Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 47 No. 01 (2026)

Complexities in Remote Sensing-Based Object Detection and Interpretation

Shalini L [1], Dr. Thirupurasundari D R [2]

[1, 2] *Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India*

Abstract:- Detection of object is a fundamental part of processing images from remote sensing. Object analysis in aerial images generally covers the process of automatically identifying and localizing specific objects in satellite or aerial images, and addresses the challenges specific to remote sensing data, such as different scales, complex backgrounds, and spectral information, as well as state-of-the-art techniques. It primarily employs the approaches of deep learning like convolutional neural networks (CNN's) to classify and detect objects of interest accurately. The exponential growth of deep convolutional neural networks has significantly improved the detection of object in remote sensing images. However, it has a substantial amount of impact on the effectiveness of detector, when the scenario is complex and the dimensions of objects changes drastically. The techniques of deep learning for detecting objects in aerial images has many uses, including pavement extraction, urban building detection, and forest fire monitoring. It also offers robust assistance for agricultural management, environmental monitoring and urban development. We examine the difficulties that both deep learning and conventional approaches encounter. The complexity of various methods is also evaluated and examined with generic datasets of remote sensing aerial images.

Keywords: *Remote Sensing, Aerial Images, Bounding Box Segmentation, Object Recognition, Deep Learning.*

1. Introduction

Remote sensing is a technique which is applied to get the target knowledge without directly interacting with it, using remote sensing framework like aerial and space-based sensors. This technology involves comprehensive observation of the Earth, where the images contain an extensive amount of information about the target and are often implemented in the fields such as disaster management, agriculture, environmental monitoring, and smart city planning [1,2,3]. The efficiency of observing and acquiring remote sensing images has improved due to growth and development of Geo-spatial sensing technologies. Things in aerial sensing images are usually compact and often close together, making the retrieval of useful information using the human eye ineffectual and prone to failures, and manual extraction of such data is an impractical effort. In remote sensing, target detection computational methods are often used [4].

Conventional remote sensing image object detection techniques mainly rely on manually created features and standard machine learning approaches. Perhaps, due to the complexity and diversity of remote sensing image data, traditional methods encounter many difficulties in processing these data. In the past few years, the rapid advancement of deep learning technologies, specifically due to the emergence of convolutional neural networks (CNNs), has increased the efficiency and scope for object detection in remote sensing applications [5, 6]. When contrasted with standard methods based on manually constructing feature, deep learning can effortlessly acquire executive meaningful characteristics through significantly stronger representation capabilities and readily respond to the intricate changes of Geo-spatial image sensing data. The detection of targets in aerial images also requires integrating the surrounding contextual data of the target to increase the precision of identification and localization. Through the use of pooling strategies and local fields of reception in ConvNets, deep learning model gathers contextual cues around the target, thereby improving the reliability and precision of target detection. Deep learning algorithm often require significant volume of training data for optimal performance. Remote sensing image data offers vast coverage along with rich temporal and geographical information, providing an excellent source for training deep learning models. This enables deep learning to effectively handle the diverse forms and intricate characteristics of target in remote sensing imagery, leading to highly precise and generalized architecture trained on huge datasets.

Detection and classification of objects in earth observation images is a very challenging task, but it has various potential applications [7-10], including autonomous driving, face recognition, pedestrian detection, and environmental monitoring. However, object detection can also support more computer vision tasks like [11-14] visual segmentation, image interpretation, activity analysis and target monitoring. Conventional object detection algorithms in remote sensing images generally involve three steps: object preprocessing, feature extraction, and classification. These techniques have good accessibility of attributes like spatial dimensions, surface pattern and intensity of objects. Even so, when it deals with intricate environments, occlusion, and scale variation, etc, traditional methods have certain limits and constraints. Similarly, traditional approaches also face some limitations by the demand of extensive geographic scope and fine spatial detail of remote sensing imagery in terms of accuracy and performance.

2. Literature Survey

In a wide range of computer vision and remote sensing tasks, deep learning has achieved an outstanding performance [15]. The small size of an object is one of the biggest obstacles in remote sensing images for detecting the target. Many researchers have proposed solutions for this difficulty by using deep learning-based object detection algorithms. Faster R-CNN is improved to support the detection of small and medium-sized herbivores in large scale imagery [16]. Additionally, a novel overlap segmentation technique is implemented to address the issues of target mismatch and missed detection. Also, the HRNet feature learning architecture is employed to localize the small-sized targets more effectively. [17] Suggested an approach for locating conventional village architectures, in which Mask R-CNN is enhanced by incorporating PAFPN and ASPP module to improve multi-scale feature representation and integration. To collect more discriminate feature information and alterations in convolution size, [18] proposed SCMask R-CNN, an aerial vehicle detection method in aerial images leveraging an modified SC-conv integrated with the ResNet101 architecture. Compared to the baseline model, this technique achieves the better detection effect on aerial targets within the DOTA dataset. [19] Deep ensemble techniques such as stacking, bagging, and boosting include various CoAtNet models trained on the SPARK datasets. By using CoAtNet, it obtained the advantages of both ConvNet and Transformer.

The meta-learner in [20] OBBStacking could handle the issues of performance disparities, calibration inconsistencies and inter-model redundancy across the neural network models during ensemble integration. Derived from the YOLOv5 architecture, a detection approach of remote sensing image (RSI-YOLO) was proposed [21], has demonstrated for its highly effective and broad applicability. The original YOLOv5 algorithm's feature extraction capabilities are enhanced by integrating a a spatial context enhancement module and semantic channel emphasis module and upgrading the PAN based feature aggregation of the neck region of network with a weighted Bi-FPN framework, thereby resulting in feature refinement and richer feature integration.

3. Methodology

The three different categories of object detection:

Object Detection (OD): OD aims to detect objects irrespective of their class category. OD algorithms typically present several possible region proposals, and then the best candidate is selected based on certain criteria.

Salient Object Detection (SOD): The concept of human attention mechanism used in the SOD algorithm to highlight and detect objects in an image or video.

Category-Specific Object Detection (COD): Multiple objects can be detected using COD. In contrast to OD and SOD, COD predicts the categorical class and location of an object in an image or video.

Traditional Approaches to Object Detection

Extraction of manual features and machine learning classifiers are the primary components of conventional target detection techniques for remote sensing imagery. These techniques offer outstanding compatibility of features, which include texture, color, and scale of objects. However, when traditional methods are applied to complex scenes, object scaling and occlusion can have some limitations. Traditional techniques also have certain disadvantages in terms of analytical precision and computational performance due to the vast coverage area of remote sensing images and high resolution [25]. Table 1 summarizes various techniques used in the traditional approach b their accuracy.

Table 1: Techniques used by the traditional approach

Technique	Image Type	Accuracy (approx.)	Datasets Used
Edge Detection	Grayscale, RGB	60-70%	AID, DOTA
Region-Based	RGB	65-75%	AID, DIOR
Classical ML	Multi spectral, RGB	70-80%	xView, DOTA
Template Matching	Grayscale, RGB	50-65%	AID
Feature Engineering (SIFT)	RGB, Multi spectral	65-75%	DIOR

Deep Learning-Based Object Detection

Convolutional Neural Networks (CNN's) serves in target localization, applying machine learning concepts to identify and analyse visual components in images. It serves as an essential element for emerging applications in computer vision. Generally, for single and multiple object detection, fast and precise results can be generated by Deep Neural Networks because CNNs support automated learning and eliminate manual engineering efforts [23].

Object detection is not possible without building a model. These object detection models are trained using hundreds of thousands of visual assets and then automatically optimize their detection accuracy. Datasets make training and refining models more efficient. Table 2 summarizes various technique used by Deep Learning methods with its accuracy.

Table 2: Techniques used by Deep Learning Methods

Technique	Image Type	Accuracy (approx.)	Datasets Used
Transformer- Based Models	RGB	80-90%	DIOR, AID
Hybrid Architectures	RGB, Multi-spectral	90-95%	DOTA, xView
Two-Stage Detectors (FRCNN)	RGB, Hyper-spectral	90-98%	DIOR, AID

Regression-Based Object Detection

Regression-based object detection is a technique that treat the process of detecting and localization of objects in an image as a regression problem. That is, rather than first generating potential object regions and then classifying them, it predicts the coordinates in bounding box of objects in an image. A good example of this approach is the "You Only Look Once" (YOLO) algorithm. The algorithm directly generates the bounding box positions and class probabilities for each target data in a single step. Table 3 summarizes various technique used by Regression-based method with its accuracy [24].

Table 3: Techniques used by Regression-based method

Technique	Image Type	Accuracy (approx.)	Datasets Used
CNN (YOLO, U-Net)	RGB, Multi spectral	85-95%	DOTA, xView
Single-Stage Detectors (YOLO)	RGB	85-95%	DOTA, xView

Dataset

Remote sensing datasets are essential for various applications, including scene classification, object detection, land use analysis, and disaster monitoring. Among the widely used datasets in aerial and satellite imagery, AID, DOTA, DIOR, and xView play a significant role in advancing remote sensing research and AI-based Geo-spatial analysis. Table 4 provides a general introduction to these datasets, highlighting their key features, applications, and importance in remote sensing.

Table 4: Dataset Description and Type

Dataset	Type	Description
AID	Object Detection High-resolution aerial images	Urban Environment 30 classes e.g., airport, farmland, forest, residential, industrial, river, parking lot.
DOTA	Satellite Object Detection Large-scale aerial image dataset	Contains images of 15 different object categories, like ships, vehicles, and buildings.
DIOR	Satellite Object Detection Optical remote sensing images.	Industrial and urban settings. 20 classes airplane, bridge, ship, windmill, vehicle, dam, train station)
XVIEW	Satellite Object Detection High-resolution aerial images	60 different types of objects, such as vehicles, ships, and buildings.

Comparison Chart

A comparison of traditional, Deep Learning, and Regression methods across various challenges is summarized in Table 6. The efficiency scores of different algorithms are compared in Figure 1. The performance, cost, and complexity are compared in Figure 2.

Table 5: Comparison chart

Challenges	Traditional	Deep Learning Method	Regression Method
Accuracy	70-80%	90-98%	85-95%
Complex Backgrounds	Difficult	Difficult	Difficult
Scale Variations	Difficult	Difficult	Difficult
Limited Labelled Data	Capable through feature engineering	Capable through feature engineering	Capable through feature engineering
Environmental Variations	Sensitive, requires precise adjustment	Sensitive, requires precise adjustment	Sensitive, requires precise adjustment

Table 6: Performance Comparison

Challenges	Traditional	Deep Learning Method	Regression Method
Accuracy	70	90	85
Complex Backgrounds	40	50	45

Challenges	Traditional	Deep Learning Method	Regression Method
Scale Variations	40	50	45
Limited Labelled Data	60	60	60
Environmental Variations	50	55	50

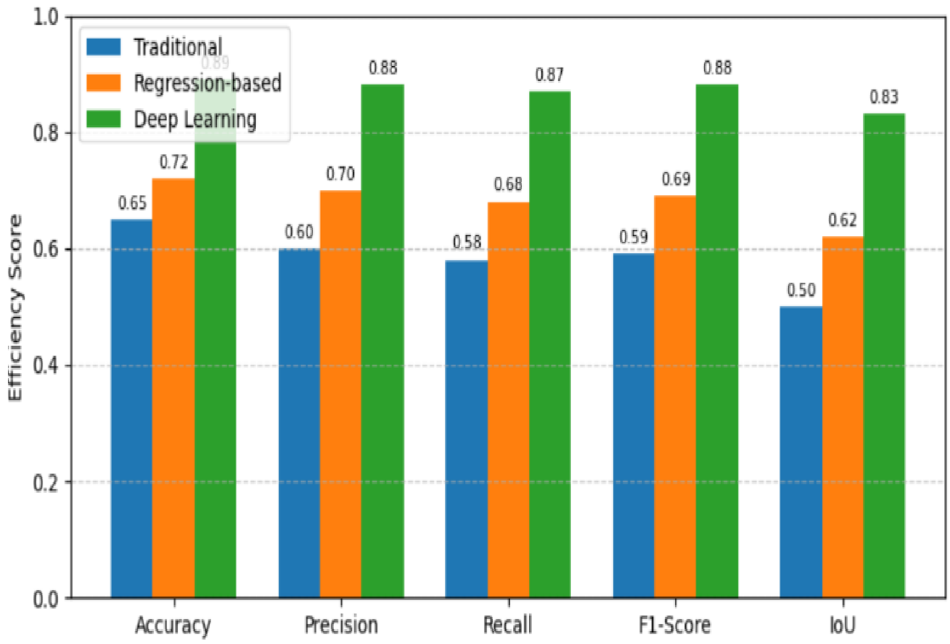

Figure 1: Efficiency Comparison

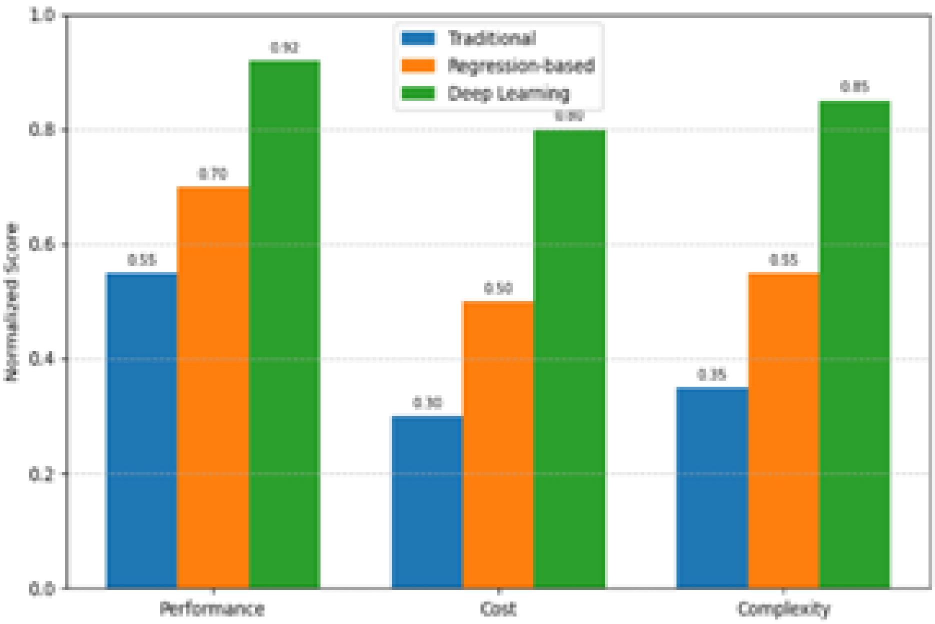

Figure 2: Performance Comparison

4. Conclusion

Advanced object detection techniques such as multi-scale feature extraction, attention mechanisms, directional object detection, and multimodal data fusion can significantly improve the quality and operational speed of remote sensing systems. These techniques address the various challenges such as scale variation, object localization, and sensor data integration. Applications in ecosystem monitoring, smart city planning, agricultural analysis, and disaster mitigation will improve decision-making, support sustainable development, and improve responses to global challenges. By leveraging these advances, remote sensing will continue to provide critical information across a wide range of sectors, ensuring better resource management and more informed planning.

Future advancement will focus on several significant technological advancement in remote sensing object detection. Deep learning frameworks, such as more efficient Convolutional Neural Networks (CNN's) and transformer-based models can be enhanced to provide better feature extraction, high performance, and accurate results, especially when dealing with challenging environments. Finally, continuous, real-time monitoring over large areas with minimal human assistance can be enabled with the help of autonomous surveillance systems using AI and machine learning which in turn improves the efficiency and scalability in the applications of urban planning and agriculture. These improvements will provide efficient autonomous remote sensing systems with more accurate and improved performance across a variety of fields.

References

[1] Boglaev, "A numerical method for solving nonlinear integro-differential equations of Fredholm type," J. Comput. Math., vol. 34, no. 3, pp. 262–284, May 2016, doi: 10.4208/jcm.1512-m2015-0241.

[2] Mao, M.; Zhao, H.; Tang, G.; Ren, J. In-Season Crop Type Detection by Combing Sentinel-1A and Sentinel-2 Imagery Based on the CNN Model. Agronomy 2023, 13, 1723.

[3] Cardama, F.J.; Heras, D.B.; Argüello, F. Consensus Techniques for Unsupervised Binary Change Detection Using Multi-Scale Segmentation Detectors for Land Cover Vegetation Images. Remote Sens. 2023, 15, 2889.

[4] Suleymanov, A.; Suleymanov, R.; Kulagin, A.; Yurkevich, M. Mercury Prediction in Urban Soils by Remote Sensing and Relief Data Using Machine Learning Techniques. Remote. Sens. 2023, 15, 3158.

[5] Tombe, R.; Viriri, S. Remote Sensing Image Scene Classification: Advances and Open Challenges. Geomatics 2023, 3, 137–155.

[6] Mou, L.; Bruzzone, L.; Zhu, X.X. Learning Spectral-Spatial-Temporal Features via a Recurrent Convolutional Neural Network for Change Detection in Multispectral Imagery. IEEE Trans. Geosci. Remote Sens. 2019, 2, 924–935.

[7] Khankeshizadeh, E.; Mohammadzadeh, A.; Moghimi, A.; Mohsenifar, A. FCD-R2U-net: Forest change detection in bi-temporal satellite images using the recurrent residual-based U-net. Earth Sci. Inform. 2022, 15, 2335–2347.

[8] Chen, Q.; Tang, S.; Yang, Q.; Fu, S. Cooper: Cooperative perception for connected autonomous vehicles based on 3d point clouds. In Proceedings of the 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), Dallas, TX, USA, 7–9 July 2019; IEEE: Piscataway Township, NJ, USA, 2019; pp. 514–524. [Google Scholar]

[9] Najibi, M.; Samangouei, P.; Chellappa, R.; Davis, L.S. Ssh: Single stage headless face detector. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4875–4884.

[10] Udayakumar, K., Revathi, M., Janani, K., & Nithiyanandam, N. (2025). Case Studies on Generative Adversarial Networks in Environmental Changes. In A. Vibhute, R. Dhanaraj, & K. Kale (Eds.), Generative Adversarial Networks for Remote Sensing (pp. 199-226). IGI Global Scientific Publishing.

[11] Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. Med. Image Anal. 2017, 42, 60–88.

[12] Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 2021, 44, 3523–3542.

[13] Vedantam, R.; Lawrence Zitnick, C.; Parikh, D. Cider: Consensus-based image description evaluation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4566–4575.

[14] Reddy, K.R.; Priya, K.H.; Neelima, N. Object Detection and Tracking—A Survey. In Proceedings of the 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, India, 12–14 December 2015; IEEE: Piscataway Township, NJ, USA, 2015; pp. 418–421.
[15] Kuehne, H.; Jhuang, H.; Garrote, E.; Poggio, T.; Serre, T. HMDB: A large video database for human motion recognition. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; IEEE: Piscataway Township, NJ, USA, 2011; pp. 2556–2563.
[16] Adegun, A.A.; Dombeu, J.V.F.; Viriri, S.; Odindi, J. State-of-the-Art Deep Learning Methods for Objects Detection in Remote Sensing Satellite Images. Sensors 2023, 23, 5849.
[17] Ma, J.; Hu, Z.; Shao, Q.; Wang, Y.; Zhou, Y.; Liu, J.; Liu, S. Detection of Large Herbivores in UAV Images: A New Method for Small Target Recognition in Large-Scale Images. Diversity 2022, 14, 624.
[18] Wang, W.; Shi, Y.; Zhang, J.; Hu, L.; Li, S.; He, D.; Liu, F. Traditional Village Building Extraction Based on Improved Mask R-CNN: A Case Study of Beijing, China. Remote Sens. 2023, 15, 2616.
[19] Wu, Q.; Feng, D.; Cao, C.; Zeng, X.; Feng, Z.; Wu, J.; Huang, Z. Improved Mask R-CNN for Aircraft Detection in Remote Sensing Images. Sensors 2021, 21, 2618.
[20] Space Object Recognition with Stacking of CoAtNets using Fusion of RGB and Depth Images,2023, IEEE Access.
[21] OBBStacking: An Ensemble Method for Remote Sensing Object Detection,2023
[22] RSI-YOLO: Object Detection Method for Remote Sensing Images Based on Improved YOLO,2023.
[23] Shengxi Gui.; Shuang Song.; Rongjun Qin.; Yang Tang. Remote Sensing Object Detection in the Deep Learning Era—A Review, 2024.
[24] Jiajia Wang.; Ran Wu3.; Xiaqiong Yu. An Algorithm of Object Detection Based on Regression Learning for Remote Sensing Images, Journal of Physics 2021doi:10.1088/1742-6596/1903/1/012039.
[25] P. Deepan.; Dr. L.R. Sudha. Object Detection in Remote Sensing Aerial Images: A Review, Volume 7, Issue 4 (July 2018).