Proceedings of CONF-MLA 2025 Symposium: Applied Artificial Intelligence Research
DOI: 10.54254/2755-2721/2025.BJ25643

Review of Remote Sensing Image Classification: Technology Evolution, Method Innovation and Future Challenges

Enpu Yu

Tianjin Chengjian University, Tianjin, China
yuenpu2020@outlook.com

Abstract. Remote sensing image classification is the core technology of environmental monitoring, urban planning, and resource management. Its goal is to accurately identify the ground cover through intelligent algorithms. In recent years, breakthroughs in deep learning technologies (such as CNN and Transformer) have significantly improved classification accuracy and efficiency, but problems such as generalization in complex scenes, high annotation costs, and insufficient real-time performance remain to be addressed. In this paper, the theoretical basis and technical evolution of remote sensing image classification are systematically reviewed. By combining with the latest research progress from 2024 to 2025, the research directions of multimodal fusion, lightweight networks, small sample learning etc. are analyzed, and the future development trends of self-supervised learning, edge computing and so on are discussed. By integrating classical methods and intelligent technologies, this paper aims to provide a comprehensive technical reference and development directions for researchers in the field of remote sensing classification.

Keywords: Remote sensing image classification, Deep learning, Multimodal fusion, Lightweight network, Small sample learning

1. Introduction

Remote sensing image classification plays a pivotal role in land cover monitoring, disaster assessment, and urban development by analyzing multispectral and high-resolution imagery. Key challenges include spectral similarity between objects, mixed pixel interference, and complex terrain variations. Traditional methods relying on manual feature extraction (e.g., spectral indices and texture descriptors) struggled with high-dimensional data and intricate scenes. The advent of deep learning, with its capacity for automated feature learning, has revolutionized the field. However, there are still three key bottlenecks in the current application of deep learning in remote sensing image classification. First, the generalization ability of models in cross regional and cross temporal applications is insufficient. Second, the computational efficiency urgently needs to be improved when facing massive remote sensing data [1]. Third, there is still significant room for improvement in model performance in small sample scenarios [2]. Notably the breakthrough progress in Hyperspectral imaging technology and multimodal data fusion methods in recent years has provided new technological paths to address these challenges. These innovative methods significantly enhance the robustness and accuracy of classification systems by fully mining complementary

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.BJ25643

information between different data sources. This article systematically reviews the evolution of remote sensing image classification from classical methods to cutting-edge technologies, with a focus on analyzing the advantages and limitations of current mainstream algorithms. Through in-depth analysis of existing technological bottlenecks, this article further explores possible future research directions, including: the application of self-supervised learning in annotated data scarcity scenarios, lightweight network architecture design, and interpretability fusion of multimodal feature knowledge. These studies not only have important theoretical value, but also provide useful guidance for engineering practice in practical remote sensing applications.

2. Theoretical overview

2.1. Technical characteristics of hyperspectral remote sensing

Hyperspectral remote sensing, a cutting-edge technology in remote sensing, operates on the principle of precise spectral measurements of surface objects [3]. It decomposes the reflected sunlight from land objects into continuous spectral bands through a spectroscopic system, with each band corresponding to a specific wavelength range, thereby obtaining spectral information of land objects in hundreds or even thousands of narrow bands. This high spectral resolution offers significant advantages, as it can capture more subtle spectral feature distinctions of ground objects, providing rich information for ground object recognition and classification.

Compared with traditional multispectral remote sensing, hyperspectral remote sensing has narrower bands and numbers of bands. Traditional multispectral systems typically have only a few to a dozen broad bands, while hyperspectral systems have dozens or even hundreds of narrow bands. In terms of band characteristics, the narrow bands of hyperspectral imaging can more accurately reflect the spectral characteristics of land cover. For example, hyperspectral imaging enables accurate differentiation of vegetation types based on their unique spectral curves [4]. In practical applications, hyperspectral remote sensing has demonstrated strong ground recognition capabilities in geological exploration, agricultural monitoring, and other fields, and has enabled the identification of ground types that are difficult to distinguish using traditional multispectral methods [4].

2.2. Deep learning framework

Deep learning frameworks for hyperspectral remote sensing image classification have evolved to leverage diverse architectural paradigms, each addressing unique challenges in spectral-spatial feature representation. Among these, Convolutional Neural Networks (CNNs), Transformer-based models, and hybrid architectures have emerged as pivotal approaches, offering complementary strengths in balancing local feature extraction, global dependency modeling, and multi-scale information fusion. The following sections elaborate on their distinct mechanisms and integration strategies.

CNNs have demonstrated powerful capabilities in hyperspectral remote sensing image classification, and their architecture design is key to achieving high-precision classification. In hyperspectral classification, CNNs usually adopts a three-dimensional convolutional structure to fully utilize the spectral and spatial information of hyperspectral data. The input layer receives hyperspectral image data, which passed through multiple convolutional and pooling layers to extract image features. The convolutional layer learns local features via kernel sliding, whereas the pooling layer reduces feature map dimensionality to lower computational complexity. Finally, the extracted

DOI: 10.54254/2755-2721/2025.BJ25643

features are mapped to different categories through a fully connected layer. U-Net and ResNet are two commonly used CNN models in hyperspectral classification. The U-Net model has an encoder decoder structure, which transfers different levels of feature information from the encoder to the decoder through skip connections, enabling better restoration of image details [5]. It performs well in small object classification and edge detection. The ResNet model introduces residual blocks to solve the problem of gradient vanishing during deep neural network training, enabling the construction of deeper network structures and the learning of more complex features [6]. Based on self-attention mechanism to model global dependency relationships, ATMformer preserves key local features through adaptive token merging [7]. U-Net combined with Transformer integrates multi-scale features in the encoding decoding structure, balancing detail restoration and global modeling [5].

3. Typical datasets and classification accuracy evaluation system

3.1. Overview of international benchmark dataset

In remote sensing image classification research, international benchmark datasets play a crucial role, providing a unified and objective standard platform for performance evaluation of different algorithms, horizontal comparison, and validation of new methods [8]. These carefully constructed datasets cover a diverse range of sensor types, spatial resolutions, spectral ranges, and terrain scenes, effectively supporting research and optimization from classical methods to deep learning models. The following provides a detailed introduction to several widely adopted and representative core datasets in this field:

The Indian Pines dataset is a classic hyperspectral dataset obtained by the AVIRIS sensor. This dataset covers the agricultural region of northwest Indiana, USA, and includes 16 different land cover categories such as corn, soybeans, wheat. The image size is 145×145 pixels, with a spectral range of 0.4-2.5 μm and a total of 224 bands.

The Salinas dataset is also a hyperspectral dataset collected by the AVIRIS sensor. It records agricultural scenes in the Salinas Valley of California, USA, including 16 types of land features such as vineyards, lettuce, broccoli. The image size is 512×217 pixels, with 224 bands and high spectral resolution, providing rich spectral information for land classification.

The UC Merced dataset is mainly used for high-resolution remote sensing image classification and is obtained by QuickBird satellite sensors. This dataset contains 21 scene categories, such as airports, beaches, forests, etc., with 100 256×256 pixel images for each category. Its scenes are mostly urban and natural landscapes, focusing on large-scale scene recognition.

The NWPU-RESSISC45 dataset is also a high-resolution remote sensing image dataset acquired from multiple satellite sensors. It covers 45 different scene categories, including industrial, agricultural, residential and other scenes, with a total of 31,500 images and a size of 256×256 pixels per image. Compared to UC Merced, NWPU-RESSISC45 has a more diverse range of scenes and can better simulate complex real-world scenarios.

These four datasets represent different dimensions of remote sensing image classification research: Indian Pines and Salinas focus on hyperspectral pixel-level classification, challenging spectral resolution and mixed pixel processing; UC Merced and NWPU-RESISC45 focus on high-resolution scene-level classification, testing the model's ability to understand the overall semantics and complex spatial layout of images. They together form a solid foundation for algorithm development and performance evaluation in this field [8].

DOI: 10.54254/2755-2721/2025.BJ25643

3.2. Classification accuracy evaluation system

Classification accuracy evaluation are important components of remote sensing image classification research, and confusion matrix, Kappa coefficient, OA, and AA are commonly used evaluation indicators.

The confusion matrix is a two-dimensional matrix, with rows representing actual categories and columns representing classification result categories [9]. The elements in the matrix represent the number of samples classified into each category. By using the confusion matrix, the correct and misclassified classification of each category in the classification results can be visually observed.

The Kappa coefficient is an accuracy evaluation index that considers the influence of random classification, and its calculation method is calculated based on the confusion matrix [9]. Firstly, calculate the actual classification accuracy and random classification accuracy, and then use the formula to calculate the Kappa coefficient. The value of Kappa coefficient ranges from -1 to 1, and the closer the value is to 1, the higher the classification accuracy.

OA (Overall Accuracy) refers to the proportion of correctly classified samples to the total number of samples, reflecting the overall accuracy of the classification results. AA (Average Accuracy) is the average classification accuracy of each category, used to measure the classifier's ability to classify each category [9].

4. Current research situation

Currently, the forefront of research is co-evolving along key paths such as model architecture innovation, deepening multimodal fusion, small-sample/semi-supervised learning optimization, and lightweight real-time deployment, resulting in multiple representative achievements. In terms of multi-scale feature extraction, Luo and Li innovatively designed a hybrid convolution module that combines dilated convolution and regular convolution, achieving a classification accuracy of 98.19% on the WHU-RS18 dataset [10]. In the field of semi-supervised learning, the CNN Transformer cross-learning framework (SCL CTNet) proposed by Dan et al. significantly improves the utilization of unlabeled data through a pseudo label filtering mechanism [11]. In the direction of multispectral fusion, Liu et al. improved YOLOv5 by introducing near-infrared bands, thereby increasing the mIoU by 2.5%, effectively verifying the effectiveness of multiband input [12].

In the field of lightweight design, the STConvNeXt developed by Liu et al. not only reduces the parameter count by 56.49%, but also achieves an accuracy improvement of 1.2-2.7% [13]. Liu et al. effectively reduced the computational cost of MobileNetV2 through dynamic segmentation point convolution (DSPC) [19]. In terms of multimodal fusion, Niu et al.'s KANFusion network innovatively integrates remote sensing images and street view images, using the Kolmogorov Arnold network to mine intermodal correlation features [14]. In response to the challenge of small sample learning, the Han et al. introduced frequency domain transformer technology and successfully achieved superpixel segmentation and high-precision classification in small sample scenes [15]. In terms of main research directions, deep learning model optimization continues to advance. Yang et al. achieved a classification accuracy of 92.83% on the UC Merced dataset through collaborative design of channel attention and feature fusion [16]. The SegNet+network developed by Hu improved segmentation accuracy by 4.41% using a multi-level feature fusion module [17]. The ATMformer proposed by Niu et al. achieved SOTA performance on the EuroSAT dataset [7]. In the field of multimodal data fusion, KANMusic network improves the recognition accuracy of urban informal residential areas (UIS) through hierarchical feature fusion. In contrast LHFNet optimizes multimodal data redundancy by combining low-frequency structural features and high-frequency

DOI: 10.54254/2755-2721/2025.BJ25643

detail features [14,17]. In terms of small sample and semi-supervised learning, Han et al.'s frequency domain alignment technique and Zhang et al.'s comparative learning method combining optimal transmission theory have respectively enhanced the discriminative ability of small sample feature expression and long tail data tail class samples [15,18]. In the direction of lightweight real-time optimization, the CSM-DSPCss Ghost network achieves parameter reduction and classification acceleration through channel switching modules, while Wu et al. use a multi-scale attention model distillation to improve feature extraction efficiency [19-20]. These research advances collectively drive the development of remote sensing interpretation technology towards higher accuracy, stronger generalization, and better real-time performance.

5. Challenge and prospect

Although significant progress has been made in model accuracy, multi-source fusion, and lightweight design in current remote sensing image interpretation research, it still faces multiple challenges. Firstly, the generalization ability in complex scenes needs to be improved, and existing models have insufficient adaptability to cross regional and cross seasonal data. Therefore, it is necessary to optimize domain adaptability through techniques such as causal reasoning. Secondly, the contradiction between high annotation costs and data privacy protection is prominent, and further exploration is needed on the collaborative optimization mechanism between sparse annotation such as HSLabeling and federated learning. Finally, edge computing deployment still has bottlenecks [21]. Although lightweight models such as STConvNeXt have achieved substantial parameter compression, how to adapt to low-power hardware and meet real-time processing needs still needs to be broken through [13].

Looking ahead to the future, research will deepen along three core directions. First, developing multimodal self-supervised pre-training techniques to improve model generalization performance through large-scale unlabeled data. The second is to construct a spatiotemporal joint modeling framework and integrate temporal remote sensing data to achieve accurate monitoring of dynamic changes in land cover. The third is to innovate the design of micro network architecture, develop ultra-lightweight classification models for IoT terminal devices, and promote the deep penetration of remote sensing interpretation technology into edge intelligent scenes. These development directions will systematically solve existing technological bottlenecks and promote the large-scale application of remote sensing image interpretation in fields such as smart cities and environmental monitoring.

6. Conclusion

This article provides a systematically reviews of the technological evolution and innovation in remote sensing image classification, with a focus on the role of deep learning techniques in improving classification accuracy and efficiency. Additionally, addressing the challenges faced by current models, such as insufficient generalization ability, high annotation costs, and poor real-time performance, this paper proposes corresponding solutions, including multimodal fusion, lightweight networks and small sample learning. Furthermore, improving model generalization ability through self-supervised pre training and spatiotemporal modeling, optimizing annotation efficiency through sparse annotation and federated learning, and promoting research on the compatibility of lightweight models with low-power hardware can facilitate the application of remote sensing technology in fields.

However, this article has certain limitations, including a lack of experimental verification, insufficient coverage of research focus in the literature, and incomplete analysis of datasets covering

DOI: 10.54254/2755-2721/2025.BJ25643

extreme weather or dynamic time series scenarios. Future research should validate theoretical conclusions through more empirical studies, expand the scope of literature coverage, and introduce more challenging datasets to evaluate model robustness. Subsequent research could focus on technical implementation (such as edge computing deployment), interdisciplinary fusion (combining causal reasoning or meta learning) and data driven innovation (using generative AI to synthesize and label data) to drive the deep penetration and large-scale transformation of remote sensing classification technology into practical application scenarios.

References

[1] Lv, F. (2020). Ensemble learning method and optimization of remote sensing image classification problem [D]. Dalian University of Technology. DOI: 10.26991/d.cnki.gdllu.2020.003640.
[2] Yan, Q. (2024). Research on classification method of hyperspectral remote sensing images for small samples [D]. China University of Geosciences. DOI: 10.27492/d.cnki.gzdzu.2024.000240.
[3] Zhang, C. M, Mu, T. K., Yan, T. Y. & Chen, Z. Y. (2018). Overview of Hyperspectral Remote Sensing Technology. Spacecraft Recovery & Remote Sensing, 39 (03): 104-114. DOI: 10.3969/j.issn.1009-8518.2018.03.012.
[4] Li, M. & Luo, H. R. (2024). Application of hyperspectral remote sensing in natural resources survey andmonitoring. Guangxi Water Resources and Hydropower Engineering, (01): 150-154. DOI: 10.16014/j.cnki.1003-1510.2024.01.033.
[5] Gou, C. L., Pang, M. & Yang, Y. (2025). Improved U-Net convolutional network application for land cover classification in remote sensing images. Bulletin of Surveying and Mapping, (03): 150-155. DOI: 10.13474/j.cnki.11-2246.2025.0326.
[6] Yu, M. Q., Ji, Q., Zhang, Z. M., Wang, W., Lin, Y. S., Yao, Y. R. & Liu, N. N. (2025). Remote sensing identification of wetland types in the flood control and storage area of Yangtze River (Anhui section) based on deep learning. Journal of Nanjing University of Information Science & Technology. DOI: 10.13878/j.cnki.jnuist.20241225003.
[7] Niu, Y, Song, Z. C., Luo, Q. Y., Chen, G. C., Ma, M. M. & Li, F. (2025) ATMformer: An Adaptive Token Merging Vision Transformer for Remote Sensing Image Scene Classification. Remote Sensing. 17(4). DOI: 10.3390/RS17040660.
[8] Zhang, X. (2021). Research on Hyperspectral Remote Sensing Image Classification Based on Three-dimensional Convolutional Neural Network. University of Chinese Academy of Sciences (Changchun Institute of Optics, Precision Mechanics and Physics), Doctor thesis. DOI: 10.27522/d.cnki.gkcgs.2021.000030.
[9] de Oliveira, W. V., Dutra, L. V. & Sant'Anna, S. J. S. (2023). A novel quality assessment index for evaluating the spatial uniformity of remote sensing image classification. Remote Sensing Letters. 15 (10), pp 1035-1046. DOI: 10.1080/2150704X.2023.2255348.
[10] Luo, H. & Li, X. F. (2025). Remote Sensing Image Classification Based on Multi-scale Feature Extraction. Computer and modernization, (03): 86-92. DOI: 10.3969/j.issn.1006-2475.2025.03.013.
[11] Dan, F. L., Lv, P. Y. & Li, M. C. (2024). A Remote Sensing Image Scene Classification Based on CNN-Transformer Semi-Supervised Cross Learning. Journal of Ningxia University (Natural Science Edition). 45(03): 325-332. DOI: 10.3969/j.issn.0253-2328.2024.03.015.
[12] Liu, Y., Yang, W. L., Guo, P. Y., Cao, L., Wang, X. H., Meng, L. & Zhao, W. D (2024). Landcover Classification Method for Multispectral Satellite Remote Sensing Imagery Based on Improved YOLOv5. Spacecraft Recovery & Remote Sensing, 45(06): 113-123. DOI: 10.3969/j.issn.1009-8518.2024.06.010.
[13] Liu, B., Zhan, C. M., Guo, C., Liu, X. B. & Ruan, S. F. (2025). Efficient remote sensing image classification using the novel STConvNeXt convolutional network. Scientific Reports. 15, 8406. DOI: 10.1038/s41598-025-92629-x.
[14] Niu, H. Y., Fan, R. Y., Chen, J.J., Xu, Z. J. & Feng, R. Y. (2025, June). Urban informal settlements interpretation via a novel multi-modal Kolmogorov-Arnold fusion network by exploring hierarchical features from remote sensing and street view images. Science of Remote Sensing. (11). 100208. DOI: 10.1016/j.srs.2025.100208.
[15] Han, Z. Y., Chen, H. & Zeng, C. (2025). Research on Automatic Classification Method of Small-Sample Remote Sensing Images Based on Frequency-Domain Transformer. Science & Technology Information, 23(04): 74-77. DOI: 10.16661/j.cnki.1672-3791.2407-5042-8862.
[16] Yang, S., Wang, X. H., Wang, X. Y. & Gu, X. P. (2024). Remote sensing image scene classification based on attention mechanism and feature fusion. Chinese Journal of Sensors and Actuators, 2024, 37(12): 2078-2083. DOI: 10.3969/j.issn.1004-1699.2024.12.009.

DOI: 10.54254/2755-2721/2025.BJ25643

[17] Hu, Y. X. (2024). Research on Semantic Segmentation Algorithm of Remote Sensing Images Based on Convolutional Neural Networks. Zhejiang Gongshang University, Master Thesis. DOI: 10.27462/d.cnki.ghzhc.2024.000147.
[18] Zhang, L., Peng, L. J., Yang, C. W. & Ding, X. (2025). Multi-expert contrastive learning for remote sensing long-tailed image classification. International Journal of Remote Sensing. 46(4), PP. 1517-1542. DOI: 10.1080/01431161.2024.2433751.
[19] Liu, J., Wu, M. Y., Li, K. X. & Liu, Y. (2025, March). A Lightweight Network Based on Dynamic Split Pointwise Convolution Strategy for Hyperspectral Remote Sensing Images Classification. Remote Sensing. 17(5), 888. DOI: 10.3390/RS17050888.
[20] Wu, Y. Y, Hao, Y. B., Liu, X., Wu. F. J., Gao, L. & Xu, C. A. (2025). Optical remote sensing scene classification based on multi-scale attention. Journal of Naval Aviation University, 40(01), 171-178+204. DOI: 10.7682/j.issn.2097-1427.2025.01.008.
[21] Lin, J. X., Yang, Z., Liu, Q., Yan, Y. L., Ghamisi, P. & Xie, W. Y. (2025). HSLabeling: Toward Efficient Labeling for Large-scale Remote Sensing Image Segmentation with Hybrid Sparse Labeling. IEEE Transactions on Image Processing. 34, pp 1864-1878. DOI: 10.1109/TIP.2025.3550039.