# Geospatial Object Detection in High-Resolution Satellite Images Using an Improved Mayfly Optimization Algorithm Based on Long Short-Term Memory

**R. Hemavathy**
Department of CSE, RV College of Engineering, Bengaluru, India
hemavathyr@rvce.edu.in

**M. R. Anala**
Department of ISE, RV College of Engineering, Bengaluru, India
analamr@rvce.edu.in

**Prarthana Himanshu Upadhyaya**
Department of ISE, RV College of Engineering, Bengaluru, India
prarthanah.is23@rvce.edu.in

**R. Prarthana**
Department of ISE, RV College of Engineering, Bengaluru, India
prarthanar.is23@rvce.edu.in (corresponding author)

*Received: 19 April 2025 | Revised: 19 July 2025 | Accepted: 27 July 2025*

 *DOI: https://doi.org/10.48084/etasr.11599*

**ABSTRACT**

**Geospatial object detection in high-resolution satellite images has several applications, including environmental monitoring, military surveillance, marine monitoring, and urban planning. High Spatial and Spectral Resolution (HSSR) satellite imagery provides detailed ground-level information; however, research on this type of remote-sensing data remains limited, while object detection using such imagery remains a challenging task. In an attempt to provide effective object detection classification, this study utilizes an Improved Mayfly Optimization Algorithm based on Long Short-Term Memory (IMO-LSTM) framework trained and tested on the Dataset for Object Detection in Aerial images (DOTA). For feature extraction, Histogram of Oriented Gradients (HOG), Local Gabor Binary Pattern Histogram Sequence (LGBPHS), and Harris Corner Detection (HCD) were utilized, followed by IMO-based feature selection to retain the most relevant features. Lastly, the LSTM network is applied to classify objects in the selected images. When compared with existing methods such as Critical Feature-Capturing Network (CFC-Net), Self-adaptive Aspect Ratio Anchor (SARA), Cropping Region Proposal Network-based Scale Folding Network (CRPN-SFNet), and Ground Sample Distance (GSD), the proposed method achieved the best performance, attaining a mean Average Precision (mAP) of 89.53%.**

***Keywords-classification; deep learning; feature selection; geo-spatial object detection; and remote sensing images***

## I. INTRODUCTION

Remote sensing technologies have advanced significantly, enabling the acquisition of High Spatial and Spectral Resolution (HSSR) satellite images that empower a wide range of geospatial applications [1, 2]. Over the years, a variety of sensing methods have been developed for capturing such images. One prominent approach is passive optical remote sensing technology, known as Very High-Resolution (VHR) remote sensing, that records three electromagnetic waves emitted by ground objects and obtains different band information of the same target in the form of scanning or photography [3]. Complementing this, Hyperspectral Imaging (HIS) integrates mapping and spectrum analysis to precisely characterize material properties and object composition [4, 5].

Within this domain, object recognition requires not only accurate localization and classification of individual objects but also a deep understanding of the unique characteristics of aerial imagery [6]. The objects of interest, such as airplanes, vehicles, and tanks, typically exhibit distinct boundaries that are largely invariant to complex background variations in High Spatial Resolution Imagery (HSRI) [7-9]. Recent research efforts have multi-view modeling approaches that consider interactions among surface normals, lighting conditions, and albedo properties [10]. Nevertheless, aerial remote sensing images differ fundamentally from conventional ground-based optical images, which often rely on horizontal, side-perspective views to capture surface detail [11].

As noted in existing literature, many challenges in remote sensing have been mitigated by the rapid growth of high-resolution satellite and aerial datasets, enabling advances in land classification, building detection, semantic segmentation, and object detection [12]. Nonetheless, several limitations persist. Many state-of-the-art algorithms still require further optimization to improve accuracy and processing speed under constrained computational environments [13], and model generalization remains limited when training data is scarce [14].

Several studies have attempted to mitigate these issues. For instance, authors in [15] proposed a Region Proposal Network (RPN)-based approach using multi-scale detectors for geospatial object detection in high-resolution satellite imagery. Their framework was evaluated on benchmark datasets such as the Dataset for Object Detection in Aerial images (DOTA) and the NWPU VHR-10 dataset, and experimental outcomes demonstrated that the model effectively captured diverse spatial patterns and achieved strong performance. However, the approach incurred high computational costs and exhibited reduced accuracy when training multiple detectors within a unified framework.

In another attempt, authors in [16] developed the Critical Feature-Capturing Network (CFC-Net) tailored for detecting arbitrarily oriented objects in remote sensing imagery. Evaluated on the University of Chinese Academy of Sciences High Resolution Aerial Object Detection Dataset (UCAS-AOD), High Resolution Ship Collections 2016 dataset (HRSC2016), and DOTA datasets, the method improved semantic feature representation and detection accuracy. Nevertheless, the approach introduced computational overhead and reduced suitability for resource-constrained applications. In another study, authors in [17] implemented a new Self-Adaptive Aspect Ratio Anchor (SARA), which was used to specifically analyze variations in object scale and aspect ratios present in remote sensing images. Using an Oriented Box Decoder (OBD) to encode object orientation, their method achieved improved performance on the DOTA dataset. However, their framework struggled to accurately capture irregular shapes and fine details, diminishing robustness in complex scenarios.

Moreover, authors in [18] implemented a low-latency framework for fast object detection in large-scale remote sensing datasets. Using DOTA for evaluation, the method improved detection speed and accuracy but remained sensitive to occlusions and small-object scenarios due to reliance on contextual region information. Authors in [19] implemented a Ground Sample Distance (GSD)-guided multi-scale reasoning method for object detection in aerial images. Their approach demonstrated improved performance on DOTA by leveraging detailed convolutional representations, but scale adaptation limitations restricted performance in dense or highly complex scenes. Lastly, authors in [20] demonstrated that combining Mayfly Optimization (MO) with Deep Learning (DL) enhances object detection and classification in surveillance video data, highlighting the algorithm's cross-domain adaptability.

Across these approaches, common limitations include difficulty capturing irregular object shapes, reduced accuracy in complex environments, and limited efficiency in resource-constrained settings. To address these gaps, our research proposes the Improved Mayfly Optimization Algorithm based on the Long Short-Term Memory (IMO-LSTM) model, which combines the IMO for the selection of optimal features and the LSTM for the classification process. The proposed model was trained and tested on the DOTA dataset and was evaluated based on the mean Average Precision (mAP) metric.

## II. METHODOLOGY

The architecture of the proposed IMO-LSTM is illustrated in Figure 1. For the model training, the DOTA dataset is employed, while Min-Max normalization is applied as the preprocessing step. Feature extraction was performed using Local Gabor Binary Pattern Histogram Sequence (LGBPHS), Histogram of Oriented Gradients (HOG), and Harris Corner Detection (HCD), while feature selection was performed using the IMO algorithm. At the last layer, classification was carried out using LSTM.

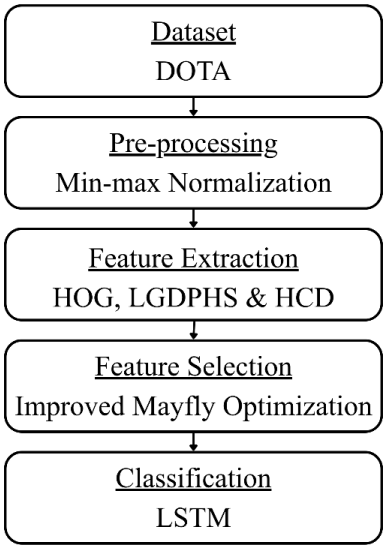


Fig. 1. Architecture of the proposed IMO-LSTM.

### *A. DOTA Dataset*

The DOTA dataset is one of the largest remote sensing datasets [21], consisting of 15 object categories, including

Baseball Diamond (BD), Plane (PL), Bridge (BR), Ground Track Field (GTF), Storage Tank (ST), Ship (SH), Tennis Court (TC), Large Vehicle (LV), Roundabout (RA), Helicopter (HC), Basketball Court (BC), Soccer Ball Field (SBF), Swimming Pool (SP), Small Vehicle (SV), and Harbor (HA).

The dataset comprises 2,806 images of various perspectives, object scales, and resolutions. Original images in the dataset are cropped into patches ranging from 800×800 to 4000×4000 pixels, and the dataset provides 188,282 annotated instances (horizontal and oriented), making it suitable for detecting both aligned and rotated objects. The dataset was partitioned into an 80% split used for training and the other 20% for testing the proposed model.

### *B. Preprocessing*

In the preprocessing stage, all images are first converted to grayscale. Then, for each image patch $I(x, y)$ each pixel intensity is scaled to the [0,1] range using the Min-Max normalization:

$$i = \frac{n_i - Min}{Max - Min} \quad (1)$$

where $n_i$ represents the numeric feature of the $i^{th}$ sample, while $Min$ and $Max$ denote the minimum and maximum values of the numeric features, respectively. Normalizing each patch independently ensures that illumination differences across patches do not bias the subsequent LGBPHS, HOG, and HCD feature extraction processes.

### *C. Feature Extraction*

#### *1) LGBPHS*

The preprocessed images are first transformed into a set of Gabor Magnitude Pictures (GMPs) using Gabor filters at multiple scales and orientations. Local Gabor Binary Pattern (LGBP) operators are then applied to each GMP to encode local texture information. Due to their invariance to monotonic grayscale changes, LGBP features are robust to illumination variations. Each GMP is partitioned into non-overlapping rectangular regions, and an LGBP histogram is computed for each region. The final LGBPHS feature vector is obtained by concatenating all regional histograms, thereby jointly encoding spatial layout and texture characteristics. This descriptor is particularly effective for capturing the structured textures commonly observed in DOTA classes such as runways, storage tanks, and sports courts.

#### *2) HOG*

The HOG descriptor is employed to capture edge and gradient-based shape information in satellite imagery. It operates on localized cells, preserving object orientation and geometric structure without being affected by photometric variations. Pixel-level gradients are computed from normalized grayscale patches, which are then divided into small spatial cells. For each cell, a one-dimensional histogram of gradient orientations is accumulated. These local histograms are subsequently normalized within larger spatial blocks to improve robustness to illumination and contrast changes. The resulting HOG descriptors effectively represent object contours and are well-suited for detecting large-scale spatial changes in remote sensing scenes.

#### *3) HCD*

HCD features are used to distinguish foreground objects from background regions. Building upon the Moravec corner detection principle, HCD evaluates grayscale intensity variations within a sliding window to identify points of significant structural change. For a normalized image patch, the Harris response function is computed based on intensity shifts in the horizontal and vertical directions. Corner points exhibit strong responses due to abrupt grayscale changes, whereas background regions produce weaker responses. These detected corner indicators represent local structural discontinuities and are effective in delineating object boundaries. The extracted HCD features are then forwarded to the feature selection stage.

### *D. Feature Selection*

After extracting the feature vectors, the Improved Mayfly Optimization Algorithm (IMOA) is applied to select the most informative subset of features.

#### *1) Mayfly Optimization Algorithm*

The original MOA is a population-based metaheuristic inspired by the mating behavior of male and female mayflies. Similar to Particle Swarm Optimization (PSO), each mayfly updates its position $p_i(t)$ and velocity $v_i(t)$ across iterations based on fitness evaluations:

$$p_i(t+1) = p_i(t) + v_i(t+1) \quad (2)$$

If the current fitness value $f(x_i)$ of the mayfly improves over its historical best fitness $f(x_{h_i})$, the velocity of a male mayfly is updated according to:

$$v_i(t+1) = g \cdot v_i(t) + \alpha_1 e^{-\beta\gamma_p^2}\left[x_{h_i} - x_i(t)\right] + \alpha_2 e^{-\beta\gamma_g^2}\left[x_g - x_i(t)\right] \quad (3)$$

where $\alpha_1$ and $\alpha_2$ are acceleration coefficients regulating attraction toward the personal and global best positions, $g$ is an inertia factor controlling the influence of the previous velocity, $x_g$ is the global best position, $\beta$ is a decay coefficient, while $\gamma_p$ and $\gamma_g$ denote Cartesian distances to the personal and global best positions. The Euclidean (Cartesian) distance between any two vectors $x_i$ and $x_j$ is:

$$\left|\left|x_i - x_j\right|\right| = \sqrt{\sum_{k=1}^{n}(x_{ik} - x_{jk})^2} \quad (4)$$

If the fitness does not improve, the male mayfly performs a random "dance" movement, and its velocity is updated as:

$$v_i(t+1) = g \cdot v_i(t) + d \cdot \gamma_1 \quad (5)$$

where $d$ is the dancing coefficient controlling random exploration and $\gamma_1 \in [-1,1]$ is a random value. Female mayflies are attracted to their corresponding male partners and update their velocity using:

$$v_i(t+1) = g \cdot v_i(t) + \alpha_3 e^{-\beta\gamma_{mf}^2}[x_i(t) - y_i(t)] \quad (6)$$

where $\alpha_3$ is an acceleration constant for female motion, $y_i(t)$ is the position of the corresponding male, and $\gamma_{mf}$ is the distance between the male and female.

If the fitness of the female mayfly $f(y_i)$ is worse than that of the male $f(x_i)$, an additional dance-based update is applied:

$$v_i(t) = g \cdot v_i(t) + fl \cdot \gamma_2 \quad (7)$$

where $fl$ is an additional dancing coefficient, and $\gamma_2 \in [-1,1]$ is a random value. Offspring generation is performed using a linear combination (crossover) of the parent positions:

$$offspring_1 = L \cdot female + (1 - L) \cdot male \quad (8)$$

$$offspring_2 = L \cdot male + (1 - L) \cdot female \quad (9)$$

where $L$ is a random variable sampled from a Gaussian distribution, controlling the contribution of each parent.

*2) Improved MOA*

The above equations indicate that, under certain conditions, mayfly velocities may fluctuate excessively, leading to inefficient exploration. To address this issue, the IMOA introduces a modified velocity update that accounts for weighted distance information when selecting the best candidate. The initial improved velocity $v_p$ formulation is given by:

$$v_p = \alpha_i e^{-\beta\gamma_j^2}(p_j - p_i) \quad (10)$$

and since this formulation may overemphasize distant candidates, it is modified as:

$$v_p = \alpha_i e^{-\frac{\beta}{\gamma_j}}(p_j - p_i) \quad (11)$$

This modification enables larger velocity updates when individuals are close (fine-grained exploitation) and smaller updates when they are far apart (controlled exploration).

In this framework, IMOA is used to select an optimal subset of features from the concatenated LGBPHS, HOG, and HCD feature vectors. Each mayfly encodes a binary feature-selection vector, where 1 indicates a selected feature and 0 indicates a discarded feature. The fitness function is defined as the classification accuracy of an LSTM trained on the selected feature subset using a validation split.

*E. Classification*

The selected features are classified using an LSTM network, where the class label is treated as the output of the LSTM, while the IMOA-selected features serve as input. LSTMs are well-suited for this task due to their ability to retain relevant patterns over long temporal dependencies while discarding irrelevant information.

Each LSTM unit consists of an input gate $i_n$, forget gate $f_n$, cell state $c_n$, and output gate $o_n$, defined as:

$$i_n = \sigma(W_{ih}h_{n-1} + W_{ia}a_t + b_i) \quad (12)$$

$$f_n = \sigma(W_{fh}h_{n-1} + W_{fa}a_t + b_f \quad (13)$$

$$c_n = f_n \times c_{n-1} +$$

$$i_n \times tanh(W_{ch}h_{n-1} + W_{ca}a_t + b_c) \quad (14)$$

$$o_n = \sigma(W_{oh}h_{n-1} + W_{oa}a_t + b_o) \quad (15)$$

where $n$ denotes the time step, $h_{n-1}$ is the hidden state from the previous time step, $a_t$ is the input feature vector at time $t$, $W$ and $b$ are weight matrices and bias vectors, $\sigma(\cdot)$ is the sigmoid activation function, and $tanh(\cdot)$ is the hyperbolic tangent function. The output of the LSTM unit is computed as:

$$h_n = o_n \times tanh(c_n) \quad (16)$$

where $h_n$ is the hidden state at time step $n$. The cell state sequence $\{c_n\}_{n=1}^{N}$ captures both short-term and long-term dependencies, enabling effective classification of geospatial objects.

## III. RESULTS AND DISCUSSION

The system requirements for the implemented research project include a Windows 10 (64-bit) operating system, an Intel Core i7 processor, and 16 GB of Random Access Memory (RAM). The effectiveness of the proposed model is evaluated using mAP.

For geospatial object detection, Average Precision (AP) is defined as the area under the precision-recall curve over the recall range [0,1] and is widely used as a performance metric. Precision $P$ and recall $R$ are used to construct the Precision-Recall Curve (PRC) and are defined as follows:

$$P = \frac{TP}{TP+FP} \quad (17)$$

$$R = \frac{TP}{TP+FN} \quad (18)$$

$$AP = \int_0^1 P(R)dR \quad (19)$$

where $TP$ denotes the number of correctly detected targets (true positives), $FP$ represents the number of incorrectly detected targets (false positives), and $FN$ denotes the number of non-detected targets (false negatives). For multi-class object detection, mAP is computed as the average AP across all object categories:

$$mAP = \frac{\sum_N^{i=1} AP_i}{N} \quad (20)$$

where $N$ is the total number of classes. All reported mAP values on the DOTA dataset were obtained using the official DOTA evaluation server.

*A. DOTA Dataset Results*

The baseline results in Table I were obtained from original publications and verified against the official DOTA leaderboard [22]. All methods were evaluated on the same test set using an identical evaluation protocol (IoU = 0.5) through the official DOTA evaluation server.

As shown in Table I, the proposed IMO-LSTM demonstrated superior performance across all 15 object categories, achieving an overall mAP of 89.53%. This corresponds to an improvement of 6.26 percentage points over RNN (83.27%), 7.55 points over GAN (81.98%), and 8.59 points over Faster R-CNN (80.94%). Notably, the proposed model achieved higher detection accuracy in structured objects

with consistent geometric patterns, including tennis courts (90.98% AP), planes (90.10% AP), basketball courts (89.62% AP), ships (89.21% AP), and storage tanks (88.57% AP). This could be attributed to the LGBPHS and HOG features ability to capture regular geometric characteristics, while the LSTM effectively models spatial dependencies in densely clustered regions.

On the other hand, in objects characterized by extreme aspect ratios or significant scale variations, the model exhibited comparatively lower absolute performance, such as in bridges (53.60% AP), roundabouts (68.71% AP), and helicopters (67.90% AP). Nevertheless, these categories still demonstrate notable improvements over baseline methods.

An illustration of the capabilities of the proposed approach is shown in Figure 2, which highlights key capabilities such as accurate detection of dense small objects without false positives, effective multi-scale detection within single images, handling of extreme aspect ratios with compact bounding boxes, and robust performance across arbitrary object orientations ranging from 0° to 360°.

TABLE I. AP COMPARISON WITH STATE-OF-THE-ART TECHNIQUES ON DOTA-V1.0 DATASET

| Methods | BD | PL | GTF | BR | LV | SV | TC | SH | ST | BC | RA | SBF | SP | HC | HA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CNN | 72.30 | 85.47 | 71.30 | 39.62 | 79.32 | 65.30 | 81.24 | 81.25 | 85.91 | 79.64 | 57.32 | 60.91 | 68.79 | 59.28 | 48.99 |
| $R^2$ CNN | 74.12 | 79.21 | 82.78 | 48.96 | 76.91 | 69.58 | 79.63 | 84.47 | 70.12 | 75.82 | 59.14 | 52.64 | 58.76 | 60.17 | 56.25 |
| Faster R-CNN | 82.20 | 86.59 | 74.56 | 40.87 | 72.21 | 70.83 | 72.24 | 78.65 | 80.99 | 82.50 | 65.56 | 56.82 | 61.49 | 61.39 | 62.21 |
| R-DFPN | 73.26 | 80.80 | 85.62 | 45.54 | 80.05 | 71.29 | 85.49 | 73.57 | 76.34 | 84.75 | 57.98 | 65.49 | 57.71 | 63.45 | 50.53 |
| RNN | 80.92 | 70.13 | 75.68 | 51.36 | 70.62 | 76.61 | 82.29 | 80.27 | 75.90 | 80.28 | 64.73 | 59.61 | 65.55 | 57.81 | 59.47 |
| GAN | 79.65 | 78.84 | 81.11 | 49.95 | 78.26 | 72.48 | 78.87 | 86.30 | 81.21 | 81.29 | 58.19 | 58.76 | 60.81 | 65.92 | 66.18 |
| Proposed | 85.96 | 90.10 | 89.35 | 53.60 | 80.26 | 79.42 | 90.98 | 89.21 | 88.57 | 89.62 | 68.71 | 67.82 | 70.28 | 67.90 | 69.96 |

The categories short names are ground field track (GTF), plane (PL), baseball diamond (BD), large vehicle (LV), ship (SH), bridge (BR), swimming pool (SP), storage tank (ST), soccer ball field (SBF), small vehicle (SV), roundabout (RA), helicopter (HC), basketball court (BC), tennis court (TC), and harbor (HA)

Fig. 2. DOTA dataset with implemented approach visualization results of detection.

### *B. Comparative Analysis*

The comparative analysis of the performance of the proposed model against the performance of models in other publications using the DOTA dataset is listed in Table II. The implemented method outperformed all other models, including CFC-Net [16], SARA [17], Cropping Region Proposal Network-based Scale Folding Network (CRPN-SFNet) [18], and GSD [19]. Specifically, IMO-LSTM achieved the highest mAP score of 89.53%, which is significantly higher than the 2nd best performance of the SARA model, achieving a score of 79.91%.

TABLE II. COMPARATIVE ANALYSIS OF THE IMPLEMENTED METHOD USING THE DOTA DATASET

| Authors | Methods | mAP (%) |
|---|---|---|
| [16] | CFC-Net | 73.50 |
| [17] | SARA | 79.91 |
| [18] | Fast inference framework | 79.54 |
| [19] | GSD | 68.28 |
| Proposed method | IMO-LSTM | 89.53 |

## IV. CONCLUSION

This study focuses on the development of an efficient feature selection strategy for object detection using a lightweight Deep Learning (DL) framework that employs an Improved Mayfly Optimization (IMO) algorithm integrated with a Long Short-Term Memory (LSTM) model to perform object detection classification.

Feature vectors are extracted from normalized remote sensing images using a combined approach based on Local Gabor Binary Pattern Histogram Sequence (LGBPHS), Histogram of Oriented Gradients (HOG), and Harris Corner Detection (HCD) techniques. The IMO algorithm is then employed to select the most relevant feature vectors, significantly enhancing classification performance. These optimized features are used as input to the LSTM classifier. The proposed IMO-LSTM framework is evaluated on the Dataset for Object Detection in Aerial images (DOTA), where it achieves a mean Average Precision (mAP) of 89.53%, outperforming existing object detection classification methods. Future work will extend the framework to support oriented bounding box detection for objects with varying orientations.

## REFERENCES

[1] S. D. Khan, L. Alarabi, and S. Basalamah, "DSMSA-Net: Deep Spatial and Multi-scale Attention Network for Road Extraction in High Spatial Resolution Satellite Images," *Arabian Journal for Science and Engineering*, vol. 48, no. 2, pp. 1907–1920, Feb. 2023, https://doi.org/10.1007/s13369-022-07082-z.

[2] D. Pan, M. Zhang, and B. Zhang, "A Generic FCN-Based Approach for the Road-Network Extraction From VHR Remote Sensing Images – Using OpenStreetMap as Benchmarks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 2662–2673, 2021, https://doi.org/10.1109/JSTARS.2021.3058347.

[3] X. Zhao, K. Zhao, S. Li, and X. Wang, "GeSANet: Geospatial-Awareness Network for VHR Remote Sensing Image Change

Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023, https://doi.org/10.1109/TGRS.2023.3272550.

[4] S. Jia, S. Zhu, Z. Wang, M. Xu, W. Wang, and Y. Guo, "Diffused Convolutional Neural Network for Hyperspectral Image Super-Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023, https://doi.org/10.1109/TGRS.2023.3250640.

[5] F. Pan, Z. Wu, Q. Liu, Y. Xu, and Z. Wei, "DCFF-Net: A Densely Connected Feature Fusion Network for Change Detection in High-Resolution Remote Sensing Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 11974–11985, 2021, https://doi.org/10.1109/JSTARS.2021.3129318.

[6] V. Chalavadi, P. Jeripothula, R. Datla, S. B. Ch, and K. M. C, "mSODANet: A network for multi-scale object detection in aerial images using hierarchical dilated convolutions," *Pattern Recognition*, vol. 126, June 2022, Art. no. 108548, https://doi.org/10.1016/j.patcog.2022.108548.

[7] I. M. Iqbal, H. Balzter, Firdaus-e-Bareen, and A. Shabbir, "Mapping Lantana camara and Leucaena leucocephala in Protected Areas of Pakistan: A Geo-Spatial Approach," *Remote Sensing*, vol. 15, no. 4, Feb. 2023, Art. no. 1020, https://doi.org/10.3390/rs15041020.

[8] R. Agarwal, S. Jindal, S. Narain, R. Kaushal, and K. Yadav, "A novel framework for fine-grained spatio-temporal change detection in satellite images," *Multimedia Tools and Applications*, vol. 83, no. 1, pp. 1241–1260, Jan. 2024, https://doi.org/10.1007/s11042-023-14705-6.

[9] Z. Dong, M. Wang, Y. Wang, Y. Liu, Y. Feng, and W. Xu, "Multi-Oriented Object Detection in High-Resolution Remote Sensing Imagery Based on Convolutional Neural Networks with Adaptive Object Orientation Features," *Remote Sensing*, vol. 14, no. 4, Feb. 2022, Art. no. 950, https://doi.org/10.3390/rs14040950.

[10] Z. Hu, P. Tao, X. Long, and H. Wang, "Shading aware DSM generation from high resolution multi-view satellite images," *Geo-spatial Information Science*, vol. 27, no. 2, pp. 398–407, Mar. 2024, https://doi.org/10.1080/10095020.2022.2125837.

[11] S. Zhang, Z. Shao, X. Huang, L. Bai, and J. Wang, "An internal-external optimized convolutional neural network for arbitrary orientated object detection from optical remote sensing images," *Geo-spatial Information Science*, vol. 24, no. 4, pp. 654–665, Oct. 2021, https://doi.org/10.1080/10095020.2021.1972772.

[12] D. Wang, X. Chen, N. Guo, H. Yi, and Y. Li, "STCD: efficient Siamese transformers-based change detection method for remote sensing images," *Geo-spatial Information Science*, vol. 27, no. 4, pp. 1192–1211, July 2024, https://doi.org/10.1080/10095020.2022.2157762.

[13] Z. Zhang, H. Zheng, J. Cao, X. Feng, and G. Xie, "FRS-Net: An Efficient Ship Detection Network for Thin-Cloud and Fog-Covered High-Resolution Optical Satellite Imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 2326–2340, 2022, https://doi.org/10.1109/JSTARS.2022.3227322.

[14] Y. Xu *et al.*, "Luojia-HSSR: A high spatial-spectral resolution remote sensing dataset for land-cover classification with a new 3D-HRNet," *Geo-spatial Information Science*, vol. 26, no. 3, pp. 289–301, July 2023, https://doi.org/10.1080/10095020.2022.2070555.

[15] S. D. Khan, L. Alarabi, and S. Basalamah, "A Unified Deep Learning Framework of Multi-scale Detectors for Geo-spatial Object Detection in High-Resolution Satellite Images," *Arabian Journal for Science and Engineering*, vol. 47, no. 8, pp. 9489–9504, Aug. 2022, https://doi.org/10.1007/s13369-021-06288-x.

[16] Q. Ming, L. Miao, Z. Zhou, and Y. Dong, "CFC-Net: A Critical Feature Capturing Network for Arbitrary-Oriented Object Detection in Remote-Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022, https://doi.org/10.1109/TGRS.2021.3095186.

[17] J.-B. Hou, X. Zhu, and X.-C. Yin, "Self-Adaptive Aspect Ratio Anchor for Oriented Object Detection in Remote Sensing Images," *Remote Sensing*, vol. 13, no. 7, Mar. 2021, Art. no. 1318, https://doi.org/10.3390/rs13071318.

[18] L. Liu *et al.*, "Object Detection in Large-Scale Remote Sensing Images With a Distributed Deep Learning Framework," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 8142–8154, 2022, https://doi.org/10.1109/JSTARS.2022.3206085.

[19] W. Li, W. Wei, and L. Zhang, "GSDet: Object Detection in Aerial Images Based on Scale Reasoning," *IEEE Transactions on Image Processing*, vol. 30, pp. 4599–4609, 2021, https://doi.org/10.1109/TIP.2021.3073319.

[20] V. Saikrishnan and M. Karthikeyan, "Mayfly Optimization with Deep Learning-based Robust Object Detection and Classification on Surveillance Videos," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11747–11752, Oct. 2023, https://doi.org/10.48084/etasr.6231.

[21] *DOTADataset*. (2021), C. T. Doloriel. [Online]. Available: https://www.kaggle.com/datasets/chandlertimm/dota-data.

[22] J. Ding and G.-S. Xia. "DOTA-v1.0 Leaderboard." DOTA: A Large-Scale Benchmark and Challenges for Object Detection in Aerial Images. [Online]. Available: https://captain-whu.github.io/DOTA/results.