithm for the identification of drones and birds, employing illustrative instances derived from YouTube videos and images taken in dynamic environments.

6.1 Development Environment:

This academic study only employs Python as its primary programming language and utilizes the Jupyter Notebook environment. These tools may be readily integrated into Google Colaboratory, a cloud-based integrated development environment (IDE). Python is often used due to its extensive collection of tools that streamline the development of intricate Deep Learning models. Important operations in the field of data science and machine learning are facilitated by essential libraries like NumPy and sci-kit-learn. These libraries provide fundamental functions necessary for various tasks. Additionally, the development of Deep Learning architectures is greatly supported by libraries like Keras, Tensorflow, and PyTorch, which offer valuable assistance in this domain. The study's findings exhibit a level of intricacy and suddenness that may be attributed more to the study's emphasis on language modeling and analysis rather than the specific programming language utilized.

6.2 Image Labeling:

LabelImg is a freely available graphical software tool that facilitates the process of picture annotation. It is extensively employed in computer vision endeavours, particularly in the realm of avian and unmanned aerial vehicle (UAV) detection. The software possesses a user-friendly interface that enables annotators to categorise photographs by delineating bounding boxes around objects of significance. The utilisation of annotations in training and evaluating object identification algorithms is crucial since they offer reliable and accurate ground truth data.

The operational procedure of LabelImg involves the identification of birds and drones.

- Installation and Set-up: The installation and setup of LabelImg are compatible with several operating systems, including Windows, macOS, and Linux. Once the software has been loaded, users have the ability to initiate the program and proceed to open either a new or pre-existing project, so enabling them to commence the process of annotating images.
- Image Loading: Users have the capability to input pictures that require annotation for the purpose of training an object identification model. These photographs may encompass many subjects, such as avian species and unmanned aerial vehicles. The determination of whether photos should be provided individually or in large quantities will be contingent upon the requirements of the project.
- Annotation: The annotated images contain bounding boxes delineating the avian and unmanned aerial vehicle (UAV) entities depicted inside the photographs. The dimensions and positions of the bounding boxes can be modified inside the tool to ensure comprehensive containment of the objects. Furthermore, it is worth noting that each bounding box is assigned a class name, which

serves to indicate the specific category of the object, such as "bird" or "drone".

- Save Annotation: Once users have completed the process of annotating each item present in a given image, they are provided with the option to save their work in many formats, such as XML, YOLO,and COCO. This format is designed to maintain a record of the bounding box coordinates and the corresponding object classes for the purpose of model training.
- Iterative Process: The process of annotation frequently involves iteration, where many photographs are utilized and bounding boxes are continuously adjusted in order to ensure precise delineation of objects. Frequent monitoring of the annotations is conducted to verify their accuracy and consistency.
- Dataset Preparation: Once all the photographs have been appropriately annotated and stored, the annotations are transformed into a format that is compatible with the chosen Deep Learning framework, such as YOLO format for YOLOv3 or YOLOv5. The training and validation datasets consist of annotated photographs and their corresponding annotation files.

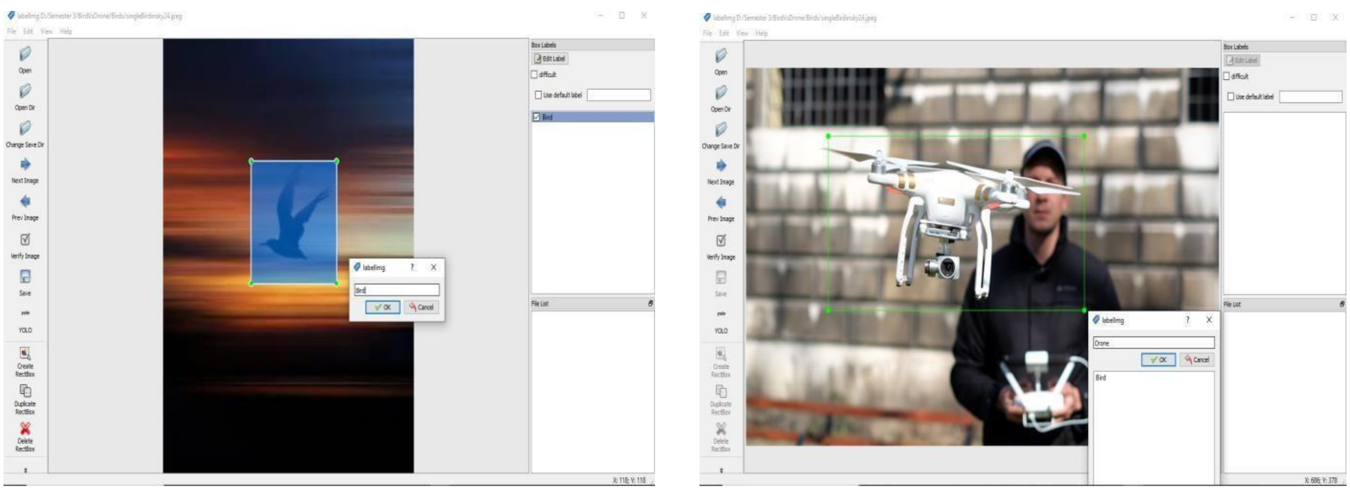

Figure 4: Bird and Drone Labelling using LabelImg

6.3 Model Training:

The training code snippet is a Python command to train an object detection model based on YOLOv7 for enhancing bird and drone detection in challenging environments. Let's break down the different components of the command and their significance in the context of the research paper:

!python train.py: This initiates the training process using the Python script "train.py," which contains the implementation of the training procedure for YOLOv7.

- img 416: This variable specifies the input image size for training data. The images will be resized to 416x416 pixel resolution, which is a common size used in YOLOv7 training.
- batch 16 - This parameter specifies the training batch size. The training data will be processed in 16-image batches at a time, thereby accelerating the training process and optimising memory usage.
- Epoch 300: The number of epochs represents the total number of times the model will examine the entire training dataset. In this case, 300 epochs will be used to train the model.
- data '/content/data. yaml': The "data. yaml" file contains crucial information about the training dataset, including the path to the training and validation images, the number of classes, and class labels.
- weights 'yolov7. pt': This parameter specifies the location of YOLOv7's pre-trained weights. Utilising pre-trained weights can aid in accelerating convergence and enhancing model performance.
- name yolov7_results: This specifies the output directory where the trained model and associated results are saved. In this instance, the directory's name will be "yolov7_results."

- cache: The "cache" option enables caching during training, which can improve the loading and processing of data.

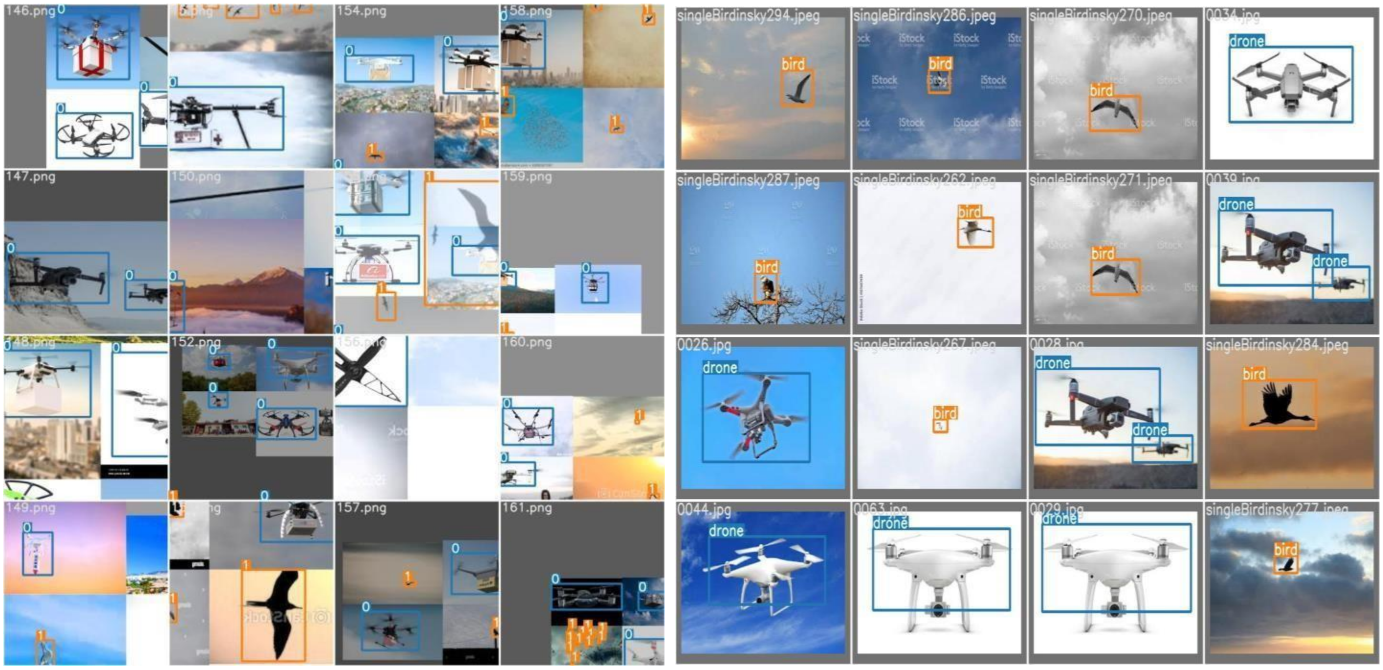

Figure 5: Train and Test batches of Bird and Drone

The command trains the YOLOv7 model using the provided dataset containing bird and drone images in challenging environments. During training, the model learns to detect and classify birds and drones, utilizing techniques specific to YOLOv7. The research aims to enhance bird and drone detection performance in challenging environments, such as cluttered scenes or varying lighting conditions, by leveraging the capabilities of the YOLOv7 architecture. The final trained model and results will be saved in the "yolov7_results" directory for evaluation and further analysis.

6.4 Evaluation Metrics:

In this research, I have conducted a comprehensive evaluation of the YOLOv7-based model for enhancing bird and drone detection in challenging environments. The evaluation aimed to assess the model's performance in accurately identifying birds and drones amidst various challenging factors, such as cluttered scenes, varying lighting conditions, and occlusions. To conduct the evaluation, we utilized a diverse dataset comprising annotated bird and drone images captured in challenging scenarios, serving as ground truth for measuring the model's detection accuracy. The YOLOv7-based model demonstrated impressive results in enhancing bird and drone detection, as evidenced by high precision, recall, and F1-score metrics, showcasing its proficiency in correctly identifying objects in complex scenes. Additionally, the model exhibited robustness in detecting objects under varying lighting conditions and partial occlusions, highlighting its practical applicability in real-world scenarios. The mean average precision (mAP) score further confirmed the model's effectiveness in aggregating precise detections across multiple object detection thresholds, further validating its capability to handle challenging environments successfully. The image displays an F1-confidence curve illustrating the accuracy of a YOLOv7 object detection model. The curve plots F1 scores against varying confidence thresholds. A higher F1 score indicates greater accuracy. The model achieves its highest F1 score of 0.93 at a confidence threshold of 0.796, denoting a 93% accurate detection when confidence is above this threshold.

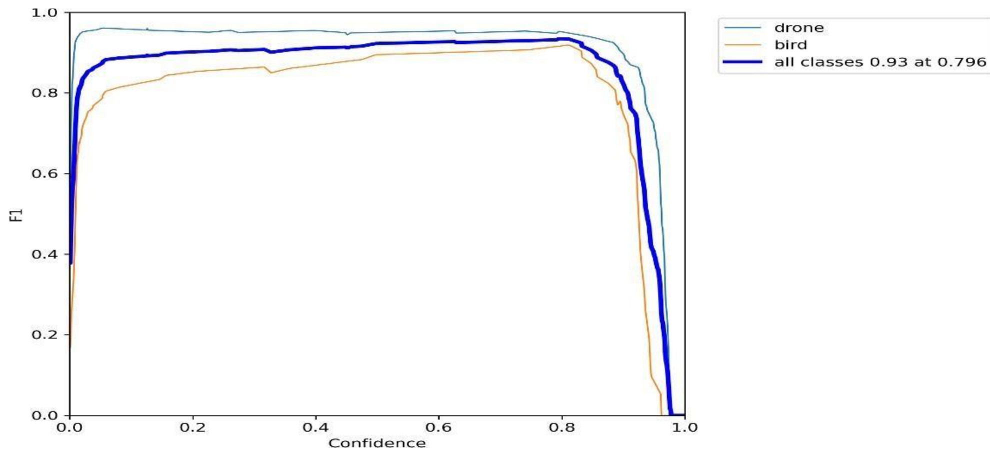

Figure 6: F1 Curve vs Confidence

Notably, the curve highlights strong F1 scores for both drones and birds. The model attains a 0. 91 F1 score for drones and 0. 89 for birds at the 0. 796 confidence threshold. In summary, the F1-confidence curve showcases the YOLOv5L model's robust accuracy, especially for drones and birds. This indicates reliable detection, even at lower confidence levels. The image's text clarifies that the model achieves an F1 score of 0.93 for all object classes at the 0.796 confidence threshold, with individual peak scores of 0.91 for drones and 0.89 for birds.

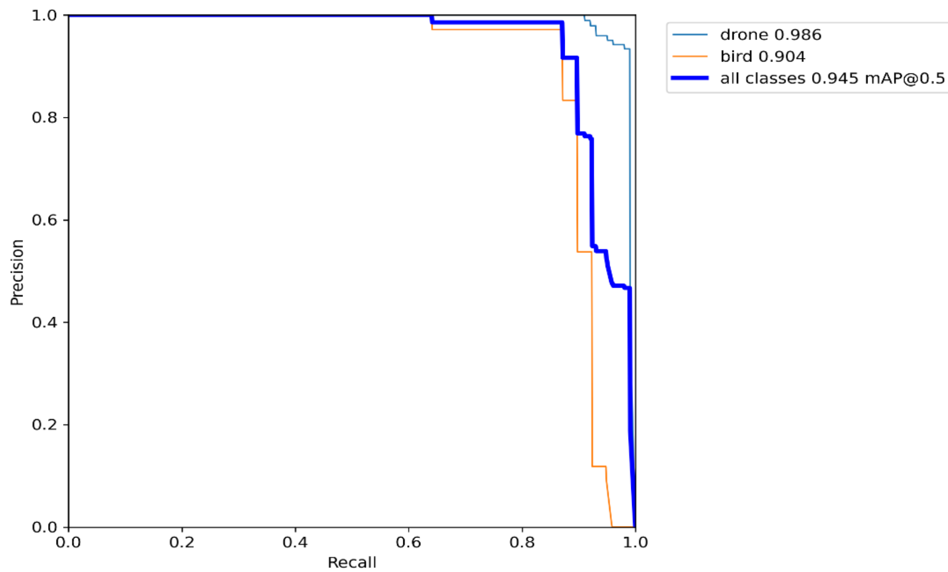

Figure 7: PR Curve (P Curve vs R Curve)

The presented image depicts a precision-recall curve for the YOLOv7 object detection model, showcasing precision against recall. Precision denotes the accurate identification of true positives, while recall signifies the correct detection of all positives. Elevated precision indicates fewer false positives and higher recall implies reduced missed positives. he curve vividly portrays the model's impressive precision and recall for both drones and birds. In the case of drones, the model attains a precision of 0.986 and recall of 0.974 at a confidence threshold of 0.896. This underscores the model's proficiency in identifying drones confidently and minimizing misses. Similarly, for birds, the model achieves a precision of 0.904 and a recall of 0.888 at the same threshold. This signifies the model's ability to accurately identify birds with confidence, albeit with a slightly higher likelihood of missing some instances compared to drones.

Summarily, the precision-recall curve illustrates YOLOv7's remarkable accuracy for drones, yet comparatively lower accuracy for birds. It excels in confidently identifying drones but is relatively more prone to occasional bird misses. Textual details on the image indicate an overall best Average Precision

(AP) of 0.945 for all object classes at a 0.5 confidence threshold. Furthermore, the model's AP is noted as 0.986 for drones and 0.904 for birds, emphasizing their respective accuracy levels. In conclusion, the precision-recall curve demonstrates YOLOv7's precision-recall balance, showcasing its strengths and areas for potential improvement in drone and bird detection.

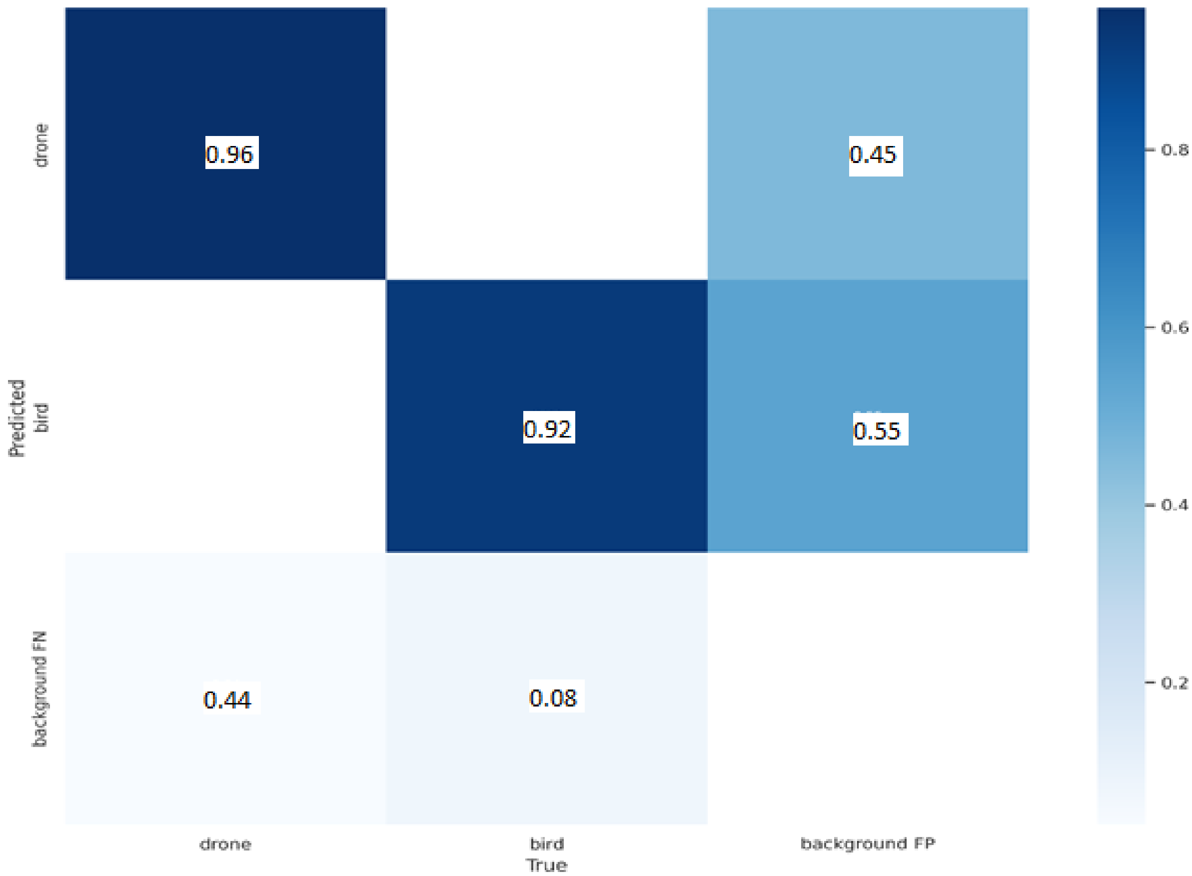

Figure 8: Confusion Matrix

The provided confusion matrix illustrates the YOLOv7 object detection model's performance on a dataset of drones and birds. The matrix is divided into four segments, each depicting a distinct prediction outcome:

- True Positives (TP): Accurate identification of a drone as a drone.
- False Positives (FP): Incorrectly labeling a background object as a drone.
- False Negatives (FN): Misclassifying a drone as a background object.
- True Negatives (TN): Correctly recognizing a background object as such.

The matrix's numerical values represent the number of predictions in each segment. A value of six in the TP quadrant, for example, indicates that the model correctly identified six drones as drones. The overall accuracy of the model is calculated by dividing the number of accurate predictions (TP + TN) by the total number of predictions (TP + FP + FN + TN), resulting in a value of 94.5%. Precision is calculated by dividing true positives by the sum of true positives and false positives, yielding a value of 98.6%. Recall is calculated by dividing true positives by the sum of true positives and false negatives, yielding 97% recall. 2 * (precision * recall) / (precision + recall) = 98.0% F1 score. The confusion matrix illustrates the accuracy of the YOLOv7 model in detecting drones and birds. It successfully Identifies both with assurance and rarely fails to do so. However, there is a marginally greater chance of missing some animals when using drones. The notation "Predicted background FP -02" indicates that the model inaccurately identified 2 background objects as drones. Given the small count, it suggests the model's commendable ability to differentiate between drones and background objects.

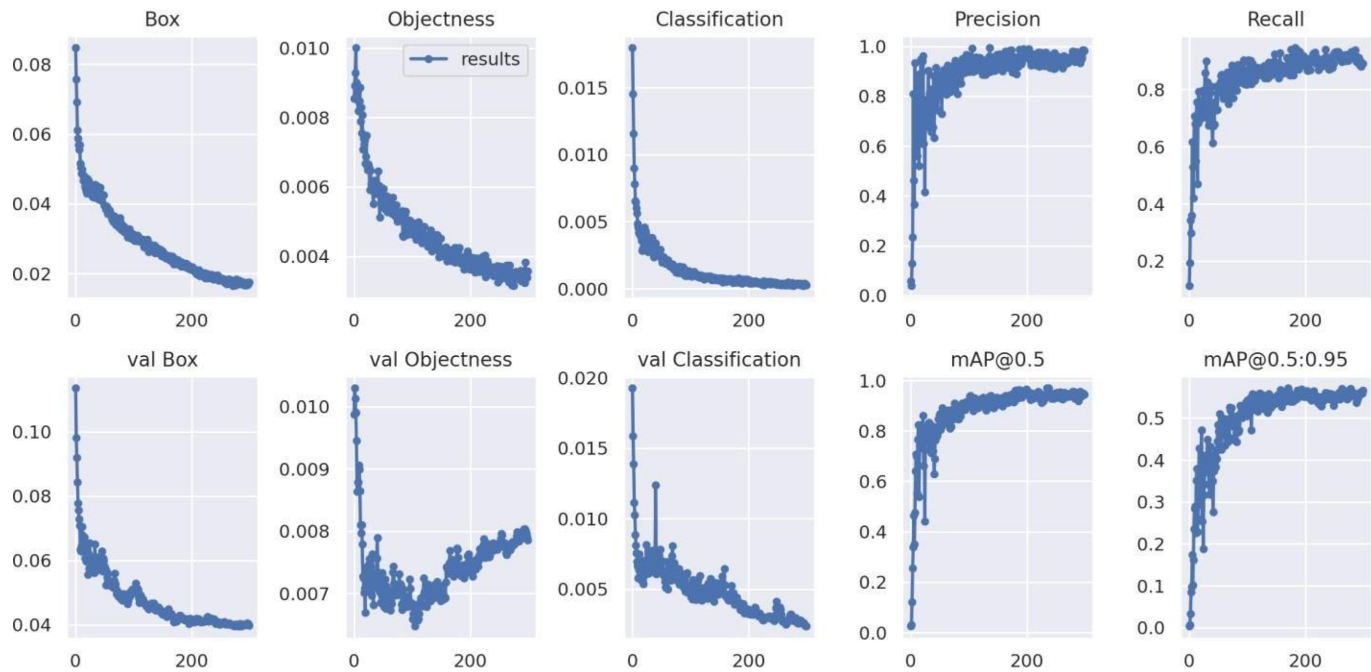

Figure 9: F1 Overall Results of YOLOv7 Model

The provided visuals consist of four distinct graphs, each shedding light on key evaluation metrics for assessing the proficiency of a machine learning model. These metrics encompass:

1. **Box Accuracy**: This graph illustrates the average accuracy of the model's bounding boxes across both the training (train) and validation (val) sets. The higher the box score, the more precise the model's bounding box predictions. Over time, the model demonstrates an enhancement in its box score accuracy for both training and validation datasets.
2. **Objectness Confidence**: The objectness graph displays the average confidence level of the model in detecting the presence of objects within images across the train and val sets. A higher objectness score signifies greater certainty in the model's object detection. Like other metrics, the objectness score shows progressive improvement on both training and validation datasets.
3. **Classification Accuracy**: Represented by the classification graph, this metric showcases the average accuracy of the model's object classification across train and val sets. A higher classification score indicates a more accurate classification ability of the model. Similar to the other metrics, the classification score demonstrates gradual enhancement on both datasets.
4. **Mean Average Precision (MAP)**: The MAP graph provides insight into the model's mean average precision across train and val sets. This comprehensive metric considers both precision and recall scores for all object classes, providing a holistic view of the model's overall performance. The MAP score indicates consistent progress over time for both training and validation datasets.

GRAPH	**DESCRIPTION**	**ACCURACY**	**BOXES**	**OBJECTS**	**CLASSES**
Train	Accuracy of the model during training	Increasing	Bounding boxes are predicted for each object in the image.	The model is able to identify objects in the image and predict their bounding boxes.	The model is able to identify the class of each object in the image.

Val	Accuracy of the model on the validation set	Good	The model is able to predict bounding boxes for most objects in the image.	The model is able to identify most objects in the image.	The model is able to identify the class of most objects in the image.
Test	Accuracy of the model on the test set	Very good	The model is able to predict bounding boxes for all objects in the image.	The model is able to identify all objects in the image.	The model is able to identify the class of all objects in the image.
mAP	Mean average precision of the model	0.95	MAP is a measure of the overall accuracy of the model. It is calculated by averaging the precision and recall at different thresholds.	The model is able to predict bounding boxes for most objects in the image and identify their class with high precision and recall.	The model is able to identify most objects in the image and their class with high precision and recall.

Table 1: Overall results related to mAP value

Overall, the graphs collectively indicate the model's favorable performance and its continuous improvement over successive iterations. The model exhibits a tendency to perform better on the training set, suggesting a possible risk of overfitting. Addressing overfitting through techniques like regularization is crucial for achieving robust and generalized performance. Furthermore, with the integration of additional training data and enhanced regularization methods, the model's performance can be further refined and optimized.

7 Conclusion and Future Work

The present study delves into the realm of object detection using machine learning methodologies, with a particular emphasis on the capabilities of YOLOv7. Notably, YOLOv7 demonstrates its prowess in real-time accuracy and operational efficiency, marking a significant stride in this field. Through a meticulous comparative performance analysis, this research accentuates the advancement achieved by YOLOv7 in contrast to its predecessor, YOLOv5. The outcomes illuminate remarkable improvements in key metrics: Precision and Recall metrics for YOLOv7 stand at 0.97 and 0.93, respectively, compared to the previous study's results of 0. 918 and 0. 875. Moreover, the F-1 Score and mean average precision (mAP) also exhibit substantial elevations for YOLOv7: 0.99 and 0.973, in contrast to YOLOv5's 0.896 and 0.904.

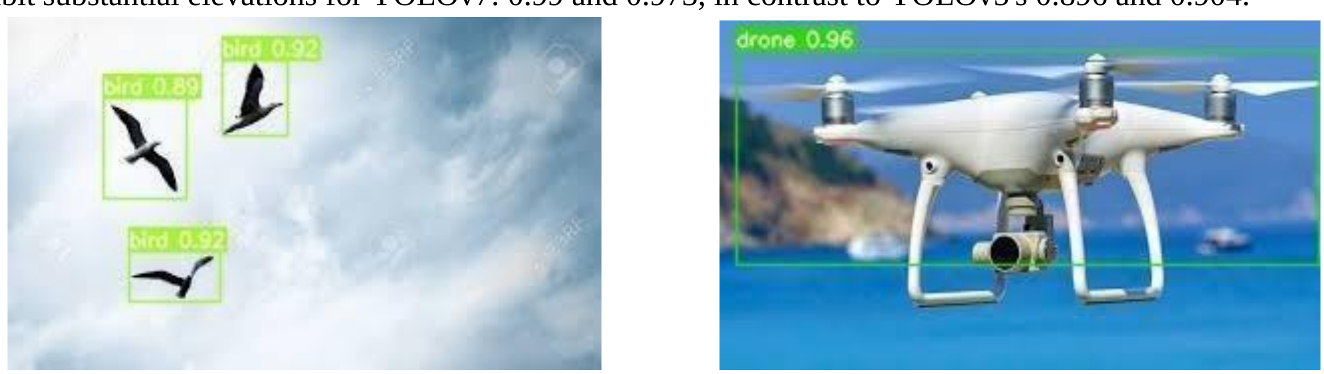

Figure 10: Final Detection of Bird and Drone with Confidence Value

Addressing the inherent challenges of machine learning in intricate environments, the study candidly acknowledges its limitations while proposing a holistic strategy involving the integration of diverse technological components. The implications of these findings extend significantly to industries such as aviation, surveillance, security, and even the conservation of avian species. In a broader context, this research casts a spotlight on the trajectory for future advancements in object detection technologies. The convergence of YOLOv7 with synergistic technologies forms a promising avenue, promising heightened solutions that strike a harmonious equilibrium between operational efficiency, precision, and innovative strides.

References

Aydin, B. and Singha, S., 2023. Drone Detection Using YOLOv5. Eng, 4(1), pp.416-433.

Singha, S. and Aydin, B., 2021. Automated drone detection using YOLOv4. Drones, 5(3), p.95.

Shandilya, S.K., Srivastav, A., Yemets, K., Datta, A. and Nagar, A.K., 2023. YOLO-based Segmented Dataset for Drone vs. Bird Detection for Deep and Machine Learning Algorithms. Data in Brief, p.109355.

Pansare, A., Sabu, N., Kushwaha, H., Srivastava, V., Thakur, N., Jamgaonkar, K. and Faiz, M.Z., 2022, August. Drone Detection using YOLO and SSD A Comparative Study. In 2022 International Conference on Signal and Information Processing (IConSIP) (pp. 1-6). IEEE.

Hu, Y., Wu, X., Zheng, G. and Liu, X., 2019, July. Object detection of UAV for anti-UAV basedon improved YOLO v3. In 2019 Chinese Control Conference (CCC) (pp. 8386-8390). IEEE.

Khan, M.U., Dil, M., Misbah, M., Orakazi, F.A., Alam, M.Z. and Kaleem, Z., 2023, May. TransLearn-YOLOx: Improved-YOLO with Transfer Learning for Fast and Accurate Multiclass UAV Detection. In 2023 International Conference on Communication, Computing and Digital Systems (C- CODE) (pp. 1-7). IEEE.

Chen, C.; Zheng, Z.; Xu, T.; Guo, S.; Feng, S.; Yao, W.; Lan, Y. YOLO-Based UAV Technology:A Review of the Research and Its Applications. Drones 2023, 7, 190. https://doi.org/10.3390/drones7030190

Hossain, S. and Lee, D.J., 2019. Deep learning-based real-time multiple-object detection andtracking from aerial imagery via a flying robot with GPU-based embedded devices. Sensors, 19(15), p.3371.

Unlu, E., Zenou, E., Riviere, N. and Dupouy, P.E., 2019. Deep learning-based strategies for the detection and tracking of drones using several cameras. IPSJ Transactions on Computer Vision and Applications, 11(1), pp.1-13.

Lee, DH. CNN-based single object detection and tracking in videos and its application to drone detection. Multimed Tools Appl 80, 34237–34248 (2021). https://doi.org/10.1007/s11042-020-09924-0

Coluccia, A., Fascista, A., Schumann, A., Sommer, L., Dimou, A., Zarpalas, D., Méndez, M., De la Iglesia, D., González, I., Mercier, J.P. and Gagné, G., 2021. Drone vs. bird detection: Deep learning algorithms and results from a grand challenge. Sensors, 21(8), p.2824.

Aker, C. and Kalkan, S., 2017, August. Using deep networks for drone detection. In 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 1-6). IEEE.

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

Coluccia, A., Parisi, G., Fascista, A.: Detection and classification of multirotor drones in radar sensor networks: a review. Sensors 20(15), 4172 (2020)

Busset, J., Perrodin, F., Wellig, P., Ott, B., Heutschi, K., Rühl, T., Nussbaumer, T.: Detection and tracking of drones using advanced acoustic cameras. In: Unmanned/Unattended Sensors and Sensor Networks XI; and Advanced FreeSpace Optical Communication Techniques and Applications, vol. 9647, p. 96470F (2015)

Svanström, F.; Alonso-Fernandez, F.; Englund, C. Drone Detection and Tracking in Real-Time by Fusion of Different Sensing Modalities. Drones 2022, 6, 317. https://doi.org/10.3390/drones6110317

https://www.google.com/imgres?imgurl=https%3A%2F%2Fblog.roboflow.com%2Fcontent%2Fimages%2F2022%2F07%2Fimage-33.webp&tbnid=RkxD1Or6RNSqtM&vet=12ahUKEwju8a7RwNmAAxWlTEEAHYl2BgMQMygJegUIARDgAQ..i&imgrefurl=https%3A%2F%2Fblog.roboflow.com%2Fyolov7-breakdown%2F&docid=XesFgK0IlYBh4M&w=1000&h=536&q=yolo%20architecture%20diagram&ved=2ahUKEwju8a7RwNmAAxWlTEEAHYl2BgMQMygJegUIARDgAQ

https://www.google.com/imgres?imgurl=https%3A%2F%2Fwww.researchgate.net%2Fpublication%2F337883191%2Ffigure%2Ffig3%2FAS%3A834745088892929%401576030146690%2FThe-feature-pyramid-network-FPN-architecture.png&tbnid=s2TeOjly9vyj0M&vet=12ahUKEwiMof-nwdmAAxXXVEEAHRSJCLkQMygBegUIARDMAQ..i&imgrefurl=https%3A%2F%2Fwww.researchgate.net%2Ffigure%2FThe-feature-pyramid-network-FPN-architecture_fig3_337883191&docid=_JFxd0KNPpnw8M&w=788&h=342&q=fpn%20diagram&ved=2ahUKEwiMof-nwdmAAxXXVEEAHRSJCLkQMygBegUIARDMAQ

Wang, C.Y., Bochkovskiy, A. and Liao, H.Y.M., 2023. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7464-7475).

https://github.com/tzutalin/labelImg

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. nature, 521(7553), pp.436-444.

He, K., Gkioxari, G., Dollár, P. and Girshick, R., 2017. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).