Journal of Computer Science and Frontier Technologies
ISSN Print: 3104-4204
ISSN Online: 3104-4212
CD ISSN: 3104-4220
https://doi.org/10.63313/JCSFT.9003

Research Progress on Urban Built-up Area Identification in High-Resolution Remote Sensing Images Using Deep Learning

Chujun Li[1,2,a], Md Gapar Md Johar[3,b,*]
[1]School of Engineering, Xi'an International University, Xi'an 710077, China
[2]School of Graduate Studies, Postgraduate Centre, Management and Science University, Shah Alam, Selangor, 40100, Malaysia
[3]Software Engineering and Digital Innovation Centre, Management and Science University, Shah Alam, Selangor, 40100, Malaysia.
Email: [a]lichujun2021@163.com [b]mdgapar@msu.edu.my

How to cite this paper: Li, C. J., & Md Johar, M. G. (2025). Research Progress on Urban Built-up Area Identification in High-Resolution Remote Sensing Images Using Deep Learning. Journal of Computer Science and Frontier Technologies, 1(1), 28 - 37. ISSN Print: 3104-4204; ISSN Online: 3104-4212.
https://doi.org/10.63313/JCSFT.9003
Published: 2025-08-05

Abstract

Against the backdrop of rapid urbanization, the urban built-up area, as the core component of urban development, requires accurate identification for urban planning, land use monitoring, and sustainable development assessment. How-ever, the complex internal features, significant scale variations, and blurred boundaries within high-resolution remote sensing images often lead to insuffi-cient accuracy with traditional identification methods. Deep learning, lever-ag-ing its powerful capabilities for automatic feature learning and semantic ex-trac-tion, has become the core technology for identifying urban built-up areas in high-resolution remote sensing images. This paper provides a comprehensive review of research on high-resolution remote sensing image-based urban built-up area identification from a deep learning perspective: Firstly, it clarifies the connotation of urban built-up areas and summarizes their identification characteristics and challenges. Secondly, it categorizes the commonly used mul-ti-source data types for identification, including high-resolution remote sensing imagery, multispectral data, and LiDAR data. Thirdly, it summarizes methodo-logical advancements in built-up area identification using deep learning-based semantic segmentation models (e.g., CNN, FCN, U-Net, Transformer). Finally, it analyzes the current problems and challenges in the research and proposes prospects and suggestions for future development directions.

Keywords

Deep Learning; High-resolution Remote Sensing Image; Urban Built-up Area; Semantic Segmentation; Image Recognition Style

1. Introduction

Urbanization is the dominant global trend, and the continuous growth of urban populations drives the ongoing expansion of urban built-up areas. Accurately delineating the extent and morphology of built-up areas has become a critical prerequisite for urban planning, ecological protection, and sustainable development. Traditional methods for identifying urban built-up areas rely on manual visual interpretation or statistical classification based on spectral features. These approaches struggle to handle the intricate details and variations of complex features within high-resolution remote sensing images, suffering from issues such as low accuracy, poor efficiency, and strong subjectivity. Since 2017, with the rapid advancement of deep learning technology, its powerful capabilities in the field of image semantic segmentation have gradually been applied to the remote sensing domain. Deep learning can autonomously learn image features from low to high levels through multi-level network architectures, enabling precise extraction from pixel-level to semantic-level information, thereby offering a new approach for identifying urban built-up areas in high-resolution remote sensing images. Currently, deep learning-based built-up area identification has become a research hotspot. However, systematic organization and trend analysis of related methods remain insufficient. Therefore, this paper reviews the research progress in deep learning-driven identification of urban built-up areas using high-resolution remote sensing images, aiming to serve as a reference for subsequent studies.

2. Connotation and Identification Features of Urban Built-up Areas

2.1. Connotation

Urban built-up areas refer to the regions within a city characterized by a concentrated distribution of industrial, transportation, commercial, residential buildings, and infrastructure at a significant scale. It serves as the primary space for urban populations and economic activities. Its core features are dominated by human-made elements, including buildings (e.g., residences, commercial complexes, factories), road networks, hardened surfaces, and ancillary facilities (e.g., parking lots, plazas). These features create a distinct contrast with natural areas (e.g., green spaces, water bodies) and agricultural zones.

2.2. Identification Features and Challenges

2.2.1. Effective Extraction and Aggregation of Multi-scale Features

High-resolution remote sensing images contain a large number of objects with vastly different sizes. For example, small cars and large areas of road, grass, buildings, etc. Extracting rich multi-scale features is the key to effectively characterizing objects of different sizes. Previous methods simply extracted multi-scale features from a certain layer of the network, lacking representation of multi-semantic hierarchical features. The U-shaped network structure eliminates explicit extraction of multi-scale features, and implicitly fuses feature maps of different resolutions through its bottom-up encoder, top-down decoder, and lateral skip connection structure. Although this approach can fuse features of different semantic levels, it lacks effective representation of multi-scale contextual information.

2.2.2. Confusion Classification Sample Mining

There is a serious class imbalance problem in high-resolution remote sensing images. For example, the number of pixels of small-sized objects such as cars only accounts for a small proportion of the entire dataset. Currently, there are few neural network models designed for class imbalance problems, and most of the work solves this problem by designing loss functions. According to different methods for defining difficult samples, the loss functions that perform well can be divided into two categories. The first category is based on weighted cross-entropy loss for the number of samples in each class, with a typical representative being weighted cross-entropy loss. This type of loss achieves balance in the loss values of various samples by increasing the loss of minority class samples while reducing the loss of majority class samples. This method assumes that the classification of minority class samples is more difficult. The other category is based on the weighted prediction results of the model, with a typical representative being focal loss (FL). This type of loss ignores the class information of samples, and divides all samples into easy samples and difficult samples according to the prediction results of the model, increasing the loss of difficult samples and reducing the loss of easy samples.

2.2.3. Difficulty Area Classification

There are a large number of objects in high-resolution remote sensing images, and due to human activities, the distribution of these objects varies greatly. For example, areas such as parks are dominated by large areas of grassland, while residential areas are distributed with a large number of objects such as cars, trees, buildings, roads, and structures. The surface conditions vary greatly between different regions. In addition, due to the extremely large size of high-resolution remote sensing images, sliding windows are often used during training to cut large images into small pictures. This results in some small pictures containing only a single type of object, while others contain complex ground information, further exacerbating the differences in the distribution of ground objects in different training images. Generally speaking, accurate segmentation of complex areas with complex ground distribution is more difficult. Previous work has mostly been targeted at processing difficult samples, ignoring regional information, and there are fewer studies on processing complex areas in high-resolution remote sensing image semantic segmentation methods.

2.2.4. Difficulty Area Classification

Due to the characteristics of the imaging principle, there is often a certain degree of distortion and deformation in the edge part of objects in high-resolution remote sensing images. However, edge features are extremely important information in semantic segmentation tasks. Currently, most of the methods for extracting edge features in deep neural networks directly use 2D convolution to learn boundary information from the intermediate layer features of the backbone network, and ultimately use this information to predict the edge of the object. Some scholars will set an edge loss function for this edge prediction result to guide the convolution kernel to learn the edge information in the intermediate layer features more explicitly. However, in addition to edge features, the intermediate layer features also contain a lot of redundant information, which is essential for semantic segmentation. This kind of edge extraction method has certain blindness and strongly relies on the guidance of edge loss for model training, making it difficult to effectively represent the boundary information of the object.

3. Data Types for Deep Learning-Based Urban Built-up Area Identification

3.1. High-Resolution Optical Remote Sensing Imagery

High-resolution remote sensing images provide basic images for large-scale and accurate identification of urban built-up areas. In terms of spatial resolution, the remote sensing image accuracy of Google Earth satellite images, QuickBird satellites, and WorldView satellites can reach 0.3~1m, providing fine ground object details for semantic segmentation and identification of urban built-up areas (Feng, et al., 2022). Recognizing remote sensing images based on deep learning can accurately identify site space. For example, Pei, H. et al. used the deep learning MSG-GCN and high-resolution images to accurately identify the boundaries of complex plant communities (Pei, et al., 2023). Based on high-resolution remote sensing images, Shao, Z. et al. used the BRRNet to extract the outlines of smaller buildings to be demolished, and the final extraction accuracy was higher than 83% (Shao, Z. et al., 2022).

Wang, et al. proposed a model called MST-DeepLabv3+ for remote sensing image classification. It's based on the DeepLabv3+and can produce better results with fewer train parameters (Wang, et al., 2024). Ismael, et al. introduced a new unsupervised domain adaptation method that employs concurrently two levels of alignment (Ismael, et al., 2024). Jia, et al. introduced a novel deep learning model that leverages the principle of band combination in remote sensing images to enhance the efficiency and accuracy of semantic segmentation (Jia, et al., 2024). This research focuses not only on advancing the segmentation capabilities of remote sensing images but also on applying this technology in urban planning and land use to foster sustainable development in smart cities.

3.2. Multispectral and Hyperspectral Data

Multispectral data encompass multiple spectral bands, enabling the differentiation of buildings from vegetation, water bodies, and other features through spectral characteristics. Hyperspectral data further capture the fine-grained spectral responses of ground objects, aiding in identifying artificial features with distinct material compositions within built-up areas. The spectral features will be affected by the diversity of the built-up area and surrounding areas, increasing image uncertainty and confusion between categories, then cause the classification accuracy reduce. In recent years, machine learning and deep learning have been rapidly developed due to their high accuracy and efficiency, and have begun to be applied to urban built-up area recognition. In terms of spectral resolution, the sources of data sets such as EuroSAT and KSC, and multispectral satellites such as Sentinel-2 and AVIRIS can be used for remote sensing classification through spectral features when identifying urban built-up areas (Feng, et al., 2022).

By extracting complementary features and similar features between features, Wang, et al. used multiple spectral information of remote sensing images to achieve accurate segmentation of complex ground objects, especially for building corners and grass edges, which can achieve better results (Wang, et al., 2022). Based on deep learning, Zhang, et al. integrated spectral information and DSM information to improve the segmentation performance of remote sensing images, achieve more accurate segmentation in identifying grassland, and deal with the fuzzy boundary problem (Zhang, et al., 2022). To overcome limitations in hyperspectral image classification, including rigid spectral-spatial coupling, computational redundancy, and noise sensitivity, Lou, M., et al. proposed HyperTransXNet, a CNN-Transformer hybrid featuring adaptive spectral-spatial fusion through three novel modules (Lou, M., et al., 2025).

3.3. LiDAR Data

LiDAR data provides 3D elevation information of ground objects, enabling effective height differentiation of buildings. This capability resolves the "same-spectrum-different-object" issue inherent in optical imagery. To address limitations in multi-modal LULC classification, Pan, X., et al. proposed AFNKA, an attribute-guided fusion network featuring knowledge-inspired attention mechanisms that leverage HSI spectral correlations and LiDAR elevation attributes for discriminative feature learning. The framework introduces two novel fusion modules to selectively aggregate HSI and LiDAR features, minimizing redundant information while preserving modality-specific physical properties. Evaluated on multi-source datasets, AFNKA outperforms state-of-the-art methods by explicitly incorporating domain knowledge into attention and fusion processes (Pan, X., et al., 2025).

To address limitations in LiDAR semantic segmentation where methods underutilize available datasets, Sanchez, J., et al. introduced multi-source training via the COLA dataset-constructed by harmonizing diverse sources with coarse labels to overcome domain discrepancies. They propose three specialized applications: COLA-DG for domain generalization, COLA-S2S for source-to-source segmentation (+5.3%), and COLA-PT for pre-training (+12%) (Sanchez, J., et al., 2023). Jiang, H., et al. presents two novel Transformer-based approaches for automating the extraction of rural multilane highway infrastructure elements from LiDAR data (Jiang, H., et al., 2025).

4. Advances in Deep Learning Methodologies for Urban Built-up Area Recognition

Deep learning achieves pixel-level recognition of built-up areas through semantic segmentation models. Its evolution can be categorized into three stages: traditional convolutional neural networks, optimization of encoder-decoder architectures, and integration with Transformers.

4.1. Fundamental Applications of Traditional Convolutional Neural Networks

In 2017, Zhao, et al. used convolutional neural networks to extract features from remote sensing images, and then used conditional random fields to model the contextual information of the high-level semantic features extracted by convolutional neural networks (Zhao, et al., 2017). Kaiser, et al. proposed a method to train neural networks using an open street view dataset with noise and only image level annotations, which achieved good performance in the network model. This method provides a new method for obtaining a large amount of manually labeled data required for supervised deep learning (Kaiser, et al., 2017).

4.2. Encoder-Decoder Architecture Optimization

Since 2018, encoder-decoder architectures such as FCN and U-Net have become mainstream approaches. Wang, et al. proposed a dual path hybrid FCN for human motion estimation, which includes two branches: appearance FCN and motion FCN, capturing static appearance features and dynamic motion features (Wang, et al., 2016). Dai, et al. proposed a region-FCN for object detection. This method proposes a position sensitive scoring map to record the position information of targets in each region, which is used to alleviate the contradiction between the translation invariance of classification models and the requirement of target detectors to be sensitive to translation changes (Dai, et al., 2016). To overcome limitations in manual and supervised lithology identification, Qin, S., et al. proposed an unsupervised semantic segmentation method combining improved FCN with superpixel segmentation to generate pseudo-labels for debris rock images from Qingshuihe-Karazha, eliminating reliance on large annotated datasets (Qin, S., et al., 2025).

To address challenges in road extraction from high-resolution remote sensing imagery, including complex urban layouts and irregular rural paths, Ding, T., et al. proposed DEDU-Net, a dual-encoder-decoder network with a Global Fusion Module

(GFM) that integrates multi-scale features through weighted context fusion for precise road segmentation (Ding, T., et al., 2024). Fang, W., et al. proposed FPS-U^{2}Net, a salient object detection network adapting U^{2}Net with a Multi-level Aggregation Module (MAM) to fuse adjacent encoder features for enriched context, trained via hybrid BCE+IoU loss for pixel- and map-level precision (Fang, W., et al., 2024). Deng, J., et al. proposed RSKAU-net, a U-Net variant integrating Residual Selective Kernel Attention (RSKA) modules for adaptive multi-scale feature extraction and Channel Attention (CA) blocks for spectral preservation (Deng, J., et al., 2024). Wang, J., et al. enhanced U-Net by integrating VGG-Net residual blocks for deeper feature extraction and an adaptive attention mechanism to prioritize distinct building types, followed by four-channel fine detection to refine edges (Wang, J., et al., 2025).

4.3. Attention Mechanisms and Multi-Scale Fusion

Since 2019, attention mechanisms and multi-scale feature fusion have emerged as key optimization strategies. Li, Y., et al. proposed AOHSR, an attention-guided network that fuses high-resolution optical images with HSIs through three specialized modules: a Spectral Reservation Module preserving spectral fidelity via channel attention, an Optical Image Guidance Module enhancing spatial details via spatial attention, and a Spatial-Spectral Recovery Module reconstructing high-resolution HSIs (Li, Y., et al., 2023). To address maritime ship classification challenges in heterogeneous remote sensing data, Tienin, B. W., et al. proposed MS3Net, an ensemble deep learning model integrating spatial attention and second-order pooling to synergistically process SAR and optical ship images through specialized convolutional branches (Tienin, B. W., et al., 2024).

Xiao, H., et al. proposed DFMNet, a novel architecture featuring a symmetrical dual-path module for multi-scale feature aggregation and a cross-attention mechanism for encoder-decoder feature fusion, effectively enhancing high-frequency detail restoration (Xiao, H., et al., 2025). To address the high cost and challenge of manual annotation for detecting dead pine trees in UAV imagery for pine wilt disease control, Wang, S. B., et al. proposed a semi-supervised learning framework (Wang, S. B., et al., 2025). Fayaz, M., et al. proposed the Dual Land Cover Attention Network (DLAN), integrating channel and spatial attention mechanisms within a CNN framework to enhance discriminative capability for diverse land cover patterns and spectral variations (Fayaz, M., et al., 2025).

4.4. Transformers and Cross-Modal Fusion

Since 2021, the integration of Transformers with CNNs has emerged as a new trend. Wang, L., et al. proposed the Swin Transformer as the backbone to extract the context information and design a novel decoder of densely connected feature aggregation module (DCFAM) to restore the resolution and produce the segmentation map. The experimental results on two remotely sensed semantic segmentation datasets demonstrate the effectiveness of the proposed scheme (Wang, et al., 2022). He, et al. proposed an ST-UNet. This network structure uses two encoders, Swin Transformer and CNN, and embeds a spatial interaction module in the Swin Transformer structure to extract pixel level cross-correlation information. Finally, a relationship enhancement module was used to fuse the characteristics between the two encoders, effectively improving the performance of the model (He, et al., 2022). Nan, et al. proposed a new AS-Unet++. First, the atrous spatial pyramid pooling (ASPP) and the squeeze-and-excitation (SE) module are added to traditional Unet, such that the sensing field can be expanded and the important features can be enhanced, which is called AS-Unet. Second, an AS-Unet++ structure is built by using different layers of AS-Unet, such that the feature extraction parts of each layer of AS-Unet are stacked together. Compared with Unet, the proposed AS-Unet++ automatically learns features at different depths and determines a depth with optimal performance (Nan, et al., 2024).

5. Discussion and Prospects

Three major bottlenecks persist in current research:

1) High Data Dependency: Model performance relies heavily on annotation quality, yet significant cross-regional variations in built-up areas (e.g., urban-rural divides, climate zones) limit model generalizability.

2) Low Small-Target Recognition Accuracy: Minor ancillary facilities (e.g., bus stops) occupy minimal pixels in high-resolution imagery, leading to frequent model oversight.

3) High Computational Costs: Large-volume high-resolution imagery and complex models (e.g., Transformers) demand substantial GPU resources, hindering practical deployment.

Future Research Directions:

1) Cross-Modal Data Fusion: Integrate optical, LiDAR, SAR, and multi-source data. Frameworks like FSFM (Zhou, T., et al., 2025) leverage dynamic feature fusion to extract complementary information, enhancing robustness in complex scenarios.

2) Lightweight Model Design: Approaches such as MST-DeepLabv3+ (Wang, Y., et al., 2024) replace backbone networks (e.g., Xception) with MobileNetV2 and integrate attention mechanisms, reducing parameters while enabling real-time application.

3) Semi-/Unsupervised Learning: Address annotation scarcity via pseudo-label generation. Frameworks like segWCD (Wu, Y., et al., 2025) utilize weakly supervised learning to minimize manual labeling dependence.

4)Boundary Refinement Optimization: Modules like Scale Adaptive Module and Position Guidance Module (Li, L., et al., 2025) incorporate morphological features of built-up areas to enhance edge recognition precision.

Acknowledgements

This research was supported by 14th Five-Year Plan for Educational Science in Shaanxi Province, project titled “Research on the Construction of Curriculum System for the Major of Computer Science and Technology in Application-oriented Undergraduate Universities Facing New Engineering Education” Grant no: SGH23Y2905.

References

[1] Feng, Q., Chen, B., Li, G., et al. (2022). A review of research on remote sensing image sample datasets. Journal of Remote Sensing, 26(4), 589–605.

[2] Pei, H., Owari, T., Tsuyuki, S., & Zhong, Y. (2023). Application of a Novel Multiscale Global Graph Convolutional Neural Network to Improve the Accuracy of Forest Type Classification Using Aerial Photographs. Remote Sensing, 15(4), 1001.

[3] Shao, Z., Tang, P., Wang, Z., Saleem, N., Yam, S., & Sommai, C. (2020). BRRNet: A Fully Convolutional Neural Network for Automatic Building Extraction From High-Resolution Remote Sensing Images. Remote Sensing, 12(6), 1050.

[4] Wang, Y., Yang, L., Liu, X., & Yan, P.. (2024). An improved semantic segmentation algorithm for high-resolution remote sensing images based on deeplabv3. Scientific Reports, 14(1).

[5] Ismael, S. F., Kayabol, K., & Aptoula, E.. (2024). Unsupervised domain adaptation for the semantic segmentation of remote sensing images via a class-aware fourier transform and a fine-grained discriminator. Digital Signal Processing, 151, 104551.

[6] Jia, P., Chen, C., & Zhang, Y. Z. L.. (2024). Semantic segmentation of deep learning remote sensing images based on band combination principle: application in urban planning and land use. Computer communications, 217(Mar.), 97-106.

[7] Feng, Q., Chen, B., Li, G., et al. (2022). A review for sample datasets of remote sensing imagery. National Remote Sensing Bulletin. 26(4), 589–605.

[8] Wang, B., & Wang, Y. (2022). Research on the current status of remote sensing image land category information acquisition technology based on deep learning. Journal of Land and Resources, 19(4), 74–80.

[9] Zhang, W., Liu, W., Sun, X., et al. (2022). High resolution remote sensing image semantic segmentation using multi-source feature adaptive fusion network. Chinese Journal of Image and Graphics, 27(8), 2516–2526.

[10] Lou, M., Zhang, S., Zhou, H. Y., Yang, S., Wu, C., & Yu, Y. (2025). TransXNet: learning both global and local dynamics with a dual dynamic token mixer for visual recognition. IEEE Transactions on Neural Networks and Learning Systems.

[11] Pan, X., Jiao, C., Yang, B., Zhu, H., & Wu, J. (2025). Attribute-guided feature fusion network with knowledge-inspired attention mechanism for multi-source remote sensing classification. Neural Networks, 187, 107332.

[12] Sanchez, J., Deschaud, J. E., & Goulette, F.. (2023). Cola: coarse-label multi-source lidar semantic segmentation for autonomous driving. IEEE Transactions on Robotics, vol. 41, pp. 1742-1754, 2025.

[13] Honglin Jiang, Hesham Elmasry, Sangwon Lim, and Karim El-Basyouny. 2025. Utilizing deep learning models and LiDAR data for automated semantic segmentation of infrastructure on multilane rural highways. Canadian Journal of Civil Engineering. 52(8): 1523-1543.

[14] Zhao, W., Du, S., Wang, Q., & Emery, W. J. (2017). Contextually guided very-high-resolution imagery classification with semantic segments. ISPRS journal of Photogrammetry and Remote Sensing, 132, 48-60.

[15] Kaiser, P., Wegner, J. D., Lucchi, A., Jaggi, M., Hofmann, T., & Schindler, K. (2017). Learning aerial image segmentation from online maps. IEEE Transactions on Geoscience and Remote Sensing, 55(11), 6054-6068.

[16] Wang, L., Qiao, Y., Tang, X., & Van Gool, L. (2016). Actionness estimation using hybrid

fully convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2708-2717).

[17] Dai, J., Li, Y., He, K., & Sun, J. (2016). R-fcn: Object detection via region-based fully convolutional networks. Advances in neural information processing systems, 29.

[18] Qin, S., Wang, Q., Zeng, Q., Ye, M., Fu, A., & Chen, G. (2025). Automatic recognition of debris rock lithology based on unsupervised semantic segmentation. Computers & Geosciences, 196, 105790.

[19] Ding, T., & Wang, X. (2024). DEDU-Net: Dual-Encoder-Decoder-U-Net for road extraction from high-resolution remote sensing images. International Journal of Remote Sensing, 45(10), 3231-3247.

[20] Fang, W., Fu, Y., & Sheng, V. S. (2024). FPS-U2Net: Combining U2Net and multi-level aggregation architecture for fire point segmentation in remote sensing images. Computers & Geosciences, 189, 105628.

[21] Deng, J., & Yang, B. (2024). Hyperspectral and multispectral image fusion via residual selective kernel attention-based U-net. International Journal of Remote Sensing, 45(5), 1699-1726.

[22] Wang, J., Cheng, L., Zheng, Y., Cui, H., & Wang, Y. (2025). Extraction of Rural Buildings with Different Main Structure Types Based on a Revised U-Net Model. Journal of Computing in Civil Engineering, 39(2), 04024058.

[23] Li, Y., Chen, S., Xu, F., Tang, H., & Xie, W. (2023). AOHSR: attention mechanism-based optical image-guided hyperspectral image super-resolution. International Journal of Remote Sensing, 44(13), 3939-3971.

[24] Tienin, B. W., Cui, G., Ukwuoma, C. C., Talla Nana, Y. A., Mba Esidang, R., & Moniz Moreira, E. Z. (2024). MS3Net: a deep ensemble learning approach for ship classification in heterogeneous remote sensing data. International Journal of Remote Sensing, 45(3), 748-771.

[25] Xiao, H., Chen, X., Luo, L., & Lin, C. (2025). A dual-path feature reuse multi-scale network for remote sensing image super-resolution. The Journal of Supercomputing, 81(1), 17.

[26] Wang, S. B., Gao, Z. M., Jin, D. H., Gong, S. M., Peng, G. L., & Yang, Z. J. (2024). AMEA-YOLO: a lightweight remote sensing vehicle detection algorithm based on attention mechanism and efficient architecture. The Journal of Supercomputing, 80(8), 11241-11260.

[27] Fayaz, M., Dang, L. M., & Moon, H. (2025). DLAN: a dual attention network for effective land cover classification in remote sensing. Knowledge-Based Systems, 113620.

[28] Wang, L., Li, R., Duan, C., Zhang, C., Meng, X., & Fang, S. (2022). A novel transformer based semantic segmentation scheme for fine-resolution remote sensing images. IEEE Geoscience and Remote Sensing Letters, 19, 1-5.

[29] He, X., Zhou, Y., Zhao, J., Zhang, D., Yao, R., & Xue, Y. (2022). Swin transformer embedding UNet for remote sensing image semantic segmentation. IEEE transactions on geoscience and remote sensing, 60, 1-15.

[30] Nan, G., Li, H., Du, H., Liu, Z., Wang, M., & Xu, S. (2024). A semantic segmentation method based on AS-Unet++ for power remote sensing of images. Sensors, 24(1), 269.

[31] Zhou, T., He, H., Wang, Y., &Liao, Y.. (2025). Improved gated recurrent units together with fusion for semantic segmentation of remote sensing images based on parallel hybrid network. Multimedia Systems, 31(2).

[32] Wang, Y., Yang, L., Liu, X., & Yan, P. (2024). An improved semantic segmentation algorithm for high-resolution remote sensing images based on DeepLabv3+. Scientific reports, 14(1), 9716.

[33] Wu, Y., Zhang, X., Zhao, X., Sun, Y., & Li, T. (2025). segWCD: A new segmentation-based weak supervision neural network for building change detection. Applied Intelligence, 55(2), 147.

[34] Li, L., Lu, N., & Qin, J. (2025). Joint-task learning framework with scale adaptive and position guidance modules for improved household rooftop photovoltaic segmentation in remote sensing image. Applied Energy, 377, 124521.