Journal of Oceanology and Limnology

https://doi.org/10.1007/s00343-026-6009-4

Research Paper

A global-local multi-scale feature extraction network for fine-grained segmentation of cyanobacterial blooms and aquatic vegetation*

Yunchang MU[1, 2, 3], Junsheng LI[2, 3, 4, **], Huan ZHAO[5], Xiaodi ZHU[2, 3, 6], Zeyan ZHANG[2, 3, 6], Fangfang ZHANG[2, 3], Shenglei WANG[2, 3]

[1] *Beijing Institute of Surveying and Mapping, Beijing 100038, China*

[2] *Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China*

[3] *International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China*

[4] *School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China*

[5] *Satellite Application Center for Ecology and Environment, Ministry of Ecology and Environment of the People's Republic of China, Beijing 100094, China*

[6] *College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China*

Received Jan. 5, 2026; accepted in principle Feb. 24, 2026; accepted for publication Apr. 23, 2026

Abstract Cyanobacterial blooms and aquatic vegetation exhibit similar spectral characteristics, which limits the discriminative capability of traditional spectral index-based methods. Moreover, existing deep learning models still suffer from omission and misclassification errors in fragmented boundary regions. This study proposes a semantic segmentation network, termed global-local multi-scale feature extraction network (GLMFENet), which is designed to jointly exploit global and local multi-scale features. Based on an encoder-decoder architecture, GLMFENet incorporates a global feature extraction module to model long-range semantic dependencies, together with a spatial attention-guided local multi-scale feature extraction module to enhance fine-grained discriminability. By synergistically integrating global contextual information with local structural features, the proposed network enables accurate differentiation between cyanobacterial blooms and aquatic vegetation. A dataset comprising 467 Sentinel-2 MSI images collected from multiple lakes across China was established for model training and validation. GLMFENet achieved a mean IoU (mIoU) of 84.6%, a mean F1-score of 91.3%, and a Kappa coefficient of 86.4% on the test set, outperforming widely adopted deep learning models such as DeepLabV3+, SegFormer, and Swin-UNet. The proposed method improves the mIoU by 10.7–19.6 percentage points compared with traditional spectral index-based methods, leading to more accurate and reliable extraction of cyanobacterial blooms and aquatic vegetation.

Keyword: cyanobacterial bloom; aquatic vegetation; remote sensing; deep learning

1 INTRODUCTION

Cyanobacterial blooms and aquatic vegetation represent two typical forms of aquatic biota in lake ecosystems, whose spatial distributions and evolutionary dynamics directly reflect the ecological structure and functional status of lakes (Pan et al., 2014). Cyanobacterial blooms are characterized by the rapid proliferation and surface accumulation of cyanobacteria, resulting in visible algal aggregates and constituting an extreme manifestation of water eutrophication (Du et al., 2024). Bloom outbreaks

* Supported by the National Key Research and Development Program of China (No. 2024YFD1700802) and the Future Star Program of the Aerospace Information Research Institute, and Scientific and Disruptive Technology Research Fund Project of Aerospace Information Research Institute, Chinese Academy of Sciences (No. 2025-AIRCAS-SDTP-16)

** Corresponding author: lijs@radi.ac.cn

can release harmful toxins, disrupt aquatic ecological balance, and pose serious threats to drinking water safety and public health (Anderson et al., 2002; Han et al., 2018; Pan et al., 2024). In contrast, aquatic vegetation constitutes a critical component of lake ecosystems, playing essential roles in carbon sequestration, oxygen production, sediment stabilization, and the suppression of excessive algal growth (Hestir et al., 2016; Wang et al., 2019). Extensive studies have demonstrated that the frequent occurrence of cyanobacterial blooms and the degradation of aquatic vegetation often exhibit pronounced co-variation, closely linked to the ecological degradation of shallow lakes undergoing a regime shift from a “macrophyte-dominated clear-water state” to an “algal-dominated turbid-water state” (Hestir et al., 2016; Phillips et al., 2016; Song et al., 2025). Consequently, refined and integrated monitoring of cyanobacterial blooms and aquatic vegetation is of great significance for lake ecosystem assessment and water environment management.

Remote sensing has become a primary technical approach for large-scale and long-term monitoring of cyanobacterial blooms and aquatic vegetation, owing to its extensive spatial coverage, high temporal resolution, and strong timeliness (Qin et al., 2025). Existing studies have predominantly relied on multispectral satellite imagery and constructed spectral indices to identify or distinguish cyanobacterial blooms and aquatic vegetation. For example, indices such as the floating algae index (FAI) (Hu, 2009), virtual-baseline floating macroAlgae height (VB-FAH) (Xing and Hu, 2016), and algae-turbidity bloom index (ATBI) (Du et al., 2024) have been widely used for the detection of cyanobacterial blooms or algae. For aquatic vegetation mapping, commonly used indices include the floating vegetation submerged index (FVSI) (Luo et al., 2017), submerged vegetation sensitive index (SVSI) (Luo et al., 2017), and the normalized difference aquatic vegetation index (NDAVI) (Villa et al., 2015). In addition, several indices, such as the aquatic vegetation index (AVI) (Luo et al., 2023) and normalized difference water index$_{4,5}$ (NDWI$_{4,5}$) (Oyama et al., 2015), have been developed to differentiate between cyanobacterial blooms and aquatic vegetation. However, these methods fundamentally depend on linear or nonlinear combinations of a limited number of spectral bands to enhance target-background contrast. Their discriminative performance is therefore highly constrained by the spectral separability between cyanobacterial blooms and aquatic vegetation. In practice, both targets exhibit highly similar spectral responses in the visible to near-infrared range, and this similarity is further compounded by variations in imaging conditions, water optical properties, and regional environmental characteristics under complex scenarios (Fang et al., 2018; Hou et al., 2022). As a result, traditional spectral index-based approaches are prone to misclassification and omission errors. Furthermore, threshold determination constitutes a fundamental bottleneck in spectral index-based methods. Fixed-threshold strategies generally exhibit poor generalization across varying imaging conditions. Although automated thresholding techniques, such as Otsu’s method (Cao et al., 2021), the gradient mode method (Hu et al., 2010), and the algal pixel growing method (Zhang et al., 2014), have been proposed, they predominantly depend on global or local pixel histograms. Consequently, these methods are highly sensitive to variations in observation conditions and water quality, which hampers the reliable identification of optimal thresholds and often results in substantial extraction errors.

Machine learning approaches, such as random forest (RF), support vector machines (SVM), and eXtreme Gradient Boosting (XGBoost), have also been applied to the extraction of cyanobacterial blooms and aquatic vegetation (Pande-Chhetri et al., 2014; Han et al., 2015; Piaser and Villa, 2022; Lu et al., 2024). However, these methods mainly rely on spectral features and often lack the ability to effectively exploit spatial structural information. Furthermore, models trained for a specific lake are often difficult to generalize to other lakes with different environmental conditions, limiting their cross-region applicability.

With the advancement of deep learning technology, semantic segmentation models based on convolutional neural networks (CNNs) have achieved significant progress in remote sensing applications. A range of representative architectures, such as U-Net (Ronneberger et al., 2015) and the DeepLab series (Chen et al., 2018), have been extensively applied to tasks including land cover classification (Zhang et al., 2018), flood and water inundation mapping (Liu et al., 2019; Zhang et al., 2025b), cloud detection in remote sensing imagery (Wieland et al., 2019), cyanobacterial blooms identification (Yan et al., 2022; Zhu et al., 2023), and aquatic vegetation mapping (Rodríguez-Garlito et al., 2023; Gao et al., 2024). These models have demonstrated clear advantages in terms of recognition accuracy and automation. However, conventional CNN-based architectures predominantly exploit local neighborhood information and exhibit limited capability in modeling global contextual relationships and long-range dependencies among spatially distant pixels. Consequently, when distinguishing between cyanobacterial blooms and aquatic vegetation, these models often focus more on local color and texture features, while underutilizing information about the overall spatial distribution patterns and broader spatial configurations across larger areas. This limitation ultimately constrains the accurate identification of both target classes. In recent years, transformer-based models (Vaswani et al., 2017) have demonstrated strong capability in modeling global contextual relationships and have been increasingly introduced into remote sensing semantic segmentation tasks. Nevertheless, these models still face challenges in fine-grained structural representation, small-scale target extraction, and precise boundary delineation (Wang et al., 2022; Cheng et al., 2023). Moreover, achieving reliable discrimination in scenarios where cyanobacterial blooms and aquatic vegetation coexist or are spatially intermixed remains difficult for traditional spectral index-based approaches. Notably, existing deep learning studies have predominantly focused on detecting either cyanobacterial blooms or aquatic vegetation as a single target class, while relatively limited attention has been devoted to their simultaneous and coordinated recognition.

In lake satellite remote sensing imagery, cyanobacterial blooms and aquatic vegetation exhibit more complex spatial morphological characteristics. As illustrated in Fig.1, both targets can form spatially continuous, large-scale patchy or belt-like structures (Fig.1a, c). At the same time, under the combined influence of wind, waves, currents, and hydrodynamic processes, their boundaries often become highly irregular, giving rise to small, fragmented patches along the margins (Fig.1b, d). Therefore, a recognition model for cyanobacterial blooms and aquatic vegetation needs to be capable of accurately extracting both large and small patches simultaneously. Furthermore, in regions where cyanobacterial blooms and aquatic vegetation coexist or are spatially intermixed, their spectral and textural differences become increasingly subtle (Fig.1e, f), rendering reliable discrimination particularly challenging. This places high demands on the feature extraction and selection capability of segmentation models. Network architectures operating at a single scale or with fixed receptive fields inherently suffer

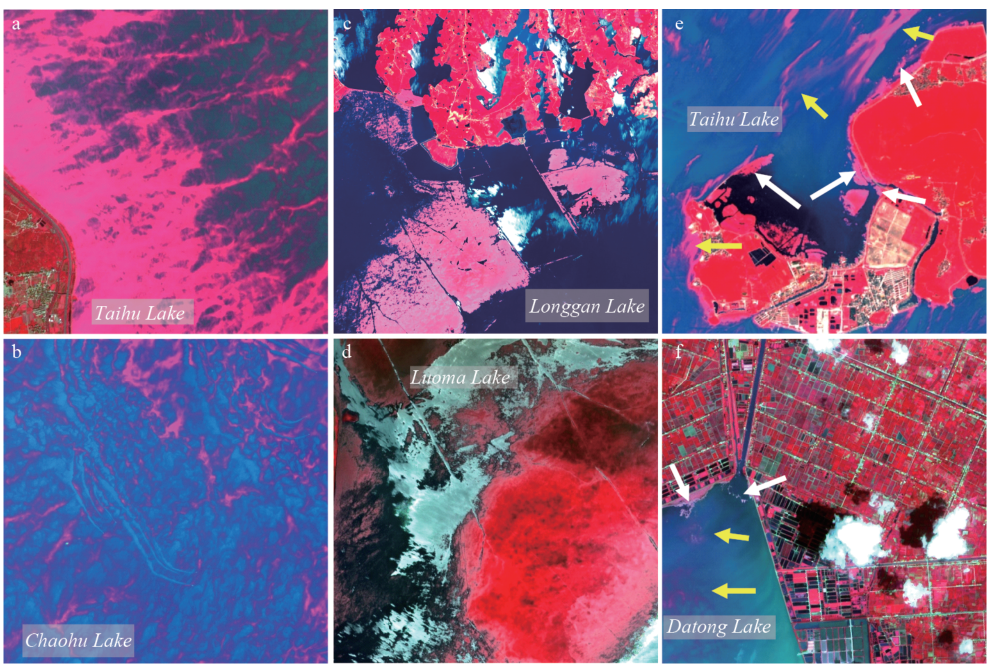

Fig.1 Representative examples illustrating the scale variability, boundary complexity, and spectral similarity of cyanobacterial blooms and aquatic vegetation in lake environments

a & b. examples of cyanobacterial blooms; c & d. examples of aquatic vegetation; e & f. examples of spatial coexistence between cyanobacterial blooms and aquatic vegetation. White arrows indicate aquatic vegetation, while yellow arrows denote cyanobacterial blooms. All images are displayed using a standard false-color composite (red: near-infrared band; green: red band; blue: green band).

from limited representational capacity, making it difficult to simultaneously capture global spatial structures and fine-grained local details. Such limitations restrict the effective fusion and joint modeling of multi-scale information (Lin et al., 2017; Liu et al., 2024; Li and Wu, 2025), which constitutes a critical technical bottleneck in the fine-grained differentiation of cyanobacterial blooms and aquatic vegetation using deep learning approaches.

To address the aforementioned challenges, this study proposes a semantic segmentation network, termed the global-local multi-scale feature extraction network (GLMFENet), which is designed to jointly extract cyanobacterial blooms and aquatic vegetation by integrating global and local multi-scale features. Built upon an encoder-decoder architecture, GLMFENet incorporates a global feature extraction (GFE) module to explicitly model long-range dependencies, together with a spatial attention-based local multi-scale feature extraction (SLMFE) module to enhance the representation of fragmented targets and complex boundaries. Through the effective fusion of global semantic context and local structural details, the proposed network enables accurate and robust extraction of cyanobacterial blooms and aquatic vegetation.

2 STUDY AREA AND DATA DESCRIPTION

2.1 Study area

To develop a robust recognition model for cyanobacterial blooms and aquatic vegetation with strong generalization capability, this study selected 33 representative lakes and reservoirs across China as study areas (Fig.2), including major inland water bodies such as Taihu Lake, Chaohu Lake, Dianchi Lake, Hulun Lake, and Gaoyou Lake. These sites span diverse geographic regions, including the eastern plains, the middle and lower reaches of the Changjiang (Yangtze) River, the southwestern plateau, and high-latitude areas of northwestern China, covering temperate, subtropical, and arid climate zones. The pronounced heterogeneity in aquatic environmental conditions across these regions ensures comprehensive representation of the major types of inland lakes and reservoirs in China.

To facilitate the analysis of different ecological states, the selected lakes were further categorized into three representative types based on long-term satellite observations during 2016–2022. The classification framework adopted in this study is based on the methodological approach proposed by

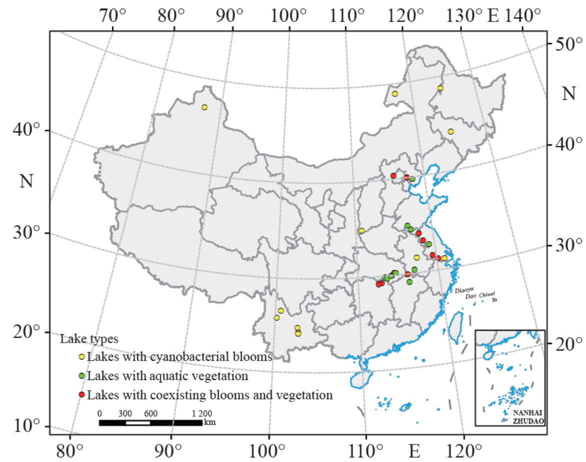

Fig.2 Spatial distribution of study lakes used for training and validation of cyanobacterial blooms and aquatic vegetation segmentation models across China

The map illustrates the geographic locations of representative lakes and reservoirs selected in this study. Yellow dots indicate lakes with cyanobacterial blooms, green dots represent lakes with aquatic vegetation, and red dots denote lakes with coexisting cyanobacterial blooms and aquatic vegetation. Map review No. GS(2024)0650.

Xu et al. (2025), where the original thresholds were modified to accommodate the specific characteristics of the study lakes.

Specifically, lakes were identified as (1) lakes with aquatic vegetation if the maximum annual coverage of aquatic vegetation exceeded 15% of the lake area. Lakes were classified as (2) lakes with cyanobacterial blooms if cyanobacterial bloom events occurred in at least two years during the study period, indicating recurring bloom activity. Lakes satisfying both criteria were categorized as (3) lakes with coexisting cyanobacterial blooms and aquatic vegetation, representing environments where these two ecological features frequently occur together. Representative examples include bloom-prone lakes such as Chaohu Lake and Dianchi Lake, vegetation-rich lakes such as Gaoyou Lake, and lakes with coexisting blooms and vegetation such as Taihu Lake and Longgan Lake, which provide challenging scenarios for distinguishing spectrally similar targets.

2.2 Satellite data

This study employs Sentinel-2 Multispectral Instrument (MSI) imagery as the primary data source. Sentinel-2 MSI offers rich spectral information across 13 bands (400–2 500 nm) with hierarchical spatial resolutions of 10, 20, and 60 m. For this study, the three 60-m resolution bands were excluded; the six 20-m spatial resolution bands were resampled to 10 m and merged with the original four 10-m bands, yielding a unified 10-band multispectral dataset with a consistent 10-m spatial resolution. This dataset is well suited for the fine-scale remote sensing identification of cyanobacterial blooms and aquatic vegetation. A total of 467 Sentinel-2 Level-1C (top-of-atmosphere reflectance, TOA reflectance) images acquired between 2016 and 2022, with minimal or no cloud contamination, were obtained via the Google Earth Engine (GEE) platform. The collected imagery spans multiple seasons and geographic regions, enabling comprehensive characterization of the spatiotemporal distribution patterns of cyanobacterial blooms and aquatic vegetation. The utilized Sentinel-2 data offer several key advantages: (1) rich spectral dimensionality, providing abundant spectral information for effective feature discrimination; (2) moderate spatial resolution, allowing reliable identification of small-scale structural features; and (3) extensive spatial coverage combined with a long temporal span, facilitating the construction of a diverse dataset with strong generalization potential. Overall, the selected Sentinel-2 dataset provides a robust and consistent foundation for model training, validation, and the evaluation of generalization performance.

2.3 Data preprocessing and sample set construction

To obtain high-quality training samples, all Sentinel-2 images were subjected to systematic preprocessing, including spatial resampling, study area cropping, water masking, and data annotation. First, all bands with a spatial resolution of 20 m were resampled to 10 m using the GEE platform to ensure spatial consistency across spectral channels. Subsequently, images were cropped based on lake boundaries, with the lake area accounting for approximately 80% of each image, thereby minimizing interference from extensive land pixels. A precise water mask was generated following established methods (Yan et al., 2022), and only pixels within lake boundaries were considered during the annotation process.

For data annotation, initial labels were generated using spectral index-based approaches following prior studies (Yan et al., 2022), where cyanobacterial blooms were delineated using a threshold on the FAI, and aquatic vegetation was extracted using a threshold on the normalized difference vegetation index (NDVI). The thresholds for both indices were determined through visual interpretation by experienced annotators, ensuring consistency between the spectral index responses and the actual distribution patterns observed in the satellite imagery. These preliminary labels were subsequently reviewed through cross-validation by multiple annotators and further refined by pixel-wise manual correction to ensure the accuracy and consistency of the final annotations for three classes: cyanobacterial blooms, aquatic vegetation, and background water. To enhance spatial diversity and mitigate class imbalance, representative regions from different areas within each lake were selectively annotated. Examples of the final ground-truth labels are shown in Fig.3, illustrating the labeling quality and classification criteria used in this study.

The final dataset comprises 467 Sentinel-2 images, including 397 images for model development (training and validation) and 70 images for independent testing. All images were cropped into 512×512 patches based on their spatial locations, with an overlap ratio of 10% between adjacent patches to preserve boundary continuity and increase sample quantity and diversity. The 397 images yielded 6 471 training and 719 validation patches, while the remaining 70 images produced 1 355 test patches. As illustrated in Supplementary Fig.S1, the constructed dataset covers a wide range of sample types, providing sufficient support for systematic evaluation of model performance, robustness, and generalization capability.

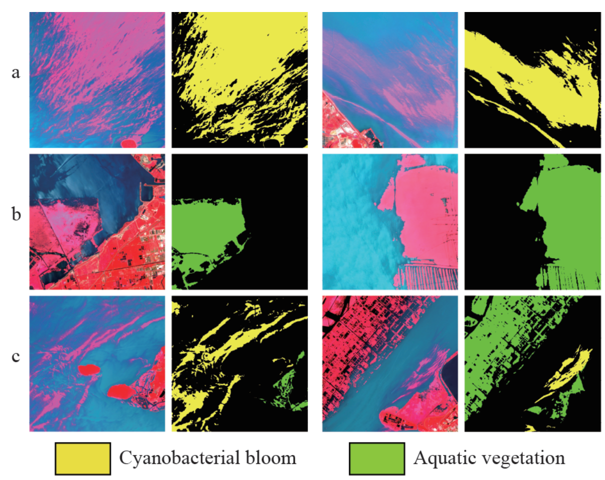

Fig.3 Examples of ground-truth annotations

a. cyanobacterial blooms; b. aquatic vegetation; c. coexisting cyanobacterial blooms and aquatic vegetation. Sentinel-2 image patches displayed using a standard false-color composite (red: near-infrared band; green: red band; blue: green band), along with the corresponding ground-truth labels (yellow: cyanobacterial blooms; green: aquatic vegetation).

3 METHOD

3.1 GLMFENet network architecture for cyanobacterial blooms and aquatic vegetation recognition

3.1.1 Overall architecture of GLMFENet

To address the challenges in effectively distinguishing and recognizing cyanobacterial blooms and aquatic vegetation under conditions of spectral similarity, significant variations in patch sizes, and fragmented local spatial structures, this paper proposes a semantic segmentation network named GLMFENet, which integrates global semantic modeling with local multi-scale feature enhancement. The overall network adopts an Encoder-Decoder architecture, utilizing a typical deep convolutional network as the backbone. By incorporating specially designed feature enhancement modules during both encoding and decoding stages, the model's ability to discriminate between cyanobacterial blooms and aquatic vegetation, as well as to delineate details, is enhanced for complex scenarios.

In the encoding stage, the network performs hierarchical convolutional feature extraction on the input multispectral remote sensing imagery, generating feature representations at progressively higher semantic levels. The spatial resolutions of these feature maps are successively reduced to 1, 1/2, 1/4, 1/8, and 1/16 of the original input, with corresponding channel dimensions of 64, 128, 256, 512, and 512, respectively. On the deepest feature map, a GFE module is introduced to model global relationships among high-level features, thereby enhancing the network's ability to capture long-range dependencies and global contextual information.

In the decoding stage, the feature maps produced by the GFE module are progressively upsampled. At each decoding level, the SLMFE module is introduced to fuse the upsampled deep global features with the corresponding shallow features. The SLMFE module consists of a multi-scale feature extraction (MSFE) submodule and a spatial attention mechanism, enabling the generation of discriminative multi-scale feature representations. This fusion process reinforces the representation of boundary regions, fine fragmented patches, and complex mixed areas, and is repeated across five decoding stages. Finally, pixel-wise classification is achieved via a 1×1 convolution followed by a Softmax activation to generate the semantic segmentation results.

Through the above architectural design, GLMFENet effectively preserves global semantic consistency while simultaneously enhancing the representation of fine-grained local spatial structures, thereby improving discrimination performance between cyanobacterial blooms and aquatic vegetation under complex scenarios. The overall framework of GLMFENet is illustrated in Fig.4a.

3.1.2 Design of the global feature extraction module

At local scales, cyanobacterial blooms and aquatic vegetation exhibit highly similar spectral responses and textural characteristics, whereas their differences become more pronounced at larger spatial scales in terms of distribution patterns and structural organization. Consequently, feature representations derived solely from the local receptive fields of convolutional operations are insufficient for effectively discriminating between these two targets. Recent studies have demonstrated that transformer-based global modeling mechanisms are capable of capturing long-range spatial dependencies and providing rich contextual information, which is particularly beneficial for semantic segmentation in complex scenes (Mou et al., 2020; Fu et al., 2021; Gao et al., 2021). Motivated by these findings, a global feature extraction module is introduced at the high-level layers of the network to enhance the modeling of global contextual information and long-range spatial dependencies.

Inspired by the transformer architecture, the GFE module performs multi-dimensional global modeling on high-level feature maps (Gao et al., 2021), enabling the effective characterization of contextual relationships across different spatial regions within water bodies. Specifically, the GFE module integrates a standard transformer encoder with two directional transformer encoders operating along the horizontal and vertical spatial axes of the 2D feature map. The horizontal axis corresponds to the width direction, modeling spatial dependencies along each row, while the vertical axis corresponds to the height direction, capturing long-range relationships along each column. The standard transformer encoder models global contextual interactions among all spatial tokens, whereas the directional transformers explicitly capture structured spatial dependencies along row-wise and column-wise directions. This directional modeling is conceptually similar to axial attention (Ho et al., 2019), which decomposes 2D global attention into sequential 1D operations along different spatial axes. Such a design allows the

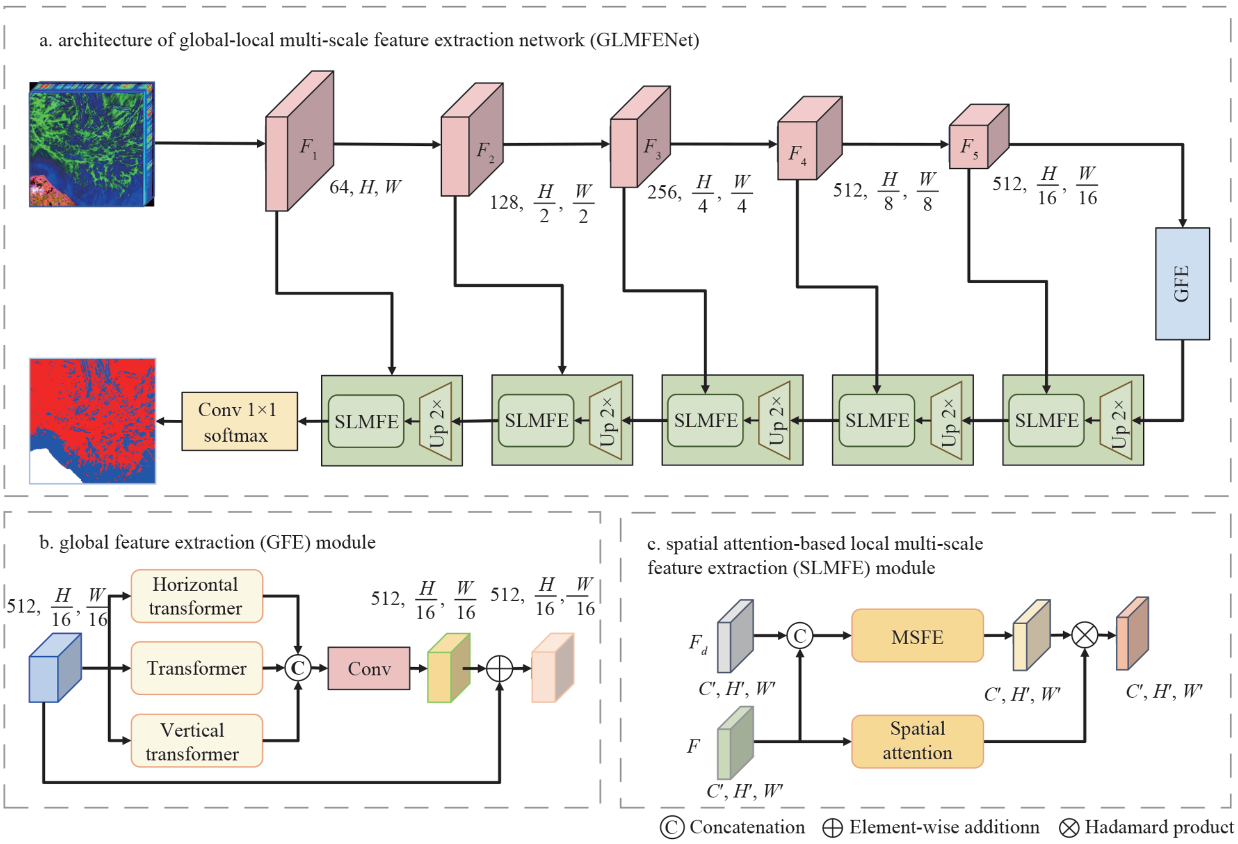

Fig.4 Overall architecture of the proposed global-local multi-scale feature extraction network (GLMFENet) for cyanobacterial blooms and aquatic vegetation recognition

a. schematic overview of the GLMFENet architecture, which adopts an encoder-decoder framework and integrates two key components: global feature extraction (GFE) module and spatial attention-based local multi-scale feature extraction (SLMFE) module; b. structural illustration of the GFE module, which employs transformer-based global modeling to capture long-range spatial dependencies and global contextual information; c. structural illustration of the SLMFE module, consisting of a multi-scale feature extraction (MSFE) submodule and a spatial attention mechanism for enhancing fine-grained local feature representation; where C, H/H', and W/W' represent the channel dimension, spatial height, and spatial width of the feature map, respectively.

network to better represent spatially dispersed or elongated patterns commonly observed in cyanobacterial blooms and aquatic vegetation, while mitigating the influence of local noise on feature discrimination.

In the multi-head self-attention mechanism, the number of attention heads is set to 4, and the dimensions of the Query, Key, and Value matrices are set to 512, which are consistent with the channel dimension of the input feature map. This configuration provides a balance between representation capacity and computational efficiency. By embedding the GFE module in the encoder stage, the network can learn more coherent and globally consistent high-level feature representations, thereby providing reliable contextual support for subsequent fine-grained segmentation in the decoding stage. Detailed descriptions of the GFE module are provided in Supplementary Text S1 and Supplementary Fig.S2.

3.1.3 Design of the spatial attention-based local multi-scale feature extraction module

While global contextual information contributes to improving the overall coherence of segmentation results, cyanobacterial blooms and aquatic vegetation in practical remote sensing imagery often exhibit highly fragmented small-scale patches, with cyanobacterial bloom boundaries being particularly irregular, variable, and visually ambiguous. Accurate characterization of local fine-grained features is therefore essential. Previous studies have demonstrated that multi-scale convolutional architectures are effective in capturing local texture and boundary information (Feng et al., 2024), and that spatial attention mechanisms can further enhance discriminative feature responses by suppressing irrelevant background regions (Yuan et al., 2020; Ni et al., 2024; Ma et al., 2025). However, most existing multi-scale attention modules are primarily designed for natural image segmentation and are not directly applicable to lake and reservoir remote sensing imagery, where targets are characterized by more fragmented fine-scale spatial patterns and stronger inter-class feature overlap.

To enhance the model's sensitivity to local spatial structures, this study designs SLMFE module (Fig.4c). The proposed SLMFE module jointly takes shallow features containing rich spatial detail information and decoder features encoding global contextual relationships as inputs. The MSFE submodule (Fig.5) is employed to model spatial structural features at different scales, followed by a scale-adaptive fusion strategy to weight multi-scale features. In addition, a spatial attention mechanism (Woo et al., 2018) is incorporated to emphasize informative regions while suppressing background interference, thereby strengthening the representation of key local features.

Specifically, the MSFE module applies convolutional operators with different kernel sizes in parallel to the input feature map, thereby capturing spatial feature representations at multiple scales. Subsequently, a multilayer perceptron (MLP) combined with a feature selection gate is employed to estimate scale-wise importance weights, enabling adaptive weighted concatenation and refined fusion of multi-scale features. This design effectively enhances the model's capability to characterize boundary structures, fragmented patches, and complex spatial patterns. Detailed implementation details of the SLMFE module are provided in Supplementary Text S2 and Supplementary Fig.S3.

Benefiting from the SLMFE module, the network is able to more accurately perceive the edge structures of cyanobacterial blooms and aquatic vegetation, as well as their transitional zones with surrounding water bodies. This leads to a reduction in misclassification and omission errors and demonstrates improved adaptability, particularly in regions characterized by complex backgrounds and small-scale targets. Through synergistic interaction with the GFE module, the proposed network achieves an effective integration of global semantic consistency and local fine-grained feature representation.

3.2 GLMFENet training and evaluation for cyanobacterial blooms and aquatic vegetation recognition

All models were implemented using the PyTorch framework. Experiments were conducted on a Windows Server 2019 system with PyTorch 2.0, and an NVIDIA GeForce RTX 3090 GPU was used to accelerate model training and inference. To ensure a consistent comparison, all experiments were performed under identical hardware and software configurations. Detailed descriptions of the training strategy and hyperparameter settings are provided in Supplementary Text S3 and Supplementary Fig.S4.

To comprehensively evaluate the recognition performance of the proposed model, several widely adopted semantic segmentation metrics were calculated on the test dataset. These metrics include the Intersection over Union (IoU) and F1-score for each class, as well as the mean IoU (mIoU), mean F1-score (Mean F1), Overall Accuracy (OA), and the Kappa coefficient. Specifically, IoU quantifies the overlap between predicted regions and ground truth labels, while mIoU represents the average IoU across all classes and provides an overall assessment of multi-class segmentation performance. The F1-score, which jointly considers precision and recall, offers improved stability under class-imbalanced conditions. OA indicates the proportion of correctly classified pixels among all pixels, whereas the

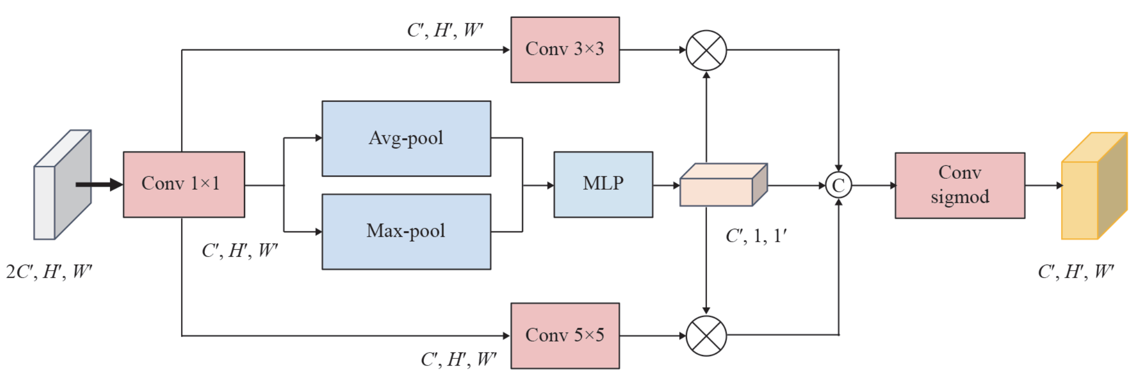

Fig.5 Architecture of the multi-scale feature extraction (MSFE) module

C', H', and W' represent the channel dimension, spatial height, and spatial width of the feature map, respectively.

Kappa coefficient measures the agreement between model predictions and reference labels after accounting for chance agreement, thereby providing a more objective and reliable evaluation. This makes Kappa particularly suitable for remote sensing applications characterized by high background proportions and class imbalance. Detailed formulations of these evaluation metrics are provided in Supplementary Text S4.

4 RESULT AND DISCUSSION

4.1 Extraction results of cyanobacterial blooms and aquatic vegetation

The proposed GLMFENet was systematically evaluated on the test dataset and compared with several representative semantic segmentation models, including DeepLabV3+ (Chen et al., 2018), DLinkNet (Zhou et al., 2018), SegNet (Badrinarayanan et al., 2017), U-Net (Ronneberger et al., 2015), SegFormer (Xie et al., 2021), and Swin-UNet (Cao et al., 2023). All comparative experiments were conducted under consistent training strategies and hardware environments to ensure the objectivity and comparability of the results. Quantitative results of different methods on evaluation metrics such as OA, mIoU, Mean F1, and Kappa are presented in Table 1.

As shown in Table 1, GLMFENet consistently achieved the highest performance across all evaluation metrics. In terms of mIoU, GLMFENet attained a value of 84.6%, exceeding the comparative methods by margins ranging from 3.3% to 7.8%. Similarly, GLMFENet achieved a Mean F1 of 91.3% and a Kappa coefficient of 86.4%, outperforming the other models by 2.0%–5.0% and 3.1%–7.6%, respectively. With respect to OA, GLMFENet reached 96.7%, showing an improvement of 0.9%–2.0% over the comparison methods. The relatively smaller differences observed in OA can be attributed to the dominance of background pixels in the dataset, which is consistent with real-world remote sensing scenarios.

At the class level, GLMFENet achieved markedly higher IoU and F1-scores for all three categories compared with other models. In particular, for the cyanobacterial blooms class, GLMFENet obtained an IoU of 85.1% and an F1-score of 92.0%, corresponding to improvements of 5.2%–13.2% and 3.2%–8.3%, respectively. For the aquatic vegetation class, the proposed method achieved an IoU of 72.4% and an F1-score of 84.0%, outperforming the other methods by 3.0%–10.7% and 2.0%–7.7%. Although aquatic vegetation remains more challenging to extract for all models due to its fragmented spatial patterns and spectral similarity to surrounding features, GLMFENet exhibits the most pronounced performance gains. These results demonstrate the superior adaptability of the proposed method in discriminating cyanobacterial blooms and aquatic vegetation under conditions of strong spectral similarity and complex spatial fragmentation.

Figure 6 illustrates the extraction performance of different models under scenarios characterized by fragmented spatial structures and pronounced multi-scale mixing, providing a qualitative assessment of model adaptability in complex lake environments.

In fragmented aquatic vegetation regions (Fig.6a), most baseline models lose fine textural information during downsampling and tend to extract only a limited number of larger patches. In contrast, GLMFENet preserves both the overall vegetation morphology and fine arc-shaped structures, demonstrating enhanced capability in

Table 1 Quantitative comparison of different semantic segmentation models for cyanobacterial blooms and aquatic vegetation extraction on the test dataset

Model	Background (IoU ↑/F1 ↑)	Cyanobacterial blooms (IoU ↑/F1 ↑)	Aquatic vegetation (IoU ↑/F1 ↑)	mIoU ↑	Mean F1 ↑	OA ↑	Kappa ↑
DeepLabV3+	94.7%/97.3%	73.1%/84.5%	69.4%/82.0%	79.1%	87.9%	95.3%	81.3%
DLinkNet	94.3%/97.1%	72.1%/83.8%	65.5%/79.1%	77.3%	86.7%	94.9%	79.3%
SegNet	94.1%/97.0%	71.9%/83.7%	64.4%/78.3%	76.8%	86.3%	94.7%	78.8%
SegFormer	94.7%/97.3%	73.5%/84.7%	69.2%/81.8%	79.2%	87.9%	95.3%	81.3%
Swin-UNet	95.3%/97.6%	79.9%/88.8%	68.6%/81.4%	81.3%	89.3%	95.8%	83.3%
U-Net	94.4%/97.1%	77.8%/87.5%	61.7%/76.3%	78.0%	87.0%	95.0%	79.9%
GLMFENet	**96.3%/98.1%**	**85.1%/92.0%**	**72.4%/84.0%**	**84.6%**	**91.3%**	**96.7%**	**86.4%**

The best results are shown in bold.; ↑ indicates that higher values represent better performance.

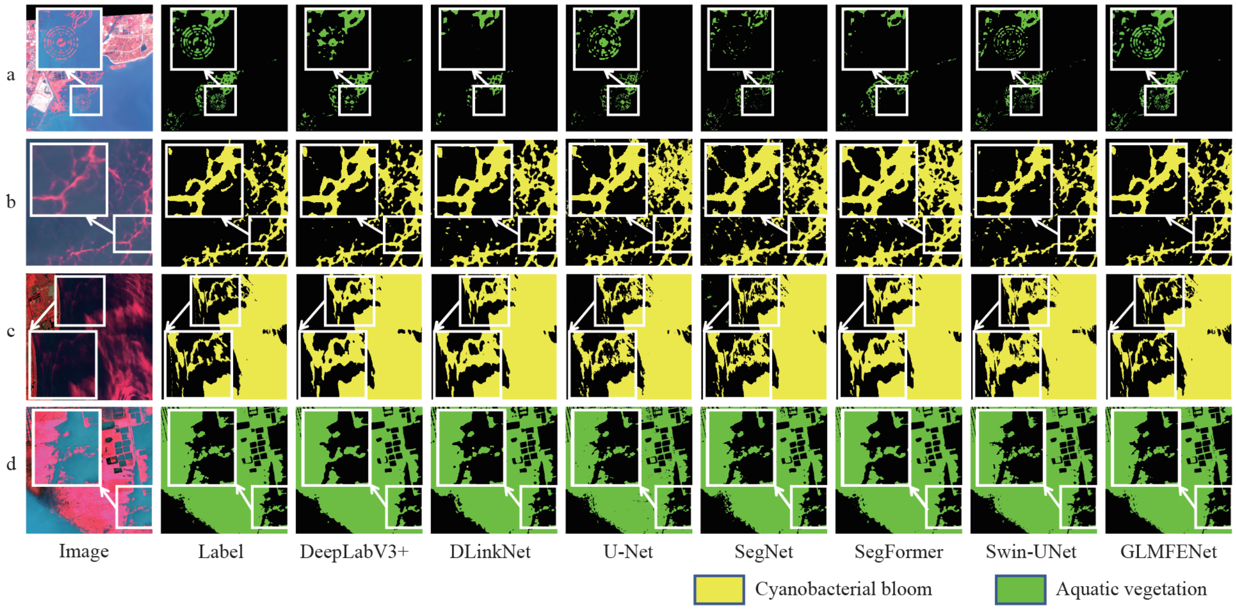

Fig.6 Visual comparison of cyanobacterial blooms and aquatic vegetation extraction results obtained by different deep learning models

a. extraction results of fine and fragmented structure of aquatic vegetation; b. extraction results of fine and fragmented structure of cyanobacterial blooms; c. extraction results of fine and fragmented features of cyanobacterial blooms in multi-scale scenes; d. extraction results of fine and fragmented features of aquatic vegetation in multi-scale scenes. The first column shows the original Sentinel-2 MSI images displayed in a standard false-color composite (red: near-infrared band; green: red band; blue: green band).

retaining local spatial details.

In areas dominated by fragmented cyanobacterial blooms (Fig.6b), the results produced by U-Net, SegNet, SegFormer, and Swin-UNet exhibit evident over-extraction or blurred boundaries. Although DeepLabV3+ better maintains the global shape of bloom regions, it suffers from substantial loss of boundary details. By comparison, GLMFENet achieves more accurate boundary delineation while effectively suppressing background noise, resulting in improved spatial consistency of the extracted bloom regions.

For scenes with pronounced scale differences between cyanobacterial blooms and aquatic vegetation (Fig.6c, d), several models show omission errors or structural fragmentation when identifying small-scale targets. In contrast, GLMFENet is able to simultaneously extract large-scale clusters and small fragmented patches, indicating strong robustness to scale variation. This performance can be attributed to the complementary roles of the GFE and SLMFE modules in capturing global contextual information and local multi-scale structural features, respectively.

Overall, GLMFENet demonstrates clear advantages in detail preservation, boundary delineation, and multi-scale target representation, exhibiting more stable and reliable adaptability across typical lake-water scenarios characterized by complex spatial structures and significant scale variability.

Furthermore, to assess model robustness under challenging and easily confused conditions, Fig.7 presents qualitative comparisons across a range of complex scenarios, including sun glint interference (Fig.7a), cloud-contaminated regions (Fig.7b), areas where cyanobacterial blooms and aquatic vegetation coexist (Fig.7c), as well as typical cases of cyanobacterial blooms misclassification (Fig.7d) and aquatic vegetation misclassification (Fig.7e). In these scenarios, transformer-based models such as SegFormer and Swin-UNet tend to misidentify clouds and highly reflective objects as target classes, and frequently confuse cyanobacterial blooms with aquatic vegetation. This behavior can be attributed to the sliding-window self-attention mechanism, which may propagate erroneous contextual cues during feature aggregation. In contrast, CNN-based models are constrained by their limited receptive fields, leading to omission and misclassification errors, particularly for small or fragmented targets.

By comparison, GLMFENet demonstrates more stable performance across all interfering conditions. Through the joint integration of global semantic constraints and local multi-scale structural features, the proposed model effectively suppresses noise-

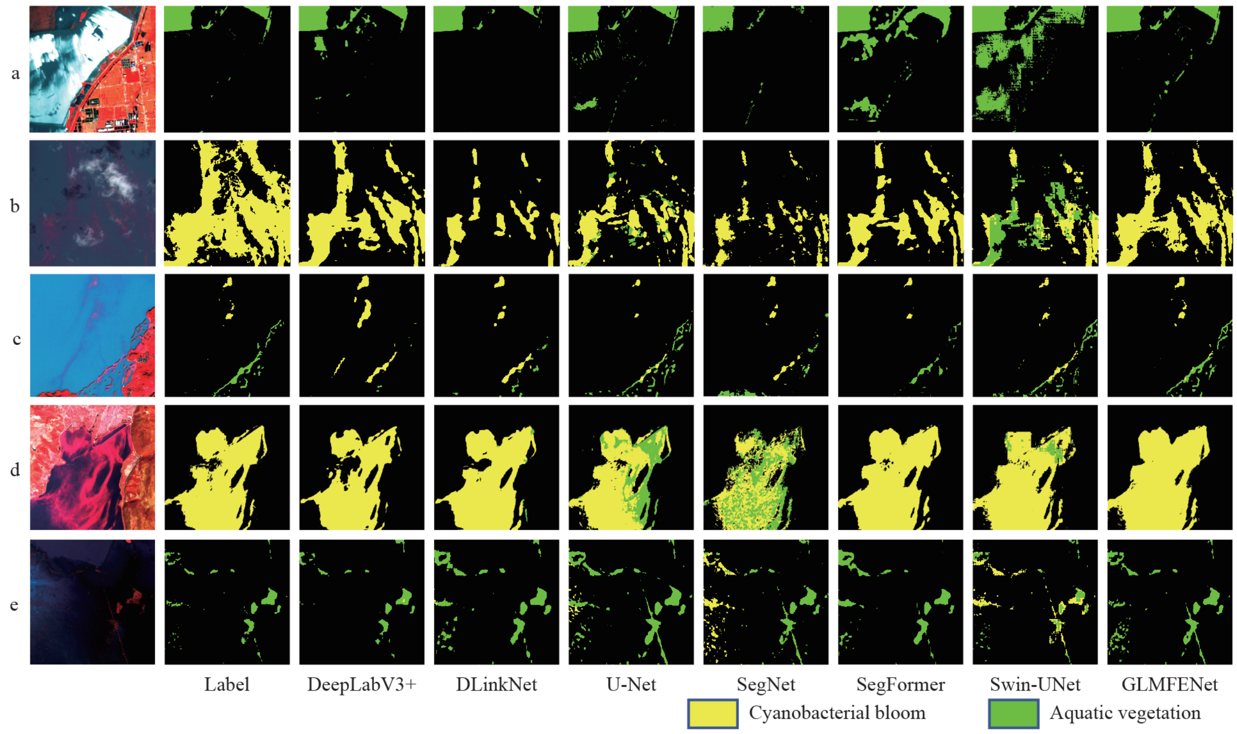

Fig.7 Visual comparison of cyanobacterial blooms and aquatic vegetation extraction results under disturbed and misclassification scenarios

a. extraction results under sun glint interference; b. extraction results under cloud-contaminated conditions; c. extraction results in regions with coexisting cyanobacterial blooms and aquatic vegetation; d. misclassification examples of cyanobacterial blooms; e. misclassification examples of aquatic vegetation. The first column shows the original Sentinel-2 MSI images displayed using a standard false-color composite (red: near-infrared band; green: red band; blue: green band).

induced responses and preserves spatial consistency, resulting in the fewest misclassification errors among the compared methods under complex environmental disturbances.

To further assess the practical advantages of GLMFENet, a comparative analysis was conducted using 15 Sentinel-2 MSI images against traditional spectral index-based methods, including the vegetation and bloom indices (VBI) and $NDWI_{4,5}$. Detailed implementation settings of these spectral index methods are provided in Supplementary Text S5 and Table S1. The selected images cover a diverse set of lakes, including Chaohu Lake, Datong Lake, Dianchi Lake, Dushan Lake, Zhaoyang Lake, Gaoyou Lake, Honghu Lake, Huanggai Lake, Hulun Lake, Taihu Lake, Weishan Lake, and Xingyun Lake, thereby representing a wide range of inland water environments.

As summarized in Table 2, GLMFENet consistently outperforms the spectral index-based approaches in terms of mIoU, Mean F1, OA, and the Kappa coefficient. The results demonstrate that GLMFENet exhibits more robust and reliable performance under complex application scenarios, achieving substantially improved accuracy compared with traditional spectral index methods.

To qualitatively compare the performance of different methods, a Sentinel-2 MSI image acquired over Taihu Lake on 9 May 2019 was selected as a representative case (Fig.8). On this date, cyanobacterial blooms were mainly distributed along

Table 2 Quantitative comparison of GLMFENet and spectral index-based methods for cyanobacterial blooms and aquatic vegetation extraction using Sentinel-2 MSI imagery

Model	mIoU ↑	OA ↑	Mean F1 ↑	Kappa ↑
VBI	67.8%	93.2%	78.9%	64.5%
$NDWI_{4,5}$	76.7%	97.0%	86.1%	81.0%
GLMFENet	87.4%	98.7%	93.0%	90.5%

↑ indicates that higher values represent better performance.

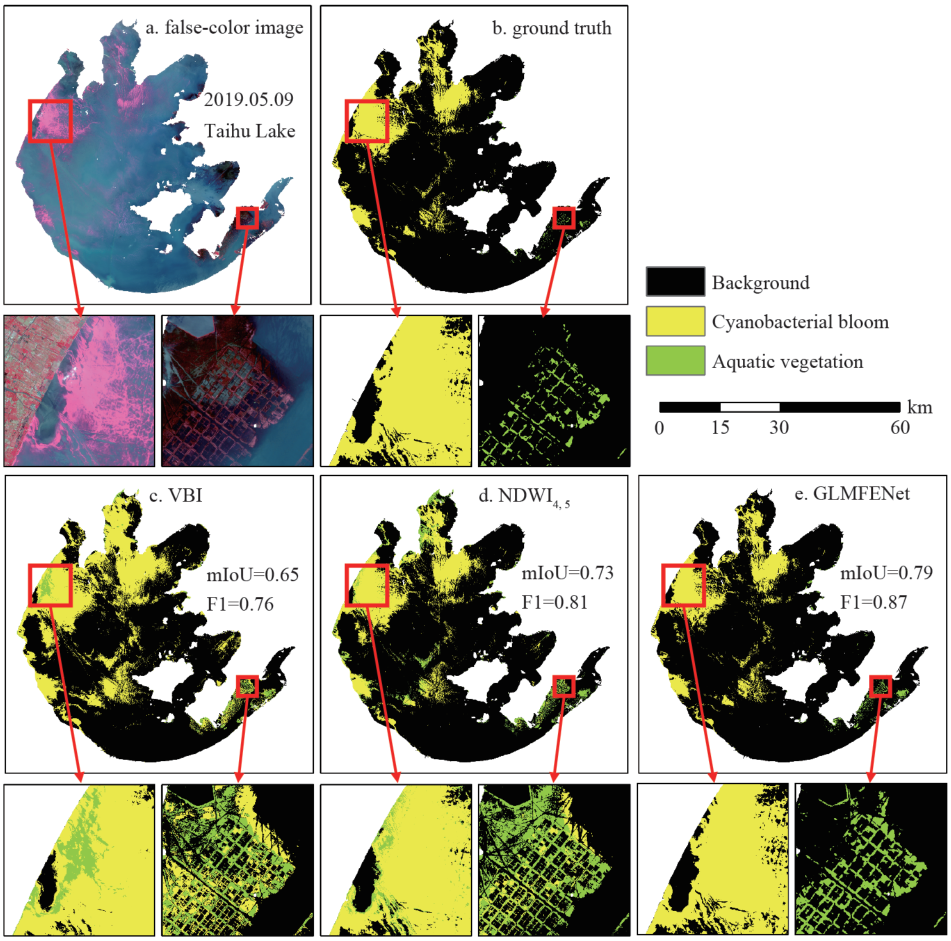

Fig.8 Visual comparison of results from automated extraction methods for cyanobacterial blooms and aquatic vegetation

a. Sentinel-2 MSI image displayed using a standard false-color composite (red: near-infrared band; green: red band; blue: green band), acquired over Taihu Lake on 9 May 2019; b. corresponding ground-truth labels (yellow: cyanobacterial blooms; green: aquatic vegetation; black: background water); c. extraction results produced by the VBI method; d. extraction results produced by the $NDWI_{4,5}$ method; e. extraction results produced by the proposed GLMFENet.

the eastern shoreline and within the northeastern bays, whereas aquatic vegetation was concentrated in East Taihu Lake and the southeastern regions. The results show that the VBI and $NDWI_{4,5}$ methods exhibit evident misclassification between the two feature types. Large bloom patches in the northeastern lake are frequently identified as aquatic vegetation, while parts of the vegetation belts in East Taihu Lake are incorrectly classified as blooms. In contrast, GLMFENet produces more accurate and spatially coherent results, with clearer boundaries and fewer omission errors. This improvement demonstrates the advantage of integrating global contextual modeling and local feature extraction for distinguishing spectrally similar targets in complex inland-lake environments.

In summary, GLMFENet consistently demonstrates clear advantages over existing deep learning models when evaluated in terms of overall accuracy, class-level discrimination, robustness under complex scenarios, and stability in the presence of anomalous conditions. By jointly leveraging transformer-based global feature enhancement and multi-scale local structural modeling, the proposed framework provides a more reliable automated solution for identifying cyanobacterial blooms and aquatic vegetation, which are characterized by strong spectral similarity, pronounced scale variability, and highly fragmented spatial structures.

4.2 Ablation study of GLMFENet components

To evaluate the effectiveness of individual components within GLMFENet and to verify the rationality of the proposed architecture, a series of ablation experiments were conducted on the Sentinel-2 MSI dataset. Specifically, the GFE module (A2), the SLMFE module (A3), and the combination of both modules (A4) were progressively introduced to quantify their respective contributions to model performance. The detailed quantitative results of the ablation experiments are summarized in Table 3, where the baseline model (A1) corresponds to the U-Net architecture.

After introducing the GFE module (Experiment A2), the mIoU increased to 82.2%, representing a gain of 4.2% over the baseline. This improvement can be attributed to the incorporation of transformer-based global modeling in the GFE module, which enhances feature representation by capturing long-range dependencies and global contextual relationships. By jointly modeling feature interactions along multiple dimensions, the GFE module enriches high-level semantic representations and improves the model's global perception capability.

When the SLMFE module was incorporated independently (Experiment A3), the mIoU increased to 80.4%, yielding an improvement of 2.4% compared with the baseline. This performance gain indicates that the SLMFE module effectively strengthens the extraction of fine-grained local features by leveraging multi-scale convolutional operations and spatial attention. As a result, the model demonstrates enhanced sensitivity to small-scale and fragmented targets, such as filamentous cyanobacterial blooms and scattered aquatic vegetation patches.

When the GFE and SLMFE modules were integrated simultaneously (Experiment A4), the model achieved its best performance, with the mIoU increasing to 84.6%, corresponding to an overall gain of 6.7% compared with the baseline. This result highlights the complementary effects of global semantic modeling and local detail enhancement. Specifically, the incorporation of global features promotes spatial consistency at the scene level, while the integration of local multi-scale features improves the accurate representation of fine-grained structures. The synergistic interaction between these two components substantially enhances the model's discrimination capability under complex lake scenarios.

Overall, the ablation results clearly demonstrate both the necessity and effectiveness of the proposed modular design in GLMFENet, with the joint optimization of global and local feature representations serving as the primary driver of the observed performance improvements.

4.3 Visualization of feature activation maps for GLMFENet

To further interpret the mechanisms of the GFE and SLMFE modules and to examine their contributions to feature representation, gradient-weighted class activation mapping (Grad-CAM) (Selvaraju et al., 2017) was used to visualize class-specific activation responses for cyanobacterial blooms and aquatic vegetation under different ablation configurations (Fig.9). In these visualizations, regions with higher activation intensities indicate areas that contribute more strongly to the classification decision, whereas low-activation regions correspond to areas with limited influence.

As shown in Fig.9b, the activation maps produced by the baseline model (Experiment A1) are diffuse and poorly localized, with weak discrimination among cyanobacterial blooms, aquatic vegetation, and surrounding water bodies. This indicates that the baseline model lacks sufficient capability to distinguish target classes without explicit global or local feature enhancement mechanisms.

After incorporating the GFE module (Experiment A2; Fig.9c), activation responses over cyanobacterial blooms and aquatic vegetation regions become noticeably stronger, while responses over background water are suppressed. Compared with the baseline model, the activation patterns reveal clearer large-scale spatial structures and more coherent object-level responses. This indicates that the GFE module effectively captures long-range spatial dependencies and strengthens the model's ability to model the global spatial organization of water-surface targets.

After incorporating the SLMFE module (Experiment A3; Fig.9d), the activation maps exhibit

Table 3 Ablation study evaluating the contributions of the GFE and SLMFE modules in GLMFENet

Experiment ID	GFE	SLMFE	mIoU ↑	OA ↑	Mean F1 ↑	Kappa ↑
A1			78.0%	95.0%	87.0%	79.9%
A2	√		82.2%	96.2%	89.9%	84.1%
A3		√	80.4%	95.6%	88.6%	82.0%
A4	√	√	84.6%	96.7%	91.3%	86.4%

↑ indicates that higher values represent better performance.

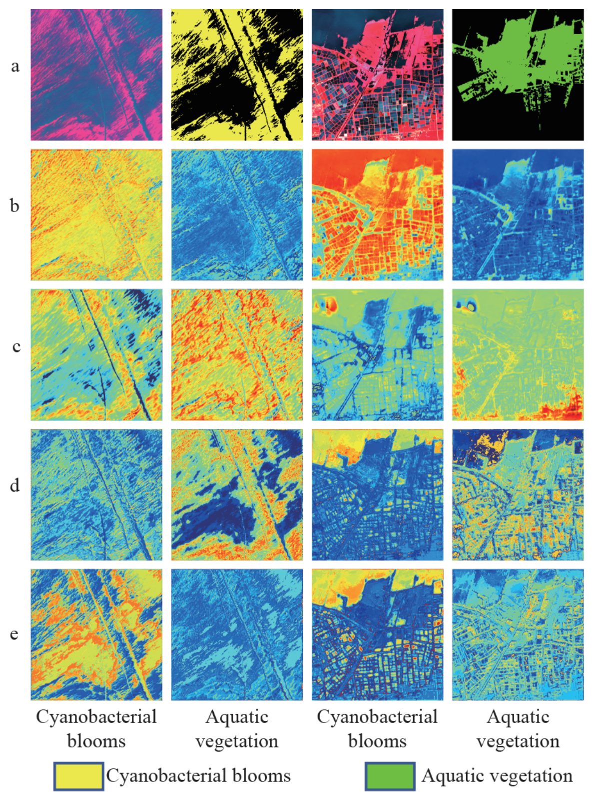

Fig.9 Visualization of feature activation maps obtained from the ablation experiments

a. the original Sentinel-2 MSI images displayed using a standard false-color composite (red: near-infrared band; green: red band; blue: green band), along with the corresponding ground-truth labels; b. feature activation maps generated by the baseline model (Experiment A1); c. feature activation maps obtained after incorporating the global feature extraction (GFE) module (Experiment A2); d. feature activation maps obtained after incorporating the spatial attention-based local multi-scale feature extraction (SLMFE) module (Experiment A3); e. feature activation maps generated by the full model with both GFE and SLMFE modules (Experiment A4).

richer texture information and more distinct boundary structures. Fragmented patches and irregular morphological patterns are more clearly highlighted, indicating that the SLMFE module improves the model's sensitivity to local spatial variations and fine-scale structural details. This improvement can be attributed to two key components of the SLMFE module. First, the spatial attention mechanism adaptively assigns higher weights to spatial locations containing distinctive local features while suppressing relatively homogeneous background regions. By emphasizing these spatially heterogeneous regions, the spatial attention mechanism enables the network to better capture fine-scale boundary structures and local morphological characteristics. Second, the MSFE submodule captures spatial structures at multiple receptive fields, allowing the network to represent both fine-scale boundary details and larger spatial patterns of bloom patches and vegetation beds. This multi-scale representation is particularly important for inland lakes, where cyanobacterial blooms and aquatic vegetation often exhibit substantial variations in patch size and spatial morphology.

When both the GFE and SLMFE modules are jointly employed (Experiment A4; Fig.9e), the activation maps exhibit the most stable and discriminative response patterns. Aquatic vegetation regions show clearer internal differentiation, while cyanobacterial blooms display highly concentrated activations with sharp boundaries and spatially coherent structures. These observations demonstrate the complementary roles of the two modules: the GFE module provides global contextual awareness and long-range dependency modeling, while the SLMFE module focuses on local structural refinement and boundary enhancement. Their integration enables the network to simultaneously capture global contextual information and fine-grained spatial details, resulting in more accurate and robust segmentation.

Overall, the visualization results are consistent with the quantitative evaluation of cyanobacterial bloom and aquatic vegetation extraction performance. From a feature-representation perspective, they provide intuitive evidence for the effectiveness of the proposed modules and offer mechanistic insights into the observed performance improvements.

4.4 Computational efficiency analysis

To evaluate the computational efficiency of the proposed method, we compared GLMFENet with several representative semantic segmentation models, including DeepLabV3+, DLinkNet, SegNet, SegFormer, Swin-UNet, and U-Net, in terms of model parameters (Params), floating-point operations (FLOPs), and inference time.

All models were evaluated under the same experimental conditions using an input patch size of 512×512 with 10 spectral channels, which is consistent with the sliding-window inference strategy adopted for large-scale remote sensing image segmentation. The inference time was measured in evaluation mode on a GPU platform after 20 warm-up iterations, followed by 100 forward passes, and the average time per image was recorded.

The quantitative comparison results are presented in Table 4. The proposed GLMFENet contains 41.28 M parameters and requires 293.54 GFLOPs, with an

Table 4 Computational efficiency comparison of different segmentation models in terms of parameter count, FLOPs, and inference time

Model	Params (M) ↓	FLOPs (G) ↓	Inference time (ms) ↓
DeepLabV3+	3.73	6.95	6.40
DLinkNet	5.82	40.10	7.18
SegNet	27.16	31.12	13.16
SegFormer	21.64	134.32	20.51
Swin-UNet	29.45	161.88	25.16
U-Net	24.90	226.24	32.57
GLMFENet	41.28	293.54	60.27

↓ indicates that lower values represent better performance.

average inference time of 60.27 ms per 512×512 patch. Although the computational cost of GLMFENet is higher than that of some lightweight models such as SegFormer and DeepLabV3+, this is mainly due to the introduction of the GFE module, which integrates standard and directional transformer encoders to capture structured global spatial dependencies.

To further assess the practical applicability of the proposed model, we estimated the inference time for a full Sentinel-2 MSI scene covering Taihu Lake, the third largest freshwater lake in China. A typical Sentinel-2 image covering Taihu Lake has a spatial size of approximately 10 196×9 897 pixels. When using a 512×512 sliding-window input, approximately 400 patches are required under a non-overlapping sliding-window setting, corresponding to a total inference time of approximately 24.1 s. In a more practical scenario with 50% overlapping sliding windows (stride=256), about 1 482 patches are required, resulting in a total inference time of approximately 89.3 s. It should be noted that the reported inference time per patch only measures the forward pass of the network and does not include image stitching or other post-processing overhead.

These results indicate that, although GLMFENet introduces additional computational overhead, its inference speed remains practical for patch-based large-scale lake monitoring using satellite imagery. Considering the significant improvement in segmentation accuracy for cyanobacterial blooms and aquatic vegetation, the proposed method achieves a favorable trade-off between computational efficiency and extraction performance, making it suitable for real-world lake ecological monitoring applications.

4.5 Thematic analysis of model performance

The superior performance of GLMFENet can be attributed to its ability to jointly model global contextual information and fine-scale spatial structures, which are both essential for distinguishing cyanobacterial blooms and aquatic vegetation in complex inland-water environments.

Traditional spectral index-based methods (e.g., VBI or $NDWI_{4,5}$ thresholds) rely mainly on single-pixel spectral characteristics. In optically complex waters, however, the spectral signatures of cyanobacterial blooms, aquatic vegetation, and turbid water often overlap. Consequently, threshold-based approaches tend to produce fragmented or erroneous extraction results, particularly in conditions with high turbidity, suspended sediments, or mixed bloom-vegetation distributions. Conventional deep learning segmentation models, which primarily rely on convolutional feature extraction, may struggle to simultaneously capture large-scale contextual information and fine-scale structural details, especially when targets exhibit fragmented spatial distributions or irregular morphology.

GLMFENet addresses these challenges by integrating two complementary modules. The GFE module captures long-range spatial dependencies and global contextual relationships, improving the recognition of large-scale bloom structures. Meanwhile, the SLMFE module enhances local structural representation through multi-scale feature extraction and spatial attention, enabling more accurate delineation of fragmented patches and complex boundaries. As a result, GLMFENet demonstrates stronger robustness under challenging environmental conditions commonly encountered in inland lakes.

4.6 On using TOA reflectance as GLMFENet input

In this study, Sentinel-2 TOA reflectance was used as the model input for semantic segmentation. This choice reflects a task-oriented engineering decision rather than a claim that TOA is physically superior to surface reflectance or bottom-of-atmosphere (SR/BOA) products. The goal of this work is to discriminate cyanobacterial blooms, aquatic vegetation, and background water by jointly exploiting multispectral responses and spatial morphological characteristics, rather than to perform quantitative water-optics inversion. In addition, some studies have conducted cyanobacterial bloom and aquatic vegetation extraction using sensor-level reflectance (e.g., TOA or near-TOA reflectance), indicating that such data can provide sufficient spectral information for reliable target discrimination without strict reliance on atmospheric correction

(Liira et al., 2010; Wang et al., 2012; Binding et al., 2021; Wynne et al., 2021; Zhang et al., 2025a). Recent advances in deep-learning-based remote sensing also demonstrate that TOA imagery can serve as an effective input representation for segmentation tasks. Several state-of-the-art frameworks, including the foundation model SatVision-TOA (Spradlin et al., 2024), near-real-time global land-use/land-cover mapping systems (Brown et al., 2022), and burned-area segmentation models (Knopp et al., 2020), have been developed directly from Sentinel-2 Level-1 TOA imagery, indicating that TOA reflectance can support robust spatial-spectral feature learning.

The use of TOA is also motivated by the inherent difficulty of atmospheric correction in aquatic environments. In water remote sensing, approximately 90% of the signal recorded over water originates from atmospheric contributions, making the retrieval of accurate water-leaving reflectance highly dependent on complex correction procedures. These uncertainties are particularly pronounced in optically complex inland waters. Comparative evaluations over Qiandao Lake show that commonly used atmospheric correction algorithms for Sentinel-2 imagery, including ACOLITE, C2RCC, iCOR, 6SV, and Sen2Cor, exhibit high levels of uncertainty (Allam et al., 2024). In addition, correction algorithms are sensitive in the near-infrared (NIR) and shortwave infrared (SWIR) bands, where aerosol contamination and high-reflectance targets may introduce additional uncertainties in the retrieval of water-leaving reflectance (Hu, 2009). Under such conditions, SR/BOA products should not necessarily be regarded as error-free inputs but rather as another processed representation that may contain processor-dependent uncertainties.

An additional advantage of TOA reflectance is its suitability for near-real-time and operational applications. TOA inputs can be obtained directly from Level-1 products after standardized preprocessing, without relying on additional atmospheric-correction workflows. This simplifies the processing pipeline and is advantageous for emergency bloom monitoring and large-scale, high-frequency lake surveillance.

Overall, the use of TOA reflectance in this study should be interpreted as a practical and task-adapted choice for segmentation under optically complex inland-water conditions. It offers a reasonable balance among input consistency, operational efficiency, and discriminative performance, although this does not imply that TOA is universally preferable for quantitative inversion tasks, where SR/BOA products remain more physically appropriate.

5 CONCLUSION

To address the challenge of high-precision automated identification of cyanobacterial blooms and aquatic vegetation under conditions of strong spectral similarity and fragmented spatial structures, this study proposed a semantic segmentation network, termed GLMFENet, which integrates global and local multi-scale features for the joint extraction of both targets. By incorporating the GFE module, GLMFENet enhances holistic scene understanding and effectively captures large-scale structural patterns. Meanwhile, the SLMFE module strengthens fine-grained representation by emphasizing multi-scale textures and boundary details.

Experiments conducted on Sentinel-2 MSI imagery from representative lakes and reservoirs across China demonstrate that GLMFENet achieves consistently superior performance in terms of mIoU, Mean F1, and Kappa coefficient, outperforming widely used models such as U-Net, DeepLabV3+, SegFormer, and Swin-UNet by 2.0%–7.8%. Ablation studies and feature activation maps visualization further confirm the complementary and synergistic contributions of the GFE and SLMFE modules to global contextual modeling and local feature representation. Compared with traditional spectral index-based approaches, GLMFENet exhibits substantially improved robustness in scenarios characterized by spectral similarity and spatial co-occurrence, resulting in markedly enhanced identification accuracy for cyanobacterial blooms and aquatic vegetation.

Overall, GLMFENet enables high-precision discrimination of cyanobacterial blooms and aquatic vegetation over extensive lake areas, providing reliable data support for eutrophication monitoring, aquatic ecosystem change assessment, and lake management applications.

6 DATA AVAILABILITY STATEMENT

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

Allam M, Meng Q Y, Elhag M et al. 2024. Atmospheric correction algorithms assessment for sentinel-2A imagery over inland waters of China: case study, Qiandao Lake. *Earth Systems and Environment*, **8**(1): 105-119, https://doi.org/10.1007/s41748-023-00366-w.

Anderson D M, Glibert P M, Burkholder J M. 2002. Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences. *Estuaries*, **25**(4): 704-726, https://doi.org/10.1007/BF02804901.

Badrinarayanan V, Kendall A, Cipolla R. 2017. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**(12): 2481-2495, https://doi.org/10.1109/TPAMI.2016.2644615.

Binding C E, Pizzolato L, Zeng C. 2021. EOLakeWatch; delivering a comprehensive suite of remote sensing algal bloom indices for enhanced monitoring of Canadian eutrophic lakes. *Ecological Indicators*, **121**: 106999, https://doi.org/10.1016/j.ecolind.2020.106999.

Brown C F, Brumby S P, Guzder-Williams B et al. 2022. Dynamic world, near real-time global 10 m land use land cover mapping. *Scientific Data*, **9**(1): 251, https://doi.org/10.1038/s41597-022-01307-4.

Cao H, Wang Y Y, Chen J et al. 2023. Swin-Unet: unet-like pure transformer for medical image segmentation. *In*: Proceedings of the European Conference on Computer Vision (ECCV) 2022. Springer, Tel Aviv, Israel. p.205-218, https://doi.org/10.1007/978-3-031-25066-8_9.

Cao M M, Qing S, Jin E et al. 2021. A spectral index for the detection of algal blooms using Sentinel-2 Multispectral Instrument (MSI) imagery: a case study of Hulun Lake, China. *International Journal of Remote Sensing*, **42**(12): 4514-4535, https://doi.org/10.1080/01431161.2021.1897186.

Chen L C, Zhu Y K, Papandreou G et al. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. *In*: Proceedings of the 15th European Conference on Computer Vision (ECCV) 2018. Springer, Munich, Germany. p.833-851, https://doi.org/10.1007/978-3-030-01234-2_49.

Cheng Y, Wang W, Ren Z P et al. 2023. Multi-scale feature fusion and transformer network for urban green space segmentation from high-resolution remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*, **124**: 103514, https://doi.org/10.1016/j.jag.2023.103514.

Du Y C, Li J S, Zhang B et al. 2024. Monitoring cyanobacterial blooms in China's large lakes based on MODIS from both Terra and Aqua satellites with a novel automatic approach. *International Journal of Applied Earth Observation and Geoinformation*, **129**: 103830, https://doi.org/10.1016/j.jag.2024.103830.

Fang C, Song K S, Shang Y X et al. 2018. Remote sensing of harmful algal blooms variability for lake hulun using adjusted FAI (AFAI) algorithm. *Journal of Environmental Informatics*, **34**(2): 108-122, https://doi.org/10.3808/JEI.201700385.

Feng Q L, Jiang Z H, Niu B W et al. 2024. Multiscale feature extraction model for remote sensing identification of erosion gullies in Northeast China's black soil region: a case study of Hailun City. *National Remote Sensing Bulletin*, **28**(12): 3147-3157, https://doi.org/10.11834/jrs.20243139. (in Chinese with English abstract)

Fu J, Liu J, Jiang J et al. 2021. Scene segmentation with dual relation-aware attention network. *IEEE Transactions on Neural Networks and Learning Systems*, **32**(6): 2547-2560, https://doi.org/10.1109/TNNLS.2020.3006524.

Gao H Y, Li R R, Shen Q et al. 2024. Deep-learning-based automatic extraction of aquatic vegetation from sentinel-2 images—a case study of lake Honghu. *Remote Sensing*, **16**(5): 867, https://doi.org/10.3390/rs16050867.

Gao L, Liu H, Yang M H et al. 2021. STransFuse: fusing swin transformer and convolutional neural network for remote sensing image semantic segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **14**: 10990-11003, https://doi.org/10.1109/JSTARS.2021.3119654.

Han M S, Li Q H, Chen H L et al. 2018. Spatial and temporal variations in cyanobacteria and microcystins in Aha Reservoir, Southwest China. *Journal of Oceanology and Limnology*, **36**(4): 1126-1131, https://doi.org/10.1007/s00343-018-7178-6.

Han X X, Chen X L, Feng L. 2015. Four decades of winter wetland changes in Poyang Lake based on Landsat observations between 1973 and 2013. *Remote Sensing of Environment*, **156**: 426-437, https://doi.org/10.1016/j.rse.2014.10.003.

Hestir E L, Schoellhamer D H, Greenberg J et al. 2016. The effect of submerged aquatic vegetation expansion on a declining turbidity trend in the Sacramento-San Joaquin River Delta. *Estuaries and Coasts*, **39**(4): 1100-1112, https://doi.org/10.1007/s12237-015-0055-z.

Ho J, Kalchbrenner N, Weissenborn D et al. 2019. Axial attention in multidimensional transformers. arXiv preprint arXiv:1912.12180. https://arxiv.org/abs/1912.12180.

Hou X J, Feng L, Dai Y H et al. 2022. Global mapping reveals increase in lacustrine algal blooms over the past decade. *Nature Geoscience*, **15**(2): 130-134, https://doi.org/10.1038/s41561-021-00887-x.

Hu C M. 2009. A novel ocean color index to detect floating algae in the global oceans. *Remote Sensing of Environment*, **113**(10): 2118-2129, https://doi.org/10.1016/j.rse.2009.05.012.

Hu C M, Lee Z, Ma R H et al. 2010. Moderate Resolution Imaging Spectroradiometer (MODIS) observations of cyanobacteria blooms in Taihu Lake, China. *Journal of Geophysical Research*: *Oceans*, **115**(C4): C04002, https://doi.org/10.1029/2009JC005511.

Knopp L, Wieland M, Rättich M et al. 2020. A deep learning approach for burned area segmentation with Sentinel-2 data. *Remote Sensing*, **12**(15): 2422, https://doi.org/10.3390/rs12152422

Li Y F, Wu G S. 2025. Multi-scale feature fusion and global context modeling for fine-grained remote sensing image segmentation. *Applied Sciences*, **15**(10): 5542, https://doi.org/10.3390/app15105542.

Liira J, Feldmann T, Mäemets H et al. 2010. Two decades of macrophyte expansion on the shores of a large shallow northern temperate lake—a retrospective series of satellite images. *Aquatic Botany*, **93**(4): 207-215, https://

doi.org/10.1016/j.aquabot.2010.08.001.

Lin T Y, Dollár P, Girshick R et al. 2017. Feature pyramid networks for object detection. *In*: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Honolulu. p.936-944.

Liu B, Li X F, Zheng G. 2019. Coastal inundation mapping from bitemporal and dual-polarization SAR imagery based on deep convolutional neural networks. *Journal of Geophysical Research*: *Oceans*, **124**(12): 9101-9113, https://doi.org/10.1029/2019JC015577.

Liu Y X, Meng Y, Deng Y P et al. 2024. Integration of CNN and transformer for high-resolution remote sensing image building extraction: a dual-stream network. *National Remote Sensing Bulletin*, **28**(11): 2943-2953, https://doi.org/10.11834/jrs.20243307. (in Chinese)

Lu L R, Luo J H, Xin Y H et al. 2024. A novel strategy for estimating biomass of submerged aquatic vegetation in lake integrating UAV and Sentinel data. *Science of the Total Environment*, **912**: 169404, https://doi.org/10.1016/j.scitotenv.2023.169404.

Luo J, Duan H T, Ma R H et al. 2017. Mapping species of submerged aquatic vegetation with multi-seasonal satellite images and considering life history information. *International Journal of Applied Earth Observation and Geoinformation*, **57**: 154-165, https://doi.org/10.1016/j.jag.2016.11.007.

Luo J H, Ni G G, Zhang Y L et al. 2023. A new technique for quantifying algal bloom, floating/emergent and submerged vegetation in eutrophic shallow lakes using Landsat imagery. *Remote Sensing of Environment*, **287**: 113480, https://doi.org/10.1016/j.rse.2023.113480.

Ma X W, Lian R R, Wu Z K et al. 2025. LOGCAN++: adaptive local-global class-aware network for semantic segmentation of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, **63**: 4404216, https://doi.org/10.1109/TGRS.2025.3541871.

Mou L C, Hua Y S, Zhu X X. 2020. Relation matters: relational context-aware fully convolutional network for semantic segmentation of high-resolution aerial images. *IEEE Transactions on Geoscience and Remote Sensing*, **58**(11): 7557-7569, https://doi.org/10.1109/TGRS.2020.2979552.

Ni Y, Liu J H, Chi W J et al. 2024. CGGLNet: semantic segmentation network for remote sensing images based on category-guided global-local feature interaction. *IEEE Transactions on Geoscience and Remote Sensing*, **62**: 5615617, https://doi.org/10.1109/TGRS.2024.3379398.

Oyama Y, Matsushita B, Fukushima T. 2015. Distinguishing surface cyanobacterial blooms and aquatic macrophytes using Landsat/TM and ETM+ shortwave infrared bands. *Remote Sensing of Environment*, **157**: 35-47, https://doi.org/10.1016/j.rse.2014.04.031.

Pan B B, Zhang J C, Feng K Y et al. 2014. Carbon storage of typical aquatic plant communities in Hungtse Lake. *Wetland Science*, **12**(4): 471-476, https://doi.org/10.13248/j.cnki.wetlandsci.2014.04.010. (in Chinese with English abstract)

Pan M, Dong J Y, Zhang Z Z et al. 2024. Changes in the ecosystem structure and function of a cyanobacteria bloom-dominated, shallow lake after ten-year eutrophication management. *Journal of Oceanology and Limnology*, **42**(6): 1726-1740, https://doi.org/10.1007/s00343-024-4062-4.

Pande-Chhetri R, Abd-Elrahman A, Jacoby C. 2014. Classification of submerged aquatic vegetation in black river using hyperspectral image analysis. *Geomatica*, **68**(3): 169-182, https://doi.org/10.5623/cig2014-302.

Phillips G, Willby N, Moss B. 2016. Submerged macrophyte decline in shallow lakes: what have we learnt in the last forty years? *Aquatic Botany*, **135**: 37-45, https://doi.org/10.1016/j.aquabot.2016.04.004.

Piaser E, Villa P. 2022. Comparing machine learning techniques for aquatic vegetation classification using Sentinel-2 data. *In*: 2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON). IEEE, Palermo, Italy. p.465-470, https://doi.org/10.1109/MELECON53508.2022.9843103.

Qin H T, Luo J H, Xu Y et al. 2025. Estimation models for pixel-scale coverage of aquatic vegetation in lakes based on landsat and sentinel data. *Journal of Remote Sensing*, **5**: 0616, https://doi.org/10.34133/remotesensing.0616.

Rodríguez-Garlito E C, Paz-Gallardo A, Plaza A. 2023. Monitoring the spatiotemporal distribution of invasive aquatic plants in the Guadiana River, Spain. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **16**: 228-241, https://doi.org/10.1109/JSTARS.2022.3225201

Ronneberger O, Fischer P, Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation. *In*: 18th International Conference on Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Springer, Munich, Germany. p.234-241, https://doi.org/10.1007/978-3-319-24574-4.

Selvaraju R R, Cogswell M, Das A et al. 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization. *In*: Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, Venice, Italy. p.618-626, https://doi.org/10.1109/ICCV.2017.74.

Song Z H, Zhu G W, Zhu M Y et al. 2025. Limiting factors on aquatic ecological health of Caizi Lake, a Changjiang River-isolated shallow lake: implications for lake restoration. *Journal of Oceanology and Limnology*, **43**(6): 1923-1937, https://doi.org/10.1007/s00343-025-4211-4.

Spradlin C S, Caraballo-Vega J A, Li J et al. 2024. SatVision-TOA: a geospatial foundation model for coarse-resolution all-sky remote sensing imagery. arXiv preprint arXiv: 2411.17000. https://arxiv.org/abs/2411.17000.

Vaswani A, Shazeer N, Parmar N et al. 2017. Attention is all you need. *In*: Proceedings of the 31st Neural Information Processing Systems. Curran Associates, Inc., Red Hook, NY. p.5998-6008.

Villa P, Bresciani M, Bolpagni R et al. 2015. A rule-based approach for mapping macrophyte communities using multi-temporal aquatic vegetation indices. *Remote Sensing of Environment*, **171**: 218-233, https://doi.org/10.1016/j.rse.2015.10.020.

Wang L, Dronova I, Gong P et al. 2012. A new time series vegetation-water index of phenological-hydrological trait across species and functional types for Poyang Lake wetland ecosystem. *Remote Sensing of Environment*, **125**: 49-63, https://doi.org/10.1016/j.rse.2012.07.003.

Wang L B, Li R, Zhang C et al. 2022. UNetFormer: a UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, **190**: 196-214, https://doi.org/10.1016/j.isprsjprs.2022.06.008

Wang S S, Gao Y N, Li Q et al. 2019. Long-term and inter-monthly dynamics of aquatic vegetation and its relation with environmental factors in Taihu Lake, China. *Science of the Total Environment*, **651**: 367-380, https://doi.org/10.1016/j.scitotenv.2018.09.216.

Wieland M, Li Y, Martinis S. 2019. Multi-sensor cloud and cloud shadow segmentation with a convolutional neural network. *Remote Sensing of Environment*, **230**: 111203, https://doi.org/10.1016/j.rse.2019.05.022.

Woo S, Park J, Lee J Y et al. 2018. CBAM: convolutional block attention module. *In*: Proceedings of the 15th European Conference on Computer Vision (ECCV) 2018. Springer, Munich, Germany. p.3-19, https://doi.org/10.1007/978-3-030-01234-2_1.

Wynne T T, Stumpf R P, Litaker R W et al. 2021. Cyanobacterial bloom phenology in Saginaw Bay from MODIS and a comparative look with western Lake Erie. *Harmful Algae*, **103**: 101999, https://doi.org/10.1016/j.hal.2021.101999.

Xie E, Wang W H, Yu Z D et al. 2021. SegFormer: simple and efficient design for semantic segmentation with transformers. *In*: Proceedings of the 35th International Conference on Neural Information Processing Systems. Curran Associates Inc., Red Hook. p.12077-12090, https://doi.org/10.5555/3540261.3541185.

Xing Q G, Hu C M. 2016. Mapping macroalgal blooms in the Yellow Sea and East China Sea using HJ-1 and Landsat data: application of a virtual baseline reflectance height technique. *Remote Sensing of Environment*, **178**: 113-126, https://doi.org/10.1016/j.rse.2016.02.065.

Xu Y, Luo J H, Duan H T et al. 2025. Satellite-based risk assessment: shifting from macrophyte- to phytoplankton-dominated states in lakes of Yangtze Plain. *Ecological Indicators*, **178**: 114096, https://doi.org/10.1016/j.ecolind.2025.114096.

Yan K, Li J S, Zhao H et al. 2022. Deep learning-based automatic extraction of cyanobacterial blooms from Sentinel-2 MSI satellite data. *Remote Sensing*, **14**(19): 4763, https://doi.org/10.3390/rs14194763.

Yuan Y H, Chen X L, Wang J D. 2020. Object-contextual representations for semantic segmentation. *In*: Proceedings of the 16th European Conference on Computer Vision (ECCV) 2020. Springer, Glasgow, UK. p.173-190, https://doi.org/10.1007/978-3-030-58539-6_11.

Zhang C, Pan X, Li H P et al. 2018. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, **140**: 133-144, https://doi.org/10.1016/j.isprsjprs.2017.07.014.

Zhang H L, Lu C, Sun D Y et al. 2025a. Automatic detection of *Cyanobacterial* blooms using multi-source optical satellite imagery: method development and application. *Ecological Indicators*, **176**: 113709, https://doi.org/10.1016/j.ecolind.2025.113709.

Zhang Y C, Ma R H, Duan H T et al. 2014. A novel algorithm to estimate algal bloom coverage to subpixel resolution in Lake Taihu. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **7**(7): 3060-3068, https://doi.org/10.1109/JSTARS.2014.2327076

Zhang Z J, Giezendanner J, Mukherjee R et al. 2025b. Assessing inundation semantic segmentation models trained on high- versus low-resolution labels using floodplanet, a manually labeled multi-sourced high-resolution flood dataset. *Journal of Remote Sensing*, **5**: 0575, https://doi.org/10.34133/remotesensing.0575

Zhou L C, Zhang C, Wu M. 2018. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. *In*: Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, Salt Lake City. p.192-1924, https://doi.org/10.1109/CVPRW.2018.00034.

Zhu S Y, Wu Y L, Ma X S. 2023. Deep learning-based algal bloom identification method from remote sensing images—take China's Chaohu Lake as an example. *Sustainability*, **15**(5): 4545, https://doi.org/10.3390/su15054545.

Electronic supplementary material

Supplementary material (Supplementary Texts S1–S5, Figs.S1–S4, and Table S1) is available in the online version of this article at https://doi.org/10.1007/s00343-026-6009-4.