



International Journal of Marketing Management

ISSN 2454 - 5007



www.ijmm.net

Email ID: editor@ijmm.net , ijmm.editor9@gmail.com

A Comprehensive Review, New Advancements, and Outstanding Problems in the Use of Deep Learning Algorithms for Object Recognition in Real-Time Drone Surveillance

¹ Pushpa Soumya, ² G.Ganga,

¹Assistant Professor, Megha Institute of Engineering & Technology for Women, Ghatkesar.

² MCA Student, Megha Institute of Engineering & Technology for Women, Ghatkesar.

Abstract:

In recent years, deep learning (DL) has become an indispensable technique for remote sensing, especially for analyzing photos taken by UAVs. Despite its widespread applicability and notable contributions, the purpose of this study is to provide a thorough overview of the topic. 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.

Keywords:

Unidentified Aircraft, "Deep learning," Both one-stage and two-stage detectors are considered.

I. 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. Aerial photography, SAR operations, environmental monitoring, defense and military, and many more uses are possible with the cameras that UAVs have. These cameras can take still photos or films from a variety of heights and perspectives. 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. However, it has nothing to do with the actual act of capturing images, whether for mapping, surveying, inspection, or surveillance. The UAV has the capability to wirelessly send the photographs to a ground station in real-time or store them onboard for future retrieval and analysis. Figure 1 shows the fundamental design of the drone

surveillance system. An onboard camera is usually the way to go in drone surveillance for capturing images. 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. 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. It was possible to maximize picture quality and gather certain sorts of information by adjusting the features and settings of the drone cameras. 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. 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. Object detection is the procedure that allows for the tracking of certain objects. When using a drone for surveillance purposes, object detection is finding and locating certain things in the footage or stills taken by the aircraft. In recent years, drone surveillance has become more important in several fields, such as agricultural, environmental monitoring, SAR, and military activities. Drones are a popular alternative for surveillance activities due to their capacity to effectively and swiftly deliver overhead images of broad 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, Outside of any kind of regulation: Difficult elements, including climate, 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.

II. OBJECT DETECTION UAV OVERVIEW

An Object: What Is It? An object is a visual representation of an element that is taken 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.

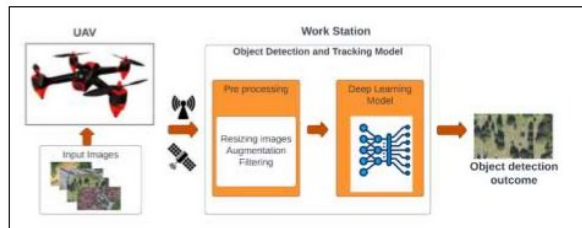


Figure 1 Sample Architecture for drone surveillance system

an 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. One possible explanation for the widespread usage of DL in data-driven and image-processing applications is this. 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).

III. 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. Since this is the case, deep learning seems to be a good strategy for drone object detection. 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. In order 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. RESEARCH CONTRIBUTION

This literature review aims to do just that by surveying the current research on object identification in drone surveillance applications of deep learning. The review will primarily contribute to the following areas:

1. In this study, we will take a look at how different deep learning algorithms and architectures handle object recognition, what their pros and cons are, and what problems and opportunities they may provide.
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. Our goal is to provide existing drone datasets with the necessary information to support more study in this field.

V. THE METHODOLOGICAL FRAMEWORK FOR LITERATURE REVIEW

We anchored the whole literature research procedure on these framed questions:

The first question is, how have the most recent and advanced object identification algorithms for drone surveillance, which are based on deep learning, changed over time? The second question is, how can we get the most out of deep learning-based object identification algorithms while using them to monitor drones? 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? Furthermore, what are the necessary future directions to progress the field?

VI. SOLUTIONS TO RESEARCH QUESTIONS

The first question is, how have the most recent and advanced object identification algorithms for drone surveillance, which are based on deep learning, changed over time? However, object recognition was mostly accomplished using more conventional computer vision methods like Haar cascades and HOG (histogram of oriented gradients) up until 2014. A change, however, occurred towards the use of deep learning approaches for object recognition around 2014, especially 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. 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. 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. Figure 2 shows the progression of the field's use of deep learning techniques from that time till the present. In the field of object identification in particular, R-CNN, Faster R-CNN, YOLO, and SSD are among the most popular deep learning techniques. These algorithms have shown state-of-the-art performance on object identification tasks, and they all rely on CNNs for feature extraction. Researchers are beginning to take an interest in other algorithms that were developed in early 2018, including CenterNet, Mask RCNN, M2Det, CPN, and FoveaBox. In general, there are three types of deep learning algorithms: one-stage,

two-stage, and advanced detectors. . Detectors for One Stage 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. Before categorizing and improving them, it suggests a collection of potential objects (or areas of interest). 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). 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 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. The R-CNN family, which comprises Fast R-CNN, Faster R-CNN, and Mask R-CNN, is one of the well-liked two-stage detector designs. Usually, these models will use a Region Proposal Network (RPN) to come up with potential objects to propose, and then another network to sort them. 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 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. After that, a tiny window is moved over the feature map and a set of predetermined anchor boxes are applied to each position to create a collection of possible object suggestions. 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. 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. Figure 1 shows a comparison of the three detectors in great detail.

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

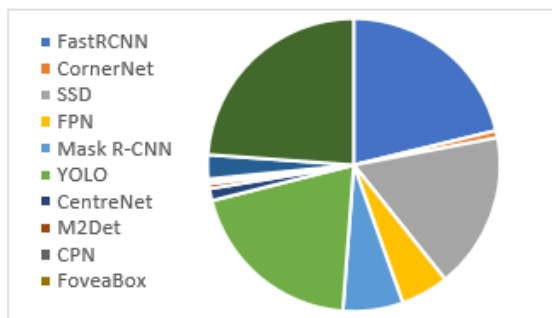


Figure 2 Relative percentage of different deep learning papers published in the UAV domain

Question 2: How can we maximize the performance of object identification algorithms for drone surveillance that are based on deep learning by optimizing their hyperparameters? Due to the time-

consuming nature of training deep neural networks, optimization is an essential part of deep learning. There are a number of optimizers that have been developed and used by researchers for deep learning in 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). Nevertheless, there are other methods for optimizing object identification models during execution, such as data augmentation, normalization, transfer learning, adjusting neural network learning rates, feature pyramid networks, and non-maximum suppression. Data augmentation is a regularization technique that involves enhancing the training data with controlled fluctuations. 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. 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. Using the R-CNN framework, the article put out a two-stage strategy for object identification, first requiring the generation of area suggestions and then categorizing these suggestions via the usage of a CNN. Even though it wasn't called "data augmentation," the authors did use a kind of data augmentation when they trained by randomly resizing and flipping the input pictures horizontally. The method contributed to the model's enhanced resilience and capacity for generalization. The use of data augmentation approaches during preprocessing has been reported in several papers to improve the object identification performance of DL models (Ottoni et al., 2023; Ruiz-Ponce et al., 2023). Achieving greater 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. 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. 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 comes to improving deep learning models for object recognition, one more optimization strategy that has been effective is the transfer learning methodology (Aytar, 2014). To do this, transfer learning makes use of pretrained models on massive datasets, which enables the transfer of task-specific information. The focus of (Chamarty, 2020) was on optimizing CNN's learning rate to get the highest possible detection accuracy. The article achieved a link between learning rate and dataset size in the range of 10-4 to 10-5. A learning rate optimization, which modifies the learning rate by modifying the direction method of multipliers, was also employed in (Na, 2022). 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). In order to increase performance in tasks such as object identification, instance segmentation, semantic segmentation, and NonMaximum Suppression, FPN combines multiscale 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.

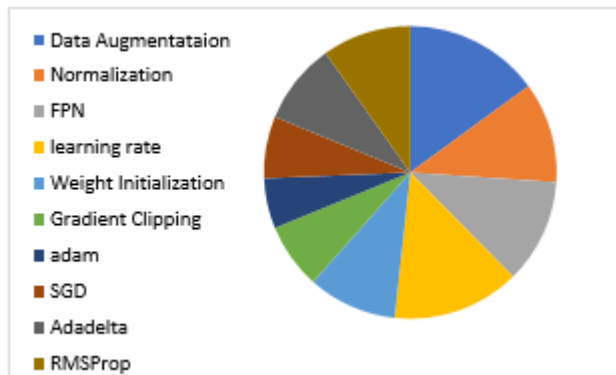


Figure 3 Count of each Optimization techniques applied on deep learning algorithm in various literatures towards object detection

Question 3: In drone surveillance, how can we enhance the effectiveness of object identification algorithms based on deep learning by using transfer learning techniques? If you're using deep learning to train object identification algorithms for drone surveillance, transfer learning is a great way to boost their performance. 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)

Question 4: What are some of the difficulties in developing algorithms for drone surveillance that use deep learning to recognize objects? Several obstacles arise during the training of object identification systems based on deep learning for use in drone surveillance: 1. Less labelled data available: Gathering and annotating data from a wide range of situations, including changes in weather, illumination, and objects in the drone's perspective, is an enormous undertaking that demands a lot of time and money. 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. Drones provide new difficulties such high altitude, changing views, occlusions, and motion blur to aerial photography and videography. Due to domain shifts caused by these variations, pretrained models may

not be able to generalize well 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 could be difficult for the model to effectively identify and locate objects in the picture if they seem tiny or show large size changes. The resolution of drone cameras is another factor that may affect how well items stand out in the final product. If you want accurate object identification results, you have to solve these problems with resolution and size.

4. Performance in real-time: In order to make quick decisions, drone surveillance applications often need object detection in real-time or near real-time. 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. 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. Interest items can be moving in complicated patterns, occluding other objects, or interacting with one another. To train a model to properly manage such dynamic situations, a wide dataset including different motion patterns and object interactions is required. Capturing the temporal information in drone surveillance films also requires careful model architecture design and temporal modeling approaches.

6. Restricted time for flying and gathering data: Due to battery capacity, drones have a restricted flight duration. This means that the quantity of data that can be acquired during each flying session is limited. 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.

VII. 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. The application of DL approaches to the analysis of photos taken by UAVs is the primary emphasis of our literature study. 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. Based on the findings of this review, First, most of the published works on object recognition using deep learning are based on convolutional neural networks (CNNs) and radial basis functions (RCNNs). However, multi/hyperspectral data is useful for specific applications, such as precision agriculture and forest-related tasks. 2. It is evident that more publicly accessible datasets specially collected by UAVs are 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. 3. GPU processing in conjunction with deep learning (DL) methods may provide fast inference solutions, allowing for efficient and speedy data processing. Nevertheless, more investigation into real-time processing using embedded systems developed for UAVs is very necessary.

VIII. REFERENCES

- [1]. Aytar, Y. (2014). Transfer learning for object category detection. PhD Thesis, 146. <https://www.robots.ox.ac.uk/~vgg/publications/2014/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/INDISCON50162.2020.00038>
- [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, 2018Decem(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 Convolutional Networks for Semantic 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 Recognition, 779–788. <https://doi.org/10.1145/3243394.3243692>

[14]. Ruiz-Ponce, P., Ortiz-Perez, D., GarciaRodriguez, J., & Kiefer, B. (2023). POSEIDON: A Data Augmentation Tool for Small Object Detection Datasets in Maritime Environments. Sensors (Basel, Switzerland), 23(7), 1–14. <https://doi.org/10.3390/s23073691>

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