ISSN: 2321-2152
IJMECE
International Journal of modern electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

Implementations of Deep Learning Techniques for Object Recognition in Continuous UAV Monitoring: A Comprehensive Analysis, Current Advancements, and Future Directions Issues

[1] Mr. P. S. S. K. Sarma, [2] Gannamani Saikumar,

[1]Associate Professor, Department of MCA, Rajamahendri Institute of Engineering & Technology. Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107

[2]Student, Department of MCA, RajamahendriInstitute of Engineering & Technology. Bhoopalapatnam, Near Pidimgoyyi,Rajahmundry,E.G.Dist.A.P. 533107.
.

Abstract:

For the purpose of analyzing photos taken by unmanned aerial vehicles (UAVs), deep learning (DL) has become an invaluable tool in remote sensing. Despite its widespread applicability and notable contributions, the purpose of this study is to provide a thorough overview of the topic. The research covers a wide range of recent methods and how they've been used to real-time drone surveillance object recognition.
Keywords: Unnamed Aerial Vehicles, Deep learning, One- stage detector, two-stage detector.

INTRODUCTION

The versatility of drones, also known as unmanned aerial vehicles (UAVs), has led to their deployment in a variety of fields where human access would be impractical or risky. UAVs have cameras that can take pictures or video from different angles and heights; they have many potential uses in fields as diverse as aerial photography, environmental monitoring, search and rescue, military, and defense. Due to the impossibility of manually following and collecting these pictures in real-time applications, automated systems that can process and analyze UAV-captured photographs are built using machine learning methods. While it is useful for tasks like mapping and surveying, it is not integral to the actual process of capturing images. The photos may be wirelessly sent to a ground station in real-time or saved on the UAV's internal memory for future use. Figure 1 shows the fundamental design of the drone surveillance system. The employment of an onboard camera is a common method for capturing images in drone surveillance. The kind of sensor used by the camera—a regular RGB camera, thermal or multispectral sensors, etc.—depends on the intended usage. While the drone is in the air, its camera records footage or stills of the landscape below. The drone will either immediately send the footage or stills to a base station or keep them for further processing. It was possible to maximize picture quality and gather certain sorts of information by adjusting the features and settings of the drone cameras. As an example, the camera at V. Ilango's Department of Computer Applications, CMR Institute of Technology, Bengaluru, may allow you to control the focus, zoom, and exposure in addition to having many modes for taking still photographs and video. The drone might have a camera on board, but it could also include other sensors and technologies like GPS and LIDAR that would aid in navigating and mapping the area being monitored. We will either keep the photographs in batches or monitor them in real time when they are taken from many regions. It is via a procedure known as object detection that certain items may be tracked. When using a drone for surveillance purposes, object detection is finding and locating certain things in the footage or stills taken by the aircraft. Agriculture, environmental monitoring, SAR, and military activities are just a few of the numerous fields that have grown to rely on drone observation. One reason drones have become so popular for surveillance is their capacity to swiftly and effectively provide overhead images of vast regions. On the other hand, quick and accurate object identification algorithms are crucial to drone surveillance's efficacy. Compared to fixed cameras, drone surveillance presents a number of challenges due to the following factors, including the fact that these UAVs are often flown at great heights: Perspective distortion, shadows, and reflections are all challenges inherent to aerial photography. Outside of any kind of regulation: Obstacles like the elements, the lighting, and the gradual transformation of the surrounding environment, Finding and following items in motion, whether they're traveling at a fast pace or undergoing a sudden change in direction, Complex and accurate object identification algorithms may be out of reach on drones due to their restricted processing power and memory compared to other computer systems.

OBJECT DETECTION UAV OVERVIEW

An Object: What Is It? An object is a visual representation of an element that is taken from a UAV. The process of object detection begins with the identification and localization of certain items, such as crops, flowers, people, weapons, etc., in order to offer information about the object's position or condition. In a nutshell, it classifies the extracted element. An offshoot of ML, Deep Learning (DL) uses a hierarchical structure to solve problems in a way that is similar to the human brain.

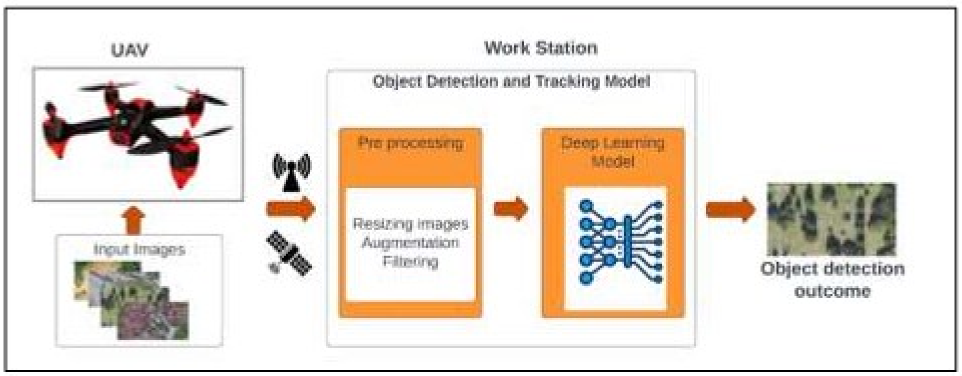

Figure 1 Sample Architecture for drone surveillance system complex and diverse range of applications.

Deep learning architectures are better at processing images and extracting features from complicated and big datasets because they use deeper combinations of input and hidden layers. Processing applications using drones, which deal with data that is often varied and difficult to handle manually, benefit greatly from its high processing capabilities. This might be one explanation for the widespread adoption of DL in data-driven and image-processing applications. A lot of testing and monitoring is still needed, even when DL gives encouraging findings. Various object detection applications that make use of Hyperspectral Imaging Sensors (HIS) to take high-resolution pictures have recently come to the forefront of this field, drawing interest in the ability to learn more about the physical and chemical characteristics of objects and landscapes through these images (Petersson et al., 2017; Signoroni et al., 2019).

RESEARCH MOTIVATION

Object identification using deep learning algorithms has great promise in many areas, such as computer vision, voice recognition, and NLP. Algorithms for deep learning can automatically sift through mountains of data in search of characteristics useful for object recognition. Because of this, deep learning is a potentially useful strategy for drone surveillance object identification. There are several obstacles that must be overcome before deep learning can reach its full potential. These include issues with data quality, scarce computer resources, and the need for strong algorithms. So, to assess the present state of the art, find research gaps, and suggest future research paths, it is essential to conduct a review of object identification in drone surveillance using deep learning. IV.

A PORTRAIT OF THE STUDY The goal of this literature review is to provide a synopsis of all the previous work on object recognition in drone surveillance using deep learning. The following are the main points that the review will cover: 1. The purpose of this study is to compare and contrast the object identification performance of different deep learning algorithms and architectures, as well as to highlight their advantages and disadvantages, possible uses, and obstacles. 2. Our goal in doing this study is to shed light on where things stand and to suggest ways forward for research into making object detection in drone surveillance more successful. 3. We strive to make accessible drone datasets that include the relevant information in order to encourage more study in this field.

THE METHODOLOGICAL FRAMEWORK FOR LITERATURE REVIEW

We anchored the whole literature research procedure on these framed questions: Question 1: How have the most recent and cutting-edge algorithms for drone surveillance object recognition, which are based on deep learning, changed over time? If we want our deep learning-based object identification algorithms to be the most effective for drone surveillance, how can we tweak their hyperparameters to get the best possible results? Question 3: In drone surveillance, how can we enhance the effectiveness of object identification algorithms based on deep learning by using transfer learning techniques? For drone surveillance, what are the difficulties in training object identification algorithms based on deep learning, and how have these difficulties been addressed in earlier research? Additionally, what are the necessary directions for the field to progress in the future? Section VI: Answers to My Research Questions Question 1: How have the most recent and cutting-edge algorithms for drone surveillance object recognition, which are based on deep learning, changed over time? Although object recognition often made use of more conventional computer vision methods like Haar cascades and HOG (histogram of oriented gradients) before 2014. On the other hand, deep convolutional neural networks (CNNs) like AlexNet, which took first place in the 2012 ImageNet Large Scale Visual Recognition Challenge, and other similar architectures began to replace traditional object recognition methods around 2014. Even though there was a lot of curiosity in using drones for

monitoring and several other uses Object identification using deep learning in drone surveillance was not yet common practice during that time. Deep learning algorithms like YOLO, SSD, and Faster R-CNN didn't gain traction for object identification in UAV surveillance and other uses until much later. Since then, deep learning algorithms have seen substantial field applications (Figure 2). A few of the most popular deep learning algorithms in the object identification domain include R-CNN, Faster R-CNN, YOLO, and SSD. These algorithms have shown state-of-the-art performance on object identification tasks, and they all rely on CNNs for feature extraction. New algorithms including CenterNet, Mask RCNN, M2Det, CPN, and FoveaBox were released in early 2018 and are gradually becoming popular among academics for their applications. The three main types of deep learning algorithms are the one-stage, two-stage, and advanced detector varieties. 1. Detectors for One Stage In a single neural network run, one-stage detectors—a subset of deep learning algorithms—can anticipate both the object's bounding boxes and its class probabilities. It begins by suggesting potential objects or areas of interest, and then it sorts and improves them. Redmon et al. (2016), Liu et al. (2016), and RetinaNet (Lin et al., 2020) are a few well-known examples of one-stage detectors. YOLO predicts the class probabilities and bounding boxes for each cell in the input picture by splitting it into a grid of cells. There is a confidence score that represents the likelihood that each anticipated bounding box includes an item. b. Twin Detector Stages Though they are very accurate and adaptable, two-stage detectors may be computationally costly and are very sensitive to the quality of candidate object suggestions; still, they constitute a strong family of object detection models. One of the most well-liked two-stage detector designs is the R-CNN family, which comprises the Fast, Faster, and Mask variants of the Region-based Convolutional Neural Network (R-CNN). A Region Proposal Network (RPN) is usually used to create potential item suggestions in these models, and then another network is used to categorize these suggestions. Feature Pyramid Network (FPN), Hybrid Task Cascade, and Cascade R-CNN are a few more well-known two-stage detectors. Proposal generation and classification are the two primary steps in the overall operation of two-stage detectors. First, given a picture as input, the model will provide a number of suggestions for potential objects to include. The usual tool for the job is a neural network called a Region Proposal Network (RPN), which can be trained with an image and then produces a collection of bounding boxes that could include objects. A feature map of the input picture is generally produced by the RPN's collection of convolutional layers. By dragging a tiny window across the feature map and adding a specified set of anchor boxes to each point, a collection of candidate object suggestions may be generated from this feature map. Second, the model determines whether each proposed item is in the forefront (containing an object) or the background (not showing an object). c. Sensitive Detectors. When compared to one-stage and two-stage detectors, advanced detectors outperform them in terms of efficiency, accuracy, or both. Detectors that have been enhanced include EfficientDet, CenterNet, YOLOv4, and DETR. With a fraction of the parameters and processing power required by earlier approaches, the EfficientDet family of object detectors developed by Google delivers state-of-the-art accuracy. To achieve the optimal balance between speed and accuracy, it employs a compound scaling method that adjusts the depth and size of the model. However, YOLOv4, the latest version of YOLO, has a number of improvements over its predecessor, such as a more robust Darknet backbone network, a novel data augmentation method called mosaic augmentation, and anchor boxes with varying sizes and aspect ratios. The three detectors are compared in detail in table 1.

Table 1. Comparison of different deep learning object detection techniques based on several performance constraints

Parameters	One Stage Detector	Two Stage Detector	Advanced Detector
Accuracy	Less	Medium	High
Speed	Faster	Slower	Faster
Model size	Smaller	Complex	Optimal
Data Volume	smaller datasets	Require large dataset	Versatile
Object size and shape	small objects	complex object shapes	multi-scale feature fusion
Training time	Less	Longer	Less

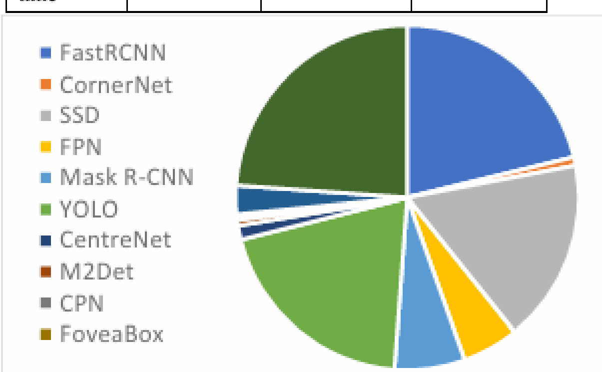

Figure 2 Relative percentage of different deep learning papers published in the UAV domain

When it comes to drone surveillance, how can we get the most out of object identification algorithms that use deep learning? Specifically, how can we optimize their hyperparameters? Due to the time-consuming nature of training deep neural networks, optimization is an essential part of deep learning. Researchers in various fields have developed optimizers for use in deep learning. Some examples include the stochastic gradient descent deep learning optimizer, the Adagrad, the mini batch stochastic gradient descent optimizers, RMSProp, and others (Cui et al., 2018; Shallue et al., 2018; Zhang et al., 2019; Xu et al., 2021). Many methods exist for optimizing object identification models during runtime, including data augmentation, normalization, transfer learning, neural network learning rate adjustment, feature pyramid networks, non-maximum suppression, and neural network training rate optimization. By introducing controlled variances to the training data, data augmentation serves as a kind of regularization. Overfitting occurs when a model becomes too dependent on its training data and loses its ability to generalize to new data. Regularization approaches assist avoid this problem. Data augmentation helps the model acquire more robust and generalizable features by giving it different instances. This, in turn, improves its object detection accuracy on unseen data and reduces overfitting. (Gorshick and colleagues, 2014) the article suggested a two-stage method for object identification utilizing the R-CNN architecture, first by creating area suggestions and then by categorizing these suggestions using a CNN. Although the authors did not use the phrase "data augmentation," they did use a method of data augmentation when training their model by randomly resizing and flipping the input pictures horizontally. The model's robustness and generalizability were both enhanced by this method. Similarly, it has been reported in several papers that data augmentation approaches were used during preprocessing to improve the object identification performance of DL models (Ottoni et al., 2023; Ruiz-Ponce et al., 2023). Improving convergence is another way to optimize deep learning models (Zhang et al., 2019). To speed up the training of the CNN model, (Ioffe & Szegedy, 2015) included batch normalization methods into the model architecture. These approaches acted as a regularizer. When normalization was used, dropout was no longer necessary, and the same accuracy could be achieved in 14 fewer cycles. To improve the efficiency of image recognition models, a preprocessing model was suggested by Koo and Cha (2017). This approach would use CNN classifiers to extract features and normalize them. The normalized picture is recognized using a fine-tuned CaffeNet model. With the use of a size-normalized picture, the CNN model was able to improve its performance from an average of 93.24% to 96.85%. When it

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

comes to improving deep learning models for object identification, one more optimization strategy that has been successful is the transfer learning methodology (Aytar, 2014). Transfer learning does this by transferring learned information from one job to another by making use of pre-trained models on massive datasets. The focus of (Chamarty, 2020) was on optimizing CNN's learning rate to get the highest possible detection accuracy. A link between learning rate and dataset size in the range of 10^-4 to 10^-5 was successfully achieved in the article. (Na, 2022) also makes use of a similar strategy, namely a learning rate optimization that modifies the learning rate by modifying the direction method of multipliers. In comparison to previous adaptive gradient approaches, the suggested methods of learning rate adjustment performed better. Objects of varying sizes may be efficiently processed using Feature Pyramid Networks (Yang et al., 2022). Object identification, instance segmentation, semantic segmentation, and Non Maximum Suppression are just a few of the tasks that may be enhanced by FPN's ability to identify and recognize objects of different sizes by incorporating multi-scale information in a feature pyramid (Song et al., 2019). The domain's optimization strategies are shown in Figure 3.

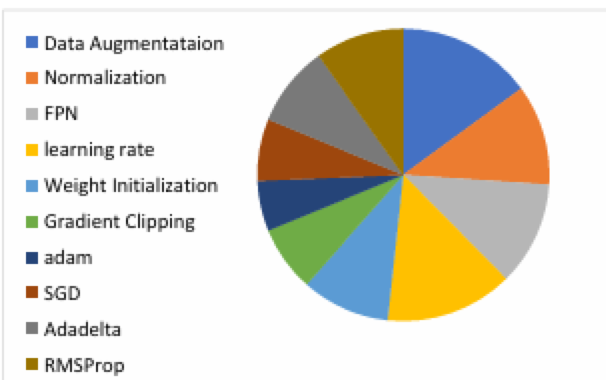

Figure 3 Count of each Optimization techniques applied on deep learning algorithm in various literatures towards object detection

Question 3: In drone surveillance, how can we enhance the effectiveness of object identification algorithms based on deep learning by using transfer learning techniques? When it comes to drone surveillance, object identification algorithms that rely on deep learning may greatly benefit from transfer learning. Drone surveillance algorithms that use deep learning for object recognition might benefit from this method's ability to conserve computing resources, speed up training, and enhance overall performance. Algorithm 1 lays out a comprehensive, step-by-step process for applying transfer learning to object identification, as seen here:

II. Use the pre-trained model as feature extractor
III. Training and fine tuning (Iterate through steps 3(i), (ii), (iii))
   i. Train the modified model, update the weights for new layers, retain the knowledge gained from previous steps.
   ii. Adjusting parameters such as learning rate, batch size, optimizer, and regularization techniques.
   iii. Asses the performance based on precision, recall and f1 score.
iv. Fine-tuning the model or adjusting hyperparameters, include re-annotating data, collecting additional data, or experimenting with different model architectures.
4. Final Output (A fine-tuned or Adapted Model)

In order to train object identification algorithms for drone surveillance, what are the challenges? Drone surveillance object identification systems trained using deep learning faces many obstacles: 1. Inadequate labeled data: It takes a lot of time and money to collect and label a dataset that includes all possible situations, weather, illumination, and item changes in the drone's vision. When labeled data is scarce, it might make it harder to train the model and reduce its generalizability to real-world scenarios. 2. Change in domain: Traditional object identification datasets are not always applicable to drone surveillance, which often makes use of unique imaging settings. Drones provide new difficulties such high altitude, changing views, occlusions, and motion blur to aerial photography and videography. Because of these variations in domains, a domain shift may occur, making it such that pre-trained models fare poorly when applied to the domain of drone surveillance. Additional training or fine-tuning may be necessary if the model has trouble properly detecting objects in these new settings. As a result of the drone's height and distance from the items of interest, drone surveillance sometimes entails identifying things at varied sizes. It might be tough for the model to recognize and locate objects effectively when they look tiny or display large size fluctuations in the scene. Another consideration is

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

that the resolution of drone cameras is often low, which might affect how well things stand out in the final product. If you want accurate object identification results, you have to solve these problems with resolution and size. In order to make decisions quickly, many applications need object detection to happen in real-time or very close to it. It may be tough to reach the appropriate speed on the limited onboard computing capabilities of drones when using deep learning-based object identification algorithms since these techniques can be computationally costly. It is rather challenging to achieve a balance between detection accuracy and real-time speed without using optimization methods such as model compression, quantization, or hardware acceleration. 5. Adapting to settings that are always changing: Drone surveillance often captures scenes that are constantly changing, with objects in motion and backdrops that are constantly shifting. Objects of interest might exhibit intricate patterns of motion, occlusions, or interactions with one another. A diversified dataset including multiple motion patterns and item interactions is necessary for training a model that can handle such dynamic scenarios efficiently. To extract the time-related data from drone surveillance footage, one must meticulously plan the model's architecture and use temporal modeling approaches. Drones have a limited length of time to fly owing to their batteries, which means that the quantity of data that can be acquired during each flying session is restricted as well. Obtaining a big enough and representative dataset becomes more difficult due to this restriction. Further restricting the dataset's variety and quantity are operational limits, privacy issues, and legislation that prohibit data collecting in some locations or under specified situations.

CONCLUSION

Although there is continuous research to dispel the notion that deep learning (DL) is a "black-box" solution, many still see it as such. Deep learning has already achieved remarkable strides in a number of remote sensing applications. We have narrowed our literature study to articles that discuss processing photos taken by UAVs using DL algorithms. Our research presents an overview of state-of-the-art methodologies and viewpoints on their application with the intention of providing a full grasp of the issue. We want to provide a comprehensive overview of the uses of DL-based methods for UAV image processing via this literature review. 1. The majority of the published papers on object recognition using deep learning focus on convolutional neural networks (CNNs) and radial basis functions (RCNNs). Nevertheless, multi-and hyperspectral data might be advantageous in some applications, such as precision agriculture and forest-related applications. Second, to improve benchmarking and training for

there is an obvious need for more publicly accessible datasets that were explicitly collected utilizing UAVs for networks. For researchers to successfully train and assess their networks, it is essential that these datasets be appropriately labeled to support supervised learning methods. The combination of GPU computing with deep learning (DL) techniques allows for efficient and fast data processing via fast inference solutions. Further investigation into real-time processing using embedded systems developed for UAVs is, however, certainly required.

REFERENCES

[1]. Aytar, Y. (2014). Transfer learning for object category detection. PhD Thesis, 146. https://www.robots.ox.ac.uk/~vgg/publications/201 4/Aytar14a/aytar14a.pdf

[2]. Chamarty, A. (2020). Fine-Tuning of Learning Rate for Improvement of Object Detection Accuracy. Proceedings - 2020 IEEE India Council International Subsections Conference, INDISCON 2020, 135–141. https://doi.org/10.1109/INDISCON50162.2020.000 38

[3]. Cui, X., Zhang, W., Tüske, Z., & Picheny, M. (2018). Evolutionary stochastic gradient descent for optimization of deep neural networks. Advances in Neural Information Processing Systems, 2018 Decem(NeurIPS), 6048–6058.

[4]. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 580–587. https://doi.org/10.1109/CVPR.2014.81

[5]. Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. 32nd International Conference on Machine Learning, ICML 2015, 1, 448–456.

[6]. Koo, K. M., & Cha, E. Y. (2017). Image recognition performance enhancements using image normalization. Human-Centric Computing and Information Sciences, 7(1). https://doi.org/10.1186/s13673-017-0114-5

[7]. Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2020). Focal Loss for Dense Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(2), 318–327. https://doi.org/10.1109/TPAMI.2018.2858826

[8]. Liu, W., Anguelov, D., Erhan, D., & Fu, C.-Y. (2016). SSD:Single Shot MultiBox Detector. European Conference on Computer Vision, 1(January 2017), 398–413. https://doi.org/10.1007/978-3-319-46448-0_2

[9]. Long, J., Darrell, T., & Berkeley, U. (2015). Fully Segmentation. CVPR, IEEE Access, 7, 3431–3440. https://doi.org/10.1109/ACCESS.2019.2908685

[10]. Na, G. S. (2022). Efficient learning rate adaptation based on hierarchical optimization approach. Neural Networks, 150, 326–335. https://doi.org/10.1016/j.neunet.2022.02.014

[11]. Ottoni, A. L. C., de Amorim, R. M., Novo, M. S., & Costa, D. B. (2023). Tuning of data augmentation hyperparameters in deep learning to building construction image classification with small datasets. International Journal of Machine Learning and Cybernetics, 14(1), 171–186. https://doi.org/10.1007/s13042-022-01555-1

[12]. Petersson, H., Gustafsson, D., & Bergström, D. (2017). Hyperspectral image analysis using deep learning - A review. 2016 6th International Conference on Image Processing Theory, Tools and Applications, IPTA 2016. https://doi.org/10.1109/IPTA.2016.7820963

[13]. Redmon, J., Santhosh, D., Ross, G., & Farhadi, A. (2016). You Only Look Once: Unified , Real-time Object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognitio, 779–788. https://doi.org/10.1145/3243394.3243692

[14]. Ruiz-Ponce, P., Ortiz-Perez, D., Garcia Rodriguez, J., & Kiefer, B. (2023). POSEIDON: A Data Augmentation Tool for Small Object Detection Datasets in Maritime Environments. Sensors (Basel, Switzerland), 23(7), https://doi.org/10.3390/s23073691

[15]. 1–14. Shallue, C. J., Lee, J., Antogniniy, J., Sohl Dickstein, J., Frostig, R., & Dahl, G. E. (2018). Measuring the effects of data parallelism on neural network training. ArXiv, 20, 1–49.

[16]. Signoroni, A., Savardi, M., Baronio, A., & Benini, S. (2019). Deep learning meets hyperspectral image analysis: A multidisciplinary review. Journal of Imaging, https://doi.org/10.3390/jimaging5050052

[17]. 5(5). Song, Y., Pan, Q. K., Gao, L., & Zhang, B. (2019). Improved non-maximum suppression for object detection using harmony search algorithm. Applied Soft Computing Journal, 81, 105478. https://doi.org/10.1016/j.asoc.2019.05.005

[18]. Xu, D., Zhang, S., Zhang, H., & Mandic, D. P. (2021). Convergence of the RMSProp deep learning method with penalty for nonconvex optimization. Neural Networks, 139, 17–23. https://doi.org/10.1016/j.neunet.2021.02.011

[19]. [19]. Yang, X., Wang, W., Wu, J., Ding, C., Ma, S., & Hou, Z. (2022). MLA-Net: Feature Pyramid Network with Multi-Level Local Attention for Object Detection. Mathematics, 10(24), 1–13. https://doi.org/10.3390/math10244789

[20]. [20]. Zhang, N., Lei, D., & Zhao, J. F. (2019). An Improved Adagrad Gradient Descent Optimization Algorithm. Proceedings 2018 Chinese Automation Congress, CAC 2018, 3, 2359–2362. https://doi.org/10.1109/CAC.2018.8623271