Article

Detection and Recognition of Drones Based on a Deep Convolutional Neural Network Using Visible Imagery

Farhad Samadzadegan [1], Farzaneh Dadrass Javan [1,2,*], Farnaz Ashtari Mahini [1] and Mehrnaz Gholamshahi [3]

[1]   School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran 14399-57131, Iran; samadz@ut.ac.ir (F.S.); f.ashtari@ut.ac.ir (F.A.M.)
[2]   Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, 7522 NB Enschede, The Netherlands
[3]   Department of Electrical and Computer Engineering, Faculty of Engineering, Kharazmi University, Tehran 15719-14911, Iran; mehrnazgholamshahi.khu@gmail.com
*   Correspondence: f.dadrassjavan@utwente.nl

Abstract: Drones are becoming increasingly popular not only for recreational purposes but also in a variety of applications in engineering, disaster management, logistics, securing airports, and others. In addition to their useful applications, an alarming concern regarding physical infrastructure security, safety, and surveillance at airports has arisen due to the potential of their use in malicious activities. In recent years, there have been many reports of the unauthorized use of various types of drones at airports and the disruption of airline operations. To address this problem, this study proposes a novel deep learning-based method for the efficient detection and recognition of two types of drones and birds. Evaluation of the proposed approach with the prepared image dataset demonstrates better efficiency compared to existing detection systems in the literature. Furthermore, drones are often confused with birds because of their physical and behavioral similarity. The proposed method is not only able to detect the presence or absence of drones in an area but also to recognize and distinguish between two types of drones, as well as distinguish them from birds. The dataset used in this work to train the network consists of 10,000 visible images containing two types of drones as multirotors, helicopters, and also birds. The proposed deep learning method can directly detect and recognize two types of drones and distinguish them from birds with an accuracy of 83%, mAP of 84%, and IoU of 81%. The values of average recall, average accuracy, and average F1-score were also reported as 84%, 83%, and 83%, respectively, in three classes.

Keywords: drone; UAV; deep learning; convolutional neural network CNN; drone image dataset; drone detection; drone recognition

1. Introduction

With the increasing development of drones and their manufacturing technologies, the number of them being used for military, commercial, and security purposes is increasing [1–3]. In recent years, the use of different types of drones has received much attention due to their efficiency in applications such as airport security, the protection of its facilities, and integration into security and surveillance systems [4–6]. On the other hand, drones can also be considered a serious threat in these security areas, and therefore, it is important to develop an efficient approach to detect types of drones in these applications [7–9]. Such technologies can be used in airport security and any military systems to prevent drone intrusion or to ensure their security [7,10,11]. Therefore, the detection, recognition, and identification of UAVs are crucial in discussing public safety and the threats posed by their existence. Detection is the process of observing the target, and this target may be suspicious and threaten the security of the target environment, recognition is the determination of the target category, and identification refers to diagnosing the type of target category. In this article, based on the physical and behavioral similarities between drones and birds, two

types of drones are detected and recognized and their distinction from birds is determined. For this purpose, various sensors can be applied such as radar [12,13], LIDAR [14], and RF-based [15,16] sensors. In addition, drone detection and recognition have also been performed using acoustic sensors [17,18] and thermal sensors [19]. However, the use of these sensors is costly and energy-consuming [12]. In addition, drone integration with these sensors is limited due to the weight and size required, and in the case of thermal imagery, sensors usually suffer from a lower resolution. However, the use of visible imagery does not have the problems associated with integrating sensors and drones, and unlike thermal sensors, it has higher resolution. However, visible imagery also has problems such as occluded areas, crowded backgrounds, and lighting problems within the image. Therefore, the solution to this problem depends on the method used to detect and recognize the drone.

In the last decade, deep learning networks have become the best model for visual processing, such as object detection and tracking [20–23]. Object detection using deep learning networks has received much attention due to its higher computational power and accuracy [24]. Among deep neural networks, convolutional neural networks (CNNs) are the best representative for object recognition [25]. These networks are powerful in feature extraction and hence have been considered and investigated more for object recognition [26–28]. They are more desirable for object recognition as they extract more features than conventional object recognition methods [29,30]. Object recognition methods are divided into two categories according to their function in examining network input. The first category includes area-based detection methods where a set of the proposed areas is first considered, and then each of these areas is classified into different object categories. The second category refers to classification and regression-based detection methods such as YOLO [31] and SSD [32] deep learning methods [33].

Due to the importance of detecting and recognizing drones for various applications, providing public safety, and problems associated with different sensors, the use of visible imagery is better due to features such as high resolution, low cost, and the ability to integrate with different types of drones. However, there are challenges such as crowded backgrounds and confusing drones with birds due to their small size in these images; therefore, it is necessary to use a suitable method to solve these challenges. YOLO Deep Learning Network is the best way to overcome these challenges due to its higher accuracy, speed, and the accurate analysis of the input images. Among the different versions of this network, the latest version has a higher speed and accuracy in detecting objects [34]. For this reason, this paper investigates UAV detection and recognition using YOLOv4 Deep Convolutional Neural Networks and visible imagery.

1.1. Drone Detection and Recognition Challenges

It is important to detect and recognize different types of drones as they can trespass into sensitive areas and pose potential threats. However, detecting and recognizing different types of drones and distinguishing them from birds is always fraught with challenges. Some of these challenges are discussed below.

1.1.1. The Resemblance of Drones and Birds

Drones can be mistaken for birds, especially at long distances, because of their similarity in behavior and physical characteristics. For some samples on the similarity of drones and birds, see Figure 1.

Figure 1. Some samples of challenges related to the resemblance of drones and birds.

1.1.2. Different Weather Conditions and Crowded Background

Problems such as the presence of drones in crowded environments, varying weather conditions, and different lighting conditions make drone detection difficult (Figure 2).

Figure 2. Some samples of challenges related to the different weather conditions and crowded background.

1.1.3. Small Size of Drones at Long Distances

The presence of drones at long ranges makes them smaller and causes problems in detection and recognition. Figure 3 illustrate this challenge.

Figure 3. Some samples of challenges related to the small size of drones at long distances.

1.1.4. Lack of Scalability

The presence of drones at close and distant distances with different resolutions poses challenges to accurately detect and recognize different types of drones (Figure 4).

Figure 4. Some samples of challenges related to lack of scalability.

Given the challenges in detecting and recognizing different types of drones and distinguishing them from birds, it is very important to use a fast and accurate method to overcome these challenges and prevent drone intrusion into critical infrastructure.

2. Related Works

In recent years, drone detection, recognition, and identification have received much attention in various applications. The concept of detection in this study means the ability to detect the presence of a drone as opposed to its absence. The concept of recognition is also the ability to detect the category to which the drone belongs. Identification is also the ability to recognize the type of drone group. In this study, the problem of drone detection was investigated using the dataset and the proposed method. According to the studies, the dataset for drone detection is obtained using active and passive sensors [35,36]. In studies related to the detection and recognition of drones using active sensors, the use of radar and LIDAR sensors is discussed [14,35,37]. Problems with both of these sensors include high costs and limited integration into small drones. In addition, the use of thermal sensors results in lower accuracy due to low spatial resolution [19], and the use of acoustic sensors in drone detection and recognition has limitations such as high cost and limited onboard use [17]. Therefore, due to the aforementioned limitations of using active sensors, visible imagery was used in the context of passive sensors that do not have the mentioned problems and do not have weight limitations when integrated into small drones.

As previously mentioned, issues such as the unpredictable movements and speed of drones, the long-distance of the drone, its close resemblance to birds, its small size, the presence of hidden areas in the images, crowded backgrounds, the inability to separate the background, the problems with light in visible images, and different weather conditions challenge drone detection and recognition.

For this reason, new methods of deep learning are used to solve the challenges based on studies. In 2001, Q et al. detected moving objects using a set of visible images with fixed background and edge tracker methods. The object is then detected by finding the edge difference in successive images [38]. In 2011, Lai et al., in a study called vision-based air collision detection system, detected drones using morphological filters to prevent airborne collisions [39]. In 2016, Ganti et al. detected drones using background subtraction and image-based methods [40]. Moreover, Li et al. proposed a new drone detection method by mounting cameras on a large variety of drones. In this work, the drone was detected by computing background motion with a perspective transformation model and detecting moving objects by foreground spatio-temporal features [41]. In 2017, Wu et al. detected the drone using visible images and image sensors. In this study, the drone is detected using a saliency map, and it is localized using a Kalman filter [42].

In these studies, the traditional method of background subtraction has been used to detect drones, which do not have the appropriate accuracy and speed compared to modern methods. This year, researchers detected drones in a set of visible images using artificial intelligence-based methods and using RPN [43], CNN, Zeiler, VGG16 [44], and YOLOv2 neural networks [45]. The limitation of these studies was the low accuracy in detecting drones, which was improved in later studies by improving the methods used. In 2018, Li et al. detected drones in video datasets by subtracting background images and classification methods based on deep learning networks. In this article, the Kalman filter is applied to moving objects for better detection [24]. In this study, the deep learning method used can improve the accuracy of diagnosis using visual information. In 2019, drone detection was performed using YOLO [46], Faster-RCNN [47], and SSD [47] methods. RCNN and SSD methods were used to detect drones in video datasets, with the RCNN method showing better accuracy. The use of the YOLOv3 deep learning network in this study has resulted in improved accuracy and precision of drone detection compared to other methods due to its lightweight architecture and appropriate depth. In 2020, drones were detected using YOLOv4 [48], YOLOv3 [21,48], YOLOv2 [20], tiny-YOLOv3 [49], Fast-RCNN [49], and SSD [48] networks and the results were compared [48]. The three models YOLOv4, YOLOv3, and SSD were compared, and, respectively, YOLOv4, YOLOv3, and SSD had the best accuracy. The YOLOv2 and YOLOv3 deep learning networks had the best accuracy.

In 2021, using a deep learning network, the challenges in drone detection were examined in more detail. This year, segmentation-based methods were used to detect drones in crowded backgrounds [50], and another study detected drones in real-time using the YOLOv3 network on NVIDIA Jetson TX2 hardware [51]. The use of this method has provided good accuracy and speed and is capable of detecting drones of various sizes. Other methods used to detect drones include Faster RCNN, SSD, YOLOv3, and DETR, whose performance was examined in a series of visible images [22]. All the methods used in this study performed well in detecting drones, but YOLOv3 provided the best precision. Researchers have also recently used YOLOv4 [52], a pruned YOLOv4 [36], RetinaNet [36], FCOS [36], and YOLOv3 [36] network in video and image datasets to achieve high accuracy in drone detection. The use of YOLOv4 in the first study provided acceptable drone detection results compared to similar studies and had better accuracy. Furthermore, in the next study, the networks used had good accuracy but good performance in detecting small and fast drones. Therefore, the pruned YOLOv4 method gave better performance compared to these methods. In 2021, Coluccia et al. identified several types of multirotors and a fixed-wing with their commercial models in video sequences. The diagnostic system in this work is associated with a warning algorithm that sounds when the drone is observed. In

this work, the standard Cascade R-CNN architecture, Faster R-CNN, YOLOv3 network, and YOLOv5 network were used to identify drones vs. birds. The discussion on detection in a variety of backgrounds with additional data also needs to be extended [53].

Based on the results of the studies, the YOLOv4 Deep Learning network presents higher accuracy and speed in detecting and recognizing drones in visible imagery than conventional methods. Therefore, this method was used to detect and classify two types of drones, such as multirotors, helicopters, and birds.

3. Materials and Methods

Due to the challenges in drone detection and recognition such as crowded background, a close resemblance to birds, smaller size of drones, longer distance, and lighting problems in the image, in this study, a deep learning-based method is proposed. The proposed drone detection and recognition process consist of four main steps, as presented in Figure 5. The first step is to prepare the data properly as the input of the proposed architecture. The second step is the network training phase which is implemented to detect and recognize two types of drones and also birds. Then, in the third step, the trained model is tested using a large variety of drone and bird datasets. Finally, the performance of the model is evaluated, and the detection and recognition process is performed on the input test data.
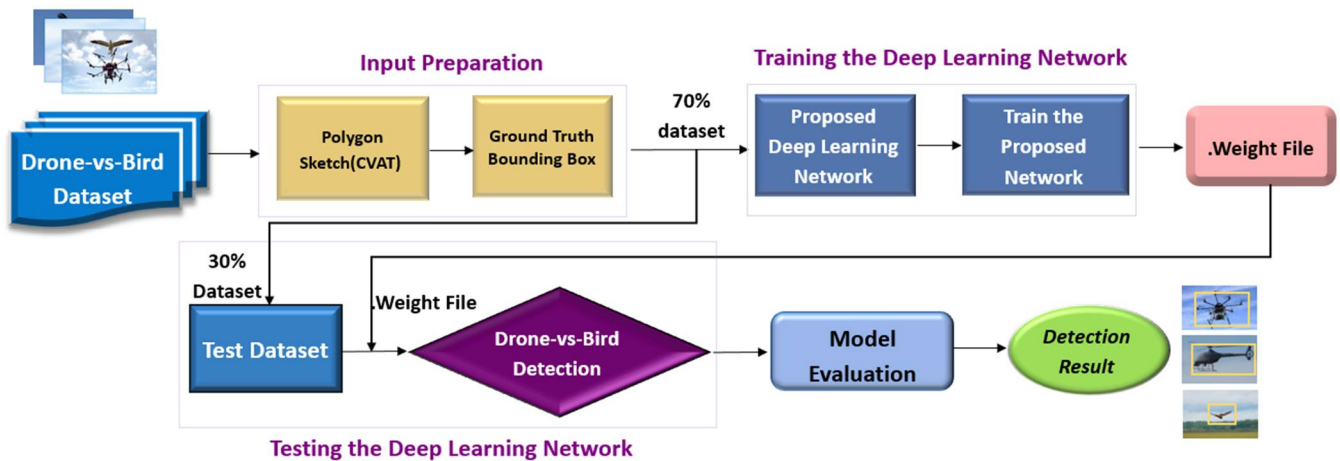

Figure 5. Proposed detection and recognition diagram using convolutional neural network.

3.1. Input Preparation

In order to train the network, a set of drone and bird visible images are prepared to be fed into the proposed network. According to Figure 6, the drone dataset used for training includes a number of multirotors, helicopters, and birds (Figure 6). In total, 70% of the images are used for training and the rest for validation.
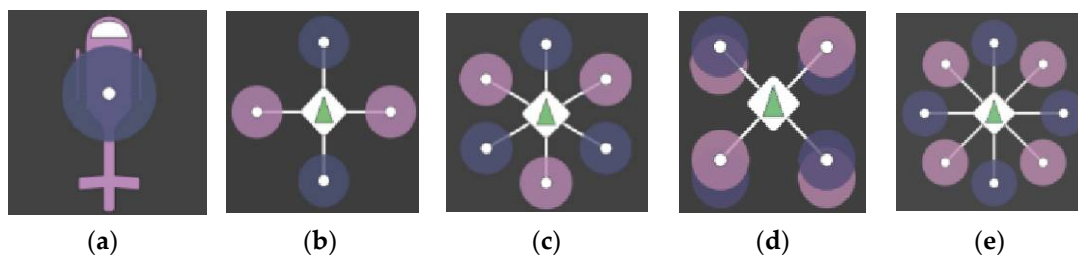

| (a) | (b) | (c) | (d) | (e) |

Figure 6. Types of drones: (**a**) Helicopter, (**b**) Quadrotor, (**c**) Hexarotor, (**d**) Octo Coax Wide, (**e**) Octorotor.

Preparation of the input data involves drawing the ground truth bounding box around the drone and converting it to the normal input format between [0, 1]. In the proposed method, as presented in Figure 7, the input includes the class number, the center coordinate of the bounding box (x, y), and its width and height (w, h) [31].
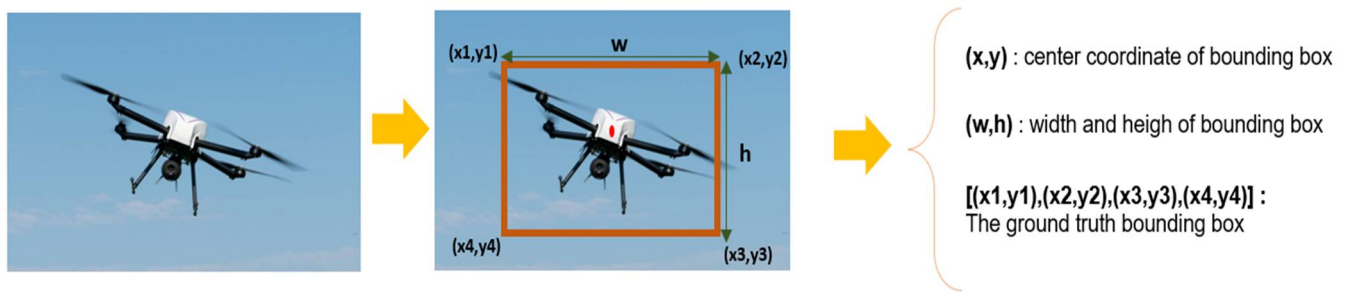
Figure 7. Ground truth bounding box sketch.

Afterwards, the normalized coordinates of the center of the bounding box containing the drone and its height and width are obtained. This information includes x_center, y_center, w, and h. The input data is then divided into two categories of training and testing. Then the bounding box information in the appropriate format is sent to the training stage and finally for the network test.

3.2. Training the Deep Learning Network

Considering the reviewed advantages of the YOLOv4 deep learning network, in this paper, it is applied to detect flying drones and birds in crowded environments. The proposed network consists of a four-section architecture as the input, backbone, neck, and head (Figure 8).
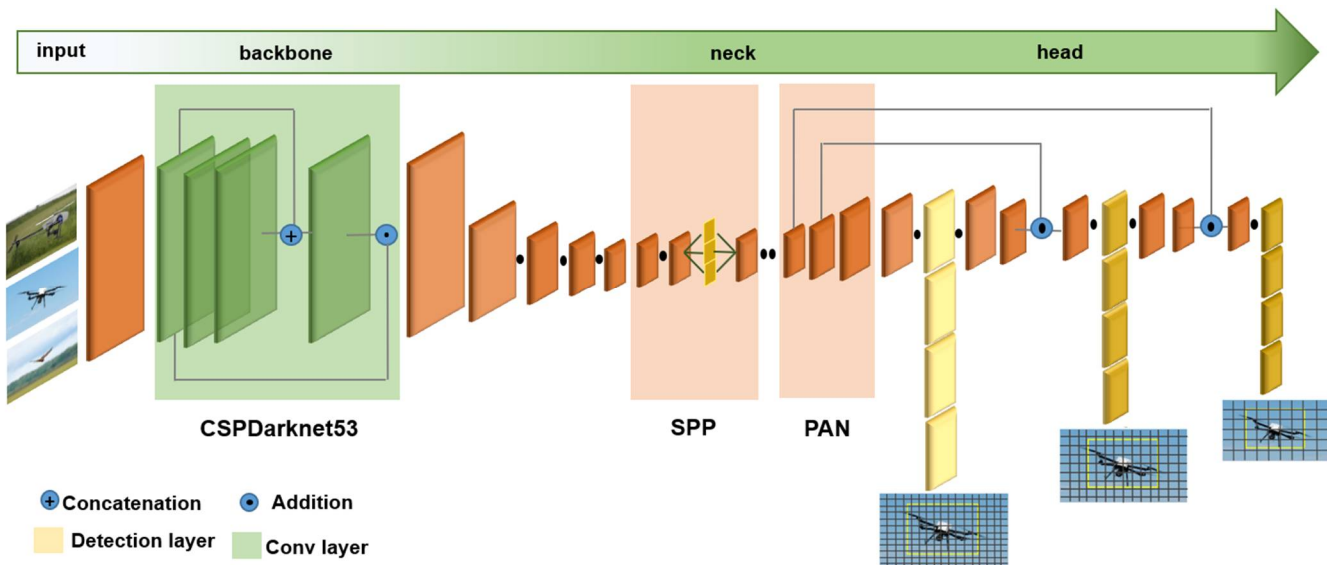

Figure 8. The architecture of the proposed deep learning network.

3.2.1. Backbone; Feature Map Extractor

In the backbone, input data which is prepared in the previous step, is introduced into the network, and feature extraction is performed on the visible imagery of drones and birds dataset. CSPDarknet53, where CSP stands for cross stage partial, is the feature extractor network used in the proposed method to extract more accurate features. This network has good accuracy and speed due to having desirable convolution layers [34].

- CSPDarknet53

The proposed method uses the CSPDarknet53 [54] feature extraction network to detect two types of drones and birds. CSPDarknet53 is a convolutional neural network that uses the Darknet53 network architecture. This feature extractor divides the basic drone feature

map into two sections while they are finally merged step by step to extract drone features from the input dataset. This stage is one of the most critical in drone and bird detection. It is obvious that better performance and more accurate feature extraction will improve the detection in terms of accuracy and speed and error reduction while detecting and recognizing drones and birds.

3.2.2. Neck; Feature Map Collector

When the feature extraction is completed, the generated feature map is introduced to the next processing step, which is the neck part in the proposed method and is a feature map collector. This part consists of two main sections as additional blocks and path aggregation blocks. In the additional blocks section, spatial pyramid pooling (SPP) and a path aggregation network (PAN) were used in the path aggregation blocks [34]. According to Figure 9, In the SPP network, the input drone and bird dataset first enter the convolutional layer, and a feature map is generated. The created feature map then goes through three integration layers with different scales of 16×256-d, 4×256-d, and 256-d [28]. Then, a one-dimensional vector is created and enters the fully connected (FC) layers. All neurons in these layers are connected to the neurons of the previous layer. The main function of the FC layers is to combine the local property in the lower layer with the local property in the upper layers. One of the advantages of using the SPP network is to improve the prediction speed of bounding boxes containing drones or birds. This network, due to having three pooling layers, can receive inputs of different sizes and have an acceptable performance [28]. Finally, the improved PAN network completes the neck step in the proposed detection and recognition method [55].
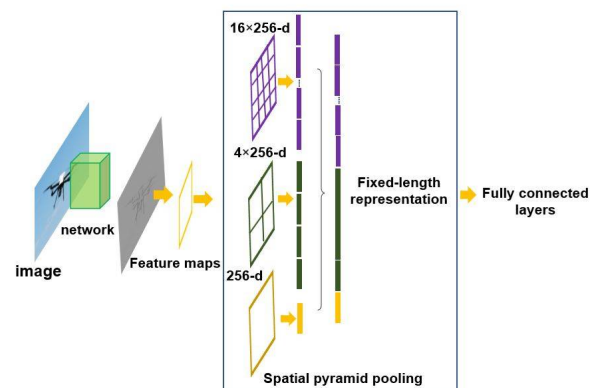

Figure 9. Spatial pyramid pooling (SPP) block.

3.2.3. Head; Detection and Recognition Results

The head stage in the proposed deep learning network consists of three main sections. First, the input drone and bird images with input parameters enter the network, and they are divided into $S \times S$ cells, in which s is determined by the network. This image enters the network, and the convolutional layers in the YOLO network are applied to each cell grid of the convolutional network. The output of the network in the last step is the class probabilities along with the bounding box, which are represented as a three-dimensional tensor with dimensions of $(5 + C) \times B \times S \times S$. The value of C indicates the number of classes and the value of B indicates the number of the predicted bounding boxes. Each drone bounding box contains the information of the center point (x, y) and the width and height of the bounding box (w, h), and the confidence score parameter. Then in the last two stages of the proposed architecture, the type of the extracted drone or whether it is a bird is predicted and classified.

To improve the detection and recognition capabilities of the proposed method, two features called bag of freebies (BOF) and bag of specials (BOS) are applied.

1. Bag of Freebies (BoF)

The bag of Freebies method is only responsible for increasing the cost of training or changing the proposed training strategy. In the proposed network, CutMix [56] and Mosaic methods for data enhancement, DropBlock regularization [57], and class label smoothing are used as the most important BoF features. Data augmentation methods are also used to increase the variety of drone and bird images and to improve the generalization of the deep learning model. For example, in this study, to overcome photometric distortions of the drone and bird dataset, methods are used to adjust brightness, color, saturation, contrast, and image noise reduction. In addition to eliminate geometric distortions and increase the generalizability, scalability, and accuracy of prediction, methods such as random rotation, scaling, cutting, and rotating images of drones or birds are considered.

Another feature of BoF is the use of Focal Loss (*FL*) [52], which is an improved version of the cross-entropy (*CE*) [58] loss function. *FL* fixes class imbalance problems and assigns more weight to misclassified examples or the object of our interest and less weight to easy examples such as background objects. Thus, focal loss can reduce the influence of simple examples and focus on hard negative examples. *FL* has an additional coefficient $(1 - p_t)$ γ to the cross-entropy loss, with a tunable focusing parameter $\gamma \geq 0$; it is presented in Equations (1) and (2).

$$CE(\rho_t) = -\log \rho_t \tag{1}$$

$$FL(\rho_t) = -(1 - \rho_t)^\gamma \log \rho_t \tag{2}$$

The proposed network uses the concept of label smoothing to create a more robust model. Label smoothing smooths hard labels and turns them into soft labels. This concept avoids overconfidence that often occurs in deep networks.

In order to network training, the inclusion of IoU loss is also considered in the proposed method. To evaluate the model quality in traditional deep learning models, the L2 concept is used to calculate the difference between the real bounding box and the predicted bounding box. One of the disadvantages of the L2 error is that it limits and minimizes the errors both in the larger and the smaller bounding boxes (Figure 10). However, using the IoU loss can provide a more accurate prediction of the bounding box error [34].
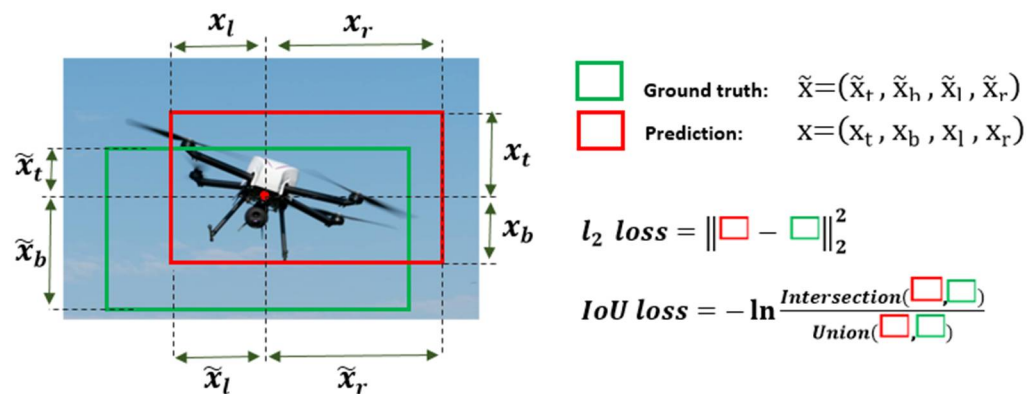

Figure 10. The concept of IoU loss in the proposed network.

2.   Bag of Specials (BoS)

BoS is a set of methods that increases the accuracy of object detection and recognition for types of drones and birds exploration, despite a small increase in the cost of inference. Several techniques have been used in BoS [34]. Some of the main techniques are the use of the Mish activity function, CSP connections path aggregation network (PAN) [34], and spatial pyramid pooling (SPP) block [28]. In the proposed detection and recognition method, the Mish activity function helps to improve the information flow in the network. This function avoids saturation and generally avoids the gradient vanishing problem on near-zero values and overfitting issues [34]. At the end, after completing the network

training process, the model weight file is created and saved to test the network with a variety of drone and bird images.

3.3. Testing the Deep Learning Network

To test the capabilities of the proposed deep learning network in the detection and recognition of drones (multirotor, helicopter) and to distinguish drones from birds in visible imagery, the generated weight file, which is the result of the training stage, is applied. The proposed technique also uses the non-maximum suppression (NMS) method to select the best bounding box containing the drone or bird from several predicted bounding boxes. This method is used to remove possible bounding boxes and select the best bounding box that contains the drone or bird. Finally, the final bounding box containing the target objects and the output parameters of the bounding box are presented.

3.4. Evaluation Metrics

To evaluate the potential of the proposed method, the IoU, precision, mAP, recall, accuracy, and F1-score are used. This evaluation strategy will give us a better understanding of how the model works.

- IoU (Intersection over Union). This evaluation metric means the degree of overlap between the predicted bounding box and the ground truth bounding box. In this study, a threshold of 0.7 was used to classify the input data. This means that if the IoU value is more than 0.7, the classification is True Positive (TP) and otherwise False Positive (FP). Using the number of these values, a complexity matrix was formed, and the rest of the evaluation metrics were calculated using it.
- Confusion matrix. This is a matrix of size n × n (n = number of classes) to show how accurate the model works [59]. The columns of this matrix represent the true class of intended objects, which in this case includes two types of drones and birds. On the other hand, the rows of this matrix represent the predicted classes by the proposed deep learning model. For a better explanation of the confusion matrix in this application, an example of the confusion matrix 2 × 2 is shown in Figure 11. The positive class is related to drones, and the negative class is related to birds. Since this study involves three classes, this matrix is generalized to a size of 3 × 3. Precision, recall, F1-score, and accuracy can be calculated using FN, TN, TP, and FP values.
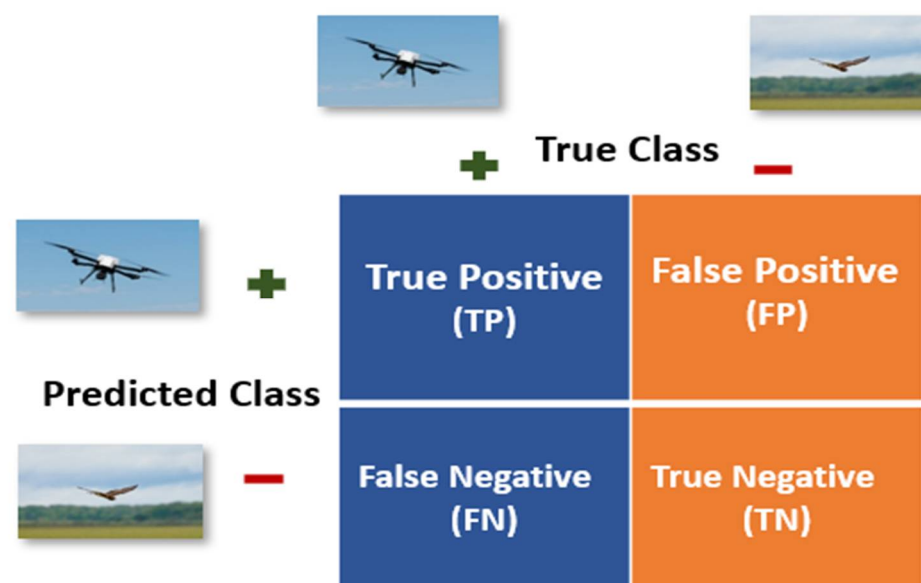

Figure 11. Sample confusion matrix in the proposed method.

- **Precision** means that among the inputs whose class is predicted to be positive, what percentage of them are actually positive class members [59]. According to Equation (3), the value of this metric is between zero and one. Precision is calculated separately for each of the classes. In this study, precision is defined in each of the multirotor, helicopter, and bird classes. For instance, the precision of the multirotor class means that of all the inputs projected as multirotor, what percentage are actually multirotor. Similarly, these criteria are defined for other classes.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

- **mAP** is determined by calculating the average precision of the multirotor, helicopter, and bird classes. In other words, the mAP evaluation metric compares the ground truth bounding box with the predicted bounding box of the targets and calculates a certain value as the score. An increase in this number indicates the more accurate performance of the proposed model in detection and recognition (Equation (4)).

$$\text{mAP} = \frac{1}{|Classes|} \sum_{c \in Classes} \frac{TP(c)}{TP(c) + FP(c)} \tag{4}$$

- **Recall** indicates the percentage of the total data in the positive class, which is predicted to be positive [59]. Similar to the concept of precision, recall is calculated separately for each class. For example, the recall in the multirotor class means that among all the entries that are multirotor, what percentage of them are correctly detected and recognized as multirotor (Equation (5)).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5}$$

- **F1-score** is the harmonic average of recall and precision and is calculated separately for each of the classes [59]. According to Equation (6), this measure performs well on unbalanced data because it considers false negative and false positive values [59].

$$\text{F1\_score} = \frac{2TP}{2TP + FP + FN} \tag{6}$$

- **Accuracy** shows the overall performance of the model [59]. Accuracy means that the proposed model correctly detects and recognizes what percentage of the data is truly positive and negative. In this study, accuracy means that the deep learning model correctly detects the percentage of the input data class (multirotor, helicopter, and bird).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

4. Experiments and Result

In order to evaluate the capability of the proposed method regarding the detection and recognition of types of drones and to distinguish them from birds, the implementation steps and the dataset are resented, and the obtained results are discussed.

4.1. Data Acquisition and Model Implementation

To begin the training phase of the network, it is necessary to prepare a dataset of drones and birds. To increase the performance, reliability, and generalizability of the network, a variety of public images and videos covering two types of multirotor and helicopter drones and a set of several bird species are used. Common to all these images is the use of a visible sensor with a resolution between 96 dpi and 300 dpi. The imaging system in this study is a digital camera. The images were taken keeping in mind the basic concepts of digital

photography such as aperture, ISO, and shutter speed settings. In addition, the collection of videos was converted into images with a frame rate of 2 FPS.

Figure 12 illustrates some sample images of multirotor and helicopter drone types. Multi-rotors include four types as Quadrotor, Hexarotor, Octo Coax Wide, and Octorotor, and the collected data covers all four types of multirotors. These images are collected in different environments with crowded backgrounds and different lighting conditions at diverse distances to evaluate the accuracy and generalizability of the proposed model in different conditions. The proposed dataset contains images where different types of moving drones. A total collection of 10,000 images covering multirotors, helicopters, and birds are collected. Approximately 70% of the collected images are used for network training and 30% for network testing. Therefore, there are 1166 images of each of the four types of multirotors (quadrotor, hexarotor, octo coax wide, and octorotor), helicopters and birds, of which a total of 7000 images are prepared for the training phase.To label the images and draw the rectangle that fits the object, the computer vision annotation tool (CVAT) is applied, and the data is divided into three classes. In this method, the multirotor is labeled in the first class, the helicopters in the second class, and the birds in the third class.
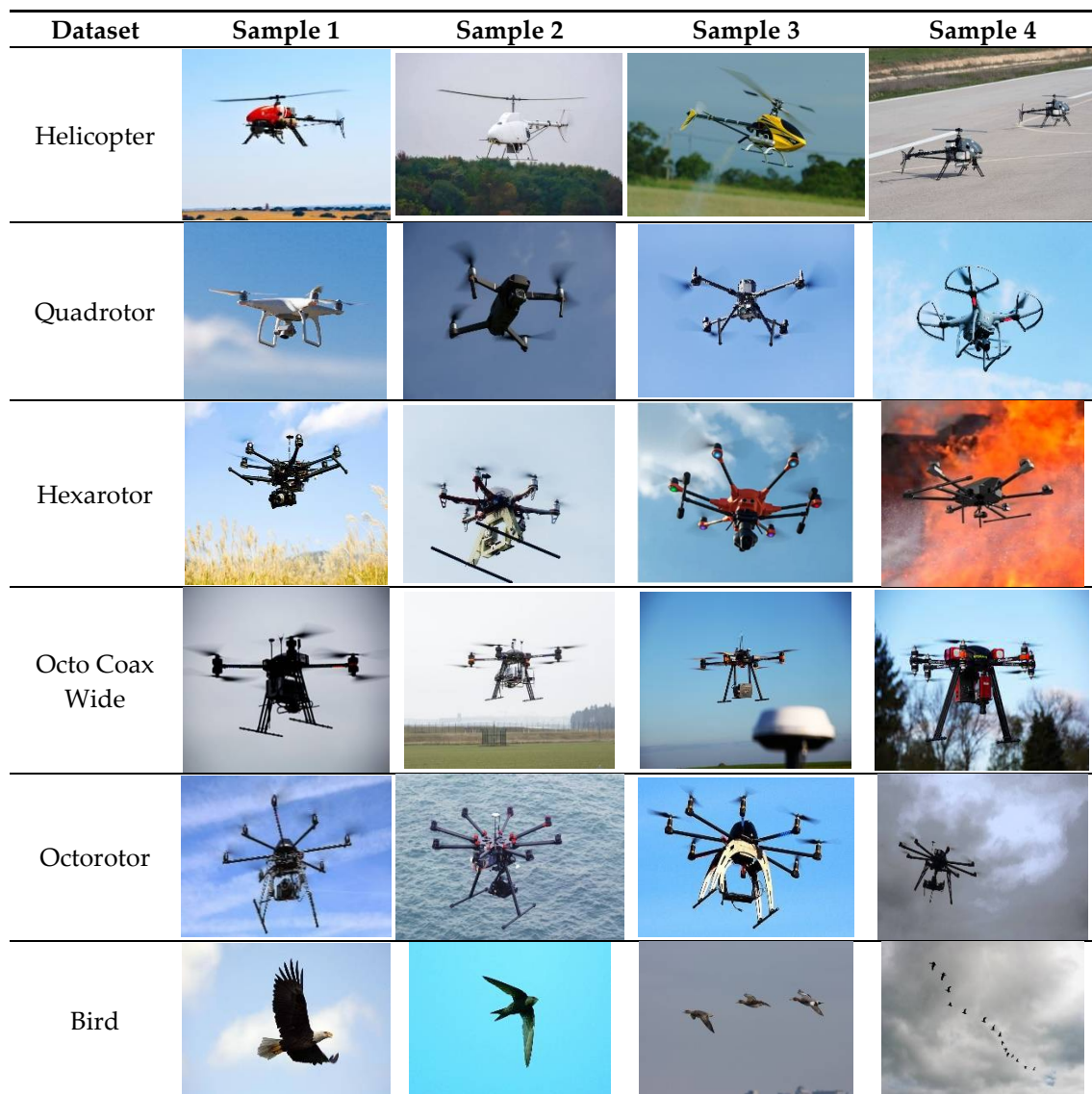

Dataset	Sample 1	Sample 2	Sample 3	Sample 4

Figure 12. The drone-vs-bird dataset.

In this study, the Darknet framework [60] and an Nvidia Geforce MX450 graphics processing unit (GPU) are used to train the network. Furthermore, CUDNN 8.2, Cuda Toolkit 10.0, and OpenCV Library version 4.0.1 are implemented to train the deep convolutional neural network technique.

In order to train the proposed CNN model, the main source code of the darknet framework is prepared, and the configuration files are modified [34]. Moreover, the number of classes in the configuration file is changed to three. In this method, there are three convolutional layers before each of the three layers of YOLO to build a high-level feature map of the drone-vs-bird images. In these three layers, filters are used to extract the features from input drone-vs-bird images. According to Equation (3) and the number of classes equal to 3, the number of filters is changed to 24 in the three convolutional layers, as is explained.

$$Filters = (number\ of\ classes + 5) \times 3 \qquad (8)$$

To start the training step of the deep learning network, the number of batches and learning rate is set to 1 and 0.0005, respectively. The subdivision is set to 64 according to the GPU type used, and the size of each of the input images is 160 × 160. The steps are changed to 16,000, 18,000 using the formula (80% maximum batches, 90% maximum batches). Finally, the model is trained with 20,000 iterations, and the weights file is saved after every 10,000 iterations. The overall view process of training the network and reducing the average loss until 0.52 after 20,000 iterations and 23 h is presented in Figure 13. To test the network, the final weight file is used, and its performance is compared using evaluation metrics.
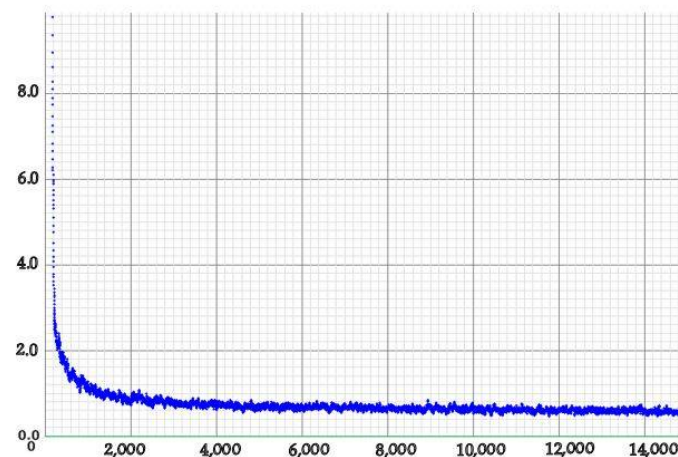

Figure 13. The average loss graph during training the network.

4.2. Evaluation of the Proposed Method

In order to present and observe the functioning of the proposed method, in this study, the confusion matrix representation is used. As presented in Figure 14, in this matrix, the columns represent the actual classes, and the rows represent the predicted classes. Based on Figure 14, it is obvious that in the proposed network, 83% of the samples that are originally taken from multirotors, are correctly detected as multirotor class. In the other two classes, the rate is 87 and 80 percent. It is also clear that the cells related to misdiagnoses have lower values in the network, and the cells related to correct diagnoses have higher values. For example, in the multirotor class, 10% of the multirotors were mistaken for a bird, and 7% were mistaken for a helicopter, while 83% of the multirotors were correctly diagnosed as multirotor. In the other two classes, it is the same, and the percentage of errors is less than the percentage of correct diagnoses.

Figure 14. The confusion matrix of the proposed method.

The proposed deep learning network is also accurately evaluated using confusion matrix, mAP, accuracy, precision, recall, and F1-score measures in the detection and recognition of the two types of drones and birds. Table 1 show the evaluation indices results of the proposed model. According to this table, the overall evaluation metrics of the model such as accuracy, mAP, and IoU reached 83%, 84%, and 81%, respectively, indicating the generalizability of the model and the possibility of a lower error rate in drone image detection and recognition of input drone images.

Table 1. Evaluation results of the models.

Dataset	Num of Images	Precision %	Recall %	F1-Score %	Accuracy %	mAP %	IoU %
Bird	1000	90	87	88	-	-	-
Helicopter	1000	86	80	83	-	-	-
Multirotor	1000	76	83	79	-	-	-
Total	3000	-	-	-	83	84	81

Evaluation metrics such as precision, recall, and F1 score are displayed in the three classes of bird, multirotor, and helicopter (Figure 15). As it appears from this figure, these evaluation matrics reached high values in the precision, recall, and F1-score.

Figure 16 illustrates some samples of the obtained results related to the detection and recognition of two types of drones and their capability in distinguishing them from birds in the proposed network. As it is apparent, the detection and recognition of drones and birds with bounding boxes and class probabilities is displayed.
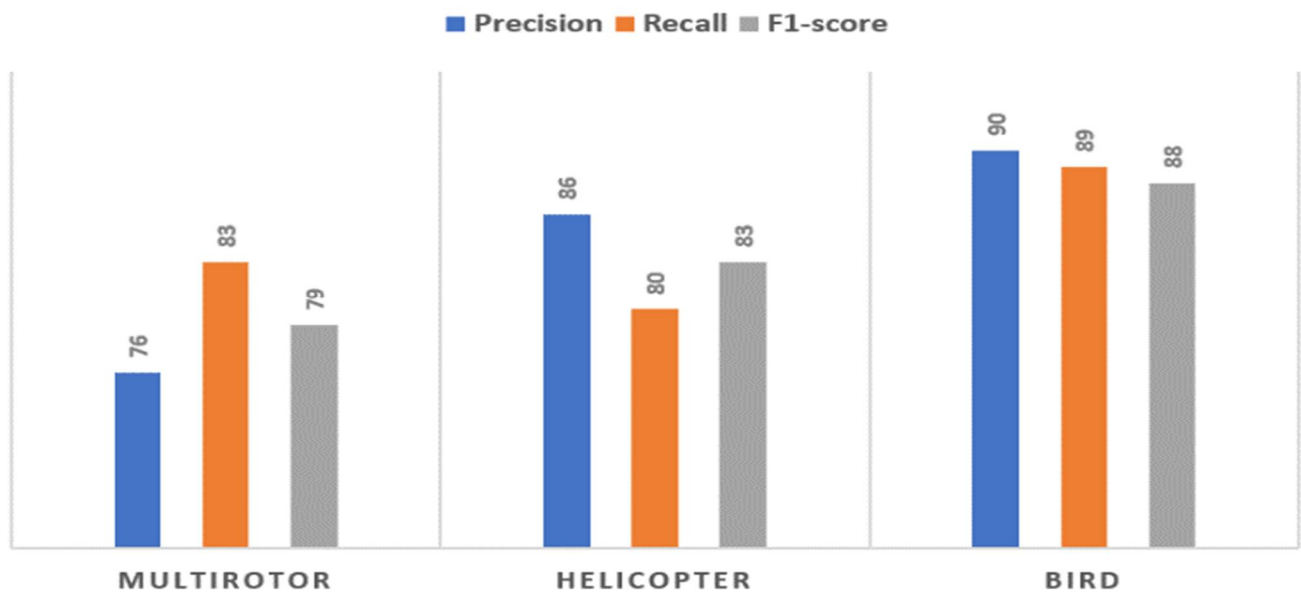
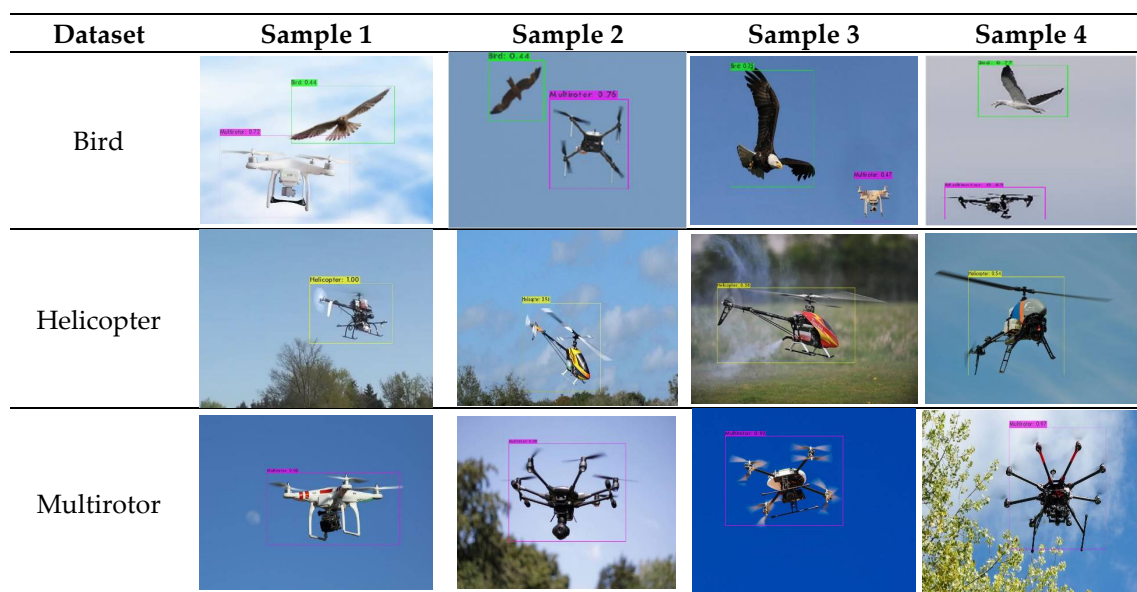**Figure 15.** Evaluation metrics of the proposed CNN network.

Dataset	Sample 1	Sample 2	Sample 3	Sample 4
Bird				
Helicopter				
Multirotor				

Figure 16. Some samples of detection and recognition results of the proposed network.

4.3. Model Evaluation in Addressing the Challenges

Drone detection and recognition always face challenges such as the inability to isolate the background, crowded backgrounds, lighting issues within the image, and the presence of occluded areas. On the other hand, the small size of the drone and its far distance caused it to be confused with the bird and reduced the accuracy of the diagnosis. The proposed convolutional neural network can overcome a variety of challenges in drone detection and recognition, such as multirotors, helicopters, and distinguishing between birds and drones even at longer ranges. As it appears from Figure 17, small drones are detected using the network in a variety of images with different lighting conditions and crowded backgrounds. In these images, drones and birds with a minimum dimension of 15×30 and a maximum dimension of 600×600 are detected and recognized.
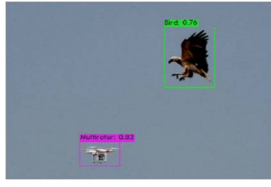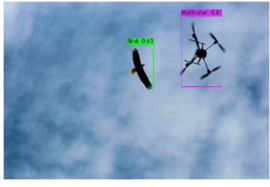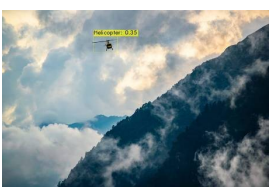
Challenge	Sample 1	Sample 2	Sample 3
(a) Confusion with bird			
(b) Crowded backgrounds			
(c) Small drone size			
(d) Lack of scalability			

Figure 17. Some samples of solving a variety of challenges in drone detection.

Some samples of drone detection and its distinction from birds in the model are presented in Figure 17. As it is apparent in the figure, the proposed model has the ability to distinguish birds and drones from each other and solve these challenges. In addition, some samples of drone detection in crowded background environments are also illustrated in this figure. This model is able to detect drones in these images. Furthermore, the third row of this figure shows the ability to detect and recognize different types of drones at longer distances. Considering the accuracy, it can be said that the implemented network is able to detect different types of drones with higher accuracy. In the last row, samples with different dimensions are detected, and higher accuracy is achieved.

Figure 18 illustrates some samples of more complex and challenging images of different drone sizes in different weather and light conditions and complex backgrounds. Based on this figure, it can be said that the network in question has the ability to detect and recognize drones in these images.

Figure 18. Some samples of small drone detection in crowded background, different weather, and lighting conditions.

5. Discussion

As presented in the evaluation section, the proposed model uses evaluation metrics such as confusion matrix, IoU, mAP, accuracy, precision, recall, and F1-score. The use of the mAP metric in this study was to determine the mean average precision of a set of diagnoses in the proposed model, reaching 84%, showing the overall performance of the proposed model in three classes. The accuracy criterion was checked to determine the correct classification of the input data into three classes and also showed the robustness and generalizability of the implemented model. In this study, we achieved an accuracy of 83%, indicating a high error of the system in classification. To determine the overlap of the predicted bounding box in the model, the IoU metric was checked against the ground truth bounding box, which reached a value of 81%, indicating that 81% of the predicted bounding boxes overlap with the ground truth bounding boxes, which is an acceptable value. In order to accurately evaluate the performance of the model, the metrics of precision, recall, and F1 score in three classes were calculated separately. The results of the model in three

separate classes are as follows: (76% precision, 83% recall, 79% F1-score) for multirotor, (86% precision, 80% recall, 83% F1-score) for helicopter, and (90% precision, 87% recall, 88% F1-score) for birds. According to the results, these evaluation criteria have desirable values in all three classes separately, which according to their definitions, indicate the proper performance of the model in all three classes separately, and it is necessary to examine them in each class.

In recent studies, deep learning methods have been used to detect and recognize drones. In 2021, Xun et al., the drone was detected using a set of visible images and the YOLOv3 deep learning network method [51]. This year, Isaac-Medina et al. detected drones using SSD, DETR, YOLOv3, and Faster RCNN in visible imagery [22]. One of the limitations of these studies is the inability to detect small objects and the inability to detect drones at long distances. Finally, Liu et al. detected drones using pruned YOLOv4, RetinaNet, FCOS, and YOLOv3 deep learning networks in video and image datasets [36]. This study improved the challenges related to small drone detection but did not address the challenges related to crowded backgrounds and the similarity between drones and birds. In addition, the drone detection problem was solved in a single class, and detection was not discussed in any of the research.

In this paper, the YOLOv4 deep learning network was used to detect and recognize target objects, which has high accuracy in long-distance small drone detection. In addition, challenges related to drone detection and recognition in environments with crowded backgrounds, hidden areas, and issues such as confusing drones with birds in visible imagery were addressed. No studies have been conducted to detect and recognize two types of drones (multirotors, helicopters).

6. Conclusions

Due to the emerging and development of the application of drones and the security threats associated with their presence in sensitive locations such as airports, drone detection and recognition has attracted much attention. Due to similar behavior and appearance of drones and birds in the sky, as well as their high speed and problems such as crowded backgrounds, the presence of hidden areas, lighting problems in the images, and the small size of drones at long distances, this paper proposes a new deep learning-based method for detecting and recognizing drones and birds to solve the problems caused by their unauthorized existence.

In this study, two types of drones and birds were extracted from videos and images. A collection of 10,000 visible images was collected. The training, testing, and evaluation of the model were performed on the collected dataset. Moreover, using the Convolutional Deep Learning Network and Nvidia Geforce MX450 Graphics Processing Unit (GPU), scores of 84% mAPs, 81% IoU, and 83% accuracy were achieved, which solved the challenges well. Future work will use other deep learning networks to compare their performance in drone-vs-bird detection, and identification will be performed in addition to detection and recognition. In addition to multi-rotors and helicopters, we also aim detect and recognize other types of drones, such as fixed-wing and VTOL. Drone detection, recognition, and localization can be performed in real-time and on onboard systems.

Author Contributions: All authors contributed to the study conception and design. F.S. contributed to supervision, reviewing, editing, and validation. F.D.J. involved in conceptualization, methodology, software, visualization, and editing the draft. F.A.M. contributed to programing, software, writing and editing. M.G. involved in software and data collection and preparation and drafting. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The public part of the database can be accessible through: https://www.kaggle.com/dasmehdixtr/drone-dataset-uav, accessed on 1 January 2021. https://www.kaggle.com/kmader/drone-videos, accessed on 1 January 2021.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shi, X.; Weige, X.; Yang, C.; Shi, Z.; Chen, J. Synthesis: Anti-Drone System with Multiple Surveillance Technologies: Architecture, Implementation, and Challenges. *IEEE Commun. Mag.* **2018**, *56*, 68–74. [CrossRef]
2. Anwar, M.; Kaleem, Z.; Jamalipour, A. Machine Learning Inspired Sound-Based Amateur Drone Detection for Public Safety Applications. *IEEE Trans. Veh. Technol.* **2019**, *68*, 2526–2534. [CrossRef]
3. Sathyamoorthy, D. A Review of Security Threats of Unmanned Aerial Vehicles and Mitigation Steps. *J. Def. Secur.* **2015**, *6*, 81–97.
4. Zwęgliński, T. The Use of Drones in Disaster Aerial Needs Reconnaissance and Damage Assessment–Three-Dimensional Modeling and Orthophoto Map Study. *Sustainability* **2020**, *12*, 6080. [CrossRef]
5. Hayeri Khyavi, M. Rescue Network: Using UAVs (Drones) in Earthquake Crisis Management. *arXiv* **2021**, arXiv:2105.07172.
6. Gomez, C.; Purdie, H. UAV-based Photogrammetry and Geocomputing for Hazards and Disaster Risk Monitoring—A Review. *Geoenviron. Disasters* **2016**, *3*, 23. [CrossRef]
7. Yaacoub, J.-P.; Noura, H.; Salman, O.; Chehab, A. Security analysis of drones systems: Attacks, limitations, and recommendations. *Internet Things* **2020**, *11*, 100218. [CrossRef]
8. Solodov, A.A.; Williams, A.D.; Al Hanaei, S.; Goddard, B. Analyzing the threat of unmanned aerial vehicles (UAV) to nuclear facilities. *Secur. J.* **2018**, *31*, 305–324. [CrossRef]
9. Pyrgies, J. The UAVs threat to airport security: Risk analysis and mitigation. *J. Airl. Airpt. Manag.* **2019**, *9*, 63. [CrossRef]
10. Shvetsova, S.V.; Shvetsov, A.V. Ensuring safety and security in employing drones at airports. *J. Transp. Secur.* **2021**, *14*, 41–53. [CrossRef]
11. Park, S.; Kim, H.; Lee, S.; Joo, H.; Kima, H. Survey on Anti-Drone Systems: Components, Designs, and Challenges. *IEEE Access* **2021**, *9*, 42635–42659. [CrossRef]
12. Drozdowicz, J.; Wielgo, M.; Samczynski, P.; Kulpa, K.; Krzonkalla, J.; Mordzonek, M.; Bryl, M.; Jakielaszek, Z. 35 GHz FMCW drone detection system. In Proceedings of the 2016 17th International Radar Symposium (IRS), Krakow, Poland, 10–12 May 2016; pp. 1–4.
13. Semkin, V.; Yin, M.; Hu, Y.; Mezzavilla, M.; Rangan, S. Drone Detection and Classification Based on Radar Cross Section Signatures. In Proceedings of the 2020 International Symposium on Antennas and Propagation (ISAP), Osaka, Japan, 25–28 January 2021; pp. 223–224.
14. de Haag, M.U.; Bartone, C.G.; Braasch, M.S. Flight-test evaluation of small form-factor LiDAR and radar sensors for sUAS detect-and-avoid applications. In Proceedings of the 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), Sacramento, CA, USA, 25–29 September 2016; pp. 1–11.
15. Nguyen, P.; Ravindranatha, M.; Nguyen, A.; Han, R.; Vu, T. Investigating cost-effective rf-based detection of drones. In Proceedings of the 2nd Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use, Singapore, 26 June 2016; pp. 17–22.
16. Basak, S.; Rajendran, S.; Pollin, S.; Scheers, B. Combined RF-based drone detection and classification. *IEEE Trans. Cogn. Commun. Netw.* **2021**, 1. [CrossRef]
17. Mezei, J.; Fiaska, V.; Molnár, A. Drone sound detection. In Proceedings of the 2015 16th IEEE International Symposium on Computational Intelligence and Informatics (CINTI), Budapest, Hungary, 19–21 November 2015; pp. 333–338.
18. Svanström, F.; Englund, C.; Alonso-Fernandez, F. Real-Time Drone Detection and Tracking With Visible, Thermal and Acoustic Sensors. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 7265–7272.
19. Andraši, P.; Radišić, T.; Muštra, M.; Ivošević, J. Night-time detection of uavs using thermal infrared camera. *Transp. Res. Procedia* **2017**, *28*, 183–190. [CrossRef]
20. Seidaliyeva, U.; Alduraibi, M.; Ilipbayeva, L.; Almagambetov, A. Detection of loaded and unloaded UAV using deep neural network. In Proceedings of the 2020 Fourth IEEE International Conference on Robotic Computing (IRC), Taichung, Taiwan, 9–11 November 2020; pp. 490–494.
21. Behera, D.K.; Raj, A.B. Drone Detection and Classification using Deep Learning. In Proceedings of the 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 13–15 May 2020; pp. 1012–1016.
22. Isaac-Medina, B.K.; Poyser, M.; Organisciak, D.; Willcocks, C.G.; Breckon, T.P.; Shum, H.P. Unmanned aerial vehicle visual detection and tracking using deep neural networks: A performance benchmark. *arXiv* **2021**, arXiv:2103.13933.
23. Liu, H.; Qu, F.; Liu, Y.; Zhao, W.; Chen, Y. A drone detection with aircraft classification based on a camera array. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, *322*, 052005. [CrossRef]
24. Ye, D.H.; Li, J.; Chen, Q.; Wachs, J.; Bouman, C. Deep learning for moving object detection and tracking from a single camera in unmanned aerial vehicles (UAVs). *Electron. Imaging* **2018**, *2018*, 4661–4666. [CrossRef]
25. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef]

26. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Processing Syst.* **2012**, *25*, 1097–1105. [CrossRef]

27. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 21–23 June 1994; pp. 580–587.

28. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef] [PubMed]

29. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Processing Syst.* **2015**, *28*, 91–99. [CrossRef]

30. Rozantsev, A.; Lepetit, V.; Fua, P. Flying objects detection from a single moving camera. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 21–23 June 1994; pp. 4128–4136.

31. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 21–23 June 1994; pp. 779–788.

32. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 21–37.

33. Zhao, Z.-Q.; Zheng, P.; Xu, S.-T.; Wu, X. Object detection with deep learning: A review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232. [CrossRef] [PubMed]

34. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.

35. Blake, W.; Burger, I. Small Drone Detection Using Airborne Weather Radar. In Proceedings of the 2021 IEEE Radar Conference (RadarConf21), Atlanta, GA, USA, 7–14 May 2021; pp. 1–4.

36. Liu, H.; Fan, K.; Ouyang, Q.; Li, N. Real-Time Small Drones Detection Based on Pruned YOLOv4. *Sensors* **2021**, *21*, 3374. [CrossRef] [PubMed]

37. De Quevedo, Á.D.; Urzaiz, F.I.; Menoyo, J.G.; López, A.A. Drone Detection with X-Band Ubiquitous Radar. In Proceedings of the 2018 19th International Radar Symposium (IRS), Bonn, Germany, 20–22 June 2018; pp. 1–10.

38. Gao, Q.; Parslow, A.; Tan, M. Object motion detection based on perceptual edge tracking. In Proceedings of the Second International Workshop on Digital and Computational Video, Tampa, FL, USA, 8–9 February 2001; pp. 78–85.

39. Lai, J.; Mejias, L.; Ford, J. Airborne Vision-Based Collision-Detection System. *J. Field Robot.* **2011**, *28*, 137–157. [CrossRef]

40. Ganti, S.R.; Kim, Y. Implementation of detection and tracking mechanism for small UAS. In Proceedings of the 2016 International Conference on Unmanned Aircraft Systems (ICUAS), Arlington, VA, USA, 7–10 June 2016; pp. 1254–1260.

41. Li, J.; Ye, D.H.; Chung, T.; Kolsch, M.; Wachs, J.; Bouman, C. Multi-target detection and tracking from a single camera in Unmanned Aerial Vehicles (UAVs). In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016; pp. 4992–4997.

42. Wu, Y.; Sui, Y.; Wang, G. Vision-Based Real-Time Aerial Object Localization and Tracking for UAV Sensing System. *IEEE Access* **2017**, *5*, 23969–23978. [CrossRef]

43. Schumann, A.; Sommer, L.; Klatte, J.; Schuchert, T.; Beyerer, J. Deep cross-domain flying object classification for robust UAV detection. In Proceedings of the 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, Italy, 29 August–1 September 2017; pp. 1–6.

44. Saqib, M.; Khan, S.D.; Sharma, N.; Blumenstein, M. A study on detecting drones using deep convolutional neural networks. In Proceedings of the 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, Italy, 29 August–1 September 2017; pp. 1–5.

45. Aker, C.; Kalkan, S. Using deep networks for drone detection. In Proceedings of the 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, Italy, 29 August–1 September 2017; pp. 1–6.

46. Suresh Arunachalam, T.; Shahana, R.; Vijayasri, R.; Kavitha, T. Flying Object Detection and Classification using Deep Neural Networks. *Int. J. Adv. Eng. Res. Sci.* **2019**, *6*, 180–183. [CrossRef]

47. Nalamati, M.; Kapoor, A.; Saqib, M.; Sharma, N.; Blumenstein, M. Drone Detection in Long-Range Surveillance Videos. In Proceedings of the 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Taipei, Taiwan, 18–21 September 2019; pp. 1–6.

48. Shi, Q.; Li, J. Objects Detection of UAV for Anti-UAV Based on YOLOv4. In Proceedings of the 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Weihai, China, 14–16 October 2020; pp. 1048–1052.

49. Kavitha, T.; Lakshmi, K. Evaluation of the Performance of Tiny YOLOv3 based Drone Detection System with Different Drone Datasets. *J. Crit. Rev.* **2020**, *7*, 835–848. [CrossRef]

50. Ashraf, M.W.; Sultani, W.; Shah, M. Dogfight: Detecting Drones from Drones Videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 7067–7076.

51. Xun, D.T.W.; Lim, Y.L.; Srigrarom, S. Drone detection using YOLOv3 with transfer learning on NVIDIA Jetson TX2. In Proceedings of the 2021 Second International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP), Bangkok, Thailand, 20–22 January 2021; pp. 1–6.

52. Singha, S.; Aydin, B. Automated Drone Detection Using YOLOv4. *Drones* **2021**, *5*, 95. [CrossRef]

53. Coluccia, A.; Fascista, A.; Schumann, A.; Sommer, L.; Dimou, A.; Zarpalas, D.; Méndez, M.; De la Iglesia, D.; González, I.; Mercier, J.-P. Drone vs. Bird Detection: Deep Learning Algorithms and Results from a Grand Challenge. *Sensors* **2021**, *21*, 2824. [CrossRef]

54. Wang, C.-Y.; Liao, H.-Y.M.; Yeh, I.-H.; Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W. CSPNet: A New Backbone that can Enhance Learning Capability of CNN. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 1571–1580.

55. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–28 June 2018; pp. 8759–8768.

56. Yun, S.; Han, D.; Oh, S.J.; Chun, S.; Choe, J.; Yoo, Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 6023–6032.

57. Ghiasi, G.; Lin, T.-Y.; Le, Q.V. Dropblock: A regularization method for convolutional networks. *arXiv* **2018**, arXiv:1810.12890.

58. De Boer, P.-T.; Kroese, D.P.; Mannor, S.; Rubinstein, R.Y. A tutorial on the cross-entropy method. *Ann. Oper. Res.* **2005**, *134*, 19–67. [CrossRef]

59. Susmaga, R. Confusion matrix visualization. In *Intelligent Information Processing and Web Mining*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 107–116.

60. Redmon, J. Darknet: Open Source Neural Networks in C. 2013–2016. Available online: http://pjreddie.com/darknet/ (accessed on 1 January 2021).