A Hybrid Approach to Enhance Drone and Bird Differentiation Using YOLOv7 and Deep Learning Classification Models

MSc Research Project
Data Analytics

Vinutha Nagaraju
Student ID: 23110686

School of Computing
National College of Ireland

Supervisor: Hamilton V. Niculescu

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Vinutha Nagaraju
Student ID:	23110686
Programme:	Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Hamilton V. Niculescu
Submission Due Date:	12/08/2024
Project Title:	A Hybrid Approach to Enhance Drone and Bird Differentiation Using YOLOv7 and Deep Learning Classification Models
Word Count:	7500
Page Count:	22

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:	Vinutha Nagaraju
Date:	16th September 2024

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):	

A Hybrid Approach to Enhance Drone and Bird Differentiation Using YOLOv7 and Deep Learning Classification Models

Vinutha Nagaraju
23110686

Abstract

This research explores a hybrid approach to enhance the accuracy of distinguishing between drones and birds, leveraging the strengths of both object detection and deep learning classification models. The motivation behind this study stems from the increasing prevalence of drones which are usually referred to as UAV(Unmanned Aerial Vehicles) in various sectors like aerial photography, delivery services and surveillance. The ease of use of drones equipped with camera is causing serious privacy issues and potential security threats, as drones can be used for reconnaissance, gathering intelligence, or even facilitating attacks This highlights the need for efficient detection system, various researches have been going on in this field, the major problem faced is distinguishing drone from similar objects like birds since they look visually similar. To address this, the study aimed to determine the most effective classification model among VGG16(Visual Geometry Group 16), ResNet18(Residual Network 18) and InceptionV3(Inception Version 3) and evaluate the impact of integrating it with YOLOv7(You Only Look Once, Version 7) on overall detection accuracy.YOLOv7 achieved a mAP(Mean Average Precision) of 0.913 with F1-score of 0.8867. VGG16 was found to be the most effective model with accuracy of 98.45% with higher precision and recall rate exceeding 90%. A hybrid pipeline was developed where YOLOv7 detects the object, and the Region of Interest (ROI) is classified by VGG16. However, the proposed method, with an accuracy of 90.38% on the test dataset, did not consistently outperform YOLOv7 alone and sometimes misclassified drones as birds. These results highlight the potential of combining detection and classification models but also indicate the need for more adaptive and iterative approaches.

1 Introduction

Drones, a general term for any unmanned vehicle, are specifically referred to by International Civil Aviation Organization(ICAO) as Unmanned Aerial Systems(UAS) Delleji et al. (2020). Drones come in various shapes and sizes. Drones provide significant advantages across various fields, enhancing agriculture with detailed crop imagery and data analysis. They improve public safety through effective search and rescue operations and disaster response, while revolutionizing construction with precise surveying. Additionally, drones support environmental conservation, streamline commercial services like package delivery, and offer unique aerial perspectives in filmmaking and journalism Björklund and

Wadströmer (2019).As drones become more prevalent in both commercial and military applications, they pose significant security challenges. Illegitimate uses of drones, such as unauthorized surveillance, reconnaissance, and even attacks, have raised serious concerns Mahdavi and Rajabi (2020).

Camera-equipped drones can easily breach secure areas, making traditional surveillance methods like CCTV insufficient. This necessitates incorporation of automated systems to enhance intrusion detection capabilities.

A drone detection system identifies, tracks, and responds to UAVs within a specific area to ensure airspace safety and prevent unauthorized activities Naveen et al. (2023). These systems use radar, RF detectors, visual cameras, and acoustic sensors—radar measures reflected waves, RF detectors capture signals between drones and controllers, cameras provide visual tracking, and acoustic sensors detect drone sounds. Machine learning and AI enhance detection by processing sensor data, reducing false positives, and improving effectiveness Taha and Shoufan (2019).

Deep learning has shown remarkable performance in object recognition and computer vision, but challenges remain in drone detection. Detecting drones in complex backgrounds or when they are small, and misclassification of birds as drones, leading to false positives, are key issues Al-Zahrani (2023). Since birds and drones often share the same airspace, it's vital to develop methods to distinguish between them. Accurate classification is essential to prevent costly false alarms and ensure dangerous drones are not mistaken for birds Mohamed and Alharbi (2023). Higher classification accuracy and lower false alarm rates make detection systems more effective across various applications.

YOLOv7 stands for "You Only Look Once, Version 7." It is a state-of-the-art object detection model that builds upon the previous versions of YOLO series, developed by a research team including Chien-Yao Wang, Alexey.YOLOv7 processes images in real time, making it highly suitable for detecting small objects like drones, even in cluttered or complex backgrounds. YOLOv7's architecture allows it to perform detection with high precision, which is crucial for distinguishing drones from other objects. ResNet18(Residual Network 18) is effective in managing issue of misclassification by leveraging its deep learning capabilities. By incorporating residual connections, it overcomes vanishing gradient problem, allowing network to learn more complex features and improve its accuracy. Inception v3 is known for speed and accuracy, making it ideal for scenarios where real-time classification is needed. VGG16(Visual Geometry Group 16), has simple and depth architecture, can capture detailed features that might be missed by other models. it also has proven performance in classification tasks Oh et al. (2019).YOLOv7 is a detection algorithm, while ResNet18, VGG16, and InceptionV3 are classification algorithms.

To address these issues, a dataset of drones and birds with various backgrounds, orientations, sizes, colors, and weather conditions was used. To overcome misclassifications, a novel approach combining detection and classification is proposed. Unlike previous studies that focused on either detection or classification, this research leverages both to improve detection accuracy and reduce false positives. Different classification algorithms, including ResNet18, VGG16, and InceptionV3, were tested, and the most effective one in terms of accuracy, precision, recall, and F1 score will be integrated with YOLOv7, a state-of-the-art object detection framework known for its speed and accuracy.The separation into detection and identification modules allows for efficient training and scalability.

This study does have certain limitations.The proposed fusion method is heavily reliant on performance of YOLOv7.

The paper is structured as follows. A literature review of various drone detection system is presented in section 2. Section 3 describes research methodology. Design specifications is described in sections 4.Section 5 provides implementation details.Section 6 evaluates experiments conducted based on results, followed by Discussion section. Section 7 presents conclusion of research and future work.

2 Related Work

Research on distinguishing drones from birds has advanced, with significant focus on machine learning and deep learning. Many studies have employed different algorithms to tackle this issue. Below is an overview of recent developments, highlighting key trends and research gaps.

2.1 Conventional Techniques of Drone Detection

Taha and Shoufan (2019) reviewed state-of-the-art technologies for drone detection and classification using machine learning, focusing on radar, visual, acoustic, and RF signals. The study found that machine learning significantly improved radar detection accuracy and reduced false positives. Visual detection methods using CNNs showed high precision, but acoustic and RF-based methods faced challenges due to environmental noise and limited datasets.

Al-Zahrani (2023) explored drone detection and classification based on factors like size, weight, and application. The paper discussed the strengths and limitations of each method, noting challenges like radar's difficulty in distinguishing drones from birds and acoustic sensors struggling in noisy environments. RF analysis faced interference, and LADAR was energy-intensive. The study suggests that combining multiple detection methods could enhance performance to counter rogue drones but lacks detailed experimental comparisons between the techniques.

Al-Emadi and Al-Senaid (2020) presents a drone detection solution using Deep Learning (CNNs) to analyze Radio Frequency (RF) signals from live drone-controller communication sessions. Using a dataset with 454 records from various drones and non-drone activities, the solution achieves 99.7% accuracy and F1 score for drone detection and 88.4% accuracy for drone identification which shows the efficiency of CNNs in detection tasks. However, reliance on a lab-generated dataset may limit performance in real-world scenarios with high RF noise or unfamiliar drones.

Björklund and Wadströmer (2019) focuses on distinguishing small drones from birds and classifying drone types using a deep learning classifier on radar measurements. Using micro-Doppler Time Velocity Diagrams (TVDs) from X-band radar data, the study achieved 97.6% accuracy and a probability of false alarm (PFA) below 0.03%. Despite the high accuracy, the data had limited variability, and the classifier was not tested on unseen targets, different backgrounds, or varied weather conditions, though it showed significant improvement over previous SVM and Boosting classifiers.

2.2 Classifying Drones and Birds Using Deep Learning

Mahdavi and Rajabi (2020) explores the effectiveness of CNN, SVM, and KNN for drone detection, aiming to address security concerns from drone misuse. Using 712 images, features were extracted using HOG for SVM and KNN, while CNN handled automatic feature extraction. CNN achieved the highest accuracy at 93%, followed by SVM at 88%, and KNN at 80%. The study concludes that CNN outperforms the other methods in accuracy and reliability, suggesting that advanced CNN architectures and hyperparameter optimization could further improve detection performance.

Oh et al. (2019) compares the performance of seven well-known CNN models: Alexnet, GoogLeNet, Inception-v3, VGG-16, ResNet-18, ResNet-50, and SqueezeNet to identify the best models to classify drone v/s bird.These models were chosen due to their prior validation in the ImageNet Large Scale Visual Recognition Competition (ILSVRC), where they demonstrated high performance in classifying 1000 labels. However, this study focuses on only three labels: drone, bird, and background. A dataset comprising 7000 images (3000 birds, 1500 drones, 2500 backgrounds) was used, with 5100 images for training and 1900 for testing.In terms of accuracy alexnet, VGG-16, Resnet- 18 has performed better than other models. Complex models like GoogLeNet and Inception-v3 showed lower accuracy and efficiency due to insufficient batch size and epochs for training.

The performance of three popular pre-trained deep learning models—VGG16, ResNet18, and InceptionV3 using transfer learning technique on a public dataset consisting of 320 images of drone and bird is evaluated in study by Mohamed and Alharbi (2023) .Transfer learning is opted due to limited dataset. Images are resized to fit the requirements of the pre-trained models. The pre-trained networks (VGG16, ResNet18, InceptionV3) are modified by freezing the first layers and adding new convolutional and max-pooling layers to fine-tune the models for the specific dataset. Performance metrics like accuracy, F-Score, precision, and recall are used to compare the effectiveness of each model. ResNet18 achieved the highest performance with an accuracy and F-Score exceeding 98%.VGG16 and InceptionV3 also demonstrated excellent performance, with accuracy and F-Score exceeding 94%.The dataset used is relatively small, which may limit the generalizability of the findings to larger and more diverse datasets.

Dale et al. (2021) tackles the challenge of distinguishing drones from birds in low-altitude airspace, where similar radar returns cause high false alarm rates. The study used 966 spectrograms (half birds, half DJI Inspire 1 drone) and tested six CNN architectures (AlexNet, GoogLeNet, SqueezeNet, Inception-v3, ResNet-18, ResNet-50) under varying SNR levels by adding Gaussian noise. All CNNs performed well at high SNR, but accuracy dropped with lower SNR. AlexNet had the highest accuracy (81.3%) at 24 dB degradation, while deeper networks like ResNet-50 and Inception-v3 saw accuracy drop to around 50%. The study suggests exploring data augmentation to improve performance at low SNRs, highlighting that shallower networks are more robust to noisy data than deeper architectures.

Ultra-Wideband (UWB) radar profiles were collected for five drones and a radio-controlled bird. Kurosaki et al. (2022) used four CNN models (AlexNet, GoogLeNet, ResNet-50, ResNet-101) for classification, applying transfer learning with pre-trained networks fine-tuned on the radar data. Using k-fold cross-validation (k=6 and k=10), all models achieved over 90% accuracy. ResNet-50 performed best, with 97.8% accuracy at k=6 and 94.1% at k=10. The study confirms that transfer learning effectively enhances CNN performance for classification tasks.

The classification algorithms such as Resnet18, VGG16, InceptionV3 has performed well in classifying drones, hence these are chosen for this research.

2.3 Drone v/s Bird Detection with Advanced Techniques

Wei Xun et al. (2021) utilized transfer learning with pre-trained YOLOv3 weights to detect drones, deploying the system on the NVIDIA Jetson TX2 for real-time application. The dataset, initially consisting of 1435 images, was expanded to 7175 images through data augmentation techniques. The trained model achieved an average accuracy of 88.9% , demonstrating robust performance in various conditions like sunny,cloudy and altitudes (30m, 40m). This study highlights the effectiveness of transfer learning and data augmentation in improving detection accuracy and the feasibility of real-time deployment on the NVIDIA Jetson TX2. YOLOv3 have also been used by Delleji et al. (2020) for better small object detection of five drone types.

Naveen et al. (2023) trained the YOLOv5 model using Google Colab, which provides access to GPUs, facilitating efficient training of the deep learning model. Synthetic data augmentation is applied to enhance the model's performance and generalization. The integration of synthetic data and the use of cloud-based platforms for efficient training are relevant strategies that can be applied to enhance the accuracy and robustness of the detection system.Different scenarios where the model performed well or not is not shown and clear evaluation of model is not documented.

The performance of YOLOv4 and YOLOv5 is evaluated by Sethu Selvi et al. (2022) for detecting drones and birds.They have Collected a custom dataset consisting of 664 drone images and 236 bird images from various internet sources.The YOLOv4 model achieved a mean Average Precision (mAP) of 97.4% and an F1-score of 98%, demonstrating its high accuracy in detecting drones and birds. Additionally, it had a detection speed of 54 frames per second (fps), making it suitable for real-time applications. On the other hand, the YOLOv5 model, while slightly less accurate with a mAP of 95% and an F1-score of 94%, outperformed YOLOv4 in terms of detection speed, achieving 77 fps. This indicates that YOLOv5, with its faster detection capabilities, offers a significant advantage for applications requiring rapid response times.

Karthikeya Nalam et al. (2022) proposes two deep learning-based approaches, YOLOv4 and Faster RCNN for this problem.The dataset includes 1097 images of drones and 1200 images of birds sourced from Kaggle and OpenImagesv6.Roboflow tool is used to annotate the images. YOLOv4 uses a dense prediction architecture with input, backbone, neck, and head segments. Faster RCNN uses a Region Proposal Network (RPN) for generating object proposals and classifying them through convolutional layers. The dataset was split into 70% for training and 30% for testing.The input data were classified using a 0.5 threshold in this analysis. YOLOv4 achieved a mean Average Precision (mAP) of 75%, demonstrating higher performance compared to Faster RCNN, which achieved an mAP of 72%.The results indicated that YOLOv4 is more effective in detecting and recognizing drones in images with challenging backgrounds and varying object sizes.Future work in this research suggests the investigating of use of YOLOv5 and other recent advancements in object detection to further enhance detection speed and accuracy.

Valaboju et al. (2023) conducted research to identify, locate, and classify drones into categories like Army, Surveillance, and Delivery drones to distinguish between potential threats and harmless drones. Images of specific military drones, delivery drones with suspended boxes, and quadcopters with cameras were used. The YOLOv5 algorithm was implemented, achieving 68% accuracy in detection. Backgrounds were sometimes mistaken for drones, suggesting a need to increase detection confidence. Future work aims to improve accuracy by reducing misidentifications, adding more military drone types, and refining classification within the Army drone category.

By using the freely available Drone-vs-Bird dataset from Kaggle, a study on comparison of YOLOv4 and YOLOv5 is performed by Jarray and Bouallègue (2023). The dataset is split into 80% for training and 20% for testing the YOLOv5 model.The model is trained using the Google Colab, which provides access to GPUs.YOLOv5 Model has achieved a mAP of 98.4%, precision of 96.9%, recall of 93.5%, and F1-score of 95%.YOLOv4 model has achieved a mAP of 94.57%.The results indicate that YOLOv5 outperforms YOLOv4 in terms of mAP, making it more efficient for drone detection.

Pansare et al. (2022) have developed a drone detection system using two deep learning-based algorithms, YOLOv5 and SSD (MobileNetSSDv2). The study uses two datasets for training and testing the algorithms. The first is an open-source dataset (Anti-UAV), and the second is created by parsing video frames from a widely used video dataset.The paper concludes that both YOLOv5 and MobileNetSSDv2 are effective for drone detection, with YOLOv5 showing superior performance in terms of precision and accuracy. Kataria and Lall (2023) have employed centroid tracking algorithm along with YOLOv5 to improve the accuracy of YOLOv5 model.

Lee et al. (2018) developed a two-part system for drone detection and identification in video frames captured by a drone's camera. Using the Haar Cascade Classifier from OpenCV, they trained on 2088 drone images and 3019 non-drone images. The detection module achieved 89% accuracy, and the identification module reached 91.6% accuracy. The system combines object detection and classification for real-time monitoring, with potential for improvement using more complex models like ResNet.The separation into detection and identification modules allows for efficient training and scalability. The modular design and practical implementation details provide a solid foundation for developing similar systems aimed at enhancing drone surveillance and security.

From the above papers, it is evident that YOLO-based models are highly effective and can handle challenging detection scenarios, which supports the use of YOLOv7 in this research which is one of the recent official version.

The research questions are as follows:

What is the best algorithm to classify drone and birds accurately among Vgg16, ResNet18 and Inceptionv3?

How does integrating YOLOv7 with classification algorithms impact the accuracy of distinguishing between drones and birds?

This research explores the effectiveness of classification algorithms for distinguishing drones and birds and their integration with YOLOv7 for object detection. By identifying the most accurate model, the study aims to enhance the detection and classification of drones and birds under various conditions, improving surveillance and security systems to better manage drone-related activities and ensure safety in complex environments.

The research objectives are as follows:

- Implementing the YOLOv7 algorithm for the detection of drone v/s bird in an image.
- Comparision of classificaton algorithm such as VGG16,ResNet18,InceptionV3 in classifying drones and birds.
- Integrating the classification model with YOLOv7 by utilizing the bounding boxes predicted by YOLOv7.
- Evaluating the pipeline model to verify if that works better than the baseline YOLOv7.

3 Methodology

The Cross-Industry Standard Process for Data Mining (CRISP-DM) framework is valuable for development of any machine learning project in a systematic way.The approach consists of six essential phases - business understanding, data understanding, data preparation, modeling, evaluation, and deployment Hayat Suhendar and Widyani (2023). All phases are followed to develop the system which distinguishes drones from birds other than the deployment stage.

3.1 Business Understanding

This phase involves understanding the project's goals and requirements from a business perspective.By implementing the proposed method of integrating detection with classification using state-of-the-art machine learning models such as YOLOv7 for object detection and classification models like ResNet18, VGG16, and InceptionV3, the project aims to achieve high detection accuracy while minimizing false positives. The ultimate goal is to develop a reliable and efficient system that can be integrated into existing security infrastructures to mitigate potential threats posed by unauthorized drones.It addresses the need for improved security measures in a rapidly evolving technological landscape, ensuring that the benefits of drone technology can be used safely and effectively.

3.2 Data Understanding

This phase involves data collection and initial data exploration.For data collection, two open-source datasets were selected from the Kaggle available at: `https://www.kaggle.com/datasets/harshwalia/birds-vs-drone-dataset` and the Mendeley Data website available at `https://data.mendeley.com/datasets/6ghdz52pd7/3`.The Kaggle dataset consists of 428 images of drones and 400 images of birds, total of 828 images. To enhance the dataset with diverse backgrounds, orientations, and lighting conditions, the second dataset was included. The second dataset is well-structured with train, test, and validation folders, following the YOLOv7 PyTorch specifications. The training images in this dataset were already segmented and augmented. However, to ensure the model learns accurate object detection features, high-quality data is essential. Therefore, only the test images from the second dataset, which are clean images of drones and birds, were considered. This consists of 889 images, providing a robust basis for training the model effectively.The images consist of drone and bird images belonging to 2 classes as shown in the figure 1. The image sizes were not uniform across the dataset which needs to be resized later before training the model.The maximum RGB value is 255 and minimum RGB value is 0 for the images shown.All the images are coloured images.

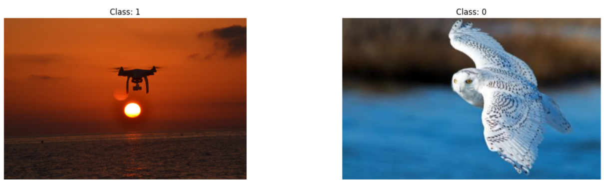

Figure 1: Representation of each class

3.3 Data Preparation

Data preparation is a crucial step to prepare the data to make it suitable for different models.The second dataset from mendeley consist of images which were extracted from videos, hence many frames were identical. To ensure diversity in the curated dataset, duplicate images were checked.Identifying duplicate images is essential to prevent overfitting, ensuring the model learns from diverse examples rather than memorizing specific ones.Additionally, removing duplicates maintains data quality and integrity, providing more accurate performance metrics and a balanced, representative dataset.

YOLOv7 needs both images and corresponding annotation files. The annotation files should have the same name as the images but with a .txt extension and contain the class and bounding box details in the format: <object-class> <x_center> <y_center> <width> <height> for each image. In order to annotate the data, common labeling tools include LabelImg, Labelbox, VGG Image Annotator (VIA) are used. However, Roboflow stands out due to its user-friendly interface, which simplifies the annotation process, and its automated annotation capabilities with pre-trained models, significantly speeding up labelling. Roboflow supports multiple export formats compatible with various machine learning frameworks Karthikeya Nalam et al. (2022). So the images are annotated using roboflow tool. During inspection of the labelled images, many irrelevant images, as shown in figure 2, were removed. Additionally, images with watermarks, as shown in figure 3, which had watermark all over the image or at random positions which was hard to crop hence those images are not considered. Watermarked images should be excluded as it can introduce noise and distortions that the model might mistakenly learn as relevant features, reducing its accuracy ans also to make the images close to reality. Finally there were 1276 of drone and bird images with annotations.Total number of images and respective annotation files are checked.The annotation files are usually produced while exporting the dataset from roboflow, hence it is made sure again that there are no null values or no missing files of annotations so that the training of model will be seamless.Also the images and labels are renamed.Renaming ensures a consistent naming convention, which aids in managing and organizing datasets while avoiding filename conflicts.

Figure 2: Example of irrelavant image

Figure 3: Watermarked images

To properly divide the data into train, test, and validation sets needed by YOLOv7, simply splitting won't ensure equal portions of drone and bird images. Therefore, dataset is first divided based on class, then randomly shuffled to avoid bias, and finally split into train, validation, and test sets in a 70:20:10 ratio.Karthikeya Nalam et al. (2022) have used 70% for training and 30% for testing,since validation data is also important which allows adjustment of hyperparameters and selection of the best model without overfitting to the training data, choice of 70:20:10 ratio is made.This ratio ensures sufficient training data to minimize bias and enough validation and test data to reduce variance and ensure reliable performance evaluation.YOLOv7 requires the train, validation, and test sets without separate folders for each class, just the images and labels respectively.

Classification algorithms require only a folder containing images for each class in the dataset.To train the classification algorithm, the data needs to be organized into separate folders for each class (drone and bird) under train, test, and validation sets. The same dataset is used to train the classification algorithm.Normalization helps stabilize and speed up the convergence of the model during training by ensuring that all input features have a similar scale Mohamed and Alharbi (2023).Normalization of images are done and On-the-fly augmentation is applied, where augmentations are applied dynamically during training, providing unlimited data variability, reducing storage needs, allows for dynamic adjustments, and improves training efficiency and model generalization as used by Naveen et al. (2023).

3.4 Modeling

The modeling phase involved training both detection model YOLOv7 and classification models like Resnet18, VGG16, InceptionV3. We have hypertuned all the models which is crucial for optimizing model performance, as it involves adjusting parameters like learning rate and epochs, optimizer to find the best combination for training Mahdavi and Rajabi (2020).

For the training of YOLOv7 model, various configurations were experimented to optimize the model. Initially, YOLOv7 was trained for 10 epochs with a high learning rate of 0.1 and the SGD optimizer, using images resized to 640x640, a batch size of 16, and pretrained weights.The learning rate determines the speed at which the model learns, and it is a small positive value typically ranging from 0 to 1.0 Wei Xun et al. (2021).So we have experimented with learning rate of 0.1, 0.01,0.001 in our study. The batch size refers to the number of training examples utilized in one iteration of the model's training process. It influences the model's learning process, with larger batch sizes providing more stable updates, while smaller batch sizes offer more frequent updates.Wei Xun et al. (2021) have used batch size of 12, in order to get more stable output the batch size of 16 was used. The Stochastic Gradient Descent (SGD) optimizer updates model parameters using only a subset of data at each iteration. This incremental approach reduces memory usage and can lead to faster convergence.Additionally, its stochastic nature helps in escaping local minima, potentially leading to better overall model performance.The accuracy was very low.Hence the epochs were increased to 15, maintaining the same learning rate and optimizer, to allow more learning while ensuring stability.But there were no much improvements. In next steps , the model was trained for 20 epochs with a reduced learning rate of 0.01 and switched to the Adam optimizer, which adapts the learning rate for each parameter, aiming for better convergence.The Adam (Adaptive Moment Estimation) optimizer is a popular optimization algorithm used in training deep learning models.Sethu Selvi et al. (2022) have also used the adam optimizer for their research which has given them the good results.But with the 0.01 learning rate the performance was poor.Hence the model was trained for 25 epochs with an even lower learning rate of 0.001,With this combination model performed well with a mean Average Precision (mAP) of 0.935.Finally, we trained the model for 25 epochs with a learning rate of 0.01 and the SGD optimizer which yielded mAP of 0.953 and is chosen as final model of yolov7 for better detection performance.

For training ResNet18, we first employed the Adam optimizer with 10 epochs with a learning rate of 0.01, and cross-entropy loss. Cross-entropy loss was chosen as it is particularly well-suited for classification tasks.It measures the performance of a classification model whose output is a probability value between 0 and 1.With 10 epochs the performance was only around 50%. Hence the training was extended to 25 epochs with a reduced learning rate of 0.001, again using the Adam optimizer and cross-entropy loss, and introduced early stopping. The lower learning rate facilitated more precise weight updates, while early stopping helped prevent overfitting by halting training when the validation performance stopped to improve. A patience value of 5 is used for early stopping meaning the training will stop if there is no improvement in validation performance for 5 consecutive epochs.It keeps track of the best validation performance observed and save the corresponding model weights as a checkpoint. The model achieved a accuracy of 93.31 % at epoch 18.

For the training of VGG16 model, Adam optimizer with learning rate of 0.01 is used over 10 epochs.Next, the training duration was extended to 25 epochs with early stopping with reduced learning rate of 0.001.But both the training yielded accuracy only upto 50%.In order to enhance the performance of VGG16 model, a custom classifier was integrated into the model architecture, including dropout and batch normalization layers. Dropout was set at a rate of 0.5, which helps prevent overfitting by randomly deactivating 50% of the neurons during each training iteration. This encourages the model to learn more robust features by not relying on specific neurons, thus improving its generalization to unseen data Al-Emadi and Al-Senaid (2020). Batch normalization was applied to standardize the inputs to each layer.This stabilization reduces the chances of the network parameters diverging during training, leading to more stable and reliable learning Mohamed and Alharbi (2023).The model was trained with the SGD optimizer, using a learning rate of 0.01 and a momentum of 0.9 to maintain steady updates with early stopping over 25 epochs.The model yielded accuracy of 97.24%.

In training the InceptionV3 model, adam optimizer with a learning rate of 0.01 for 10 epochs, leveraging pretrained weights is used. This setup achieved an accuracy of 50.39% indicating room for improvement. Next,the learning rate was lowered to 0.001 since lower learning rate allowed for finer adjustments to the model parameters, promoting more stable and gradual learning Sethu Selvi et al. (2022). The training was extended to 20 epochs, the model's performance improved dramatically, achieving an accuracy of 96.85%. Additionally, the extended training period provided the model with more opportunities to learn and refine the features from the data. This combination of a lower learning rate and longer training period proved to be highly effective, resulting in a increase in accuracy.

3.5 Evaluation Metrics

To evaluate the models, we will use several key metrics and visual tools. The primary metric for the YOLOv7 model is mean Average Precision (mAP), which measures precision and recall across different classes and thresholds, using an IoU threshold of 0.5

for correct bounding box predictions. We'll also use a confusion matrix to detail true positives, false positives, true negatives, and false negatives, helping to identify any bias in the model's classifications. Precision, recall, and the F1 score will be calculated to assess the accuracy of predictions, while overall accuracy will determine the proportion of correct predictions Karthikeya Nalam et al. (2022).

Additionally, we'll analyze training and testing accuracy to check for overfitting. ROC curves will visualize trade-offs between true positive and false positive rates. Training and validation loss curves, along with learning curves, will provide insights into the model's learning progress and performance over time. These metrics and tools ensure a comprehensive evaluation of the model's ability to detect and classify drones and birds.

4 Design Specification

The purpose of this research is to develop a drone v/s bird detection and classification system that leverages YOLOv7 model for detecting objects and VGG16 classifier for classifying. This system is designed to accurately identify objects like drones and birds within images. Among various deep learning models evaluated, VGG16 had the best performance in terms of accuracy, F1-score and classification results.As a result, it has been selected for final implementation. The system architecture comprises modules like: YOLOv7 Model, Image Cropping Module, VGG16 Classification Model.Figure 4 shows System Architecture for drone v/s bird detection and classification.

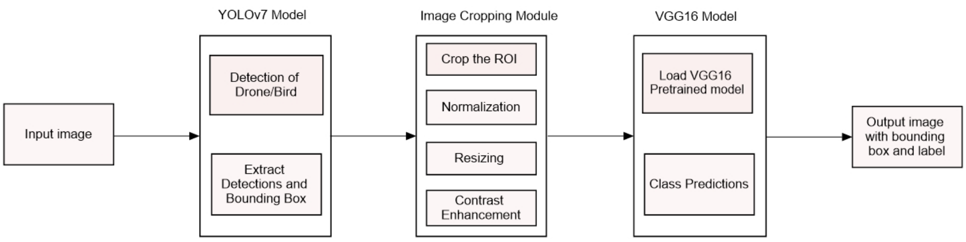

Figure 4: System Architecture

Image where drone or bird needs to be detected is provided to input of YOLOv7 along with the confidence threshold which is taken as 0.50.YOLOv7 model is utilized for object detection. YOLOv7 architecture comprises several key components, including a backbone, neck, and head.Backbone, typically a convolutional neural network (CNN), extracts essential features from input image. Neck, often using a Feature Pyramid Network (FPN) or Path Aggregation Network (PAN), merges features at different scales to enhance model's ability to detect objects of various sizes. The head, which consists of multiple detection layers, predicts bounding boxes, object classes, and confidence scores simultaneously for multiple objects within image. During inference, YOLOv7 processes input image in a single forward pass, generating predictions for bounding boxes, class probabilities, and confidence scores. Non-maximum suppression (NMS) is then applied to filter out redundant boxes, ensuring precise and efficient object detection.Model outputs bounding boxes, class labels, and confidence scores for each detected object. These bounding boxes are crucial for locating regions of interest (ROIs) within image, enabling precise cropping of detected objects.

Following detection, image cropping module uses bounding box coordinates provided by YOLOv7 to crop each detected object from the original image. This step ensures that only relevant parts of the image are prepared for further processing.VGG16 consists of 16 layers, primarily using small 3x3 convolutional filters stacked in depth, which allows it to capture intricate features in images. The architecture ends with fully connected layers for classification, making it powerful for image recognition tasks.Normalization of the image using the ImageNet mean and standard deviation values to match the training conditions of the VGG16 model is done along with resizing to 224*224.Image contrast of the cropped image part is also enhanced using Contrast Limited Adaptive Histogram Equalization(CLAHE). CLAHE is applied to enhance contrast without affecting colors, making features more pronounced. This preprocessing step improves the visual quality of the image, helping the VGG16 model to accurately classify objects. Each cropped image is then ready to be input into the VGG16 classification model.

The VGG16 model is responsible for classifying the cropped object images into pre-defined categories, like bird/drone. The pre-trained VGG16 model is loaded and used to classify the preprocessed images.

The classified images along with the predicted class labels are displayed in the provided input image. Visualization tools like Matplotlib are used to display the classified images.

The system first loads the YOLOv7 and VGG16 models. YOLOv7 processes the input image to detect objects and create bounding boxes. These are then used to crop the objects, which are further classified by VGG16. The classified images and their predicted labels are displayed using Matplotlib. This setup provides an efficient and accurate framework for object detection and classification using YOLOv7 and VGG16.

5 Implementation

5.1 Environment and Framework

The model was trained using Google Colab Pro, which provided enhanced computational resources with upto 32 GB of RAM and 100 GB of disk space. Colab Pro was chosen to utilize a high-performance GPU for efficient training and to avoid frequent disconnections that occur with the standard Google Colab. Python was used as the programming language. The primary framework for model development was PyTorch, which allowed for flexibility and ease of use.

5.2 Data Preparation

Data preparation involved many steps as explained further.Imagehash library is used to identify duplicates where it produces a unique fixed-size string called hash, based on the visual content. If a matching hash is found for any image, the image is flagged as a duplicate. Hence identified duplicate images like as shown in figure 5 are removed from the dataset.

Figure 5: Duplicate images

To annotate the data using roboflow, the class names should be provided along with color of bounding box for each class as shown in figure 6. The annotation of images are done using automatic labelling where it identifies drone or birds in a image and draws the bounding box.Once all the images are auto-labeled, we need to manually verify the predicted class for each image. This process is time-consuming but essential to prevent the model from learning incorrectly from mislabeled images, which can reduce performance. Although Roboflow offers options for preprocessing and augmenting data, it was used only for labeling in this case.

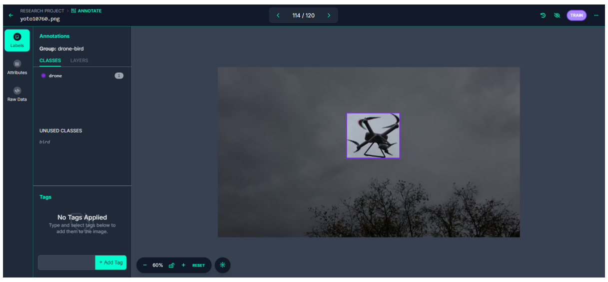

Figure 6: Labelling using roboflow

The corresponding annotaton file for image shown in figure 6 looks like as in figure 7
.

```
1 0.514453125 0.40208333333333335 0.17578125 0.2652777777777778
```

Figure 7: Annotation file

For the models ResNet18, VGG16 and InceptionV3, the images are normalized accordingly to match ImageNet statistics since they are pre-trained on the ImageNet dataset. The normalization parameters applied are mean values of [0.485, 0.456, 0.406] and standard deviation values of [0.229, 0.224, 0.225].Data Augmentation was also applied only on the training data since validation and test sets should remain unchanged to provide a consistent and unbiased evaluation of the model's performance,The applied variations are RandomResizedCrop which resizes images to 224x224 pixels as expected by ResNet18 and VGG16.For InceptionV3 images are resized to 299x299, RandomHorizontalFlip to flips the images horizontally to introduce variability, RandomRotation which rotates images randomly within a range of -10 to +10 degrees, and ColorJitter which changes the brightness, contrast, saturation, and hue of the images to simulate different lighting conditions.The example of augmented images can be shown be figure 8.

Figure 8: Augmented images

5.3 Modeling

The YOLOv7 repository was downloaded and establishment of requirements necessary to execute YOLOv7 from the official repository on GitHub were made. The dataset was uploaded to the data folder within YOLOv7, and the number of classes and class names were defined in the customyolov7.yaml file. Pre-downloaded YOLOv7 weights from Google Drive were used for training the model on various configurations to achieve optimal performance. Post-training, the model was tested on a test dataset using test.py and evaluated on new images with detect.py. Results, including bounding boxes and class predictions with confidence levels, were saved in the run folder. A detection threshold of 0.50 was set for identifying drones or birds in images as used by Naveen et al. (2023) which retains only the most relevant box for each detected object.

For data visualization in the implementation, the libraries Matplotlib and Seaborn were utilized. Matplotlib and Seaborn are Python libraries for data visualization. Matplotlib offers versatile, customizable plots, while Seaborn builds on it, simplifying the creation of aesthetically pleasing statistical graphics. Pretrained models from torchvision such as resnet18, VGG16_Weights, and Inception_V3_Weights were utilized to train the classification models.The final fully connected layer in the respective models have been changed to 2 output features for classifying drones and birds. The models are saved to drive to further use it.To evaluate the models built various evaluation metrics from Scikit-learn was employed to generate metrics such as confusion matrix, classification report, ROC curve, AUC, precision-recall curve, and average precision score.The better performing model was found to be VGG16 among classification algorithms.

The implementation of the pipeline involves execution of detect.py from YOLOv7 to identify objects in images, generating bounding boxes and class predictions with confidence levels.The region of interest(ROI) is obtained using the bounding box details from YOLOv7.Applied Transformations are applied including resizing, normalization, and contrast enhancement to the cropped images. Pretrained VGG16 model is used to further predict the class the object belongs i.e, drone or bird. Bounding boxes and class labels on the original image based on predictions is obtained as the output.

The results and findings of the models built will be discussed in further sections.

6 Evaluation

This section outlines the experiments conducted in the study, which aimed to enhance drone versus bird detection by leveraging the strengths of YOLOv7 combined with various classification models. The study compared and analyzed the performance of YOLOv7 alongside different classification algorithms, like VGG16, ResNet18, and InceptionV3, to evaluate how effectively integrating these classification algorithms with YOLOv7 improves the accuracy and reliability of distinguishing between drones and birds.

6.1 Experiment 1

As the first experiment, the YOLOv7 model was implemented with various configurations including changing of learning rate, optimizer and epochs. With the learning rate of 0.01, SGD optimizer and for 25 epochs, the model had mAP@0.5 value as 0.953.The model was tested on test dataset which yielded a accuracy of 0.913.The F1-Score was 0.8867.Figure 9 represents the YOLOv7 results following model training on the dataset.

The X-axis represents the number of epochs, while the Y-axis corresponds to the metrics indicated at the top of each graph. Key graphs to focus on include Precision, Recall, mAP@0.5, mAP@0.5:0.95, and val_classification. The mAP@0.5 metric indicates the mean average precision at an IoU threshold of 0.5, while mAP@0.5:0.95 reflects the mean average precision across varying IoU thresholds, ranging from 0.5 to 0.95. The IoU threshold mentioned is the value used to evaluate the overlap between predicted and ground truth bounding boxes.

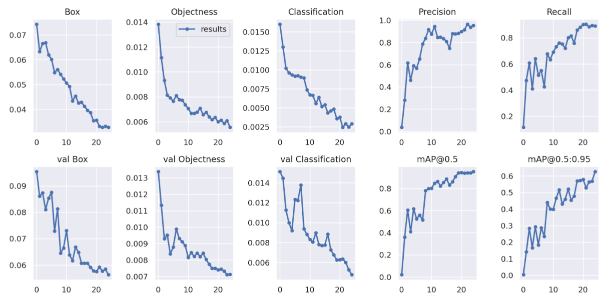

Figure 9: YOLOv7 Training Results

The confusion matrix representing the true labels on the x-axis with the predicted labels on the y-axis is shown in figure 10.

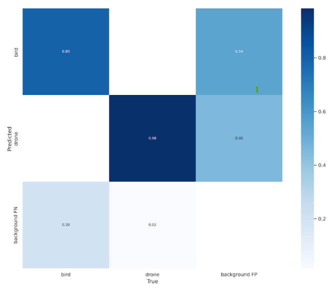

Figure 10: Confusion Matrix - YOLOv7

Figure 11 shows detection of drone and bird in a single image by the YOLOv7 model.It also shows the confidence level of the object detected which is 80% or higher in this image which shows the model is performing good on the unseen image.

Figure 11: Detection by YOLOv7

6.2 Experiment 2

In the second experiment, the performance of three classification algorithms was evaluated. ResNet18, VGG16, InceptionV3 are the three models considered. As already explained in the modeling, hyperparameter tuning was used to select the best fit parameters. Resnet18 had a training accuracy of 93.31% and test accuracy of 93.02%. VGG16 had training accuracy of 97.24% and test accuracy of 98.45%. The testing accuracy usually expected to be lower than the training accuracy, but here it could be due to the regularization effects of dropout and batch normalization introduced during training of VGG16 to help the model generalize well.InceptionV3 had training accuracy of 96.85% and testing accuracy of 86.05% which is quite low compared to training accuracy. Table 1 shows the accuracy, precision, recall, and f1 score for all three selected models after hyperparameter tuning. It can be observed from the table that VGG16 gives the best results.VGG16 has achieved improved F1 score and has high precision and recall rates along with accuracy.Hence for integration with YOLOv7, VGG16 was chosen.

Table 1: Performance metrics for different classifiers

Classifier	**Class**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
ResNet18	Bird	93.02%	91%	95%	93%
	Drone		95%	91%	
VGG16	Bird	98.45%	97%	95%	96%
	Drone		95%	97%	
InceptionV3	Bird	86.05%	85%	88%	86%
	Drone		87%	85%	

The confusion matrix presented in figure 12 offers evaluation of the VGG16 model's classification performance on drone and birds.

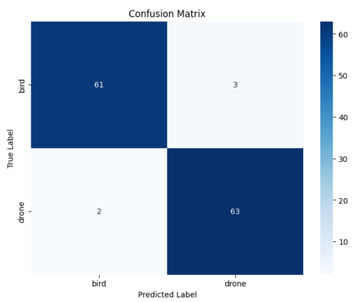

Figure 12: Confusion Matrix - VGG16

The learning curves for VGG16 model is shown in the figure 13 which helps in understanding how training and validation losses or the accuracies is varying with the epochs.

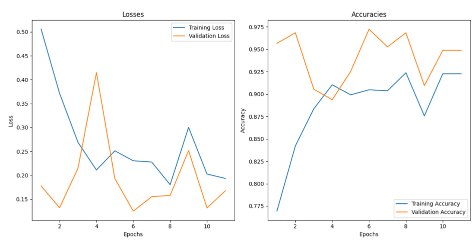

Figure 13: Learning Curves for VGG16

The Receiver Operating Characteristic Curve(ROC) is shown in figure 14. The area under ROC curve value is 0.96 indicating the model is performing well with high true positive rate and a low false positive rate.

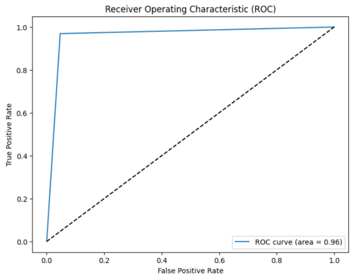

Figure 14: ROC Curve

6.3 Experiment 3

This experiment involved evaluation of pipeline of YOLOv7 with VGG16. The accuracy obtained is 90.38% when evaluated on the test dataset.The classification report including precsion,recall and F1-score for each class is shown in the figure 15.

Classification Report:

	precision	recall	f1-score	support
bird	0.86	1.00	0.93	93
drone	1.00	0.76	0.86	63
accuracy			0.90	156
macro avg	0.93	0.88	0.90	156
weighted avg	0.92	0.90	0.90	156

Figure 15: Classification Report - YOLOv7 with VGG16

The confusion matrix and ROC curve is shown in figures 16 and 17 respectively.

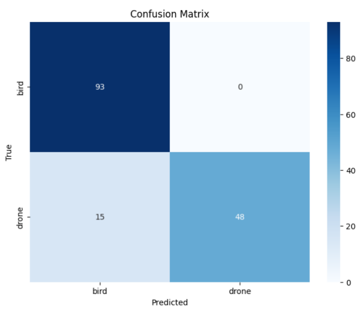

Figure 16: Confusion Matrix

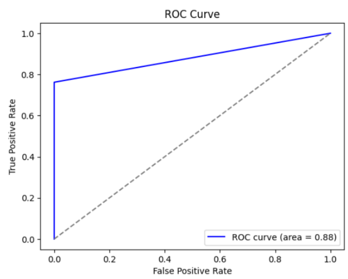

Figure 17: ROC Curve

6.4 Experiment 4

As next steps, the evaluation of the developed pipeline is performed on some new images other than training or test to check the pipeline performance on predicting drone and bird.The figure 18, figure 19, figure 20, figure 21 shows some of the scenarios where the model performed well.

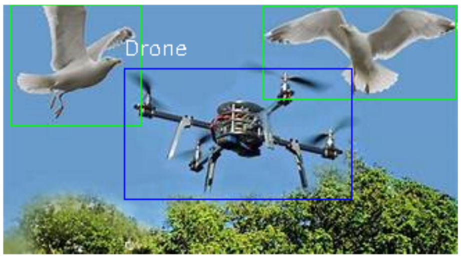

Figure 18: Drone-multiple birds detection

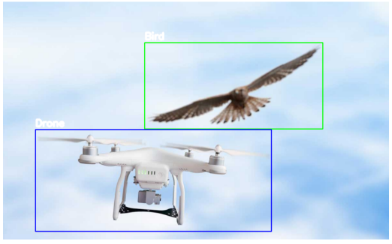

Figure 19: Drone-bird detection

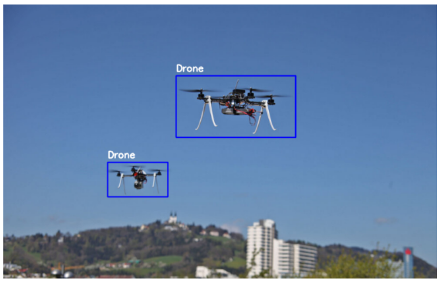

Figure 20: Mutiple drone detection

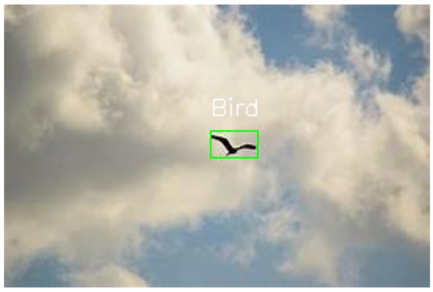

Figure 21: Single bird detection

The figure 22 and figure 23 shows where the pipeline predicted wrongly compared to original prediction by YOLOv7.

Figure 22: Prediction - YOLOv7

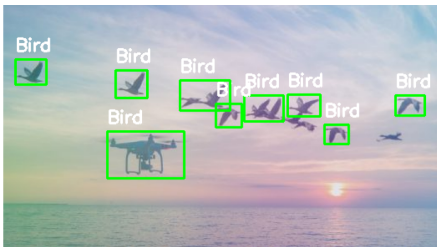

Figure 23: Prediction - YOLOv7 with VGG16

6.5 Discussion

Initially, efforts were made to integrate YOLOv7 with classification algorithms by using ResNet18 as a feature extractor to enhance the YOLOv7 model. This involved attempting to replace the backbone of YOLOv7 with ResNet18, excluding its classification layer.However, due to time constraints and architectural complexities along with a lack of available references this approach proved challenging. It required a deeper understanding and more extensive experimentation with architectural modifications. Later, the focus shifted to leveraging the same concept of combining classification algorithms with YOLOv7 by extracting Region of Interest (ROI), followed by classification with models like ResNet18.

To pursue the second approach, the YOLOv7 model was implemented as the baseline. From figure 9, it is evident that there is increasing trend in precision, recall and mAP values. The increasing trend in precision indicates that the model is getting better at making accurate predictions with fewer false positives.The increasing recall shows that the model is improving its ability to detect objects that are present. The increasing mAP@0.5 indicates that the model is achieving higher precision and recall overall as training progresses.A decreasing validation classification loss suggests that the model's ability to classify objects correctly on unseen data is improving. It is also crucial to observe that both training and validation metrics are improving, indicating that the model is not overfitting and is likely to generalize well to unseen data.

Figure 10 shows model performs quite well in identifying drones with 98% accuracy and birds with 80% accuracy. And it is evident that misclassification between drone and bird is significantly reduced. However, the confusion between objects and the background is a significant issue, as evidenced by the matrix,it is identifying background as drone and bird.This could be due to similarities in the appearance between drone/bird and certain elements of the background in the dataset, leading the model to struggle in distinguishing them.

The YOLOv7 model has shown better performance than it's lower version, YOLOv3, as indicated by a higher mean average precision (mAP). In particular, the YOLOv7 model achieved an impressive mAP of 95.3%, significantly surpassing the YOLOv3 model, which recorded an mAP of 88.9%, according to previous research Wei Xun et al. (2021).

Even though our model has high mAP value and good F1- score,it has not reached the high levels of mAP and F1-scores observed in the research using YOLOv4 and YOLOv5. YOLOv5 model achieved a high mean Average Precision (mAP) of 98.4% and F1-score of 95% Jarray and Bouallègue (2023) .YOLOv4 model demonstrated superior performance with a mAP of 97.4% and an F1-score of 98% Sethu Selvi et al. (2022). These differences could be attributed to variations in the datasets, model performance on specific dataset and other hyperparameters settings.

VGG16 was found to be having good accuracy of 98.45% and F1-Score of 96%. Figure 12 suggests that VGG16 model is highly effective at distinguishing between bird and drone with only a few classifications. The high accuracy, combined with strong recall and precision for both classes, indicates that the model is suited for accurate identification of these categories.

In figure 13 the overall trends suggest that the model is learning effectively, as evidenced by the decreasing training loss and increasing training accuracy. The fluctuations in validation loss and accuracy, especially the spike in validation loss around the 4th and 6th epochs, suggests potential overfitting. This means that the model fits the training data well but struggles with validation data, possibly due to differences in data distribution. The eventual stabilization of both validation loss and accuracy indicates that the model has found a balance and is performing well on both the training and validation sets.

ResNet18 was identified as the top performer among the three models with accuracy and F-Score exceeding 98% in research Mohamed and Alharbi (2023).In our research we found VGG16 as the best performing in terms of accuracy and F1-score.

VGG16 model with it's better performance answers our first research question of what is the best algorithm to classify drone and birds.

As the next steps on the research the YOLOv7 model was pipelined with VGG16 along with the image cropping module explained before. After testing the model on test dataset. The classification report in the figure 15 suggests the model has strong precision of 0.86 and recall of 1 for the bird class, resulting in an F1-score of 0.93. However, while the drone class exhibits perfect precision of 1.00, its recall is lower at 0.76, leading to a slightly lower F1-score of 0.86. Figure 16 presents the confusion matrix where it shows all the birds are being predicted properly without confusing them with drones, as there are no false positives. but some percentage of drones are predicted as bird by the pipeline. The figure 18 shows the detection of drone and multiple birds by the pipeline model.Figure 19 shows the detection of single drone and bird.Figure 20 shows the detection of multiple drones and figure 21 shows detection of single bird by the pipeline of YOLOv7 with VGG16.But figure 23 shows the pipeline models predicted drone as a bird when there are multiple birds and a drone which was predicted correctly by YOLOV7 in shown in figure 22. The test accuracy of the pipeline model is 90.38% which is litte less compared to the YOLOv7 test accuracy of 91.3%.In figure 22 we can observe that some birds are not detected by YOLOv7 itself, in this cases the pipeline also misses as already mentioned earlier that the proposed method is reliant on YOLOv7 which is the limitation.

Based on the discussions, it appears that the proposed pipeline method did not perform compared to the baseline model, YOLOv7. This conclusion effectively addresses the second research question.

But if we consider individual performance of the detection and classification models.The detection module has achieved mean average precision of 91.3% and VGG16 achieved classification performance of 98.45% which is higher than the accuracies achieved by Lee et al. (2018). As they have suggested the complex models could yield good results than the basic detection and CNN models. In their research they have considered multi class for drones but we have taken only drone and bird which could be extended further.

7 Conclusion and Future Work

This research aimed to enhance the accuracy of distinguishing between drones and birds by integrating advanced object detection and classification algorithms. The primary research question focused on comparing algorithms such as VGG16, ResNet18, and InceptionV3 for drone and bird classification to identify the better-performing model, and the other question focused on the evaluation of how integrating these algorithms with YOLOv7 could enhance detection accuracy. The objectives were to implement YOLOv7 for object detection, assess the performance of the classification models, and develop a pipeline that combines detection and classification to improve overall accuracy.

VGG16 was found to be the best-performing classification algorithm, and it was demonstrated that YOLOv7 is highly effective in detecting drones and birds. However, when integrated into a combined pipeline, the expected improvements in detection accuracy were not consistently achieved. Also acquiring new images that featured both drones and birds in same frame proved challenging, which limited testing to only a few scenarios.The integrated pipeline was able to perform well in identifying single bird, multiple drones and birds and drone together in various backgrounds. It struggled to identify drone when there are multiple birds. Our proposed system was successfully good at reducing false alarm by not identifying bird as a drone.Because false alarms could lead to unnecessary actions, like triggering security protocols, scrambling defense systems, or diverting attention from actual threats when implemented on real-time.At same time some drones were identified as bird leading to false negative leading to missed detections of potentially critical threats.It's essential for the system to minimize this error.

Future work could involve exploring alternative frameworks like EfficientDet or Detection Transformer(DETR) for detection. Additionally, developing more sophisticated integration strategies, possibly through ensemble methods or custom architectures, may lead to better synergy between detection and classification stages.

The research could be expanded to include multi-class classification and testing the models on larger and more diverse data would also help assess their robustness and generalizability in real-world applications and it could be extended to videos rather than images to make it more suitable for real- time. Moreover, addressing limitation of reliance on initial detection by using feedback loops for iterative refinement, is another promising area for future work. By pursuing these more accurate, reliable, and robust systems for drone detection and classification could be developed, ultimately contributing to enhanced surveillance and security in diverse environments.

References

Al-Emadi, S. and Al-Senaid, F. (2020). Drone detection approach based on radio-frequency using convolutional neural network, *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, pp. 29–34.

Al-Zahrani, A. Y. (2023). Unauthorized drones: Classifications and detection techniques, *2023 IEEE 11th International Conference on Systems and Control (ICSC)*, pp. 548–555.

Björklund, S. and Wadströmer, N. (2019). Target detection and classification of small drones by deep learning on radar micro-doppler, *2019 International Radar Conference (RADAR)*, pp. 1–6.

Dale, H., Baker, C., Antoniou, M., Jahangir, M. and Atkinson, G. (2021). A comparison of convolutional neural networks for low snr radar classification of drones, *2021 IEEE Radar Conference (RadarConf21)*, pp. 1–5.

Delleji, T., Fekih, H. and Chtourou, Z. (2020). Deep learning-based approach for detection and classification of micro/mini drones, *2020 4th International Conference on Advanced Systems and Emergent Technologies* $(IC_ASET), pp.$ $332 - -337$.

Hayat Suhendar, M. T. and Widyani, Y. (2023). Machine learning application development guidelines using crisp-dm and scrum concept, *2023 IEEE International Conference on Data and Software Engineering (ICoDSE)*, pp. 168–173.

Jarray, R. and Bouallègue, S. (2023). Vision-based drone recognition using a yolov5-like deep learning network, *2023 IEEE International Conference on Artificial Intelligence Green Energy (ICAIGE)*, pp. 1–6.

Karthikeya Nalam, V. S., Amar Koushik Tanniru, V. S., Posani, A. and Suneetha, M. (2022). Detection and recognition of drones using deep convolution neural networks, *2022 IEEE 6th Conference on Information and Communication Technology (CICT)*, pp. 1–5.

Kataria, M. and Lall, B. (2023). Tracking aided drone bird classification using yolo and lstm, *2023 IEEE International Conference on Image Processing Challenges and Workshops (ICIPCW)*, pp. 1–5.

Kurosaki, M., Kawaguchi, D., Ogawa, K., Nakamura, R. and Hadama, H. (2022). Comparison of drone classification accuracy for cnn models using uwb radar, *2022 Asia-Pacific Microwave Conference (APMC)*, pp. 907–909.

Lee, D., Gyu La, W. and Kim, H. (2018). Drone detection and identification system using artificial intelligence, *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 1131–1133.

Mahdavi, F. and Rajabi, R. (2020). Drone detection using convolutional neural networks, *2020 6th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)*, pp. 1–5.

Mohamed, T. M. and Alharbi, I. M. (2023). Efficient drones-birds classification using transfer learning, *2023 4th International Conference on Artificial Intelligence, Robotics and Control (AIRC)*, pp. 15–19.

Naveen, Menon, H. P., Vinitha, V., Vishnuraj, K., Satheesh, A. and Nikhil, A. (2023). A study on yolov5 for drone detection with google colab training, *2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, pp. 1576–1580.

Oh, H. M., Lee, H. and Kim, M. Y. (2019). Comparing convolutional neural network(cnn) models for machine learning-based drone and bird classification of anti-drone system, *2019 19th International Conference on Control, Automation and Systems (ICCAS)*, pp. 87–90.

Pansare, A., Sabu, N., Kushwaha, H., Srivastava, V., Thakur, N., Jamgaonkar, K. and Faiz, M. Z. (2022). Drone detection using yolo and ssd a comparative study, *2022 International Conference on Signal and Information Processing (IConSIP)*, pp. 1–6.

Sethu Selvi, S., Pavithraa, S., Dharini, R. and Chaitra, E. (2022). A deep learning approach to classify drones and birds, *2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon)*, pp. 1–5.

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

Valaboju, R., Vaishnavi, Harshitha, C., Kallam, A. R. and Babu, B. S. (2023). Drone detection and classification using computer vision, *2023 7th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1320–1328.

Wei Xun, D. T., Lim, Y. L. and Srigrarom, S. (2021). Drone detection using yolov3 with transfer learning on nvidia jetson tx2, *2021 Second International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP)*, pp. 1–6.