National College of Ireland

Micro-UAV Detection using Mask R-CNN

MSc Research Project
Data Analytics

Shubham Rajabhau Maske
Student ID: x19232551

School of Computing
National College of Ireland

Supervisor: Dr. Hicham Rifai

National College of Ireland
Project Submission Sheet
School of Computing

National College of Ireland

Student Name:	Shubham Rajabhau Maske
Student ID:	x19232551
Programme:	Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Dr. Hicham Rifai
Submission Due Date:	16/08/2021
Project Title:	Micro-UAV Detection using Mask R-CNN
Word Count:	5919
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:	Shubham Rajabhau Maske
Date:	15th August 2021

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):	

Micro-UAV Detection using Mask R-CNN

Shubham Rajabhau Maske
x19232551

Abstract

With the advancement in production of micro-UAV, they have become cheap and easy to operate. While the widespread use of micro-UAVs has provided many benefits to all sectors of society, it has also presented a significant danger to personal, public, and military security. Micro-UAVs are difficult for conventional air-defence systems to identify as they are small in size and have low flying altitude. The proposed research aims in identifying micro-UAVs by implementing deep learning technique. The method presented is a deep learning algorithm called Mask R-CNN, which is a notion for object identification and will be utilized in micro-UAV detection. Publicly available dataset named, Det-Fly, is used in this research. The model is evaluated using mean Average Precision (mAP) as well as validation loss, bounding box loss and classification loss are graphically plotted. A very good results are obtained from the implemented model with mAP value of 72.10%.

Keywords- Mask R-CNN, Convolutional Neural Network, Micro-UAV, Segmentation, Deep Learning

1 Introduction

Given the rapid advancement of autonomous vehicles and the technology required to build them, the number of micro-UAV developed for defence, commercial, or entertainment reasons has increased significantly over time. Micro-UAVs are also extensively employed in precision agriculture, where they provide rapid mapping and scanning of farms for crop variety and phenology, crop dusting or spraying for weed and pest control, irrigation system monitoring, and animal observation. Micro-UAVs may also be used for facility healthcare services, package delivery, media and sports, and Internet networking via ad-hoc base stations. In addition to the beneficial applications of micro-UAV, a rising number of illicit use of micro-UAV is being identified on a worldwide basis. Micro-UAVs have previously been accused of violating public privacy and endangering the security of vital infrastructure such as nuclear power plants and airports. Recent examples include drone strikes on a US military facility near Iraq's Erbil airport [1]. In similar case, six drones were also aimed at the Riyadh Oil Refinery in Riyadh, Saudi Arabia's capital [2]. Likewise, the use of unregistered drones in Singapore and Dubai's airport traffic zones caused flight delays and disturbances [3] [4]. As a consequence, misusing micro-UAVs poses

[1] https://www.reuters.com/article/iraq-security-int/explosives-laden-drone-targets-u-s-forces-at-iraqs-erbil-airport-idUSKBN2C210G

[2] https://www.reuters.com/article/saudi-security-yemen-int-idUSKBN2BB19Q

[3] https://www.reuters.com/article/us-singapore-airport-drones-idUSKCN1TK1P2

[4] https://www.reuters.com/article/us-emirates-airport-drones-idUSKCN1Q40P5

serious security, privacy, and safety concerns. Despite the fact that numerous national and international rules regulate the use of micro-UAVs, they have yet to be created using different technologies that might help with compliance. Hence, a critical need exists for real-time detection and countermeasures against drones.

Researchers are increasingly focused on the visual identification of micro UAVs due to their crucial role in many key applications. Visual identification of hostile micro unmanned aerial vehicles (UAVs) is a critical technique for building civilian UAV defensive systems. The detection of micro-UAVs can be divided into two categories: 1) ground-to-air detection, which involves placing cameras on the ground to identify flying UAVs. 2) air-to-air, in which a flying UAV detects other flying UAVs using its onboard cameras. Micro-UAV detection, often known as anti-drone system, is a technology-based method for determining the presence or position of drones. Numerous detection techniques, including acoustic, radar, radio frequency, and vision, have been suggested for detecting micro-UAVs, as described in the section 2. Deep learning has been a scientific spotlight in recent years because to its phenomenal performance in object identification and computer vision. Deep learning-based approaches to identifying UAVs, as well as vision-based approaches, are gaining momentum in recent years. Among these, methods based on convolutional neural networks (CNNs) have improved object detection significantly. Region based CNN (R-CNN) employs a two-stage detection approach, beginning with selective search to generate a large number of region suggestions, followed by CNN feature extraction from each proposal, and finally, classification and modification of each area's bounding box.

Mask R-CNN is a part of R-CNN family. It is an extension of Faster R-CNN with an additional mask object. Researchers in (Dongye and Liu; 2020) have made use of Mask R-CNN to detect the damaged pavement. Similarly in (Malbog; 2019), authors used Mask R-CNN to detect the pedestrian cross-walk. In (Wang et al.; 2020), researchers proposed a model based on Mask R-CNN that will be able to detect disease spot on different fruits. The usage of a UAV for power-line inspection utilizing a Deep Learning system for detection is explored in this article (Vemula and Frye; 2020). Here, the authors used Mask R-CNN and deployed it on the UAV to identify the power lines.

The dataset used in this research have micro-UAV images that were captured from camera mounted on other micro-UAV. These images were taken in various background scenarios such as clear sky, night, field, cities and much more. Due to its small size, the algorithms may sometimes mistook the bird for an UAV which makes its challenging task. As seen above, Mask R-CNN have provided good performance in object detection. Its capability to detect and mask the image is the reason this method was selected in this research. More details on the Mask R-CNN can be found in section 3 and section 4.

1.1 Research Question and Research Objective

To what extent Mask R-CNN can identify micro-UAV in given images?

The objective here is to evaluate how accurate the Mask R-CNN can identify micro-UAV in the images. There are two scenarios in detection of micro-UAV: first is ground-to-air where the camera placed on the ground station detects the airborne UAV; second is air-to-air where an airborne UAV uses its mounted camera to detect other micro-UAV. Images captured in air-to air scenario makes UAV detection complex due to factors such as viewing angles, background of target UAV, change in shape, scale and color as the camera is flying dynamically. This research uses the images of UAV captured in air-to-

air scenario. Mask R-CNN has shown very good performance in many object detection tasks as shown earlier. The images used to train the model have been taken in various background scenarios which makes it more complex to identify the UAV. A significant part of the target UAV pictures are small in the dataset with regard to the total image size. More than half of the images have picture dimensions that are less than 5% of the total image size (Zheng et al.; 2021). Despite these complexities, the motivation here is to utilize the Mask R-CNN in detecting UAVs.

Rest of the research is divided as: Section 2 describes all the related work performed in this field. Section 3 explains the research method. Section 4 will show the implemented design for this research. The implementation of the methodology is explained in section 5. The evaluation performed is shown in section 6. This research is concluded in section 7 and suggests future work.

2 Related Work

With the proliferation of commercial micro-UAVs, the number of detection mechanisms has increased proportionately. Among these, acoustic characteristics, micro-doppler effect, radio frequency, and vision-based detection techniques are addressed below.

2.1 Detection using Acoustic features

Micro-UAV detection based on acoustic characteristics is identifying, monitoring, or differentiating the drone based on its propeller noise. The article (Yang et al.; 2019) presents a machine learning-based UAV detection framework with multiple auditory nodes, as well as an empirically tuned configuration for implementation. A control unit and numerous listening nodes make up this system. The Short-Time Fourier Transform (STFT) and Mel-frequency Cepstral Coefficients (MFCC) are two characteristics that the authors have utilized. For MFCC and STFT, CNN and SVM were used. The STFT-SVM configuration yielded the best results.

Authors in (Kim et al.; 2020) proposed a technique for extracting and identifying acoustic characteristics of CNN-based unmanned aerial vehicles. STFT is utilized to convert one-dimensional data to two-dimensional features in this study. CNN received the raw data after it was pre-processed using 30,711 drum samples, 38,907 tiny fan samples, and 45,493 UAV samples. In this research, a high detection rate was obtained with a low false detection rate.

Even if identification by acoustic characteristics is regarded a feasible option, disturbances in the air may interfere with acoustic tracking, effectively drowning out the relatively quiet sound produced by the micro-UAV's blades or propeller.

2.2 Detection based on Micro-Doppler effect

The Doppler effect is defined as the audible change in frequency between sound waves reflected from a moving target and a stationary source. The Micro-Doppler effect is based on doppler modulation caused by the target's interior motion, such as a micro-UAV. Radar sensors can work in any environment and at any time, and they offer the added benefit of long-range detection (Li et al.; 2020). The author took 100 samples from the signal components, with ambient clutter accounting for 40% of the sample and drone

accounting for 60%. The authors were able to obtain an accuracy of 85% by utilizing 70% of the data as training and the remainder as test.

The micro-doppler effect has been interpreted by the authors (Gérard et al.; 2021) in five ways: The micro-doppler effect has been interpreted in the literature in five ways: weighted spectrum of the signal (WSP), cepstrum (CP), time signal after range compression x(t), cadence velocity diagram (CVD), and spectrogram (SG).

The disadvantage is that radar signals will be obstructed by walls, buildings, and other obstacles, which are all too common in civilian settings.

2.3 Detection using Radio Frequency

Studies showed that the micro-UAV devices have a unique radio frequency (RF) signature as a consequence of its circuitry design and modulation technique (Ezuma et al.; 2019). The authors illustrated how to identify and classify micro-UAVs using the RF footprints of the signals sent from the operator to the micro-UAV. The detection stage divides raw signals into frames and converts them to the wavelet domain in order to remove signal noise and reduce the quantity of data to examine. The Naive Bayes machine learning technique is used to look for the presence of a micro-UAV in each frame. The authors also suggested a technique in which the time signal frequency is first converted to energy-time frequency, and then the energy trajectory is calculated. With a signal-to-noise ratio (SNR) of 15dB and the KNN classifier, accuracy of more than 80% was obtained.

Another research (Allahham et al.; 2020), proposes a novel method for drone network detection, categorization, feature selection, and recognition based on the DroneRF dataset. This dataset contains the radio frequency (RF) signatures of a variety of drones operating in a variety of flight modes, including on, off, orbiting, traveling, and video shooting, as well as existing RF activities conducted in the absence of drones. The authors here tested three multi-channel 1DCNN models in order to perform three classifications: recognizing the existence of a drone; identifying the drone and determining its kind; and identifying the drone, determining its type, and determining its flying mode. They evaluated the proposed multi-channel 1DCNN model using accuracy, recall, precision, error, and F1 measures, and obtained excellent results.

In the paper (Akter et al.; 2020), authors have developed a Convolutional Neural Network model on the same dataset, as explained above, i.e. the DroneRF dataset. For RF-based UAV classification, they proposed a cost-effective sequential CNN approach. For assessing RF signal correlations and extracting higher level features, this suggested architecture employs nine 1D convolution layers with hyper-parameters. Authors (Al-Emadi and Al-Senaid; 2020), on the other hand, utilized five 1D convolution layers in their CNN design, with an average 1D pooling layer added after each layer. Same dataset, DroneRF, was utilised.

Researchers (Alipour-Fanid et al.; 2020) has noted that the majority of current market UAVs feature WiFi interfaces and may be controlled using portable devices. As a result, the authors have developed a system based on machine learning techniques for detecting and recognizing delay aware drones and their operating modes over encrypted WiFi data. In the architecture presented here, encrypted traffic is treated as a time series and statistical characteristics based on packet size and inter-arrival time are retrieved. With an accuracy of 85.7% percent to 95.2 percent, the suggested approach identified and tagged UAV in 15 to 35 seconds.

2.4 Vision based detection

Vision-based UAV detection refers to the identification of micro-UAVs utilizing pictures captured by a camera placed on some other micro-UAV or on the ground station. Deep learning models have gained popularity in the segmentation and classification of images.

2.4.1 Micro-UAV detection based on YOLO models

In real-time object identification, the Object Detection Algorithm YOLO (You Only Look Once) is often used. YOLO is a quick and efficient method that applies a single neural network to the whole picture, splits it into regions, and calculates the bounding boxes and probabilities for each (Hassan et al.; 2019).

The YOLO model and two neural networks, ResNet and DenseNet, are used to design and implement a method for identifying low-altitude micro-UAVs in the paper (Yuan et al.; 2020). After implementing the model, the detection time for micro-UAV was 0.02 seconds. In general, regression-based object detection predicts groupings of items rather than specific areas of individual objects in a single run of the algorithm. Researchers (Seidaliyeva et al.; 2020) utilized YOLOv2 to identify empty and loaded micro-UAV ie. without and with cargo UAVs respectively. The object detection model presented in this article has 23 convolutional layers. Two methods were utilized to evaluate detection accuracy and speed: mAP (mean Average Precision) and FPS (Frames per Second).

YOLOv3 differs from takes an incremental approach which distinguishes itself from YOLOv2 . The YOLOv2 and YOLOv3 algorithms are utilized and executed by the authors in (Hassan et al.; 2019).The accuracy achieved by YOLOv3 and YOLOv2 are 95.2% and 92.10% whereas the mAP achieved was 0.2538 and 0.2053 respectively. Researches in (Hu et al.; 2019) states that as micro-UAV are small in size, YOLOv3 cannot be used to counter them directly. In this study, an improved YOLOv3 is presented in order to enable effective real-time detection of UAVs. Instead of using three-scale feature maps, this improved method uses four-scale feature maps, which allows it to detect delicate characteristics like textures and contours in low-light situations. The results of this paper's assessment verified that the suggested improved YOLOv3 exceeded the current YOLOv3.

Researchers in the article (Shi and Li; 2020) proposed a recognition method for low-altitude micro-UAV identification based on the YOLOv4 model and demonstrated its usage to detect low-altitude UAV objects for the first time. The study compares the training results of YOLOv4, traditional YOLOv3, and SSD networks. Based on the evaluation performed in this paper, concludes that the YOLOv4 provided better results, in terms of accuracy and recall, than the other compared models.

2.4.2 Detection using CNN models

The invention of the Convolutional Neural Network (CNN) was a turning point in the area of image-based data processing. The purpose of study performed in (Nalamati et al.; 2019) is to identify a UAV in a brief video clip including birds. From these video clips, a total of 8771 frames in JPG format were extracted. Following the training of the models, it was discovered that the Faster R-CNN using Resnet-101 as the basic architecture outscored the SSD model in accurately identifying micro-UAVs. The micro-UAV was detected in this study, (Mahdavi and Rajabi; 2020), using three classification techniques: CNN, SVM, and nearest neighbour. Researchers in this paper have suggest a new CNN

model which utilises 3 max-pooling layers, 2 fully connected layers and 3 convolutional layers. According to the results in this research, the CNN topped the KNN and SVM in detecting the micro-UAV.

As per the researchers in (Zheng et al.; 2021), ground-to-air micro-UAV identification has gained increasing research focus in recent years, the far more challenging air-to-air issue has remained mostly unresolved to date. Therefore, to tackle this challenge, the authors here created a new dataset, called as Det-Fly, which has about 13,000 images of micro-UAV that were captured from the camera mounted on other micro-UAV while airborne. Eight different algorithms including SSD, RetinaNet, YOLOv3, RefineDet, Faster R-CNN, FPN, Cascade R-CNN, and Grid R-CNN are used to identify which algorithms works best in detecting micro-UAV. When it came to average precision in detecting micro-UAV, Grid R-CNN was the best of the eight algorithms, while RefineDet had the lowest value.

2.5 Summary

The related work in detecting micro-UAV was explained above. Researchers have used various methods to detect the UAV. Th vision based approach shows promising results in detecting the micro-UAV. In the paper (Zheng et al.; 2021), researchers have concluded that the Grid R-CNN and Cascade R-CNN performed better in detecting the UAV compared to other 6 algorithms they evaluated. Given the performance of R-CNN based algorithms, Mask R-CNN method is chosen in this study for detecting micro-UAV. To my knowledge, this is the first time a micro-UAV has been detected using Mask-RCNN on the dataset used in this research. Table 1 provides the summary of the related papers referred in this study.

Table 1: Summary details of related work for micro-UAV detection

Author(s)	Objectives	Research Design	Keywords	Findings
(Yang et al.; 2019)	Detecting micro-UAV using multiple acoustic nodes	MFCC and STFT were used with SVM and CNN.	audio detection, UAV detection	STFT-SVM model showed the best performance
(Kim et al.; 2020)	Using CNN to detect micro-UAV's fan noise in able to identify it.	Evaluating detection performance using CNN on the features extracted using STFT	STFT, CNN, Deep Learning	The detection rate achieved is 99.74 %.
(Li et al.; 2020)	Using micro-doppler effect to detect micro-UAV	35GHz continuous wave radar to measure UAV data	UAV, micro-doppler, detection	accuracy can approach to 85%.

(Gérard et al.; 2021)	To determine best format of micro-doppler signature to detect micro-UAV	Comparing micro-UAV classification outcomes with different micro-doppler signatures.	spectrum, micro-doppler	Recommended to use spectrum issued from long observations as it classifies better.
(Ezuma et al.; 2019)	To detect micro-UAV using RF signals transferred from the controller to the micro-UAV	Splits signals into frames and transformed into wavelet domain. Used Naive Bayes method to classify micro-UAV signal from other.	radio frequency, naive bayes, micro-UAV	Average accuracy of 96.3% is achieved.
(Allahham et al.; 2020)	Deep learning for UAV detection	Implemented multi-channel one dimensional CNN	Machine learning, Deep learning	Improved accuracies of 96.4%
(Akter et al.; 2020)	Identifying different types of drones	Using sequential CNN to learn different scales feature map of RF Signals of drones	CNN, UAV detection and Identification, RF signal.	92.5% of average classification rate and 93.5% of F1 score was achieved.
(Al-Emadi and Al-Senaid; 2020)	Using Deep Learning to identify micro-UAV based on its RF Signals	CNN was used to classify micro-UAV's RF signals	RF, Deep Learning, CNN	Accuracy of 99.7% was achieved.
(Alipour-Fanid et al.; 2020)	Framework to detect micro-UAV using encrypted Wi-Fi traffic	Features extraction from packet size and inter-arrival time of Wi-Fi traffic	Wi-Fi traffic, machine learning	Accuracy between 85.8% and 95.2% was achieved with 0.15-0.35s detection rate.

(Hassan et al.; 2019)	Creating own dataset of micro-UAV images	YOLOv2 an YOLOv3 was utilized to detect micro-UAV	YOLOv2, YOLOv3	YOLOv3 otperformed YOLOv2 in detection of micro-UAV.
(Yuan et al.; 2020)	Detection of UAV from images captured in low cost camera	Implemented YOLO model wit ResNet and DenseNet	YOLO, UAV	The findings indicate that the method can partially solve the issue of low-altitude tiny UAV detection.
(Seidaliyeva et al.; 2020)	Perform loaded and unloaded UAV detection	Used single stage YOLOv2	Loaded UAV, Unloaded UAV, YOLOv2	mAP of 74.97% as achieved.
(Nalamati et al.; 2019)	Detect micro-UAV in surveillance videos	Using Faster R-CNN for detection	micro-UAV, Deep learning, Faster R-CNN	Faster R-CNN with base ResNet-101 performed better than SSD.
(Mahdavi and Rajabi; 2020)	Detection of drone using different algorithms	Implemented CNN, SVM and nearest neighbour	CNN, drone, SVM	CNN performed best than SVM and nearest neighbour in detecting drones.
(Hu et al.; 2019)	To improve UAV detection	Using YOLOv3 for detection purpose	YOLOv3, single stage detector	Proposed model achieved good accuracy in detecting better micro-UAV.
(Shi and Li; 2020)	New recognition method to detect low-altitude UAV	Implementation of YOLOv4	micro-UAV, detection	YOLOv4 performed better than SSD and YOLOv3 in detecting micro-UAV
(Zheng et al.; 2021)	Air to air micro-UAV detection on custom dataset	Comparison of 8 algorithms on this dataset	micro-UAV, Det-Fly, detection	Grid R-CNN performed best among the 8 algorithms in detecting micro-UAV.

3 Methodology

For the identification of a micro-UAV, the study includes the CRISP-DM data analytics approach as shown in figure 1.

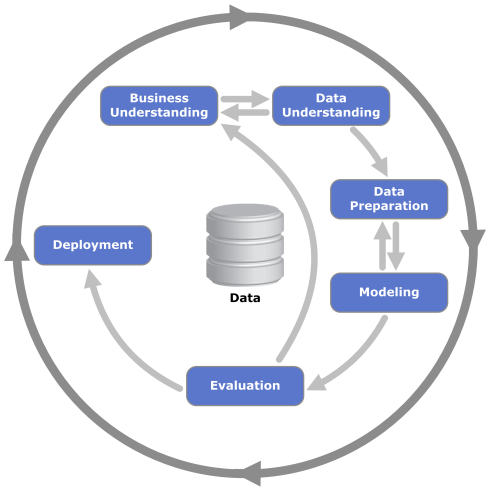

Figure 1: CRISP-DM [5]

Detecting micro-UAV from the camera mounted on another micro-UAV is a challenging task as there many objects that are of similar size as that of micro-UAV such as birds. As seen in the related work, deep learning has proven to be very effective in object detection. CNN based algorithms have improved the object detection rate than the traditional machine learning algorithms. Mask R-CNN has been used in many object detection tasks as seen in (Dongye and Liu; 2020); (Malbog; 2019); (Wang et al.; 2020) and (Vemula and Frye; 2020). Given its good ability in object detection, Mask R-CNN was chosen in this research for detecting micro-UAV. Any model in object detection is trained better with help of transfer learning. Here, in this research, we will utilize the weights of MS COCO dataset that was trained on Mask R-CNN. Figure 2 depicts the research methodology process flow, which includes associated actions that are carried out throughout the research project.

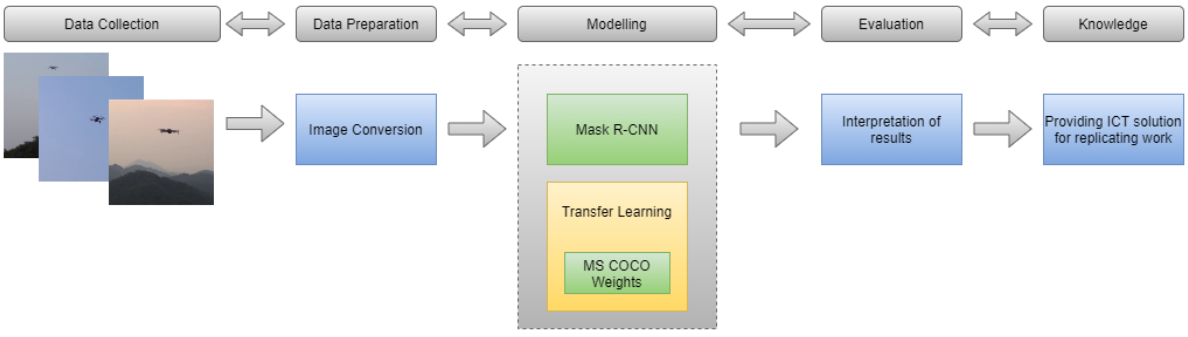

Figure 2: Project Design Process Flow

[5] https://www.datasciencecentral.com/profiles/blogs/crisp-dm-a-standard-methodology-to-ensure-a-good-outcome

3.1 Business Understanding

Unmanned aerial vehicles (UAVs) are being utilized in a variety of applications and are proving to be a very useful technology. The number of UAVs is growing rapidly as UAV technology advances, and their use for outdoor missions has become common. However, as technology develops, not only is there a risk of privacy invasion via illegal UAV photography, but also of terrorism. Micro-UAV detection would aid the government in implementing required actions. Additionally, it assists military personnel in securing the nation's border from UAV attacks. UAV detection will also aid in rapid airborne movement by avoiding collision between UAVs. As seen in the related work, most of the UAV detection methods are based on ground-to-air scenario where images of UAV are captured from a camera installed on ground station. However, in this research the air-to-air scenario is considered where a camera is mounted on an airborne UAV to capture the images of the target airborne UAV. Mask R-CNN will be implemented in detection of UAV in such images.

3.2 Data Collection

The research is based on the dataset made available by the researchers in (Zheng et al.; 2021). This dataset contains 13000 images which is made available publicly on github [6]. When downloading these images, 12120 were successfully downloaded whereas the rest of the images were not been able to download due to error. Therefore, this research will be utilizing 12120 images of micro-UAV. All of these images in this dataset are of resolution 3840 x 2160 pixels. Each image has its respective XML file which contain image information such as image dimensions, bounding box. These images were captured from a camera mounted on an airborne micro-UAV with different background scenarios such as forest, sky, city, clear sky, and night environment. Figure 3 shows images captured in different background scenarios. The type of drone used in this research is shown in figure 4.

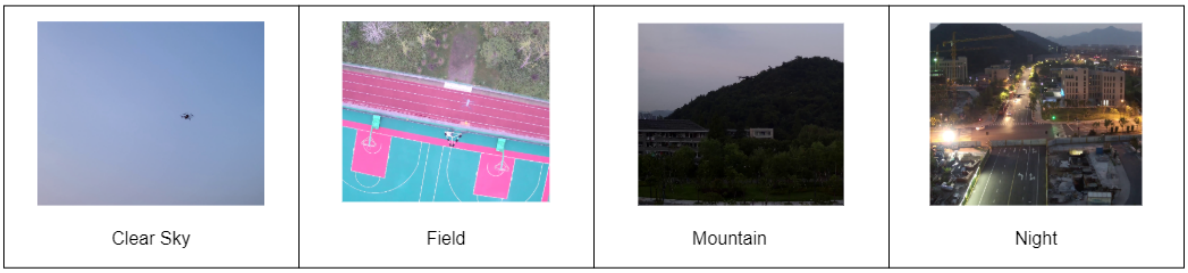

Figure 3: Micro-UAV images with different background

3.3 Data Preparation

Once the UAV images were download to personal drive, they were uploaded to Google Drive so that it can be integrated with Google Colab. Respective XML files were also uploaded on Google Drive. The dimension of the micro-UAV images is 3840 x 2160 pixels.

[6] `https://github.com/Jake-WU/Det-Fly`

Figure 4: A DJI M210 with XT2 camera used to capture target micro-UAV (DJI Mavic). Image referred from (Zheng et al.; 2021)

Training the model on such high dimensions will utilize more time and resources. Hence, these images are converted to 1024 x 1024 dimensions. Images are distributed as: 80% training and 20% testing i.e. from the 12120 images, 9696 images are used for training the model whereas 2424 images are used for testing purpose. As mentioned earlier, each image file has an associated XML file, and from these XML file the bounding box details were extracted. All the images have their respective XML files hence, the need of data cleaning did not arise.

3.4 Modelling

3.4.1 Mask R-CNN

Deep-learning models have shown their importance to the identification of micro UAVs as observed in the literature review. Mask R-CNN is a member of the R-CNN family. Mask R-CNN is a Faster R-CNN variation, which, in addition to the existing classification and bounding regression divisions, includes a division for predicting segmentation masks for each region of interest (RoI). The mask portion of Mask R-CNN is a small Fully Convolutional Network (FCN) added to each RoI that predicts a pixel-by-pixel segmentation mask. According to the algorithm's creators (He et al.; 2017), the Faster R-CNN platform, which supports a wide range of modular architectural designs, simplifies the integration and training of Mask R-CNN.

Mask R-CNN is divided into two stages. These two stages are connect by the backbone network, ResNet-101. In the first stage, a small neural network, Region Proposal Network (RPN), searches the whole feature map and suggests areas that may contain objects. To scan the feature map in an efficient way, anchors are used. Anchors are a collection of boxes with preset positions and scales in relation to images. Individual anchors are given ground-truth classes and bounding boxes based on an IoU value. Due to the fact that anchors with varying scales correspond to various levels of the feature map, RPN utilizes these anchors to determine where on the feature map the object can be found and the size of its bounding box.

In the second stage, a different neural network uses the proposal regions provided by the first stage and assigns them to various particular parts of feature map levels, scans

these parts and produces multi-categorical object classes, bounding boxes and masks.

3.4.2 Transfer Learning

In this research, transfer learning approach has been utilised to train the model efficiently. The Mask R-CNN has been trained on MS COCO dataset [7]. These pre-trained weights will be used in training our model and thus making use of transfer learning. The concept of transfer learning is shown as below.

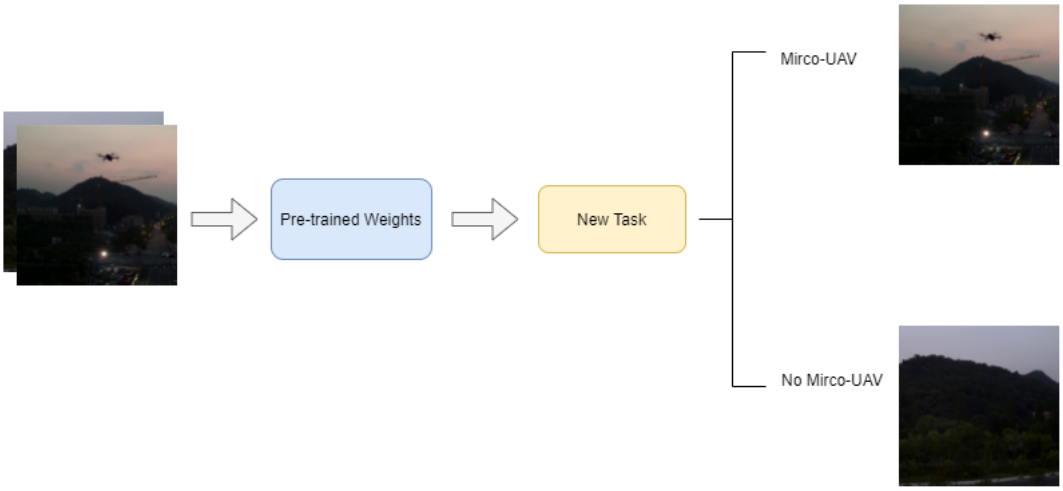

Figure 5: Transfer learning in the research

3.5 Evaluation

The mean absolute precision, or mAP, is often used to assess a model's performance on an object identification task. We are predicting bounding boxes in order to evaluate if a prediction is correct or incorrect depending on the extent to which the predicted and true bounding boxes overlap. The calculation is done by dividing the overlapping area by the total area of the two bounding boxes which is called as intersection over union (IoU). The typical prediction value of a symmetrical bounding box would be 1. If the IoU is higher than 0.5, it is common to assume a good bounding box prediction. Hence, the IoU threshold set in this research is 0.5. Precision refers to the percentage of correctly predicted bounding boxes from all anticipated bounding boxes. Recall is the proportion of accurately predicted bounding boxes across all objects in the image. The recall percentage increases as we make more predictions, but accuracy decreases or becomes unpredictable when we begin generating false positive predictions. A curve or line can be generated by plotting recall along with precision. By increasing the point's on this curve or line we can compute the average precision. The mean of this AP over all the images in the dataset is termed as mean average precision i.e. mAP. Below is the formula of mAP 6, here, Q represents the number of queries and AveP(q) is the average precision (AP) for a given query, q [8].

[7] https://github.com/matterport/Mask_RCNN

[8] https://towardsdatascience.com/breaking-down-mean-average-precision-map-ae462f623a52

$$\text{MAP} = \frac{\sum_{q=1}^{Q} \text{AveP(q)}}{Q}$$

Figure 6: Mean Average Precision

Evaluation results of the implemented model is provided in section 6.

4 Design Specification

Figure 7 shows the architecture of Mask R-CNN. Below are details of its components.

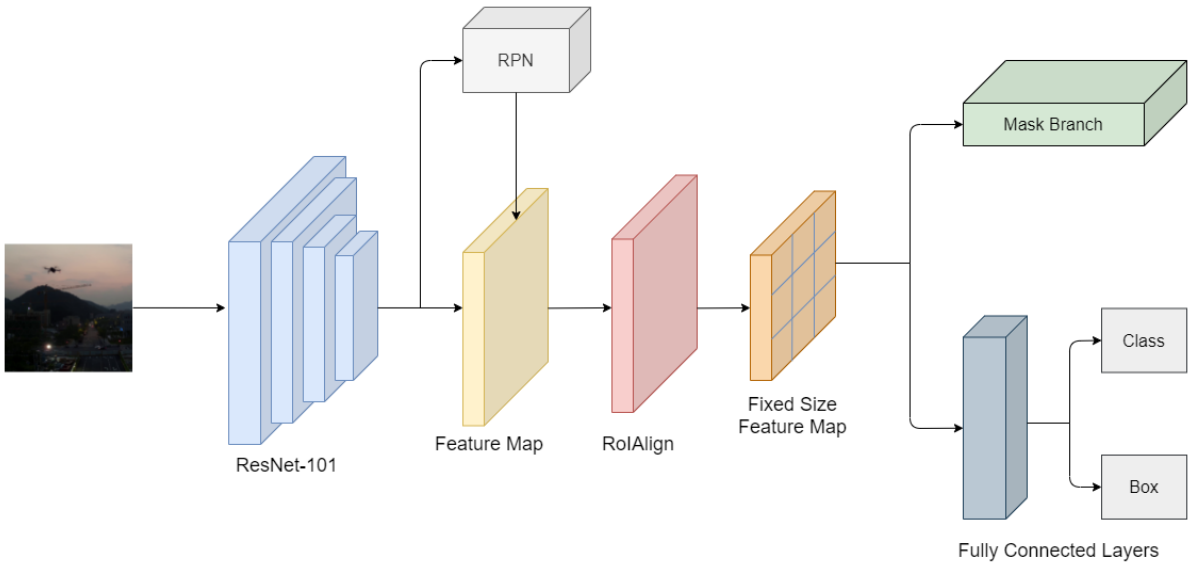

Figure 7: Mask R-CNN architecture

- **Residual Network (ResNet) 101:**

  ResNet 50 and ResNet 101 are supported by the Mask R-CNN model. In this research, ResNet-101 is used as backbone convolutional architecture. As the name indicates, this architecture has 101 layers and had proven to perform better than ResNet-50.

- **Region Proposal Network (RPN):**

  The region proposal network (RPN) is a lightweight neural network that reads every FPN top-to-bottom and suggests regions that may contain items. It does this by placing nine anchor boxes over the picture. Anchors are a group of boxes that have predetermined positions and scales in relation to pictures. Individual anchors are allocated ground-truth classes and bounding boundaries. Due to the fact that anchors with varying scales correspond to various levels of the feature map, RPN utilizes these anchors to determine which level of the feature map should include an object and the size of its bounding box.

- **RoIAlign:**

  The RoI Align is the tweaked version of RoIPool. RoIPool quantizes the floating number RoI to the feature map's discrete granularity, then subdivides it into spatial bins, which are subsequently quantized, and then aggregates the feature values represented by each bin. This has a significant detrimental impact on pixel-accurate mask prediction. Therefore, in RoIAlign it eliminates RoIPool's severe quantization and aligns the retrieved features with the input correctly.

Rather than generating a single defined bounding box, the RoI Align produces multiple warped bounding boxes in a given dimension. As the warped features are fed into FC layers for classification, the softmax classifier is employed to form the final classification result. The regression model is then utilized to enhance the prediction of the bounding box. The Mask classifier is made up of two CNNs and each one produces a binary mask for every ROI. The Mask Classifier provides the network with the ability to generate masks for each class without negatively impacting other classes.

5 Implementation

This section will cover the details of the implemented project. It outlines the initial setup, data modification processes, and model implementation, as well as the tools utilized.

5.1 Environment Setup

The model was trained on Google Colab Pro which provided 25.46 GB of RAM and 110GB of disk space. Colab Pro version was utilized so that high performance GPU can be utilized while training the dataset. To execute the Mask R-CNN utility, tensorflow version 1.15.2 was utilized and the keras version used is 2.2.5. Python programming language was used for the implementation.

5.2 Data and MRCNN Library Setup

The dataset used in this research is stored on google drive. This drive is mounted on google colab and the data is extracted during run time. Similarly, the mask-rcnn library, which is publicly available on github [9], is mounted on colab using git clone command. This mrcnn library contains several functions and classes that is used in this research. A custom function is created to extract bounding box of each image. The dataset is divided in training and testing set with ratio of 80:20 respectively. 9696 images were used for training purpose whereas 2424 images were used for testing purpose.

5.3 Transfer Learning and Model Implementation

For transfer learning, weights of MS COCO dataset, trained on Mask R-CNN, is utilized. This weights are loaded into the model before performing the train operation. The Mask R-CNN model will be utilized with default specifications, with the exception that output layers that are distinct to class will be deleted to allow for the definition and training of new output layers. This is accomplished by using the 'exclude' option and defining

[9] `https://github.com/matterport/Mask_RCNN/tree/master/mrcnn`

the required output layers to exclude from the model once it has been loaded. These excludes are the output layers for the classification label, masks and bounding boxes. The creators of the algorithm Mask R-CNN, used Stochastic Gradient Descent (SGD) optimizer in their implementation which will also be used in this research as it accelerates gradient vectors in correct direction which results in quicker convergence.

As we are identifying micro-UAV in the given images, there are total of two classes used in this research: one is background class and second is micro-UAV. The batch size is set as 2 whereas the number of steps per epoch is set as 4848. The steps per epoch is calculated as number of training images divided by batch size. ResNet-101 is used as the backbone convolutional layer with [4, 8, 16, 32, 64] as the backbone strides. The maximum image dimension is set to 1024 pixels whereas the minimum is set to 800 pixels. The number of GPU used in this research is 1. During training the data, the number of images used per GPU is set to 2. Similarly, during inference mode, the number of images passed to the GPU are set to 1. The learning rate while training the model is set 0.001 with weight decay set to 0.0001. The Evaluation and test results are discussed in the next section.

6 Evaluation

After the model have been successfully implemented, it is essential to evaluate their performance. The mrcnn library provides a inbuilt function named *compute_ap* to calculate the average precision (AP) of the given model. Here, we will be using mean average precision (mAP) to evaluate the entire model. So to calculate the mAP, a new function is developed where the list of APs are collected and then the mean value of these APs is calculated thus providing us the mAP. Here, the model was trained on 6 epochs, with 4848 steps per epoch, and then the model was evaluated. Figures 8 9 10 shows the training and validation loss, classification loss and bounding box loss.

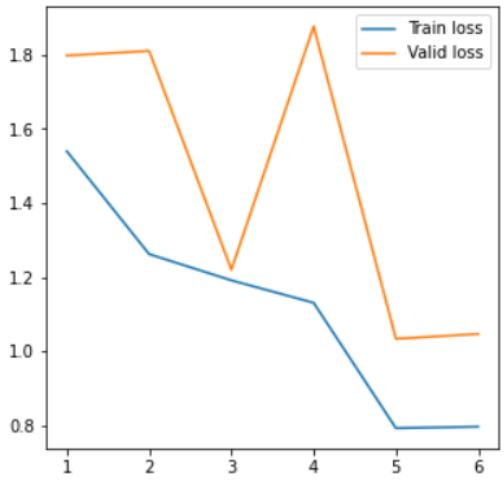

Figure 8: Training and Validation Loss

From the training and validation loss graph, figure 8, it can be seen that loss value dropped after 4th epoch. Similarly for mask r-cnn classification loss, figure 9, the loss value dropped after 4th epoch but after 5th epoch the classification loss on validation

data increased. And for the bounding box loss, figure 10, the loss value too dropped after 4th epoch.

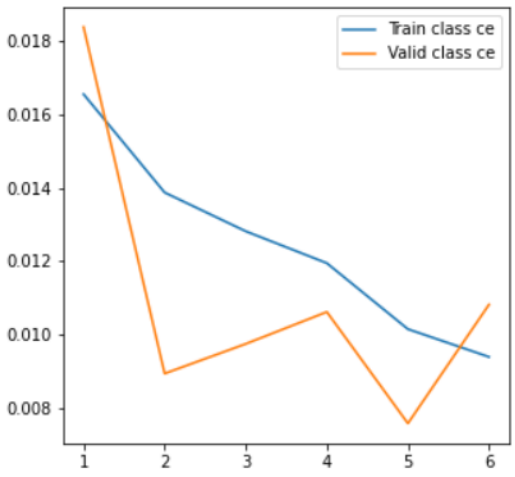

Figure 9: MRCNN Class Loss

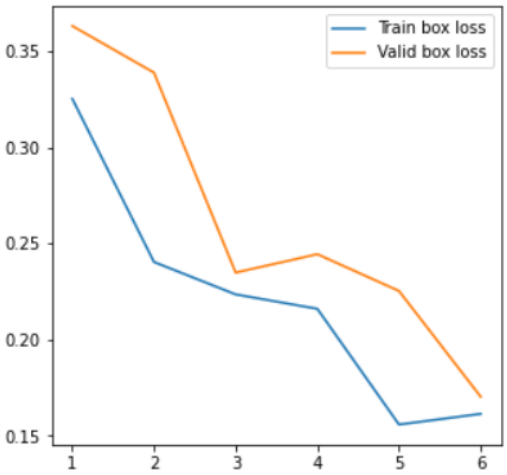

Figure 10: MRCNN Bounding Box Loss

From the loss figures we can see that for epoch 5 the model had the minimum loss among others. Therefore, epoch 5 was selected while performing the inference operation. The weights generated for epoch 5 is loaded into the model and mAP was calculated. We received mAP 72.10% for test images whereas 72.6% mAP for training images. Random images were also tested to detect for the presence of micro-UAV.

Figure 11 12 shows the ground truth and predictions for these random images. It can be seen that the Mask R-CNN is able to predict the micro-UAV in clear background as well as in the dark background.

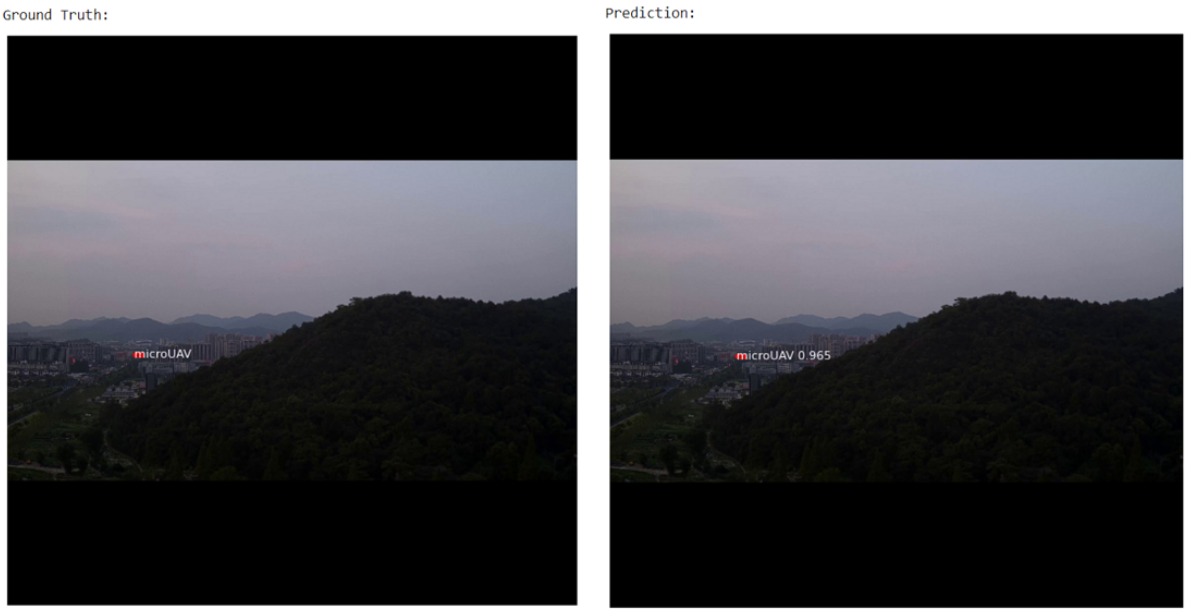

Figure 11: Test Image 1

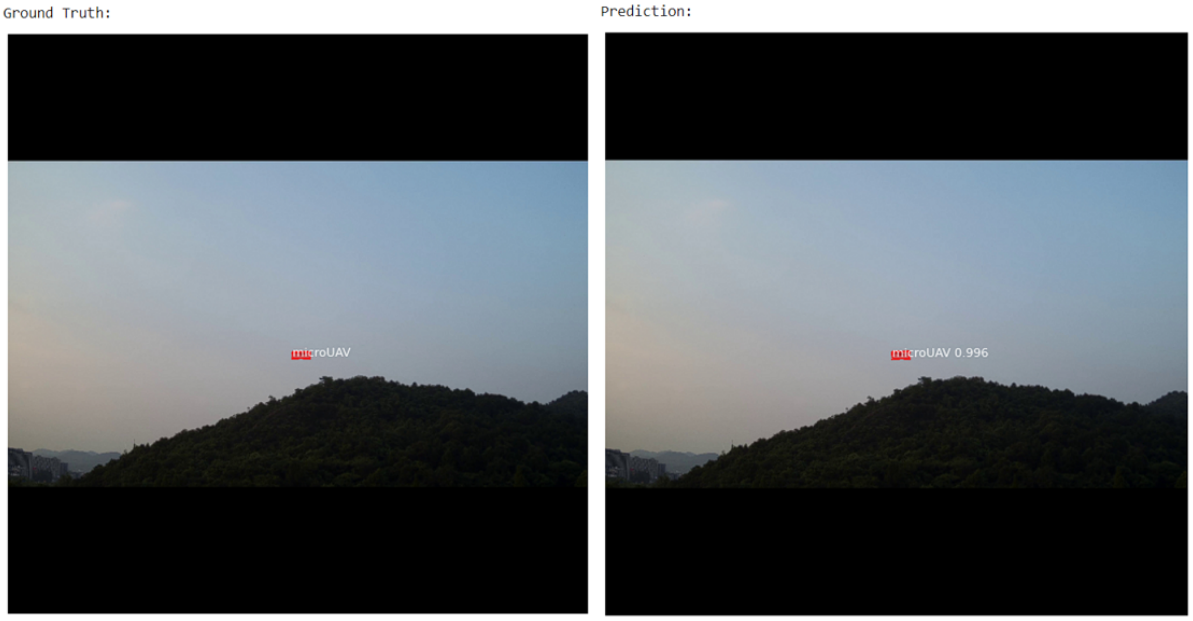

Figure 12: Test Image 2

6.1 Discussion

The objective of this research is to check the performance of Mask R-CNN in detecting micro-UAV in images that were captured from other micro-UAV in different background scenarios. This research utilizes the dataset that were created by the authors in (Zheng et al.; 2021) where these researchers compared eight algorithms on this dataset. In this research, the model was trained on 6 epochs where mean average precision achieved on test set is 72.10% whereas the mAP achieved for training data is 72.60%. As shown in the Evaluation section, the model was able to detect the micro-UAV in even darker

conditions.

The training time required for the model for each epoch was also calculated. The model was first trained using 3 epochs however the mAP achieved was low. Hence the model was then trained on 6 epochs where for 5th epoch a good mAP was achieved. Due to high resolution of the images as well as large number of images to train, the training period of the implemented model is little longer. It took around 6 hours to train the model for 6 epochs where each epoch had 4848 steps. The implemented model did not achieve the mAP score as that achieved by the best performing model i.e. Grid R-CNN in (Zheng et al.; 2021) where mAP of 82.4% was achieved. However, the implemented model can perform better by more fine tuning it as well as changing few specifications. Overall, the model was able to perform well in detecting the micro-UAV in the given set of images.

7 Conclusion and Future Work

The purpose of this research was to identify micro-UAV in given set of images. These images were captured in air-to-air scenario where a camera is mounted on a UAV to identify target UAV. This increases complexity in detecting the micro-UAV as images are captured in various angles and background scenes. Mask R-CNN was used as the model to detect the UAV in these images. Transfer learning approach was followed for improving the detection process. The dataset used in this research is publicly made available on github. The implemented model achieved mAP of 72.10% on the test set of images. Although the mAP achieved is not equivalent or greater than that achieved by Grid R-CNN in (Zheng et al.; 2021), the implemented model performs good while detecting micro-UAV in the images as seen in test images in evaluation. This model can assist many government officials in detecting illegal use of micro-UAV. This model can also be implemented in path planning for fast aerial maneuvers and collision avoidance between UAVs.

The model can be more fine tuned in the future work as well as it can be implemented on other complex UAV datasets. The images here contain single UAV in each snap however the model can be implemented on images having multiple various type of UAVs. The model performance will improve when implemented on multiple datasets.

Acknowledgement

I am grateful for the excellent assistance and valuable feedback from my supervisor Dr. Hicham Rifai throughout my research. I sincerely appreciate him for continuously encouraging me to improve my research and motivating me. It was great experience to work under his guidance.

References

Akter, R., Doan, V. S., Tunze, G. B., Lee, J. M. and Kim, D. S. (2020). Rf-based uav surveillance system: A sequential convolution neural networks approach, *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 555–558.

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.

Alipour-Fanid, A., Dabaghchian, M., Wang, N., Wang, P., Zhao, L. and Zeng, K. (2020). Machine learning-based delay-aware uav detection and operation mode identification over encrypted wi-fi traffic, *IEEE Transactions on Information Forensics and Security* **15**: 2346–2360.

Allahham, M. S., Khattab, T. and Mohamed, A. (2020). Deep learning for rf-based drone detection and identification: A multi-channel 1-d convolutional neural networks approach, *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, pp. 112–117.

Dongye, C.-l. and Liu, H. (2020). A pavement disease detection method based on the improved mask r-cnn, *2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT)*, pp. 619–623.

Ezuma, M., Erden, F., Anjinappa, C. K., Ozdemir, O. and Guvenc, I. (2019). Micro-uav detection and classification from rf fingerprints using machine learning techniques, *2019 IEEE Aerospace Conference*, pp. 1–13.

Gérard, J., Tomasik, J., Morisseau, C., Rimmel, A. and Vieillard, G. (2021). Micro-doppler signal representation for drone classification by deep learning, *2020 28th European Signal Processing Conference (EUSIPCO)*, pp. 1561–1565.

Hassan, S. A., Rahim, T. and Shin, S. Y. (2019). Real-time uav detection based on deep learning network, *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 630–632.

He, K., Gkioxari, G., Dollár, P. and Girshick, R. (2017). Mask r-cnn, *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988.

Hu, Y., Wu, X., Zheng, G. and Liu, X. (2019). Object detection of uav for anti-uav based on improved yolo v3, *2019 Chinese Control Conference (CCC)*, pp. 8386–8390.

Kim, B., Jang, B., Lee, D. and Im, S. (2020). Cnn-based uav detection with short time fourier transformed acoustic features, *2020 International Conference on Electronics, Information, and Communication (ICEIC)*, pp. 1–3.

Li, S., Chai, Y., Guo, M. and Liu, Y. (2020). Research on detection method of uav based on micro-doppler effect, *2020 39th Chinese Control Conference (CCC)*, pp. 3118–3122.

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.

Malbog, M. A. (2019). Mask r-cnn for pedestrian crosswalk detection and instance segmentation, *2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, pp. 1–5.

Nalamati, M., Kapoor, A., Saqib, M., Sharma, N. and Blumenstein, M. (2019). Drone detection in long-range surveillance videos, *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6.

Seidaliyeva, U., Alduraibi, M., Ilipbayeva, L. and Almagambetov, A. (2020). Detection of loaded and unloaded uav using deep neural network, *2020 Fourth IEEE International Conference on Robotic Computing (IRC)*, pp. 490–494.

Shi, Q. and Li, J. (2020). Objects detection of uav for anti-uav based on yolov4, *2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT*, pp. 1048–1052.

Vemula, S. and Frye, M. (2020). Mask r-cnn powerline detector: A deep learning approach with applications to a uav, *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*, pp. 1–6.

Wang, H., Mou, Q., Yue, Y. and Zhao, H. (2020). Research on detection technology of various fruit disease spots based on mask r-cnn, *2020 IEEE International Conference on Mechatronics and Automation (ICMA)*, pp. 1083–1087.

Yang, B., Matson, E. T., Smith, A. H., Dietz, J. E. and Gallagher, J. C. (2019). Uav detection system with multiple acoustic nodes using machine learning models, *2019 Third IEEE International Conference on Robotic Computing (IRC)*, pp. 493–498.

Yuan, X., Xia, J., Wu, J., Shi, J. and Deng, L. (2020). Low altitude small uav detection based on yolo model, *2020 39th Chinese Control Conference (CCC)*, pp. 7362–7366.

Zheng, Y., Chen, Z., Lv, D., Li, Z., Lan, Z. and Zhao, S. (2021). Air-to-air visual detection of micro-uavs: An experimental evaluation of deep learning, *IEEE Robotics and Automation Letters* **6**(2): 1020–1027.