UAV ADVANCED YOLOV5 MODEL COMPARING WITH YOLOV4 USING FORMULATE DEEP LEARNING TECHNIQUES TO OBJECT DETECTION AND RECOGNITION

Dr. PL. Chithra[1]*, Ms. J. Kavila[2]

Abstract:
Nowadays Drones are used in various places, which increases the advancement in the innovation of drones. According to that drones were modified with various design segments. In general, it is used for security and surveillance. In this article, modified YOLOv5 network with an additional layer neck -PANet is presented. It is presented to recognize four types of drones (multirotor, fixed-wing, helicopters, and Vertical Take-Off and Landing (VTOLs) to differentiate them from birds with a set of 1000 visible images. In this network, more effective and detailed semantic features were extracted by changing the number of convolutional multi-layers. The performance of the basic YOLOv5 network was also evaluated on the same dataset. The proposed model performs well 5% with the existing art of YOLOv3, YOLOv4 basic models. In YOLOv5 will be exponentially more comfortable for object detection.

Keywords: convolutional neural network CNN; drone detection, drone recognition, drone image dataset; YOLOV4, YOLOv5 deep learning;

[1]*Professor, Department of Computer Science, University of Madras, Chennai – 600025,
E-mail: chitrasp2001@yahoo.com
[2]Research Scholar, Department of Computer Science, University of Madras, Chennai Chennai- 600 025,
E-mail: kavija40@gmail.com

***Correspondence Author: -** Dr. PL. Chithra
*Professor, Department of Computer Science, University of Madras, Chennai – 600025,
E-mail: chitrasp2001@yahoo.com

DOI: - 10.48047/ecb/2023.12.si5a.0155

1. Introduction:

A drone refers to any aerial vehicle that receives remote commands from a pilot or relies on software for autonomous flight. Many drones display features like cameras for collecting visual data and propellers for stabilizing their flight patterns. Drone technology which includes videography[9], Search and rescue, agriculture, and transportation. Originally developed for the military and aerospace industries. Nowadays, drones have found their way into the mainstream because of their enhanced safety and efficiency. These robotic UAVs operate without a pilot on board with different levels of autonomy. Drone detection technologies include the use of various sensors such as radar (radio detection and ranging[9]), Lidar (Light Detection and Ranging[10]), acoustic[11], and thermal sensors[12]. In these methods, first, the drone's presence or absence in the scene is checked and then the drone type recognition process is performed. The application of these types of sensors has always been associated with problems such as higher costs and higher energy consumption[15]. However, the negligent and the malicious use of these flying vehicles poses a great threat to public safety in sensitive areas such as government buildings, power plants, and refineries. Recent advances in deep convolutional neural networks and the appearance of more improved hardware make it possible to use visual information to recognize objects with higher accuracy and speed[17]. Unlike conventional drone detection technologies, the nature of deep learning networks is to perform drone recognition simultaneously. In this work, mAP Value calculates, increases by 0.5% upto their progressive manner.

An Overview of YOLO Training Procedures:
The procedures taken to train a model are just as important as any factor to the end performance of an object detection system, although they are often less discussed. Let's talk about two main training procedures in YOLOv5:

- **Data Augmentation**: Data augmentation makes transformations to the base training data to expose the model to a wider range of semantic variation than the training set in isolation.
- **Loss Calculations**: YOLO [19] calculates a total loss function from the GIoU, obj, and class losses functions. These functions can be carefully constructed to maximize the objective of mean average precision.

1.1. Challenges in Drone Recognition:

Drone recognition is always very difficult task. Some of the important challenges in this regard are discussed.

1.2. Confusion of Drones and Birds

Due to the physical characteristics of drones, they can easily be confused with birds in human eyes. This problem is more challenging when using drones in maritime areas, forest areas due to the presence of more birds. The similarity between drones and birds and their distinction from each other is shown in Figure 1.

Fig 1. Challenges related to confusion with birds in drone recognition.

1.3. Challenges related to the presence of small drones at different scales.

Drone recognition is always fraught with challenges. For this reason, it is necessary to use a fast, accurate, robust, efficient method to overcome the challenges and to correctly recognize drones[1]. However, the challenges of crowded backgrounds, hidden areas, and surveys of multiple drone types have not yet been addressed [16]. In 2022, Samadzadegan et al. detected and recognized drones using YOLOv4 Deep Networks.

2. Material and Methods:

Due to the Challenges in drone detection and recognition such as crowded background, a close resemblance to birds, smallest size of drones[2], longer distance, crowded background and lighting problems in the image, in this study, a deep learning-based method is proposed.

CNN Architecture:
The CNN architecture is complicated when compared to the MLP architecture. There are different types of additional layers and operations in the CNN architecture[. It is very easy for working image formation, filters and so on.

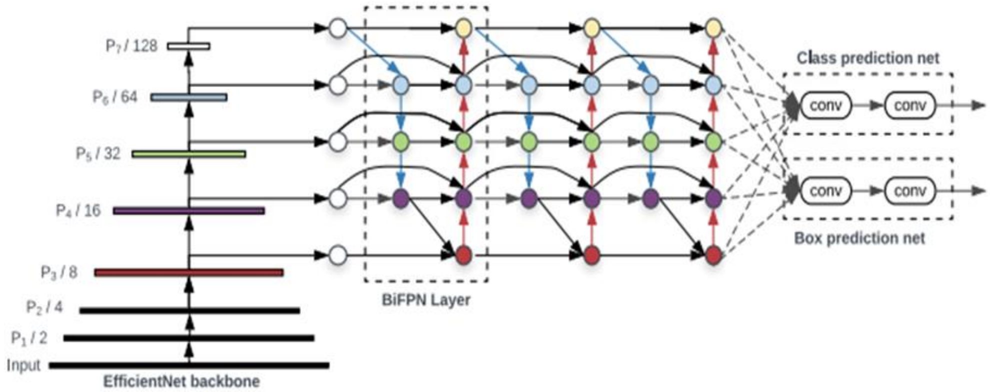


Figure 2: CNN Architecture

The proposed system facilitates path following based on visual information. A drone flies on predetermined path waypoints while recording the current front view and yaw of the drone[7]. To construct a robust system is controlled this drone to generate more data. Secondly, the recorded yaw is transfomed to relative yaw commands. Finally, train a neural networks combined of a CNN with a regressor in order to control the drone of the path using only visual information[6].

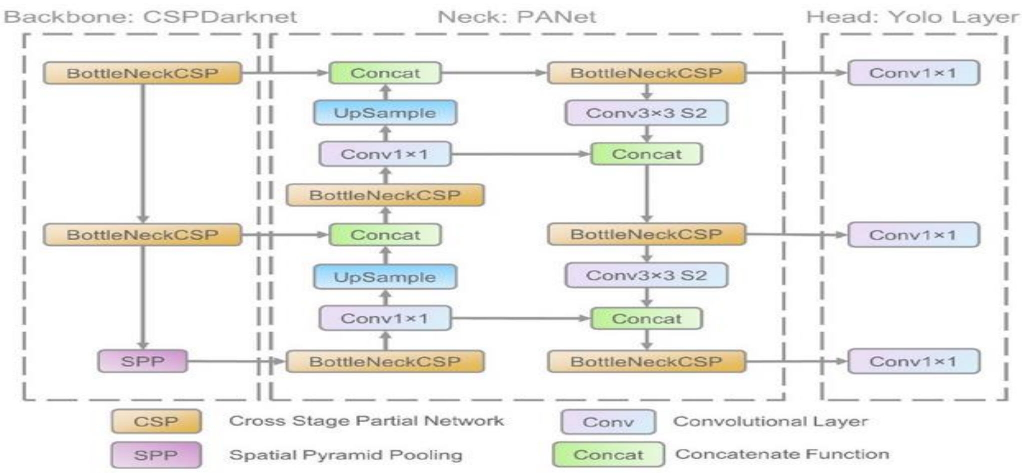


Fig. 3: YOLOv5 architecture.

High-level architecture for single-stage object detectors

There are two types of object detection models: two-stage object detectors and single-stage object detectors[12]. Single-stage object detectors (like YOLO) architecture are composed of three components: **Backbone**, a **Neck** and a **Head** to make dense predictions as shown in the figure bellow.

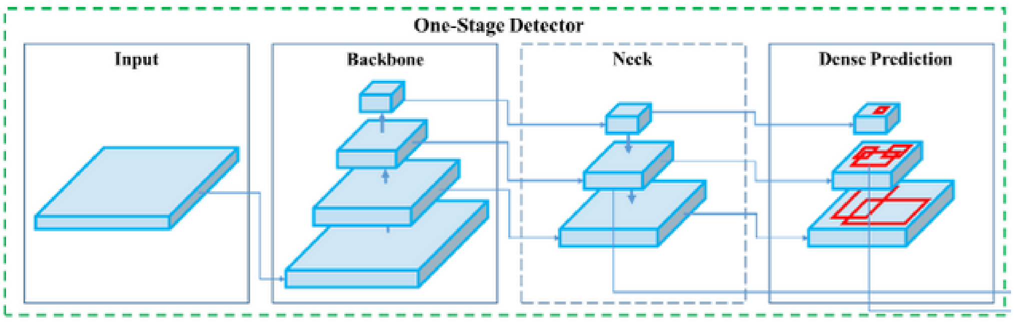


Fig. 4: Single-Stage Detector Architecture [1]

Model backbone: The backbone is a pre-trained network used to extract rich feature representation for images. This helps **reducing the spatial resolution** of the image and **increasing its feature** (channel) **resolution**.

Model Neck: The model neck is used to extract feature pyramids. This helps the model to generalize well to objects on different sizes and scales.

Model Head: The model head is used to perform the final stage operations. It applies anchor boxes on feature maps and renders the final output: **classes**, **objectness, scores** in addtionaly **bounding boxes**.

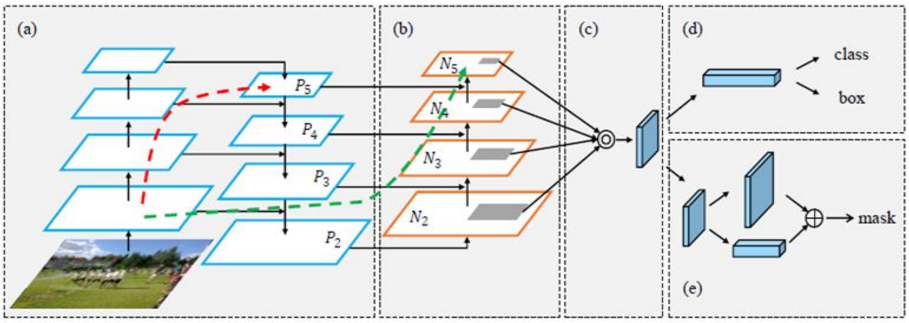


Fig. 5: Frame Work for PANet

YOLOv5 Architecture
It is also good to mention that YOLOv5 was released with five different sizes:

- **n** for extra small (nano) size model.
- **s** for small size model.
- **m** for medium size model.
- **l** for large size model

Neck of YOLOv5
YOLOv5 brought two major changes to the model neck. First a variant of **Spatial Pyramid Pooling** (SPP) has been used, and the **Path Aggregation Network** (PANet) has been modified by incorporating the **Bottle Neck CSP** in its architecture.

Path Aggregation Network (PANet)
PANet is a feature pyramid network, it has been used in previous version of YOLO (YOLOv4) to improve information flow and to help in the proper localization of pixels in the task of mask prediction. In YOLOv5 this network has been modified by applying the CSP Net strategy to it as shown in the network's architecture figure.

Spatial Pyramid Pooling (SPP)
SPP block [4] performs an aggregation of the information that receives from the inputs and returns a fixed length output. Thus it has the advantage of significantly increasing the receptive field and segregating the most relevant context features without lowering the speed of the network. This block has been used in previous versions of YOLO (yolov3 and yolov4) to separate the most important features from the backbone, however in YOLOv5(6.0/6.1) **SPPF** has been used , which is just another variant of the SPP block, to **improve the speed of the network**

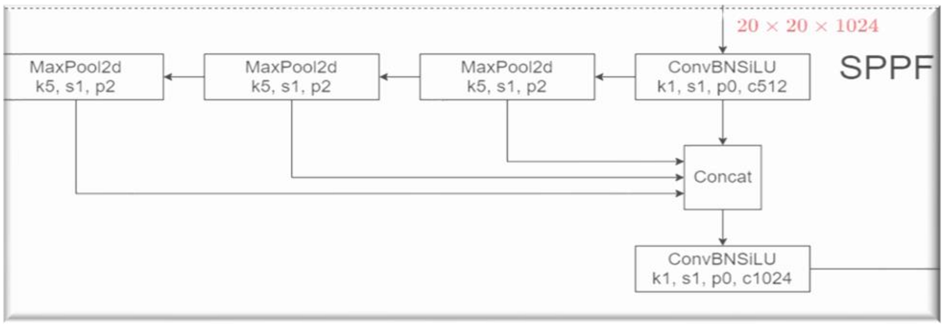
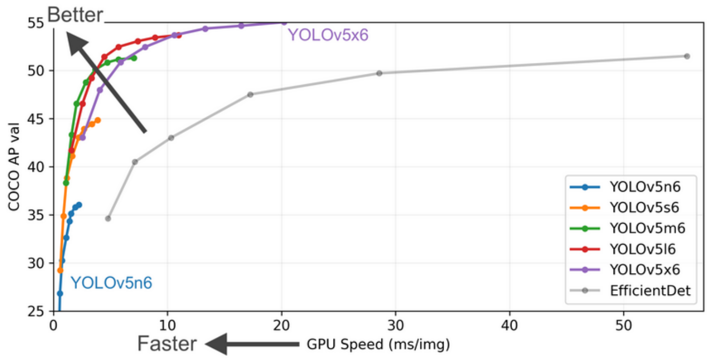


Fig. 6: Structure of the SPPF block

.

Head of the network
YOLOv5 uses the same head as YOLOv3 and YOLOv4. It is composed from three convolution layers that predicts the location of the bounding boxes (x, y, height, width), the scores and the objects classes.



Fig, 7: The YOLOv5 implementation has been done in Pytorch.

There is no difference between the five models in terms of operations used except for the number of layers and parameters.

3. Train the Networks: According to the reviewed advantages of the YOLOv5 deep learning network is new technology to detect flying drones and birds in crowded environment, The training, testing, evaluaton of the model were performed on the collected dataset. Moreover, using the Convolutional Deep Learning Network and Nvida Jetson Jetpack used. In this paper. In this paper, Graphics Processing Unit (GPU), scores of 99% mAP, 82.2%, Precision 55.5%, Accuracy and Recall 95% were achieved, which solved the challenges well.

Other improvements

In addition to what have been stated above, there are still some minor improvements that have been added to YOLOv5[19] and that are worth mentioning

The Focus Layer : replaced the three first layers of the network. It helped reducing the number of parameters, the number of FLOPS and the CUDA memory while improving the speed of the forward and backward passes with minor effects on the mAP (mean Average Precision).

Eliminating Grid Sensitivity: It was hard for the previous versions of YOLO versions to detect bounding boxes on image corners mainly due to the equations used to predict the bounding boxes, but the new equations presented above helped solving this problem by expanding the range of the center point offset from **(0-1) to (-0.5,1.5)** therefore the **offset can be easily 1 or 0** (coordinates can be in the image's edge) as shown in the image in the left. Also the **height and width** scaling ratios **were unbounded** in the previous equations which may lead to training instabilities but now this problem has been reduced as shown in the figure on the right.

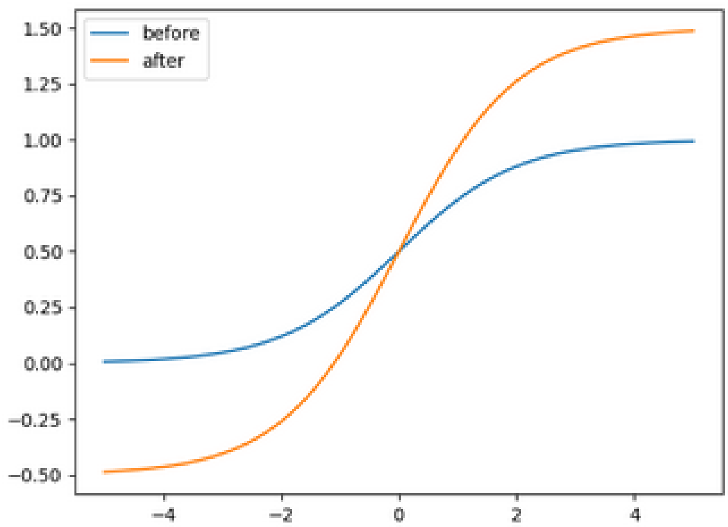


Fig. 8: Comparison of previous and advanced.

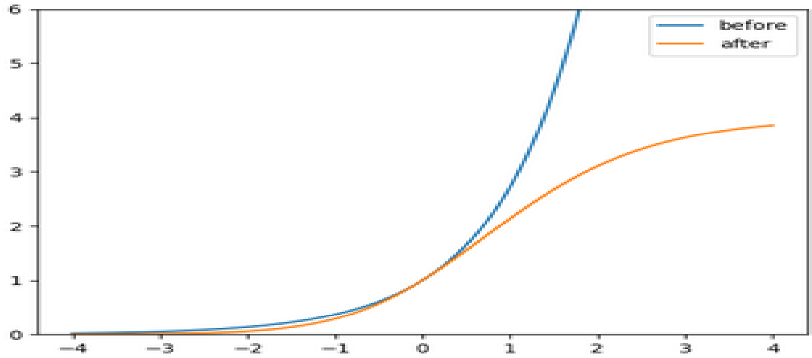


Fig. 9: Variation of 0.5 level

Eliminating grid sensitivity problem.

The running environment: The YOLOv5 is implemented in **Pytorch** giving more flexibility to control the encoded operations.

4. Results & Discussion:

Due to the emergence and development of drone application and the security threats associated with their presence in sensitive locations such as airports, drone detection and recognition have attracted much attention. This work going to develop our object detection with recognition in greater evaluation.

Dataset:

Table 1: Dataset Evaluation

Dataset	No. of images	Precision	Recall %	F1-Score%	Accuracy%	mAP%	IoU%
Birds	**1000**	**91**	**95**	**89**			
Drone	**1000**	**90**	**90**	**84**			
Person	**1000**	**80**	**93**	**80**			
Object	**1000**	**80**	**90**	**85**			
Total		**85**	**92**	**85**	**99**	**89(0.5)** Error rectified	**85**

Table 2: Analysis of data

No. of Set	Evaluation of images		
	Data analysis	**No. Of Images**	**Percentage**
Sample1	Training	31	74
	Validation	07	16
	Testing	04	10
	Total	42	100
Sample2	Training	30	71
	Validation	08	19
	Testing	04	10
	Total	42	100
Sample 3	Training	44	70
	Validation	12	20
	Testing	08	10
	Total	64	100
Sample4	Training	132	87
	Validation	12	08
	Testing	08	05
	Total	152	100

A Combination of persons, birds, objects from roboflow1.4 version have been tested with 64 images by 11 classes, 3 instances for the experimental implementation as sample 1, 2,3,4. In this Proposed system the latest stable version of NVIDIA's Jetson Jetpack used. 4 frames persecon on the Jetson Nano 2GB (with swap memory). 6 frames oer second on the Jetson Nano 4GB, 10 frames per second o the Jetson Xavier NX (single instance), 15 frames per second on the Jetson Xavier NX (2 instance cluster; see below). These results were obtained using while operating in a client-server context (so there is some minor network latency involved) and a 416x416 model.
https://universe.roboflow.com/object-detection/object-detection-smzhr
Provided by a Roboflow user License: CC BY 4.0
drones and birds detection > 2023-02-07 4:05pm
https://universe.roboflow.com/kavilajunom/drones-and-birds-detection
Provided by a Roboflow user License: CC BY 4.0
object detection - v1 2023-02-14 10:44pm

This dataset was exported via roboflow.com on February 14, 2023 at 5:23 PM GMT
* To export, train, and deploy computer vision models
* To use active learning to improve your dataset over time for state of the art Computer Vision training notebooks you can use with this dataset, visit https://github.com/roboflow/notebooks

To find over 100k other datasets and pre-trained models, visit https://universe.roboflow.com
The dataset includes 152 images.
Object- are annotated in YOLO v5 PyTorch format.
The following pre-processing was applied to each image:
* Auto-orientation of pixel data (with EXIF-orientation stripping)
* Resize to 640x640 (Stretch)

The following augmentation was applied to create 3 versions of each source image:
* 50% probability of horizontal flip

In graphical representation to justify the training, testing, valuation of the model. To describing the four categories of drone models were evaluated. In future of this work will be generate more exclusive YOLO version will be proceed to implement.

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[9]. 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. 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, 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.

Drone detection, recognition, and localization can be performed in real-time, onboard systems. In this terminology mAPs, 88% Precision 72% and accuracy 99% 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.

5. Conclusions

To wrap up what have been covered in this article, the key changes in YOLOv5 that didn't exist in previous version are: applying the CSP Net to the Darknet53 backbone, the integration of the **Focus layer** to the CSP-Darknet53 backbone, replacing the **SPP** block by the **SPPF** block in the model neck and **applying the CSP Net strategy on the PANet model**. YOLOv5 and YOLOv4 tackled the problem of grid sensitivity and **can now detect easily bounding boxes having center points in the edges**. Finally, YOLOv5 is **lighter and faster** than previous versions.

Due to the emergence and development of drone application and the security threats associated with their presence in sensitive locations such as airports, drone detection and recognition has attracted much attention. This work extracted two types of drones and birds from videos and images. The training, testing, and evaluation of the model were performed on the collected dataset. In addition to multi-rotors and helicopters, aim to detect and recognize other types of drones, such as fixed-wing and VTOL, in real-time and on-board systems.

Acknowledgment:

Our Sincere thanks to the University Of Madras. Department of Computer Science, Chennai, Tamil Nadu, India.

References :

1. Shi, X.; Weige, X.; Yang, C.; Shi, Z.; Chen, J. Synthesis: Anti-Drone https://www.kaggle.com/ kmader/drone-videos 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. Zw ˛egli ´nski, 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 recom-mendations. 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. [Cross Ref]
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. BOCHKOVSKIY, Alexey, WANG, Chien-Yao, et LIAO, Hong-Yuan Mark. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934, 2020.
12. XU, Renjie, LIN, Haifeng, LU, Kangjie, et al. A forest fire detection system based on ensemble learning. Forests, 2021, vol. 12, no 2, p. 217.
13. WANG, Chien-Yao, LIAO, Hong-Yuan Mark, WU, Yueh-Hua, et al. CSPNet: A new backbone that can enhance learning capability of CNN. In : Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020. p. 390-391.
14. HE, Kaiming, ZHANG, Xiangyu, REN, Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 2015, vol. 37, no 9, p. 1904-1916.
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. 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.
19. 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]
20. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. Nature 2015, 521, 436–444. [CrossRef] Aerospace 2022, 9, 31 19 of 20