International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-5.86

Autonomous Drone Navigation using Deep Learning

[1]Doma Arvind Kumar,[2]Kalla Anil,[3]Durgam Nikhil,[4]Gadicharla Hima Sai Girish,[5]B.Raju,

[1,2,3,4] Student,Department of ECE,Narsimha Reddy Engineering College,Misammaguda(V),Kompally-500100,Telangana State,India.
[5] Assistant Professor,Department of ECE,Narsimha Reddy Engineering College,Misammaguda(V),Kompally-500100,Telangana State,India.

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. Object identification in real-time drone surveillance is the focus of this work, which provides a comprehensive overview of current methods and their practical uses.

Keywords: Unnamed Aerial Vehicles, Deep learning, One- stage detector, two-stage detector.

INTRODUCTION

Drones, also known as Unmanned Aerial Vehicles (UAVs), have found several uses due to their capacity to reach inaccessible or hazardous locations that people would find impossible to reach on foot. 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. 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. Depending on the task at hand, the camera might use thermal or multispectral sensors instead of a conventional RGB camera. 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. To get the best possible images and to record certain kinds of data, the drone cameras may have their features and settings tweaked. The camera may, for instance, provide a variety of modes for taking still photos or video, as well as customizable settings for things like focus, zoom, and exposure. Aside from the camera, the drone may also include additional sensors and technologies like GPS and LIDAR that aid in navigating and mapping the area being monitored. Either real-time tracking or bulk storage of the photos taken from various locations is planned.

Object detection is the procedure that allows for the tracking of certain objects. Object detection is a crucial part of drone surveillance. It involves finding and pinpointing certain items or characteristics in the drone's video or photos. Agriculture, environmental monitoring, SAR, and military activities are just a few of the numerous fields that have grown to rely on drone observation. Drones have become a popular alternative for surveillance activities due to their capacity to provide overhead views of broad regions in a rapid and effective manner. Drone surveillance can only be as good as the object identification algorithms used to monitor them. Drone surveillance presents unique challenges as compared to fixed cameras due to the following factors: the high altitudes at which these UAVs are often flown; and View from above: shadows and reflections, as well as the challenge of correcting for perspective distortion, Uncontrolled environmental factors: Difficult elements including weather, illumination, and environmental changes throughout time, Finding and following items in motion, whether they're traveling at a fast pace or undergoing a sudden change in direction, The complexity and precision of object identification algorithms might be constrained by the limited processing power and memory of drones, in comparison to other computer systems.

OBJECT DETECTION UAV OVERVIEW

Objects: What Are They? A visually represented object is one that is retrieved 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.

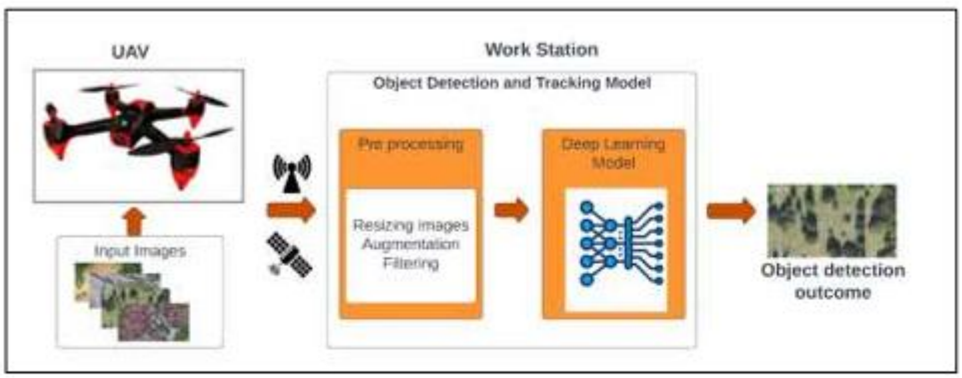

Figure 1 Sample Architecture for drone surveillance system

extensive and varied set of uses. 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. Even if DL shows promise, a lot more testing and observation is still needed. 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

Object identification using deep learning algorithms has great promise in many areas, such as computer vision, voice recognition, and NLP. As a result of their ability to autonomously learn from massive datasets, deep learning algorithms may extract object detection-relevant information. 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. To assess the existing 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.

RESEARCH CONTRIBUTION

Using deep learning for object recognition in drone surveillance, this study aims to provide a thorough summary of the current research. The following are the main points that the review will cover:
The purpose of this study is to compare and contrast the object recognition capabilities 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 following questions served as the basis for the whole literature evaluation process: In the field of drone surveillance, what are the most recent and advanced object identification algorithms that use deep learning? How have these algorithms changed over the years? Second Question: How can we get the most out of deep learning-based object identification algorithms while using drones for surveillance? Asking3, "How can we enhance the performance of deep learning-based object detection algorithms in drone surveillance by leveraging transfer learning techniques?" Question 4: How have researchers dealt with the difficulties of training object identification algorithms for drone surveillance that are based on deep learning in past studies? Additionally, what are the necessary directions for the field to progress in the future?

SOLUTIONS TO RESEARCH QUESTIONS

In the field of drone surveillance, what are the most recent and advanced object identification algorithms that use deep learning? How have these algorithms changed over the years? 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. At the time, there was a lot of talk about potential uses for drones in the surveillance industry, but few people were utilizing deep learning to identify objects in the sky. 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). A few of the most popular deep learning algorithms in the object identification domain include R-CNN, Faster R-CNN, YOLO, and SSD. Using convolutional neural networks (CNNs) as their foundation, these algorithms have shown top-tier performance on object identification tasks. Researchers are beginning to take an interest in other algorithms that were developed in early 2018, including CenterNet, Mask RCNN, M2Det, CPN, and FoveaBox.

There are three main types of deep learning algorithms: one-stage, two-stage, and advanced detectors. a. One-Stage Detectors: A subset of deep learning algorithms, one-stage detectors allow for the direct prediction of object bounding boxes and class probabilities with just a single neural network run. Before sorting and improving them, it suggests a collection of potential objects, also called areas 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. YOLO predicts the class probabilities and bounding boxes for each cell in the input picture by splitting it into a grid of cells. A confidence score is assigned to each anticipated bounding box that indicates the likelihood that the 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. The R-CNN family of two-stage detector architectures boasts various well-liked models, such as Fast R-CNN, Faster R-CNN, and Mask 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. Some more well-known two-stage detectors include Hybrid Task Cascade, Cascade R-CNN, and Feature Pyramid Network (FPN). In a nutshell, two-stage detectors function by generating proposals and then classifying them. The first step is for the model to take the input picture and come up with a list of potential objects to include. A common neural network for this task is the Region Proposal Network (RPN), which, given an image as input, generates a collection of bounding boxes that may or may not 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 over 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 next step is for the model to determine whether each of the proposed objects is in the forefront (containing an item) or the background (containing nothing). The next section is about advanced detectors. When compared to one-stage and two-stage detectors, advanced detectors outperform them in terms of efficiency, accuracy, or both. A few examples of enhanced detectors 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. The model's size and depth are optimized using a compound scaling approach, which strikes the optimal balance between efficiency and accuracy. 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. 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

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

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

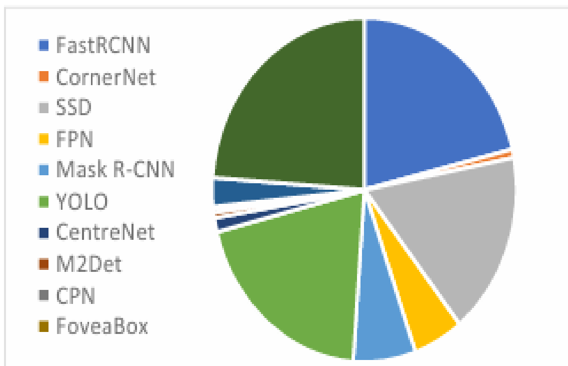

Figure 2 Relative percentage of different deep learning papers published in the UAV domain

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. 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. Data augmentation is a regularization technique that involves enhancing the training data with controlled fluctuations. When a model becomes too tailored to its training data and can't generalize to new cases, a phenomenon known as overfitting, regularization methods come in handy. 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.Using the R-CNN framework, Girshick et al. (2014) laid up a two-stage process for object detection: first, they suggested regions to include, and then they utilized CNNs to categorize them. 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 photos horizontally. The model's resilience and generalizability were both enhanced by this method. In a similar vein, a number of publications have boosted DL model performance in object recognition by using data augmentation strategies 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). 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. By using normalization, we were able to attain the same accuracy in 14 fewer cycles, without the necessity of 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. When it 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. Improving CNN's learning rate to get the best possible detection accuracy was the main focus of (Chamarty, 2020). A link between learning rate and dataset size in the range of 10^-4 to 10^-5 was successfully achieved in the article. A learning rate optimization that modifies the learning rate by modifying the direction method of multipliers was utilized in (Na, 2022), which is similar to the approach described here. 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). 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). The optimization strategies employed in the domain are detailed in Figure 3.

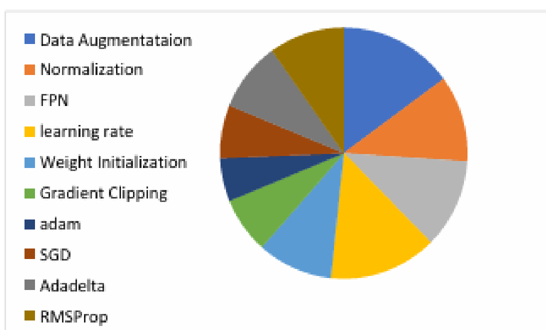

Figure 3 Count of each Optimization techniques applied on deep learning algorithm in various literatures towards object detection

Asking3, "How can we enhance the performance of deep learning-based object detection algorithms in drone surveillance by leveraging 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. Algorithm 1 lays out a comprehensive, step-by-step process for applying transfer learning to object identification, as seen here:

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

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
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)

When it comes to drone surveillance, what are the difficulties of training algorithms to recognize objects using deep learning? Several obstacles arise during the training of object identification systems based on deep learning for use in drone surveillance: 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. 2. Change in domain: In contrast to more conventional object identification datasets, drone surveillance often makes use of unique imaging situations. 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. Additional training or fine-tuning may be necessary if the model has trouble properly detecting items 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 could be difficult for the model to effectively identify and locate objects in the picture if they seem tiny or show large size changes. Furthermore, the quality and visibility of the objects in the recorded movies or photos may be compromised due to the low resolution of the drone cameras. Obtaining trustworthy object detection outcomes requires addressing these issues with size and resolution. 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.

Optimization methods such as model compression, quantization, or hardware acceleration could be necessary to strike a balance between detection accuracy and real-time speed. 5. Changing with the times: Scenes captured by drones during surveillance missions are often dynamic, with objects in motion and backdrops constantly shifting. Interest items can be moving in complicated patterns, occluding other objects, or interacting with one another. A varied dataset including different motion patterns and item interactions is necessary for training a model that can properly handle such dynamic scenarios. Capturing the temporal information in drone surveillance films also requires careful model architecture design and temporal modeling approaches. 5. Drones' restricted flight duration is caused by their battery capacity, which in turn limits the quantity of data that can be acquired during each flying session. Obtaining a big enough and representative dataset becomes more difficult due to this restriction. Furthermore, data gathering may be limited in certain places or under specific situations due to rules, privacy concerns, or operational restrictions, which further limits the dataset's variety and extent.

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. Artificial intelligence (DL) has already achieved great strides in remote sensing for a variety of uses. Our literature evaluation is devoted entirely to the application of DL approaches to the analysis of UAV-captured pictures. 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. it is determined from this review that: Most published materials on object detection using deep learning use convolutional neural networks (CNNs) and radial basis functions (RNFs). However, multi- or hyperspectral data is useful for some applications, such as precision agriculture and forest-related tasks. Additionally, additional publicly available datasets obtained using UAVs are clearly needed to improve network training and benchmarking. Researchers will be able to train and assess their networks more efficiently with appropriately labeled datasets that 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. Still, further investigation into UAV-specific embedded systems for real-time processing is 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