Vol.15, Issue No 2, 2025

Adaptive Deep Learning Methods for Object Detection in Drone Surveillance Systems

[1]KOPPARTHI DURGA GANESH, [2] BOLLA RAJESH,[3] KANURI ADI SATYA DURGA CHARAN, [4] MOTURI SAI SUBRAHMANYAM, [5]Dr. B.V.S. VARMA,

[1234]Student, Department of CSE, DNR College of Engineering & Technology, Balusumudi, Bhimavaram, India.

[5] Professor, Department of CSE, DNR College of Engineering & Technology, Balusumudi, Bhimavaram, India.

Abstract:

In recent years, deep learning (DL) has become an indispensable technique for remote sensing, especially for analyzing photos taken by UAVs. 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. Aerial vehicles without names, deep learning, one-stage detectors, and two-stage detectors are all terms that relate to this topic.

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. 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. You can see the drone surveillance system's fundamental design in Figure 1. An onboard camera is usually the way to go in drone surveillance for capturing images. The application dictates the sort of sensor used by the camera; it might be a regular RGB camera or it could include thermal or

multispectral sensors. While the drone is in the air, its camera records footage or stills of the landscape below. Either the footage or stills will be sent live to a base station or saved on the drone itself for further analysis at a later time. Drone cameras might have their features and settings tweaked to provide the best possible images and to record certain data. 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. Along with 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. We will either keep the photographs in batches or monitor them in real time when they are taken from many regions. 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. On the other hand, quick and accurate object identification algorithms are crucial to drone surveillance's efficacy. Given the aforementioned factors, drone identification in surveillance is more difficult than with fixed cameras, especially when these UAVs are flown at great heights: Perspective distortion, shadows, and reflections are all challenges inherent to aerial photography. Unregulated setting: Obstacles like the elements, the lighting, and the gradual transformation of the surrounding environment, something in motion: seeing and following something that could be traveling at a fast pace or making a quick U-turn, 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 item is anything that may be graphically represented by elements that are retrieved from a UAV. Finding and recognizing these elements—crops, flowers, people, weapons, etc.—and then localizing them to offer data about their position or condition—or, to put it more simply, classifying the extracted element—is what object detection is all about. Deep Learning (DL) developed out of ML, which attempts to simulate the human brain's hierarchical organization in order to solve problems.

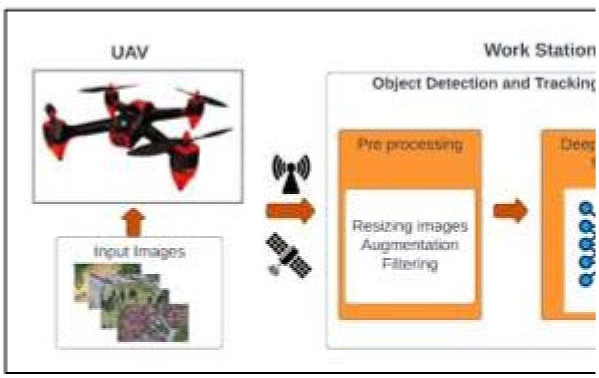

Figure 1 Sample Architecture for drone surveillance system

wide variety of difficult 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. 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

Computer vision, audio recognition, and natural language processing are just a few areas where deep learning techniques have shown great promise for object identification. Algorithms for deep learning can automatically sift through mountains of data in search of characteristics useful for object recognition. Thus, deep learning seems to be a viable method for drone object identification in surveillance. 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. 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.

RESEARCH CONTRIBUTION

Using deep learning for object recognition in drone surveillance, this study aims to provide a thorough summary of the current research. 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. Our goal in doing this evaluation is to provide light on the present state of the art and suggest avenues for future research that might enhance the efficacy of object detection in drone surveillance. 3. We seek to add the associated facts to existing drone databases so that future study in this field might be facilitated.

THE METHODOLOGICAL FRAMEWORK FOR LITERATURE REVIEW

Based on the following questions, the full literature review process was carried out: 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? Question 2: For drone surveillance, how can we get the most out of object detection systems that use deep learning? 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? In addition, what are the necessary steps to move the field forward 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. During that time, there was a lot of talk about drones having potential uses beyond surveillance, but few people were utilizing deep learning to identify objects in the sky. 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. A great deal of progress has been made in the area since then because to deep learning algorithms (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. 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. One-Stage, Two-Stage, and Advanced Detectors are the three main types of deep learning algorithms. a. Single-Stage Mechanics One-stage detectors are a subset of object detection algorithms that use deep learning to provide predictions about object bounding boxes and class probabilities using only one neural network run. It begins by suggesting potential objects or areas of interest, and then it sorts and improves them. A few well-known one-stage detectors include RetinaNet, SSD (Single Shot Detector), and YOLO (You Only Look Once) (Redmon et al., 2016; Liu et al., 2016; Redmon et al., 2020). 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. A confidence score indicates the likelihood that a given anticipated bounding box includes an item. b. Detectors with Two Stages Highly accurate and adaptable, two-stage detectors are a potent family of object identification models; yet, they are computationally costly and quality-sensitive when proposing candidate objects. Among the many well-liked two-stage detector designs, one may find the R-CNN family, which comprises 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. 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. 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. section c. "Advanced Detectors" Object detection models that excel in efficiency and/or accuracy are known as advanced detectors. These models surpass one-stage and two-stage detectors. Improved detectors include EfficientDet, CenterNet, YOLOv4, and DETR, to name a few. Google's EfficientDet family of object detectors outperforms state-of-the-art algorithms with much fewer parameters and compute requirements. To achieve the optimal balance between speed and accuracy, it employs a compound scaling method that adjusts the depth and size of the model. 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. Listing 1. Various performance metrics used to compare various deep learning object identification methods

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

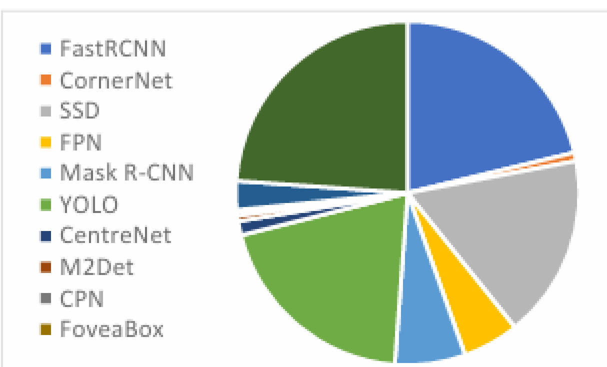

Figure 2 Relative percentage of different deep learning papers published in the UAV domain

Question 2: For drone surveillance, how can we get the most out of object detection systems that use deep learning? Due to the time-consuming nature of training deep neural networks, optimization is an essential part of deep learning. For deep learning, researchers have developed a number of optimizers; for instance, RMSProp, Adagrad, mini batch stochastic gradient descent optimizers, stochastic gradient descent deep learning optimizer, and others (Cui et al., 2018; Shallue et al., 2018; Zhang et al., 2019; Xu et al., 2021). 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. 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. In order to improve the model's object detection accuracy on unseen data and decrease overfitting, data augmentation provides various instances. This enables the model to acquire more robust and generalizable characteristics. 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. 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 method contributed to the model's enhanced resilience and capacity for generalization. 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). 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. You may get the same accuracy in only 14 fewer cycles with the help of normalization, which also gets rid of the requirement to dropout. By using feature extraction and normalization applied to CNN classifier, a pretreatment model was suggested by (Koo & Cha, 2017) to improve the performance of image recognition models. 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%. One such optimization strategy that has been useful for improving deep learning models in object identification is the transfer learning methodology (Aytar, 2014). To do this, transfer learning makes use of pre-trained models on massive datasets, which enables the transfer of task-specific information. In order to achieve the highest possible detection accuracy, the authors of (Chamarty, 2020) focused on optimizing the CNN learning rate. Using a learning rate optimization that modifies the learning rate by modifying the direction method of multipliers, the article was able to achieve a link between learning rate and dataset size ranging from 10^-4 to 10^-5. A similar approach was employed in (Na, 2022). The suggested techniques of learning rate adjustment outperformed competing adaptive gradient approaches. 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 domain's optimization strategies are detailed in Figure 3.

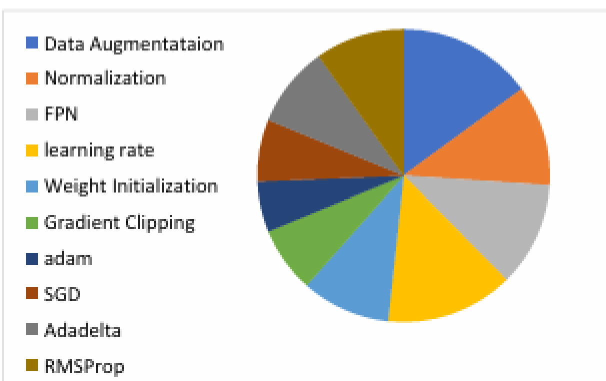

Figure 3 Count of each Optimization techniques applied on deep learning algorithm

approaching object detection Q3 in a number of different publications. When it comes to drone surveillance, how can we make the most of transfer learning approaches to make deep learning-based object identification algorithms work better? 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
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? 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. 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. Additional training or fine-tuning may be necessary if the model has trouble properly detecting objects in these new settings. 3. Object size and resolution: Depending on the drone's height and distance from the target items, drone surveillance may identify things at different 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 for applications to make decisions quickly, object detection in real-time or near real-time is typically necessary. 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. 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. 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. Complex motion patterns, occlusions, or interactions between items may be shown by objects of interest. 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. 6. Data collection and restricted flight duration: Drones' limited flight time is caused by their battery capacity, which in turn limits the quantity of data that can be acquired during each flying session. Because of this restriction, collecting a big enough and representative dataset is difficult. 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

Despite continued efforts to dispel this myth, deep learning (DL) is still often seen as a "black-box" answer to many issues. Deep learning (DL) has already achieved great strides in remote sensing for a variety of uses. 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. 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 fields. 2. More publicly accessible datasets that are explicitly collected using UAVs are clearly needed to improve network training and benchmarking. In order to train and assess networks efficiently, researchers need 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/public ations/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/INDISCON50162.2 020.000 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