

Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

Use of Deep Learning Algorithms for Object Recognition in Live Drone Observation: A Comprehensive Analysis, Current Advancements, and Future Directions Issues

[1] MS. MUBEEN BEGUM,
[2] MOHAMMED ATEEQ UR RAHMAN, [3] MOHAMMAD AQUIB AMANULLAH, [4] SYED OWAIS,
[1] Assistant Professor, Department of AIML, Lords Institute of Engineering & Technology.
[234] Student, Department of AIML, Lords Institute of Engineering & Technology.

Abstract:

For the purpose of analyzing photos taken by unmanned aerial vehicles (UAVs), deep learning (DL) has become an invaluable tool in remote sensing. This review seeks to provide a thorough grasp of the topic, despite the fact that it has produced notable contributions across many applications. Included in the paper is a comprehensive review of current methods and their real-world uses for object recognition in drone surveillance in real-time. Aerial vehicles without names, deep learning, one-stage detectors, and two-stage detectors are all terms that come up while discussing this topic.

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. Unmanned Aerial Vehicles (UAVs) include 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. In real-time applications, human monitoring and picture capture are not feasible; hence, machine learning methods are extensively used to construct an automated system capable of processing and analyzing UAV-captured photos. However, it has no hand in the actual act of capturing images, unlike in mapping, surveying, inspection, or surveillance. 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. Drone surveillance often makes use of a camera that is built into the drone itself to record images. Depending on the individual application, the camera might employ a regular RGB camera or other kinds of sensors like thermal or multispectral ones. While in flight, the camera records footage or stills of the area below from a bird's-eye view. 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 camera isn't the only sensor or technology that the drone may be outfitted with; GPS and LIDAR, for example, could 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. The term "object detection" describes the steps used in drone surveillance to single out and pinpoint certain objects or characteristics in the recorded footage. 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. Drone identification is more difficult than using fixed cameras for a number of reasons, including the fact that these UAVs are flown at great heights: View from above: shadows and reflections, as well as the challenge of correcting for perspective distortion, Environment without control: Difficult elements, including climate, illumination,

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

and environmental changes throughout time, Things in motion: Identifying and following targets that could be traveling at fast speeds or undergoing quick directional changes, Drones' limited memory and processing capacity means that object identification algorithms can't be as complicated or accurate as they might be on other computer systems.

OBJECT DETECTION UAV OVERVIEW

An Object: What Is It? An item is anything that may be graphically represented by elements that are retrieved from a UAV. To put it simply, object detection is the process of identifying and localizing objects—whether they be crops, flowers, people, weapons, etc.—in order to offer information about their position or status. Deep Learning (DL) developed out of ML, which attempts to simulate the human brain's hierarchical organization in order to solve problems.

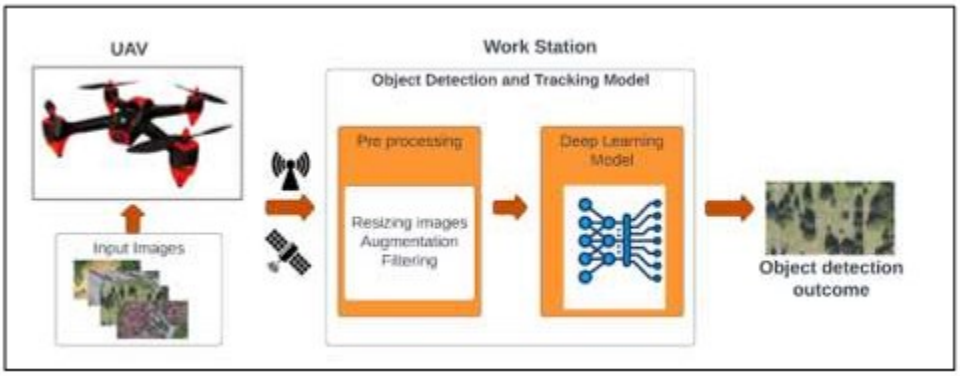


Figure 1 Sample Architecture for drone surveillance system complex and diverse range of applications.

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 could explain why DL apps have found a home in so many data-driven and image-processing domains. Many more observations and inspections are necessary, despite the fact that DL does produce encouraging findings. Hyperspectral imaging sensors (HIS) have recently attracted interest for their ability to capture high-resolution images of the earth's surface. These images can reveal information about the physical and chemical properties of objects and terrain, which is fascinating for object detection applications (Petersson et al., 2017; Signoroni et al., 2019).

RESEARCH MOTIVATION

Computer vision, audio recognition, and natural language processing are just a few areas where deep learning techniques have shown great promise for object identification. In order to accomplish object recognition tasks, deep learning algorithms can autonomously sift through mountains of data in search of useful attributes. Thus, deep learning seems to be a viable method for drone object identification in surveillance. Limited computational resources, poor data quality, and the need for strong algorithms are just a few of the many obstacles that must be overcome before deep learning can reach its full potential. Evaluating the present state-of-the-art, identifying research gaps, and proposing future research paths in object identification using deep learning in drone surveillance is, therefore, of the utmost importance. Contribution to Research (IV) This literature review aims to do just that by surveying the current research on object identification

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

in drone surveillance applications of deep learning. The review will primarily contribute to the following areas: 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. It is our hope that this evaluation will shed light on the present state of the art and point the way for future research into how to make object detection in drone surveillance more successful. 3. Our goal is to add the relevant features to existing drone datasets so that future study in this field is easier.

THE METHODOLOGICAL FRAMEWORK FOR LITERATURE REVIEW

The questions that guided the whole literature evaluation process were: 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? In order to get the most out of deep learning-based object identification algorithms used for drone surveillance, how can we optimize their hyperparameters? 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: How to Answer 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? Though object recognition sometimes made use of more conventional computer vision methods like Haar cascades and HOG (histogram of oriented gradients) before 2014. But beginning in 2014, there was a movement toward object recognition utilizing deep learning methods. This was especially true with the creation of deep convolutional neural networks (CNNs) such as the AlexNet architecture, which took home the prize at the 2012 ImageNet Large Scale Visual Recognition Challenge. Object recognition utilizing deep learning in drone surveillance was not yet common practice, however there was definitely interest in employing drones for surveillance and other purposes around that time. In the years that followed, object recognition tasks in a variety of UAV applications, including surveillance, began to see the widespread use of deep learning algorithms like YOLO, SSD, and Faster R-CNN. Since then, deep learning algorithms have seen substantial field applications (Figure 2). It was noted that R-CNN, Faster R-CNN, YOLO, and SSD are among the most popular deep learning algorithms in the object identification field. All of these algorithms hit top-tier results on object identification tasks, and they employ convolutional neural networks (CNNs) as their foundation 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. There are three main types of deep learning algorithms: one-stage, two-stage, and advanced detectors. a. Single-Stage Mechanics An example of an object detection approach in deep learning is the one-stage detector, which uses a single neural network pass to directly predict the bounding boxes and class probabilities of objects. Prior to identifying and honing down on a collection of potential objects, it suggests a number of zones of interest. Redmon et al. (2016), Liu et al. (2016), and RetinaNet (Lin et al., 2020) are a few well-known examples of one-stage detectors. To make YOLO function, we first divide the input picture into cells on a grid. Then, for each cell, we forecast the class probabilities and bounding boxes. 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). It is common practice for these models to use a Region Proposal Network (RPN) for the generation of potential object suggestions, and a second network for the classification of these proposals. Feature Pyramid Network (FPN), Hybrid Task Cascade, and Cascade R-CNN are a few more well-known two-stage detectors. In a nutshell, two-stage detectors function by generating proposals and then classifying them. First, given a picture as input, the model will

International Journal of Engineering Research and Science & Technology

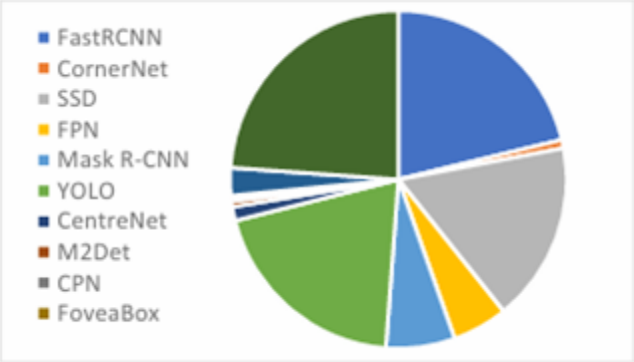
ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

provide a number of suggestions for potential objects to include. An RPN, or Region Proposal Network, is a common tool for this task; it accepts an image as input and produces a collection of bounding boxes that may include objects. In most cases, the RPN will generate an input picture feature map using a series 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. The second step is for the model to determine whether the proposed objects are foreground or background depending on whether or not they include an item. c. Detectors with Advanced Power. Object detection models that excel in efficiency and/or accuracy are known as advanced detectors. These models surpass one-stage and two-stage detectors. Some examples of detectors that have been enhanced are 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. The model's size and depth are optimized using a compound scaling approach, which strikes the optimal balance between efficiency and accuracy. In contrast, YOLOv4 incorporates a new data augmentation approach called mosaic augmentation, uses anchor boxes with varying aspect ratios and sizes, and modifies the Darknet backbone network with additional layers. Table 1 displays a comprehensive comparison of the three detectors.

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

Parameters	One Stage Detector	Two Stage Detector	Advance 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-sca feature fusion
Training time	Less	Longer	Less



For drone surveillance, how can we get the most out of object identification algorithms that use deep learning by optimizing their hyperparameters? Since training deep neural networks might take several hours or even days to finish, optimization is likely an essential part of deep learning. Many optimizers have been developed and implemented by researchers for use in deep learning across various domains. Some examples include the stochastic gradient descent deep learning optimizer, the mini batch stochastic gradient descent optimizers, Adagrad, RMSProp, and Zhang et al. (2019). Data augmentation, normalization, transfer learning, non-maximum suppression, feature pyramid networks, and adjusting the learning rate of neural networks are some of the ways that object detection models may be optimized during model execution. Data augmentation is a regularization technique that involves enhancing the training data with controlled fluctuations. To avoid overfitting—when a model becomes too dependent on its training data and unable to generalize to new instances—regularization approaches are useful. Data augmentation improves the model's object detection accuracy on unseen data by avoiding overfitting and encouraging the model to acquire more robust and generalizable characteristics by supplying various instances. In a 2014 study, Girshick et al. The article suggested a two-stage method for object identification utilizing the R-CNN framework; first, it would generate region recommendations; and second, it would utilize a CNN to categorize these proposals. Even though it wasn't called "data augmentation," the a2uthors did use a kind of data augmentation when they trained by randomly resizing and flipping the input pictures horizontally. Thanks to this method, the model's robustness and generalizability were both

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

enhanced. Also, in order to improve the object identification performance of DL models, a number of studies have used data augmentation approaches during preprocessing (Ottoni et al., 2023; Ruiz-Ponce et al., 2023). Improving convergence is another way to optimize deep learning models (Zhang et al., 2019). By including batch normalization approaches into their model architecture, the authors of (Ioffe & Szegedy, 2015) were able to speed up the CNN model training process. With normalization, we were able to attain the same accuracy in 14 fewer cycles, and we didn't even need to dropout. 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. To identify the normalized picture, a fine-tuned CaffeNet model is used. While the CNN model's average classification results were 93.24 percent, they improved to 96.85 percent when size-normalized images were used. One such optimization strategy that has been useful for improving deep learning models in object identification 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. The suggested techniques of learning rate adjustment outperformed competing adaptive gradient approaches. Feature Pyramid Networks can efficiently deal with objects of varying sizes (Yang et al., 2022). Improved performance in tasks like object identification, instance segmentation, semantic segmentation, and Non Maximum Suppression may be achieved using FPN, which combines multi-scale information in a feature pyramid. This allows for the detection and recognition of objects of varied sizes (Song et al., 2019). You can see a complete rundown of all the optimization methods used by the domain in Figure 3.

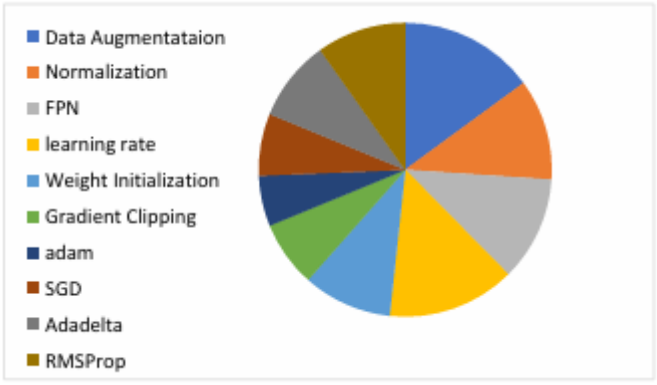


Figure 3 Count of each Optimization techniques applied on deep learning algorithm in various literatures towards object detection

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. In drone surveillance, this method may enhance the efficiency of object identification systems based on deep learning while saving computing resources and speeding up training. Here's one possible use of transfer learning to object recognition; Algorithm 1 lays out the process in great detail:

Algorithm 1: Step by step process of transfer learning for object detection

1. Select a pre trained model.
2. Choose the input data extracted through drone.
3. Transfer Learning Process
   I. Load the pre-trained model and freeze the early layers

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

II. Use the pre-trained model as feature extrac
III. Training and fine tuning (Iterate through 3(i), (ii), (iii))
  i. Train the modified model, update weights for new layers, retain the knowl gained from previous steps.
  ii. Adjusting parameters such as learning batch size, optimizer, and regulariz techniques.
  iii. Asses the performance based on preci recall and f1 score.
iv. Fine-tuning the model or adju hyperparameters, include re-annotating collecting additional data, or experime 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 most common challenges? There are a number of obstacles to overcome when training algorithms for drone surveillance that use deep learning to recognize objects: 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. The training process and the model's generalizability to real-world situations might be impacted by the scarcity of labeled data. Second, there is a change in the domain: standard object identification datasets and drone surveillance frequently use different imaging settings. High altitude, changing views, occlusions, and motion blur are some of the special difficulties that drones bring 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. The model could need further training or fine-tuning to correctly recognize items in these new settings. 3. Object size and sharpness: Depending on the drone's height and distance from the target items, drone surveillance may identify things at different sizes. It might be difficult for the model to recognize and locate objects effectively when they look tiny or display large size fluctuations in the scene. Another consideration is 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. Drones often have limited processing capabilities, making it tough to implement computationally complex object identification algorithms based on deep learning. Finding the sweet spot between real-time speed and detection accuracy could call for optimization strategies like model compression, quantization, or hardware acceleration. 5. Adjusting to fluctuating environments: Drone surveillance often captures landscapes that are in a constant state of flux, with objects in motion and backdrops 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. There may be limitations on the places or situations in which data may be collected due to rules, privacy concerns, or operational limits, which further reduces the dataset's variety and quantity. Section VII: Overall 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 (DL) has contributed significantly to remote sensing in a number of different ways. attention this literature review, we zero attention on the application of DL approaches for UAV image processing in particular. In pursuit of this goal, our research will survey current methods and provide opinions on how they may be used in order to give a thorough grasp of the topic. Our goal in compiling this literature review is to provide a comprehensive overview of the uses of DL-based methods for UAV image processing. According to this review's findings: 1. While most published works on object recognition using deep learning focus on convolutional neural networks (CNNs) and recurrent convolutional neural networks (RCNNs), multi-and hyperspectral data might be useful in some applications, such as precision agriculture and forest-related research. Having more publicly accessible datasets that are particularly collected with UAVs is essential for improving network training and benchmarking. Researchers will be able to train and assess their networks more

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

efficiently with appropriately labeled datasets that support supervised learning methods. 3. By combining GPU processing with deep learning (DL) techniques, fast inference solutions may be provided, allowing for efficient and speedy data processing. But further investigation into real-time processing using embedded systems tailored to UAVs is still 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]. 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]. 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