remote sensing

MDPI

Article

MESA-Net: A Multi-Directional Edge-Aware Network with Scale Adaptation for Water Body Segmentation in Karst Landscapes

Bo Song [1,2], Zhiyong Zhang [1], Bo Li [1], Zhili Chen [2,3,4,5], Yun Chen [1,*], Tao Yue [1], Jianwu Jiang [1], Zhen Cao [1], Xing Zhang [1] and Qingyang Wang [1]

1 College of Geomatics and Geoinformation, Guilin University of Technology, Guilin 541004, China; bosong@glut.edu.cn (B.S.); 2120242221@glut.edu.cn (Z.Z.); 2120242157@glut.edu.cn (B.L.); yuetao@glut.edu.cn (T.Y.); fengbuxi@glut.edu.cn (J.J.); caozhen@glut.edu.cn (Z.C.); zhangxing@glut.edu.cn (X.Z.); qingyang@glut.edu.cn (Q.W.)
2 College of Environmental Science and Engineering, Guilin University of Technology, Guilin 541006, China; 2017075@glut.edu.cn
3 Guangxi Key Laboratory of Environmental Pollution Control Theory and Technology, Guilin University of Technology, Guilin 541006, China
4 University Engineering Research Center of Watershed Protection and Green Development, Guilin University of Technology, Guilin 541006, China
5 Key Laboratory of Carbon Emission and Pollutant Collaborative Control, Education Department of Guangxi Zhuang Autonomous Region, Guilin University of Technology, Guilin 541006, China
* Correspondence: chenyun@glut.edu.cn

Highlights

What are the main findings?

- A multi-directional edge-aware network with scale adaptation, termed MESA-Net, is proposed to address the challenges of fragmented small water bodies, elongated structures, and inaccurate boundary delineation in complex karst landscapes.
- Three key modules (CAFF, OGLMB, and DGHEF) significantly enhance the ability to detect small water bodies, model structural continuity, and represent boundaries, leading to superior results compared with representative CNN-based, Transformer-based, and Mamba-based methods on our self-constructed karst water body dataset, named LJ-Water. Furthermore, its effectiveness is further validated on public datasets with similar characteristics, including Water-CD and LoveDA.

What are the implications of the main findings?

- This framework demonstrates exceptional robustness when dealing with the segmentation of small-scale, elongated water bodies and variations in distribution across datasets.
- The integration of the Cross-Scale Adaptive Feature Fusion Module, Directional Gradient Histogram Edge-Guided Fusion Module, and Omni-directional Global-Local Feature Mamba Block provides a new perspective for water body segmentation in karst landscapes.

Academic Editor: Zhenwei Shi

Received: 17 April 2026
Revised: 2 June 2026
Accepted: 4 June 2026
Published: 5 June 2026

Abstract

Satellite remote sensing imagery has become an essential resource for large-scale surface water monitoring. Nevertheless, in karst regions, the elongated and fragmented morphology of water bodies, along with terrain shadows and vegetation interference, still leads to limitations in existing methods for small water body detection and accurate boundary delineation. To overcome the aforementioned issues, this paper proposes MESA-Net, a CNN–Mamba hybrid segmentation network for water body extraction in complex karst terrain. The network employs ResNet-18 as an encoder to extract shallow-level features.

 https://doi.org/10.3390/rs18111865

The decoder primarily consists of three modules: the Cross-Scale Adaptive Feature Fusion (CAFF) module, the Directional Gradient Histogram Edge-Guided Fusion (DGHEF) module, and the Omni-directional Global-Local Mamba Block (OGLMB). Among these, the CAFF module enhances the detection capability for small-scale water bodies by performing cross-scale feature fusion and dynamic weight allocation on the feature outputs from each level of the encoder. The OGLMB integrates an omnidirectional state space model with an 8-directional scanning mechanism and cross-attention guidance, effectively enhancing the ability to represent the structural continuity and global consistency of water bodies. The DGHEF utilizes directional gradient histograms to explicitly model multi-directional boundary information of water bodies, and combines this with a boundary guidance mechanism to enhance the representation of water body boundary features whilst suppressing spurious responses. In addition, the LJ-Water dataset has been constructed for the Lijiang River Basin in Guangxi, which is based on Sentinel-2 imagery. To validate the effectiveness and generalization capability of the method, comparative experiments were conducted on the self-built LJ-Water dataset as well as the publicly available Water-CD and LoveDA datasets. Experimental results demonstrate that MESA-Net consistently outperforms representative CNN-based, Transformer-based, and Mamba-based segmentation networks. On the LJ-Water dataset, it achieves 84.59% IoU and 91.65% F1, whilst on the Water-CD dataset, it attains 92.15% IoU and 95.91% F1, and 69.83% IoU and 82.24% F1 on the LoveDA dataset. Relative to the strongest baseline method, the proposed model achieved IoU gains of 1.51%, 2.34%, and 1.73% on the three datasets, respectively. In summary, MESA-Net demonstrates superior water segmentation performance under complex background conditions.

Keywords: water body segmentation; satellite imagery; deep learning; karst landscapes; Mamba

1. Introduction

As one of the key natural elements of the Earth's surface, water plays a vital role in climate regulation, ecological balance, and human production and daily life [1–3]. However, due to the complexity of the surface environment and a variety of interfering factors, the extraction of information regarding water bodies still faces numerous challenges. Consequently, achieving high-precision segmentation and dynamic monitoring of water bodies is of particular importance, as it helps to improve water resource management and mitigate associated risks.

Satellite remote sensing is capable of providing stable and comprehensive data support for water body monitoring. Currently, multi-source remote sensing imagery from Landsat, Sentinel, Gaofen, and other sources has been widely used in studies on water body segmentation [4,5]. As a critical component of remote sensing data interpretation, water body segmentation is of great importance for accurately identifying the extent of water bodies. However, water bodies in karst landscapes typically take on elongated and fragmented spatial forms, and are affected by topographical shadows and vegetation cover, which increases the difficulty of segmentation [6–8].

Traditional methods for water body segmentation rely primarily on the spectral characteristics and indices of remote sensing imagery, using thresholds to distinguish between water body and non-water body [9,10]. For instance, methods based on the Normalized Difference Water Index (NDWI) and its improved variants are widely used for water body segmentation in multi-source remote sensing imagery [11,12]. In addition, the researchers incorporated multispectral data and spatial information, employing strategies such as regional growth, edge detection, and object-oriented classification to improve segmentation performance [13,14]. However, the spectral response of water bodies varies significantly across different environments and is easily affected by factors such as shading and vegetation cover. Traditional methods are often sensitive to threshold settings, leading to misclassifications and missed classifications in complex surface conditions, such as bare rock and dark-colored features in karst regions. This limits their generalizability [15].

In recent years, the rapid development of deep learning technology has overcome the limitations of traditional methods. It has gradually become the primary approach for water-ground segmentation using multi-source data [16–20]. For example, Li et al. [21] proposed DeepUNet, a deep U-Net architecture based on fully convolutional networks, for achieving pixel-level land-sea segmentation [22]. Wang et al. [23] proposed a U-Net-based convolutional neural network, TSAE-UNet, which achieves accurate detection of water bodies across multiple scenes and time phases by incorporating space–time feature extraction and attention mechanisms. However, the aforementioned methods rely primarily on local convolutional operations, making it difficult to fully extract deep global features of water bodies in complex scenes, which consequently limits segmentation performance to some extent. To address this issue, Wang et al. [24] proposed MSLWENet, a deep learning network for lake water body extraction. By introducing multi-scale feature fusion and feature enhancement mechanisms, it effectively enhances the model's ability to characterize water bodies and improve segmentation accuracy in complex backgrounds. Furthermore, Zhang et al. [25] utilized rich convolutional neural network features to achieve fine-grained segmentation of water bodies in high-resolution aerial and satellite imagery, which further validates the effectiveness of deep feature representation in water body extraction tasks [26]. Although methods based on CNN have made significant progress in the task of water segmentation, their inherent local receptive fields limit their ability to capture long-range dependencies and global contextual information [27,28].

Recently, Transformer models have gradually been adopted in the field of remote sensing image segmentation due to their superior global feature modeling capabilities [29–32]. For instance, Kang et al. [33] proposed WaterFormer, a coupled model that integrates Transformers with convolutional neural networks, achieving effective segmentation of water bodies in optical remote sensing imagery by combining global and local features. Subsequently, Wang et al. [34] designed MHNet, a mask-guided hybrid network that enhances feature representation in key regions by introducing a masking mechanism, thereby improving the robustness of water body segmentation. Furthermore, Tian et al. [35] developed WB-Former, which deeply integrates CNNs and Transformers for water body segmentation in complex scenes, significantly enhancing the model's adaptability to complex backgrounds. Transformers still exhibit high computational complexity when processing high-resolution remote sensing imagery.

The Mamba architecture, which is based on state-space models, has been increasingly applied to semantic segmentation tasks due to its efficient modeling of long sequences [5,36–38]. The original Mamba model was used for processing one-dimensional data, such as in natural language processing (NLP) [39]. Liu et al. [40] designed a two-dimensional state-space model (2D-SSM) to enable Mamba to process two-dimensional image data. Based on this, Cai et al. [41] proposed CM-UNet++, a multi-layer information optimization network which, through the introduction of multi-scale feature fusion and structural improvements, effectively enhanced the accuracy of urban water body extraction from high-resolution remote sensing imagery. Li et al. [42] improved the accuracy of small water body detection through transfer learning between Sentinel-2 and PlanetScope, whilst the introduction of V-Mamba further enhanced the model's ability to capture long-range dependencies and multi-scale spatial features. Liu et al. [43] proposed the SWD-Net dataset to provide a unified benchmark for small water body detection, and Mamba U-Net further improved the performance of high-resolution remote sensing water body segmentation by enhancing the modelling of long-range dependencies. Hong et al. [44] combined the red-blue normalized difference index with a state-space model for sea ice segmentation, effectively improving the accuracy of segmentation at the ice–water boundary. Furthermore, Liu et al. [45] proposed HyMambaNet, which combines state-space models with multi-scale features, thereby enhancing the model's ability to characterize water bodies in complex scenes whilst maintaining computational efficiency.

However, existing methods still suffer from shortcomings such as the failure to detect small-scale water bodies and inaccurate boundary identification in karst landscapes. To address this, this paper proposes MESA-Net, a hybrid CNN–Mamba segmentation network designed for the segmentation of water bodies in complex karst landscapes. Within the MESA-Net framework, we introduce a Cross-Scale Adaptive Feature Fusion (CAFF) module to enhance the model's ability to extract discriminative features for small, fragmented water bodies. An Omnidirectional Global–Local Mamba Block (OGLMB) is embedded in the decoder to improve the model's capacity to capture the multi-directional structure and global consistency of water bodies. Furthermore, considering that water bodies in karst landscapes are influenced by topography and often exhibit a certain degree of directionality, we have additionally designed a Directional Gradient Histogram Edge-Guided Fusion Module (DGHEF) to capture multi-directional boundary information of water bodies, thus enhancing the model's ability to delineate water body boundaries in complex backgrounds.

In summary, the main contributions of this paper are as follows:

(1) A CNN–Mamba hybrid segmentation network, named MESA-Net, has been proposed for the segmentation of water bodies in complex karst landscapes. Whilst ensuring low computational complexity, the network achieves a synergistic representation of both the local details and global structure of water bodies, significantly improving segmentation performance in complex scenarios.

(2) To address the challenge of identifying small water bodies in karst regions, where water bodies vary greatly in size and are distributed in a fragmented manner, a Cross-scale Adaptive Feature Fusion (CAFF) module has been designed. By employing cross-scale feature fusion and a dynamic weight allocation mechanism, this module enhances the network's ability to detect small-scale water body targets.

(3) To effectively capture the long-range spatial dependencies arising from the elongated structures and multi-directional extent of karst water bodies, an Omni-directional Global-Local Mamba Block (OGLMB) with an 8-directional scanning mechanism has been designed, significantly enhancing the network's ability to capture the structural continuity and global consistency of water bodies extending in multiple directions.

(4) To address the issue of blurred and fragmented water boundaries caused by mountain shadows and vegetation occlusion, a Directional Gradient Histogram Edge-Guided Fusion (DGHEF) module has been designed. This module utilises a gradient direction histogram to explicitly model boundary information, whilst incorporating an edge-guided mechanism to enhance the representation of multi-directional boundary features. It effectively suppresses false responses caused by shadows and occlusion, thereby further improving the accuracy of water segmentation in complex scenes.

2. Materials and Methods

2.1. Overall Architecture

The MESA-Net proposed in this paper adopts an encoder–decoder architecture, as shown in Figure 1. It is composed of three main parts: (1) a lightweight CNN-based encoder network—ResNet18; (2) a decoder network comprising an adaptive cross-scale feature aggregation module, an omnidirectional Mamba module for global and local features, and a directional gradient histogram edge-guided fusion module; (3) a residual stem (Residual Stem) based on residual blocks.

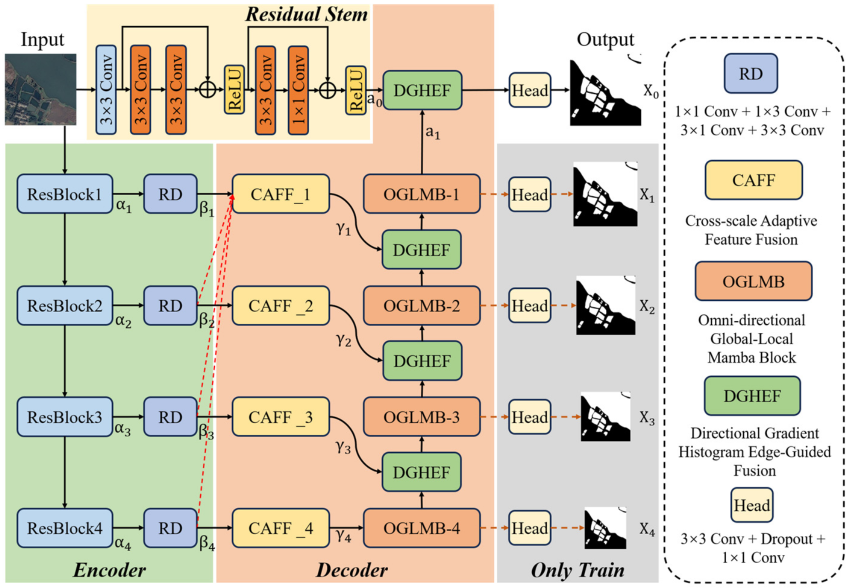

Figure 1. Overall structure of the network, where only the input stream for CAFF_1 is shown here; the input streams for CAFF_2–4 are the same as that for CAFF_1.

In particular, the encoder comprises four lightweight residual convolutional layers, ResBlock1–4, which progressively extract local features of the water body through downsampling. Furthermore, to unify the number of channels across different scales and reduce the computational cost of the model, we have designed an RD (Reduction) convolutional module. Through a series of convolutional operations within this module, the outputs α_1–α_4 from ResBlock1–4 are unified into 64 channels.

The decoder is organized into four stages corresponding to the encoder layers in a vertical manner. Specifically, $CAFF_1$–$CAFF_4$ perform spatial alignment on the multi-scale encoder features β_1–β_4 and adaptively learn fusion weights, thereby effectively integrating contextual information of karst water bodies across different scales. This process enhances the network's perception of large-area water regions and improves robustness against interference from mountain shadows. The output γ_4 of $CAFF_4$ is further fed into $OGLMB_4$ for feature refinement, strengthening the interaction between global and local context information. Subsequently, the outputs γ_1–γ_3 from $CAFF_1$–$CAFF_3$, together with the outputs of $OGLMB_2$–$OGLMB_4$, are jointly fed into the DGHEF module. By incorporating differentiable HOG-based gradient priors and a gating mechanism, the model effectively enhances boundary representation capability.

In addition, water bodies in karst terrains typically exhibit elongated and fragmented spatial structures. During repeated downsampling and upsampling operations, a large amount of fine-grained information, such as boundaries and textures, may be lost. To address this issue, an additional local feature extraction branch, called the Residual Stem, is introduced. Specifically, the original image is simultaneously fed into the Residual Stem, where fine-grained water-body features are extracted via a 3 × 3 convolution followed by two residual blocks.

Finally, the output a_0 from the Residual Stem and the output a_1 from the OGLMB module are further fused through the DGHEF module, which effectively restores the model's ability to preserve fine-grained water body features. The fused features and the outputs from the four decoding stages are fed into the segmentation head to generate five prediction maps, denoted as X_0–X_4. Among them, X_1–X_4 are used during the training stage for auxiliary supervision, while only the final prediction map X_0 is produced during the testing stage.

2.2. Cross-Scale Adaptive Feature Fusion (CAFF)

In karst terrains, water bodies typically exhibit significant scale variations and spatial fragmentation. Small-scale water bodies are often distributed as discrete patches and tend to present weak response signals in complex backgrounds, making them highly susceptible to being overlooked. To overcome this challenge, a Cross-scale Adaptive Feature Fusion module (CAFF) is proposed. By performing feature fusion across different encoding levels and incorporating a dynamic weight allocation mechanism, the proposed module effectively strengthens the network's capability to perceive small-scale water-body targets. The overall structure of the CAFF module is illustrated in Figure 2.

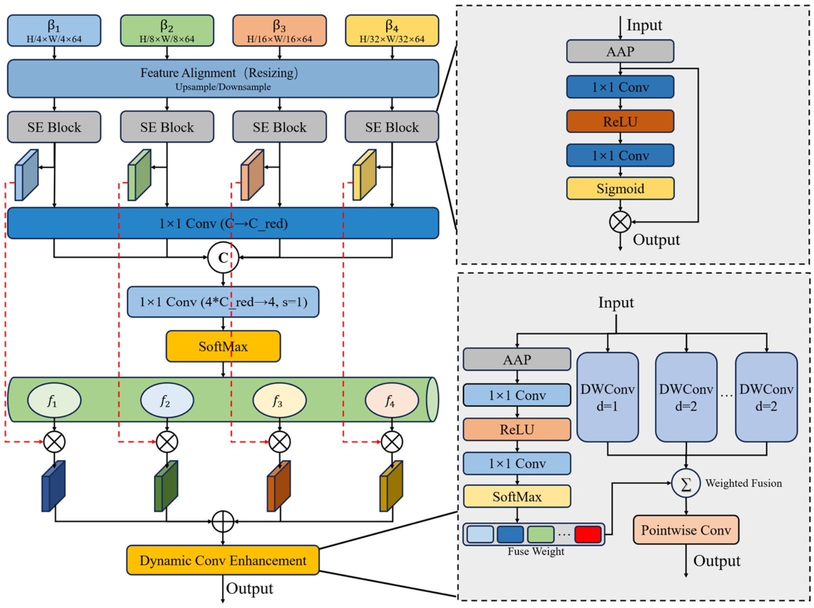

Figure 2. Structure of the CAFF.

Concretely, the proposed module takes the multi-scale feature maps from the four encoder stages, denoted as β_1–β_4, as input. Spatial alignment across different scales is achieved through upsampling and downsampling operations, thereby establishing a unified multi-scale feature representation. Bilinear interpolation was used for upsampling, whilst max-pooling was employed for downsampling during cross-scale feature alignment. Building upon this, to enhance feature discriminative power, a channel attention mechanism is introduced to adaptively recalibrate features at different scales, which strengthens responses related to water bodies while suppressing redundant background information. This can be expressed as follows:

$$\tilde{X}_i^l = v_c\left(\mathcal{R}_i^l(\beta_i)\right) \otimes \mathcal{R}_i^l(\beta_i),\ i = 1,2,3,4/l = 1,2,3,4 \tag{1}$$

where $\mathcal{R}_i^l(\cdot)$ denotes a scale alignment operation, $v_c(\cdot)$ denotes the SE attention mapping function, l is the target level, $\otimes$ represents element-wise multiplication, β_i denotes the output of the encoder for the corresponding layer.

Subsequently, considering the varying importance of multi-scale features at each spatial location, a Softmax-normalized dynamic weighting mechanism is designed to perform pixel-wise weighted fusion across different scales, thereby enabling adaptive aggregation of cross-scale information. The process can be expressed as follows:

$$F^l = \sum_{i=1}^{4} \frac{\exp\left(\psi_i(\tilde{X}_i^l)\right)}{\sum_{j=1}^{4} \exp\left(\psi_j(\tilde{X}_j^l)\right)} \odot \tilde{X}_i^l \tag{2}$$

where F^l denotes the fused feature at level l, i denotes the index of the input feature scale, j is the summation index used for normalization, $\exp(\cdot)$ denotes the exponential function, $\psi_i(\cdot)$ and $\psi_{\mathrm{j}}(\cdot)$ denotes learnable weight generation functions with different indices.

Furthermore, a multi-branch dynamic convolution structure is introduced to capture contextual information of multi-scale features by enlarging the receptive field. An input-driven weight generation mechanism is employed to adaptively weight each branch, thereby further enhancing the representation of local details in the fused features. Finally, channel fusion is performed using pointwise convolution, producing enhanced feature representations that incorporate both multi-scale semantic information and fine-grained local details, which improves the network's capability to perceive small-scale water bodies. This process can be defined as:

$$Y^l = \sum_{m=1}^{M} \alpha_m\left(F^l\right) \otimes f_m\left(F^l\right), \sum_{m=1}^{M} \alpha_m = 1 \tag{3}$$

where $f_m(\cdot)$ denotes the depthwise separable convolution branch, $\alpha_m(\cdot)$ denotes the branch weights generated adaptively.

2.3. Omni-Directional Global-Local Mamba Block (OGLMB)

In karst regions, water body structures often exhibit elongated and anisotropic distribution characteristics. Existing models struggle to capture long-range spatial dependencies without sacrificing fine-grained features. To address this issue, this paper proposes an omnidirectional global–local feature Mamba module with an eight-directional scanning mechanism (OGLMB), which enables efficient modeling of complex water body structures and significantly enhances the representation of structural continuity and global consistency of water bodies. As illustrated in Figure 3, the OGLMB mainly consists of two branches: a global branch and a local branch.

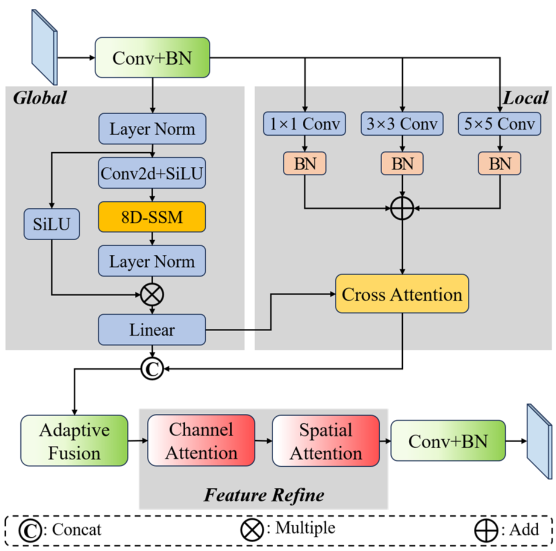

Figure 3. Structure of the OGLMB.

The global branch employs a Mamba module with an eight-directional scanning mechanism to extract global water body features from eight directions. Specifically, the input features are first transformed into a sequential representation. Subsequently, the 8D-SSM performs bidirectional traversal of the feature maps along horizontal, vertical, and diagonal directions, thereby significantly enhancing the network's ability to capture the anisotropic spatial structures of water bodies. Finally, the features are reshaped back to their original form for output. This process can be formulated as:

$$X_G = \lambda_1 X + \mu(\{SSM_k(S_k(LN(X)))\}_{k=1}^{8}) \tag{4}$$

where $LN(\cdot)$ denotes layer normalization, $S_k(\cdot)$ denotes the scanning operation along the k-th direction, $SSM_k(\cdot)$ denotes the state-space modeling along the corresponding direction, $\mu(\cdot)$ indicates the rearrangement and fusion of multi-directional features, X denotes the input feature, λ_1 is a learnable residual scaling parameter. X_G denotes the output of the global branch.

On this basis, to compensate for the insufficient representation of local details in purely global modeling, the module introduces a local branch constructed using multi-scale convolutions with kernel sizes $k = (1,3,5)$. This design enables the extraction of fine-grained spatial information under different receptive fields. The process can be formulated as:

$$X_L = \frac{1}{N}\sum_{i=1}^{N}\left(f_{k_i}(\varphi(X))\right),\ k_i \in \{1,3,5\} \tag{5}$$

where $\varphi(\cdot)$ represents the input projection operation, $f_{k_i}(\cdot)$ represents the convolution with the kernel k_i, N represents the number of branches, X_L represents the output of the local branch.

Furthermore, a cross-branch interaction mechanism is introduced, in which the local features are treated as queries, while the global features serve as keys and values for cross-attention modeling. This design enables deep fusion of global and local information, allowing local details to adaptively acquire complementary information from the global context. The process can be expressed as:

$$X_C = \sigma(\frac{(X_L W_Q)(X_G W_K)}{\sqrt{d}})(X_G W_V) \tag{6}$$

where $\sigma(\cdot)$ denotes the Softmax, W_Q, W_K and W_V denote the trainable projection matrix, d denotes the dimension of the query and the key.

In addition, to further enhance the network's representation capability for water body features, an adaptive gating strategy is adopted to dynamically adjust the contribution weights of the two types of features. This process can be formulated as:

$$\begin{cases} (W_G, W_L) = \sigma(\psi(X_G \odot X_C)) \\ X_f = W_G \otimes X_G + W_L \otimes X_C \end{cases},\ W_G = 1 - W_L \tag{7}$$

where $\psi(\cdot)$ represents the gated weight generation function composed of convolutional layers, $\odot$ represents feature merging, W_G and W_L represent the adaptive weights for the global branch and the local branch.

Finally, to further refine the multi-scale features of the water body and highlight the key characteristics of the water regions, the fused features are further enhanced through channel-space attention. This process can be expressed as:

$$Y = \lambda_2 X + \rho(CAB(X_f)) \tag{8}$$

https://doi.org/10.3390/rs18111865

where $CAB(\cdot)$ denotes channel-spatial attention, $\rho(\cdot)$ denotes the output mapping, λ_2 is the learnable residual scaling coefficient.

2.4. Directional Gradient Histogram Edge-Guided Fusion (DGHEF)

To address the issue that water body boundaries in karst terrains are prone to blurring and fragmentation due to mountain shadows and vegetation occlusion, this paper proposes a Directional Gradient Histogram Edge-Guided Fusion module (DGHEF). The module explicitly models multi-directional boundary information of water bodies using directional gradient histograms and incorporates an edge-guided mechanism to enhance the network's representation of water boundary features. The overall architecture of the module is illustrated in Figure 4.

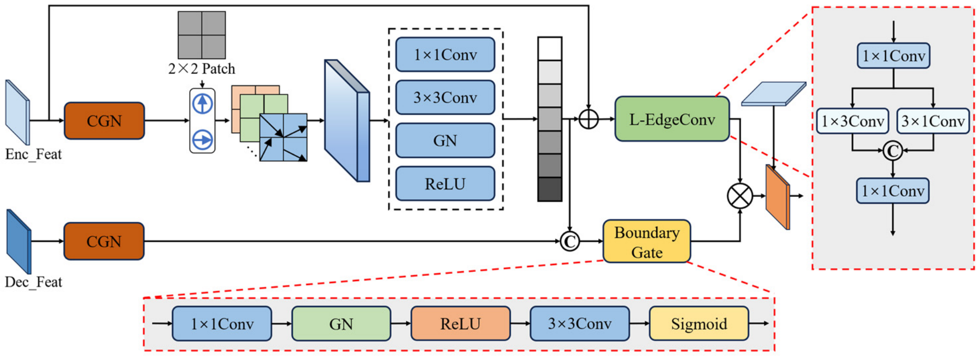

Figure 4. Structure of the DGHEF.

First, the shallow features from the encoder E and the decoder features D are linearly projected to obtain a unified representation:

$$\begin{cases} E' = \varphi_{\mathrm{e}}(E) \\ D' = \varphi_{\mathrm{d}}(D) \end{cases} \tag{9}$$

where $\varphi_{\mathrm{e}}(\cdot)$ and $\varphi_{\mathrm{d}}(\cdot)$ denote the feature projection function consisting of a 1×1 convolution and normalization.

Meanwhile, considering that shallow features contain rich boundary information, the Sobel operator is employed to compute the gradient magnitude and orientation of the feature E'. Based on this, multi-directional gradient histograms are constructed to achieve explicit modeling of boundary orientation information. This process can be formulated as:

$$H = \varphi\left(\mathrm{Norm}\left(\sum_{p\in\Omega} M(E')\cdot\delta(b(E') - k)\right)_{k=1}^{K}\right) \tag{10}$$

where $M(\cdot)$ represents the gradient magnitude computed based on the Sobel operator, $b(\cdot)$ denotes the bin corresponding to the pixel orientation, $\delta(\cdot)$ represents the indicator function used for assignment, Ω represents the cell region, K denotes the number of bins.

Subsequently, the gradient orientation features are projected into a high-dimensional feature space and fused with the original features. A lightweight edge convolution structure is then employed to enhance the feature responses along the horizontal and vertical directions, respectively. This operation further strengthens the model's representation of water body directional information. This process can be expressed as:

$$E_{\mathrm{edge}} = \theta(f_h(\theta(E' + \partial H)) \odot f_v(\theta(E' + \partial H))) \tag{11}$$

where $\theta(\cdot)$ denotes 1×1 Conv, $f_h(\cdot)$ denotes horizontal depthwise convolution, $f_v(\cdot)$ denotes vertical depth convolution, ∂ denotes the learnable factor.

In addition, to suppress pseudo-edge responses caused by shadows or noise, a boundary-guided gating mechanism based on decoder semantic information is designed to adaptively filter boundary features. This mechanism highlights real water body boundaries while suppressing interference information. Finally, residual connections are introduced to preserve shallow, detailed information, enabling fine-grained characterization of complex boundaries. This process can be formulated as:

$$X = E_{edge} \odot \sigma\left(\psi\left(\left[D', H\right]\right)\right) + E' \tag{12}$$

3. Results

3.1. Dataset

To validate the effectiveness and robustness of MESA-Net, a dedicated dataset for water body segmentation in karst terrains, named LJ-Water, is constructed in this study. Furthermore, to comprehensively evaluate the performance of the proposed network, two publicly available datasets with similar water body characteristics (Water-CD and LoveDA) are additionally selected for comparative experiments. The three datasets are compared in terms of data source, spatial resolution, categories, image size, and the number of images, as summarized in Table 1. The following subsections provide a detailed description of each dataset.

Table 1. Parameters of the experimental dataset.

Dataset	LJ-Water	Water-CD	LoveDA
Data sources	Sentinel-2	Sentinel-2	Google Earth
Spatial resolution (m)	10	10	0.3
Annotation category	2	2	7 (Original) 2 (Used)
Image size	256 × 256	512 × 512	1024 × 1024
Total image number	1930	1149	4191
Tasks	Water body Segmentation	Water body Segmentation	Object classification

3.1.1. LJ-Water Dataset

Currently, most datasets used for semantic segmentation research are designed for general purposes, while dedicated datasets for water body segmentation remain relatively scarce. Moreover, datasets specifically targeting water bodies in karst terrains are almost nonexistent. Therefore, this study constructs a novel dataset based on Sentinel-2 imagery for karst water bodies in the Lijiang River Basin, Guilin, Guangxi, China, as illustrated in Figure 5. Specifically, multi-temporal Sentinel-2 images within the study area are first collected and preprocessed, including cloud removal, band registration, and cropping to the region of interest. Subsequently, water body regions are annotated at the pixel level through visual interpretation of remote sensing images, followed by manual cross-validation to ensure label quality. After data cleaning and patch extraction, a total of 1930 image samples with a size of 256 × 256 are obtained. The LJ-Water dataset utilizes the R (red), G (green), and NIR (near-infrared) bands from Sentinel-2 imagery as input to effectively characterize the spectral differences between water bodies and background land cover.

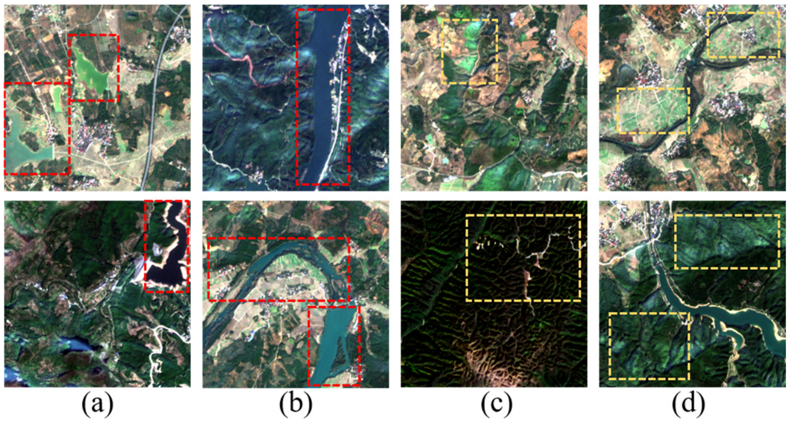

Figure 5. Examples of the LJ-Water dataset. (**a**,**b**) represent scenarios with varying water body shapes, sizes, and albedos; (**c**,**d**) represent scenarios where spectral confusion is caused by factors such as mountain shadows and vegetation.

In addition, compared with the Water-CD and LoveDA datasets, the LJ-Water dataset is more challenging in two aspects: (1) water bodies in different regions exhibit significant variations in scale, reflectance, and shape, as shown in Figure 5a,b; and (2) under the influence of karst terrain, the image background becomes more complex, including factors such as mountain shadows and vegetation, as illustrated in Figure 5c,d.

3.1.2. Water-CD Dataset

Water-CD is a multi-purpose dataset designed for water body segmentation and change detection [46]. As illustrated in Figure 6, the dataset consists of multiple Sentinel-2 satellite images with a spatial resolution of 10 m. The study areas cover seasonal water bodies in the Yangtze River Basin in China and multiple regions along the Jamuna River in South Asia. These images include three bands: R (red), G (green), and NIR (near-infrared). In this study, images with annotation errors and those without water bodies are removed. The filtered dataset contains a total of 1149 images with a size of 512 $\times$ 512. In addition, data augmentation is performed through rotation as well as horizontal and vertical flipping.

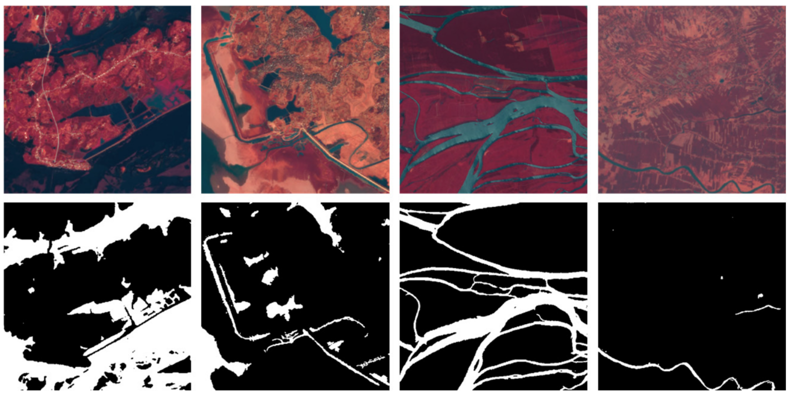

Figure 6. Examples of the Water-CD dataset.

3.1.3. LoveDA Dataset

The LoveDA dataset consists of 5987 images collected from three cities: Nanjing, Changzhou, and Wuhan [47]. As shown in Figure 7, each image is sourced from Google Earth with a spatial resolution of 0.3 m and includes three bands: R (red), G (green), and B (blue). The annotation labels cover seven land cover categories, including background, buildings, roads, water bodies, bare soil, forest, and farmland. According to the requirements of our task, these seven categories are reclassified into two classes, namely water and non-water. Similarly, after data filtering and augmentation, the resulting dataset contains a total of 4191 images with a size of 1024 × 1024.

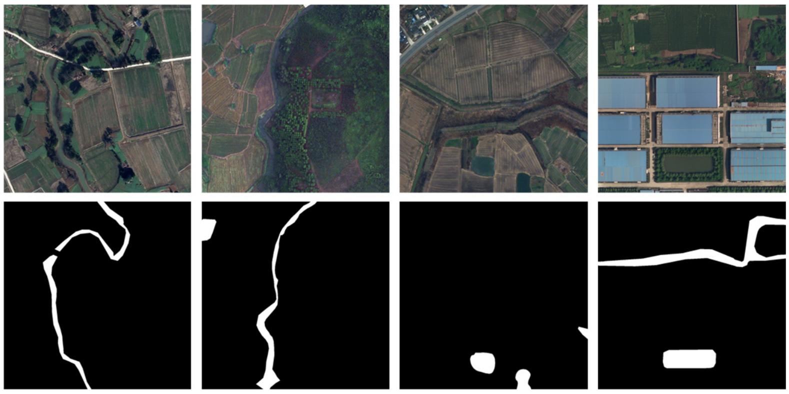

Figure 7. Examples of the LoveDA dataset.

3.2. Experimental Environment

3.2.1. Implementation Details

Experiments are conducted on the Water-CD, LoveDA, and LJ-Water datasets. Each dataset is randomly split into training, validation, and test sets with a ratio of 7:1.5:1.5 using a fixed random seed. All experiments are implemented based on PyTorch 2.1.1 and Python 3.10 on a single NVIDIA GeForce RTX 4060 GPU with 8 GB memory (NVIDIA Corporation, Santa Clara, CA, USA). Regarding training strategies, the Adam optimizer is adopted. The ResNet-18 encoder was initialized using ImageNet pre-trained weights. The initial learning rate is set to 0.0005, and a polynomial decay strategy is employed to dynamically adjust the learning rate for improved training performance. The batch size is set to 4, and the model is trained for 100 epochs.

3.2.2. Evaluation Metrics

To quantitatively compare the results, four commonly used metrics are adopted in this study: Intersection over Union (IoU), F1 score, Precision (P), and Recall (R). IoU is used to assess the spatial overlap consistency between the predicted water body regions and the ground truth masks. The F1 score provides a comprehensive measure of the model's accuracy and completeness in water body segmentation, representing the harmonic mean of Precision and Recall. The specific formulas are as follows:

$$\mathrm{IoU} = \frac{T_P}{T_P + F_P + F_N} \tag{13}$$

$$\mathrm{Precision} = \frac{T_P}{T_P + F_P} \tag{14}$$

$$\mathrm{Recall} = \frac{T_P}{T_P + F_N} \tag{15}$$

$$\mathrm{F1} = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{16}$$

where T_P denotes the number of pixels correctly predicted as water, F_P denotes the number of non-water pixels incorrectly predicted as water, F_N denotes the number of actual water pixels incorrectly predicted as background.

3.3. Comparison with Representative Segmentation Methods

To validate the effectiveness of the proposed method, seven advanced segmentation approaches and three water-specific segmentation approaches are selected for comparative experiments, including (1) CNN-based methods, such as U-Net [22], DeepLabV3+ [48], SFFNet [49], Res-UNet++ [50], MECNet [25], and BiSeNet [51]; (2) Transformer-based methods, such as MF_Seg [52] and CTCFNet [53]; (3) Mamba-based methods, such as AfaMamba [54] and MF-Mamba [55].

3.3.1. Experimental Results on the LJ-Water Dataset

To validate the effectiveness of the proposed method, experiments are first conducted on the constructed LJ-Water dataset. As shown in Table 2, MESA-Net achieves significant performance improvements over all comparative methods, reaching 84.59% IoU and 91.65% F1, respectively. Compared with the second-best segmentation network, DeepLabV3+, the proposed method improves IoU by 1.51% and F1 by 0.89%. Furthermore, compared with advanced Transformer- or Mamba-based networks such as CTCFNet, AfaMamba, and MF-Mamba, MESA-Net demonstrates stronger robustness under complex background conditions. This indicates that relying solely on global modeling or long-range dependency modeling remains insufficient to fully address the challenges of fragmented water boundaries, elongated targets, and severe background interference in karst regions.

Table 2. Quantitative comparison results for the LJ-Water dataset.

Purpose	Method	Type	Backbone	IoU	P	R	F1
DL	Unet	C	–	83.00	93.01	88.52	90.71
	DeepLabV3+	C	VGG16	83.08	88.66	92.95	90.76
	SFFNet	C	ConvNext	81.13	88.47	90.72	89.58
	Res-Unet++	C	–	82.96	91.51	89.88	90.68
	CTCFNet	C + T	–	81.77	93.12	87.03	89.97
	AfaMamba	C + M	ResNet18	76.81	82.97	91.19	86.89
	MF-Mamba	C + M	HRNetV2	81.68	89.83	90.00	89.92
WS	BiSeNet	C	ResNet18	77.72	89.14	85.85	87.46
	MECNet	C	MFConvNet	78.55	86.41	89.61	87.99
	MF_Seg	C + T	MiT-B5	68.60	88.11	75.59	81.37
	MESA-Net (Ours)	C + M	ResNet18	84.59	90.93	92.39	91.65

Notes: DL and WS denote advanced general-purpose segmentation networks and water-specific segmentation networks, respectively; C, C + T, and C + M represent convolution-based architectures, hybrid convolution–Transformer architectures, and hybrid convolution–Mamba architectures, respectively; VGG16 denotes the 16-layer VGG network; and MiT-b5 denotes Mix Transformer (b5).

In addition, compared with dedicated water body segmentation networks, although BiSeNet, MECNet, and MF_Seg are specifically designed for water segmentation tasks, they achieve only 77.72%, 78.55%, and 68.60% IoU on the LJ-Water dataset, respectively, which are substantially inferior to the proposed method. This indicates that these water segmentation approaches suffer notable performance degradation when confronted with complex backgrounds in karst regions, such as mountain shadows and vegetation occlusion, leading to evident false positives and missed detections.

To better evaluate the segmentation performance of the model under complex karst terrain scenarios, five representative examples from the LJ-Water dataset are selected for visualization, as shown in Figure 8. As illustrated in Figure 8a, water bodies in karst terrains often exhibit irregular boundary patterns extending in multiple directions, which impose higher demands on the model's boundary representation capability. The proposed method demonstrates superior performance in extracting such irregularly extended water bodies compared with other approaches. As shown in Figure 8b, the red box contains two small and scattered water bodies. Most comparative methods suffer from missed detections, particularly SFFNet, MF_Seg, and MF-Mamba, whereas the proposed method successfully identifies these water bodies and accurately delineates their boundary structures. As illustrated in Figure 8c–e, in scenarios involving narrow and meandering rivers, all competing models exhibit varying degrees of discontinuity or over-segmentation, while MESA-Net maintains good continuity of water bodies and remains robust against complex background interference, such as mountains and shadows. In summary, MESA-Net demonstrates strong robustness to background interference, superior capability in preserving the continuity of elongated water bodies, and enhanced performance in detecting small-scale water bodies in complex karst terrains.

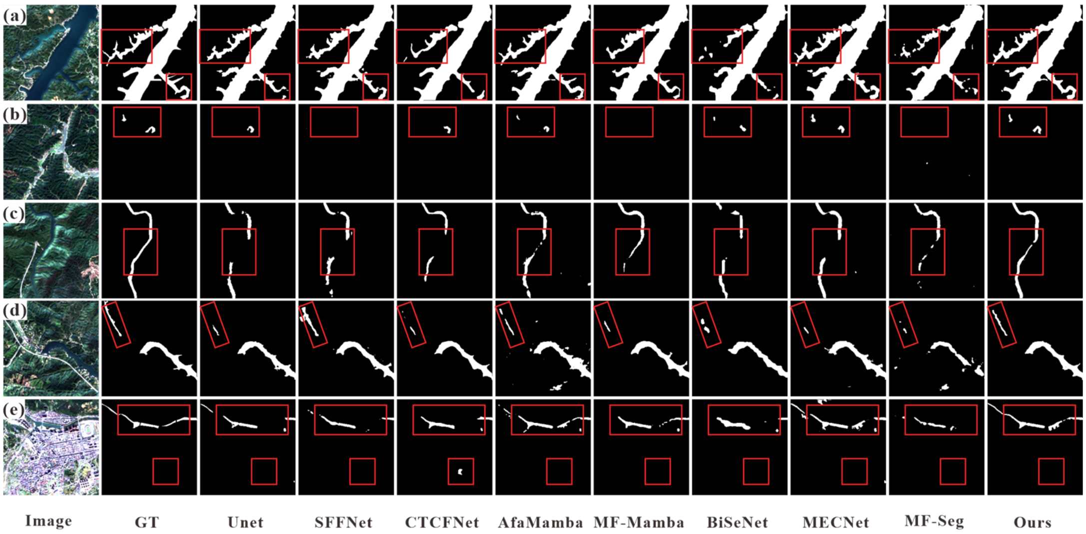

Figure 8. Visualization of segmentation results on the LJ-Water dataset. (**a**–**e**) represent five different scenarios.

3.3.2. Experimental Results on the Water-CD Dataset

To further validate the effectiveness and robustness of the proposed method, comparative experiments are conducted on the Water-CD dataset, and the results are presented in Table 3. Compared with the second-best compared segmentation network, DeepLabV3+, the proposed method improves IoU by 2.34% and F1 by 1.30%, demonstrating the superior performance of MESA-Net for water body segmentation on this dataset. Compared with the Transformer-based network CTCFNet, MESA-Net also shows significant advantages in both IoU and F1. When dealing with the Water-CD dataset, which contains a large number of small water bodies and fine-grained boundaries, Transformer-based models tend to overlook these fine-scale features, resulting in reduced segmentation accuracy. This is mainly because such models place greater emphasis on global feature modeling.

In addition, compared with dedicated water body segmentation networks, MESA-Net achieves superior performance over BiSeNet, MECNet, and MF_Seg. Among them, MECNet is the best-performing baseline, achieving 89.61% IoU and 94.52% F1, whereas the proposed method reaches 92.15% and 95.91%, yielding improvements of 2.54% in IoU and 1.39% in F1. Meanwhile, BiSeNet and MF_Seg obtain IoU scores of 87.58% and 88.52%, respectively, which are still lower than those of the proposed method. This indicates that there remains a noticeable performance gap among different dedicated water body segmentation models.

Table 3. Quantitative comparison results for the Water-CD dataset.

Purpose	Method	Type	Backbone	IoU	P	R	F1
DL	Unet	C	–	88.31	91.58	95.61	93.55
	DeepLabV3+	C	VGG16	89.81	92.95	96.33	94.76
	SFFNet	C	ConvNext	89.57	91.80	97.30	94.47
	Res-Unet++	C	–	87.88	90.12	96.95	93.41
	CTCFNet	C + T	–	88.42	92.80	94.80	93.79
	AfaMamba	C + M	ResNet18	89.74	90.73	98.79	94.59
	MF-Mamba	C + M	HRNetV2	86.74	88.18	98.15	92.90
WS	BiSeNet	C	ResNet18	87.58	92.39	94.39	93.38
	MECNet	C	MFConvNet	89.61	93.04	96.05	94.52
	MF_Seg	C + T	MiT-B5	88.52	91.87	95.83	93.81
	MESA-Net (Ours)	C + M	ResNet18	92.15	93.76	98.17	95.91

To better illustrate the segmentation performance of different networks on water bodies of varying scales, five representative examples from the Water-CD dataset are selected for visualization, as shown in Figure 9. Specifically, as shown in Figure 9a,b, most CNN-based methods are able to extract the main regions of water bodies; however, they exhibit limitations in maintaining the continuity of slender water bodies and in distinguishing spectrally similar land cover, resulting in fragmented predictions and misclassification. In contrast, Transformer-based and Mamba-based methods such as CTCFNet, AfaMamba, and MF-Mamba, due to their strong global feature modeling capability, tend to produce more false predictions. As highlighted in the red boxes in Figure 9c,d, these misclassifications become more pronounced. Furthermore, as shown in Figure 9e, most methods fail to detect the small water bodies within the red box and produce incorrect predictions, whereas the proposed method is still able to stably segment these small targets. In summary, compared with the competing models, the proposed method is capable of maintaining better water body continuity while effectively suppressing misclassification caused by spectrally similar land cover. The experimental results demonstrate that MESA-Net achieves superior performance in water body segmentation under complex backgrounds, particularly in scenarios involving both intricate environments and small-scale water bodies in the Water-CD dataset.

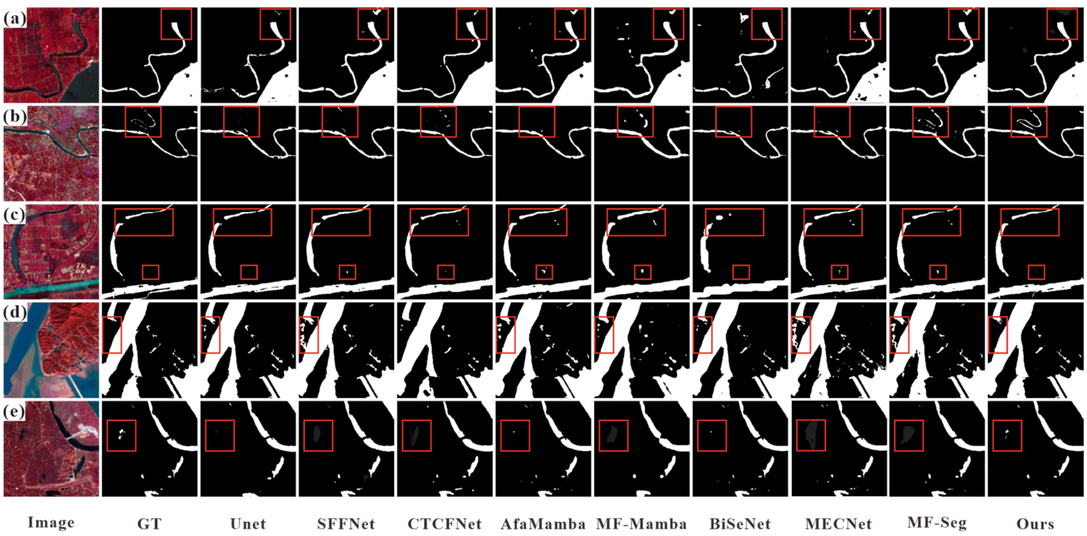

Figure 9. Visualization of segmentation results on the Water-CD dataset. (**a**–**e**) represent five different scenarios.

3.3.3. Experimental Results on the LoveDA Dataset

To further validate the effectiveness and robustness of the proposed method, comparative experiments are also conducted on the LoveDA dataset, and the results are presented in Table 4. The proposed method still achieves the best performance on this dataset, with an IoU of 69.83% and an F1 score of 82.24%. Compared with the second-best compared segmentation network, the proposed method improves IoU by 1.73% and F1 by 1.22%, further demonstrating its effectiveness and robustness across different temporal and scene conditions. Since the LoveDA dataset contains a large number of elongated water bodies, Transformer-based and Mamba-based segmentation networks such as CTCFNet, AfaMamba, and MF-Mamba perform worse than CNN-based methods in this scenario. Nevertheless, the proposed method, by incorporating a residual stem, a boundary refinement module, and omnidirectional Mamba for long-range modeling, still achieves strong segmentation performance. MESA-Net achieves Precision and Recall of 83.09% and 81.40%, respectively, demonstrating that it is capable of adequately identifying water bodies whilst minimizing false positives, thereby exhibiting superior predictive balance.

Table 4. Quantitative comparison results for the LoveDA dataset.

Purpose	Method	Type	Backbone	IoU	P	R	F1
DL	Unet	C	–	65.48	79.77	78.52	79.14
	DeepLabV3+	C	VGG16	62.68	84.12	71.09	77.06
	SFFNet	C	ConvNext	68.07	86.21	76.39	81.00
	Res-Unet++	C	–	68.10	79.29	82.83	81.02
	CTCFNet	C + T	–	65.79	83.34	75.76	79.37
	AfaMamba	C + M	ResNet18	66.97	86.41	74.86	80.22
	MF-Mamba	C + M	HRNetV2	65.94	85.40	74.31	79.47
WS	BiSeNet	C	ResNet18	63.55	84.77	71.74	77.72
	MECNet	C	MFConvNet	65.28	82.89	75.45	78.99
	MF_Seg	C + T	MiT-B5	63.33	84.56	71.61	77.55
	MESA-Net (Ours)	C + M	ResNet18	69.83	83.09	81.40	82.24

In addition, compared with dedicated water body segmentation networks, MESA-Net also outperforms BiSeNet, MECNet, and MF_Seg on the LoveDA dataset. Among them, the best-performing baseline, MECNet, achieves 65.28% IoU and 78.99% F1, whereas the proposed method reaches 69.83% and 82.24%, yielding improvements of 4.55% and 3.25%, respectively, indicating a more pronounced advantage. Meanwhile, BiSeNet and MF_Seg achieve IoU scores of 63.55% and 63.33%, respectively, reflecting relatively lower overall performance. These results demonstrate that the proposed method is not only effective for karst water body scenarios but also generalizes well to water segmentation tasks in high-resolution and complex surface environments. This can be attributed to the targeted design of the proposed method in three aspects: boundary enhancement, small-object perception, and continuity modeling of elongated structures, enabling the model to better preserve water body integrity and reduce both false positives and missed detections under complex background interference.

To better evaluate the performance of the model, five representative prediction results from the LoveDA dataset are also presented for comparison, as shown in Figure 10. As illustrated in Figure 10a, only the CNN-based SFFNet and MECNet successfully extract the slender water bodies within the red box; however, due to the spectral similarity between farmland and water, slight misclassification occurs. In contrast, the proposed method demonstrates more accurate segmentation without erroneous predictions. As shown in

Figure 10b,c, for small water body segmentation, the CNN-based U-Net and the proposed method achieve relatively better results, while other comparative methods suffer from both false positives and missed detections. In Figure 10d, the water bodies located within farmland appear dark green, making their spectral characteristics similar to vegetation, and thus most comparative methods fail to identify them effectively. Furthermore, as shown in Figure 10e, due to the high spectral similarity between roads and water bodies, some models incorrectly classify roads as water. The proposed method performs well in both scenarios, further demonstrating the strong robustness of MESA-Net against interference and its capability in extracting slender water bodies. Overall, under complex scenarios involving spectral similarity and elongated water bodies, the proposed method achieves higher segmentation accuracy and robustness. Moreover, it effectively preserves the structural integrity of water bodies under complex background interference, while reducing both false positives and missed detections.

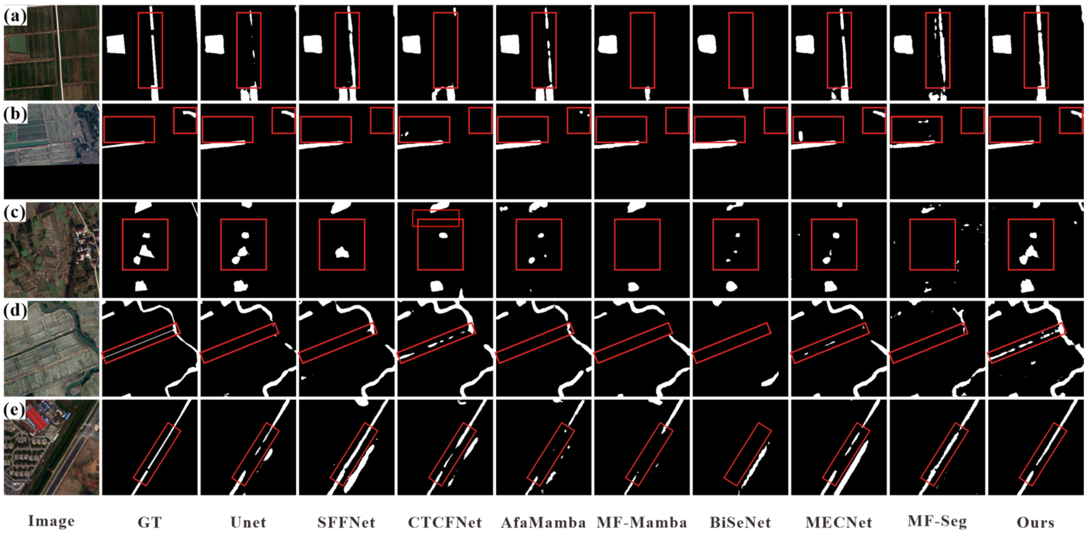

Figure 10. Visualization of segmentation results on the LoveDA dataset. (**a**–**e**) represent five different scenarios.

3.4. Cross-Validation on Different Datasets

To further evaluate the generalization capability of the proposed model under different data distribution conditions, cross-dataset experiments are conducted in this study. Specifically, LJ-Water, Water-CD, and LoveDA are each used as training sets, while the remaining datasets are used for testing, forming multiple cross-domain experimental settings. The results are presented in Table 5. Under cross-dataset evaluation scenarios, all methods exhibit performance degradation to varying degrees, which is mainly attributed to differences in imaging conditions, spatial resolution, and land cover distribution across datasets. However, compared with other methods, the proposed MESA-Net demonstrates superior stability and robustness in most cross-domain settings. For example, under the setting of training on LJ-Water and testing on Water-CD, MESA-Net achieves an IoU of 88.70% and an F1 score of 94.01%, significantly outperforming other methods. Under the setting of training on LoveDA and testing on Water-CD, although the overall accuracy decreases, the proposed method still maintains relatively strong performance. In contrast, other methods suffer more severe performance degradation in cross-dataset scenarios. For instance, Res-UNet++ achieves high accuracy on the in-domain dataset but shows a dramatic drop in IoU from 85.78% to 29.89% during cross-domain testing, indicating high sensitivity to distribution shifts. In addition, the Transformer-based water segmentation network MF_Seg exhibits unstable performance under cross-domain settings, with IoU mostly below 50%, suggesting that its global modeling capability is insufficient for effective transfer under distribution shifts.

Table 5. Cross-validation on different datasets. Specifically, we train the compared methods on one dataset and test them on the other two datasets to evaluate their generalization performance.

	Case		1		2		3		4		5		6	
	Train		LJ-Water		LJ-Water		Water-CD		Water-CD		LoveDA		LoveDA	
	Test		Water-CD		LoveDA		LJ-Water		LoveDA		LJ-Water		Water-CD	
Purpose	**Method**	**Type**	**IoU**	**F1**	**IoU**	**F1**	**IoU**	**F1**	**IoU**	**F1**	**IoU**	**F1**	**IoU**	**F1**
DL	Unet	C	78.11	87.71	63.29	75.82	69.39	78.40	52.28	64.89	63.50	72.23	66.14	78.91
	DeepLabV3+	C	67.17	80.36	59.79	72.15	72.94	81.94	46.23	58.93	59.63	67.80	52.54	68.88
	SFFNet	C	78.99	88.26	58.71	71.12	70.92	79.98	44.89	56.98	58.82	67.02	42.52	59.67
	Res-Unet++	C	85.78	92.34	59.83	72.20	71.74	80.88	45.72	58.03	61.25	69.54	29.89	46.02
	CTCFNet	C + T	88.23	93.74	57.38	69.95	74.82	83.69	50.97	62.83	68.49	77.64	63.59	77.75
	AfaMamba	C + M	77.59	87.38	53.60	66.32	68.84	77.95	46.84	59.18	54.82	62.95	63.37	77.58
	MF-Mamba	C + M	77.53	87.35	57.19	69.80	70.29	79.42	49.25	61.29	36.17	46.79	30.68	46.96
WS	BiSeNet	C	79.41	88.52	49.73	61.77	71.28	80.39	46.18	58.84	63.47	72.17	61.96	76.52
	MECNet	C	85.45	92.15	60.47	72.90	75.20	84.02	43.69	55.87	61.92	70.28	77.98	87.63
	MF_Seg	C + T	47.24	64.17	20.44	33.94	52.25	60.49	32.68	44.36	31.07	40.82	42.54	59.69
	MESA-Net (Ours)	C + M	88.70	94.01	61.20	73.73	78.36	86.92	54.39	66.52	69.30	78.33	76.85	86.91

Further analysis of the generalization differences across datasets reveals that model performance is closely related to the consistency of data sources. When the training and testing datasets share similar sensor types and spectral characteristics, such as LJ-Water and Water-CD, the model can more easily learn transferable features, resulting in smaller performance degradation. In contrast, when there are significant differences between data sources such as LoveDA and Water-CD, variations in spatial resolution and imaging mechanisms limit the model's generalization ability, leading to more pronounced performance decline.

Overall, although performance degradation is inevitable in cross-dataset evaluation, the proposed method maintains strong stability and relatively small performance fluctuations under different data distribution conditions, demonstrating competitive cross-region generalization capability in complex remote sensing scenarios. This finding is consistent with previous studies, which report that although cross-domain performance typically decreases, robust models can still retain comparative advantages.

3.5. Ablation Experiments

To validate the contribution of each module to the proposed MESA-Net for water body segmentation, ablation experiments are conducted on four components: Residual Stem, OGLMB, CAFF, and DGHEF. Specifically, eight groups of module ablation experiments are performed on the LJ-Water dataset, as shown in Table 6. From No. 1, it can be observed that the baseline model ResNet18 achieves an IoU of 58.34% and an F1 score of 73.69% on the LJ-Water dataset. Furthermore, as shown in No. 2–No. 4 and No. 8, with the progressive incorporation of Residual Stem, OGLMB, CAFF, and DGHEF, the model performance improves steadily. When all four modules are integrated, the model achieves the best performance. In addition, removal-based ablation experiments are conducted, where each module is removed individually from the full model. The results are presented in No. 4–No. 7 of Table 6. It can be observed that removing any single module leads to a noticeable decline in performance, indicating that each component contributes positively to the overall model. Among them, the removal of CAFF or DGHEF results in more significant performance degradation, highlighting the critical role of multi-scale feature fusion and boundary enhancement in water body segmentation under complex scenarios. Meanwhile, OGLMB and Residual Stem further enhance the model's capability in capturing global structural information and fine-grained details. Overall, the four modules exhibit strong complementarity, jointly contributing to the performance improvement of the proposed model.

Table 6. Ablation experiment results for the module.

No.	Baseline	Residual Stem	OGLMB	CAFF	DGHEF	LJ-Water	
						IoU	F1
1	√	×	×	×	×	58.34	73.69
2	√	√	×	×	×	77.07	87.05
3	√	√	√	×	×	78.06	87.68
4	√	√	√	√	×	81.81	90.00
5	√	×	√	√	√	83.44	90.97
6	√	√	×	√	√	83.74	91.15
7	√	√	√	×	√	81.98	90.11
8	√	√	√	√	√	84.59	91.65

Notes: "√" indicates that the relevant module is included, "×" indicates that the relevant module is not included.

To further validate the effectiveness of each module, five groups of ablation results are selected for visual comparison, as shown in Figure 11. As illustrated in Figure 11a,b, the baseline model produces relatively scattered responses and obvious low-confidence transition regions around water boundaries, indicating insufficient discrimination between water bodies and complex backgrounds. The comparison between No. 4 and No. 8 further shows that, after introducing DGHEF, the boundary responses become noticeably sharper and more continuous, while the ambiguous transition regions are significantly reduced. In particular, irregular and elongated boundary structures are better preserved in No. 8, demonstrating enhanced capability in fine-grained boundary representation under complex karst scenes. The final predictions of MESA-Net are highly consistent with the ground truth (GT) in both shape and extent, and the background is effectively suppressed to low responses. Furthermore, as shown in Figure 11d, the baseline model is almost incapable of effectively responding to small water bodies, producing only sparse local activations. With the incorporation of modules such as CAFF and OGLMB, small water bodies are progressively activated and form continuous structures, indicating a significant improvement in capturing fine-grained features. In addition, as shown in Figure 11e, in complex background scenarios (including buildings, bare land, and shadows), the baseline model assigns low confidence to water regions. As more modules are introduced, the confidence of water regions increases while background interference is gradually suppressed.

A comparison among different module combinations (such as No. 4–No. 7) further reveals the complementary roles of each component. As shown in Figure 11c, configurations No. 4–No. 6 perform well in small-object detection but still suffer from background noise. In contrast, as shown in Figure 11c,d, No. 7 is more effective in suppressing false positives but is less sensitive to small water bodies and boundary details. The final model (MESA-Net) integrates the strengths of all modules, achieving the best consistency and discriminative capability across large-scale water bodies, small elongated structures, and complex background scenarios.

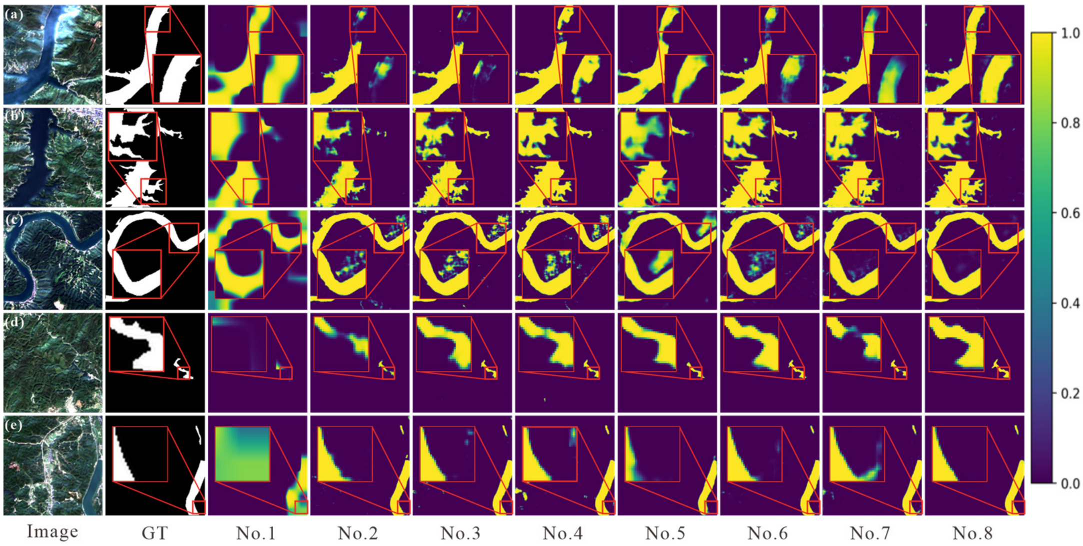

Figure 11. Visualization results of different ablation configurations, where No. 1–No. 8 correspond to No. 1–No. 8 in Table 6. The color gradient from deep blue to yellow indicates an increasing Confidence level. (**a**–**e**) represent five different scenarios.

3.6. Results of Traditional Indexing Methods

To further validate the necessity of employing a deep learning framework in complex karst terrains, four representative traditional water extraction indices, including NDWI [11], MNDWI [12], AWEI [10], and WRI [56], were selected for comparison experiments on the LJ-Water dataset. For NDWI, MNDWI, and AWEI, the segmentation threshold was uniformly set to 0, where pixels with index values greater than 0 were classified as water bodies. For WRI, the threshold was set to 1 after threshold tuning, where pixels with values greater than 1 were regarded as water bodies.

As shown in Table 7, traditional index-based methods exhibit relatively limited performance on the LJ-Water dataset. Among them, NDWI achieves the best performance, with 65.05% IoU and 76.00% F1, while MNDWI and WRI obtain even lower results. Although AWEI is designed to suppress shadow interference, its segmentation accuracy remains unsatisfactory under complex karst conditions. This is mainly because water bodies in karst regions are often affected by mountain shadows, vegetation occlusion, bare rocks, and spectrally similar dark objects, which makes fixed-threshold spectral indices prone to false positives and missed detections.

Table 7. Comparison of traditional index results on the LJ-Water dataset.

Index	LJ-Water	
	IoU	**F1**
NDWI	65.05	76.00
MNDWI	54.48	65.64
AWEI	62.90	73.61
WRI	56.40	67.36
MESA-Net (Ours)	84.59	91.65

In contrast, the proposed MESA-Net achieves 84.59% IoU and 91.65% F1 on the same dataset, significantly outperforming all traditional index-based methods. Compared with the best-performing traditional method (NDWI), MESA-Net improves IoU and F1 by 19.54% and 15.65%, respectively. These results demonstrate that simple threshold-based spectral indices are insufficient for accurately segmenting fragmented, elongated, and small-scale water bodies in complex karst terrains. Benefiting from multi-scale feature fusion, global-local dependency modeling, and edge-aware enhancement mechanisms, the proposed method is more capable of suppressing background interference while preserving water-body continuity and boundary details.

In addition, we have selected five groups of prediction results for comparison, as shown in Figure 12. As shown in Figure 12a, the spectral characteristics of water bodies and mountain vegetation are highly similar. The traditional index method exhibits discontinuities when extracting elongated water bodies. In contrast, the proposed method successfully segments the water bodies in a relatively complete manner. As shown in Figure 12b, there is a bridge in the river; MNDWI, AWEI, and WRI incorrectly identify the bridge as water. Although NDWI avoids the bridge to some extent, the segmentation of the boundary region between the water and the bridge remains unsatisfactory. MESA-Net effectively avoids false positives for the bridge and demonstrates good boundary extraction performance. As shown in Figure 12c, traditional methods exhibit severe false segmentation in built-up areas, whereas the proposed method effectively mitigates this phenomenon. As shown in Figure 12d,e, tributaries extending from the main, elongated river channel penetrate deep into the valley. Traditional index-based methods all exhibit varying degrees of omission when segmenting these tributaries. MESA-Net successfully segments the tributary within the red box. Overall, in karst terrain, the traditional water body index method is prone to missing segmentation and incorrect segmentation, and its segmentation effect is inferior compared to the proposed method and other representative deep learning methods.

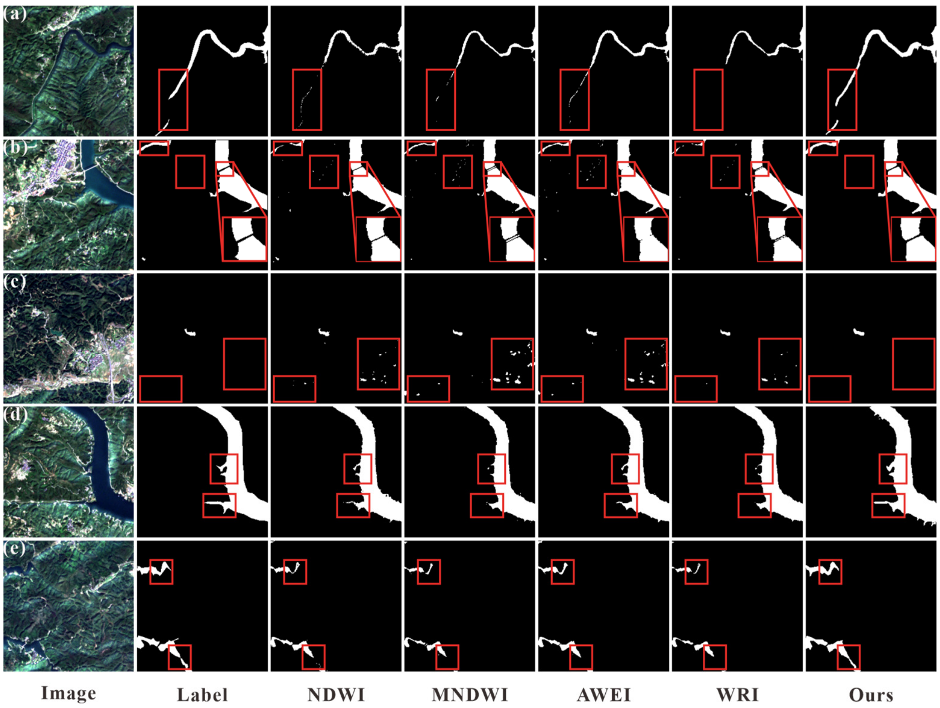

Figure 12. Visualization of traditional index comparisons. (**a**–**e**) represent five different scenarios.

3.7. Comparison of Parameters

To further investigate the model complexity, comparative experiments are conducted under the same experimental settings. Two key aspects are evaluated, namely the number of parameters (Params) and floating-point operations (FLOPs).

As shown in Figure 13a, the proposed method requires 7.18 G FLOPs, which is slightly higher than AfaMamba (7.01 G FLOPs). Although AfaMamba has a lower computational cost, its water segmentation performance is inferior to that of the proposed method. Traditional CNN-based methods, such as U-Net and DeepLabV3+, exhibit significantly higher computational complexity, reaching 81.7 G and 45.68 G FLOPs, respectively. In addition, Transformer-based methods such as MF_Seg and CTCFNet also incur high computational cost due to their complex architectures, with FLOPs of 81.87 G and 111.16 G, respectively. Compared with dedicated water segmentation models such as BiSeNet and MECNet, the proposed model achieves clear advantages in terms of FLOPs. As shown in Figure 13b, in terms of parameter scale, Transformer-based methods such as CTCFNet have a relatively large number of parameters, reaching 68.64 M. Res-UNet++ introduces residual connections on top of UNet++, which significantly reduces the parameter size to 4.06 M. Mamba-based methods, such as AfaMamba and MF-Mamba, incorporate state space models, resulting in parameter sizes of 13.48 M and 11.27 M, respectively. In comparison, the proposed model still maintains advantages over dedicated water segmentation networks in terms of parameter efficiency.

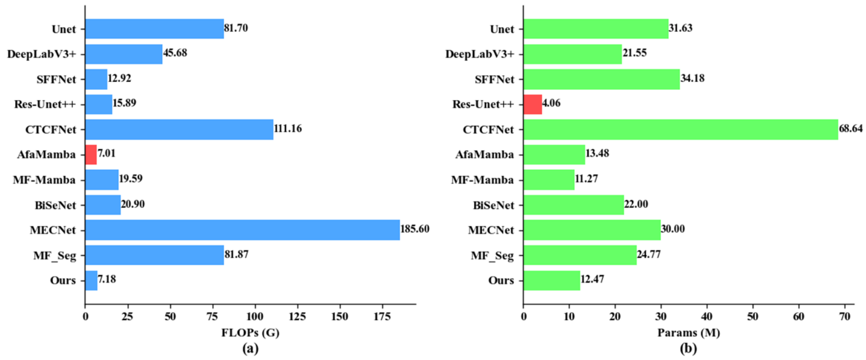

Figure 13. Comparison of Model Parameters. (**a**) represents the comparison of floating-point operations. (**b**) represents the comparison of parameters.

Overall, the proposed model achieves a better balance among parameter size, computational cost, and segmentation performance. Specifically, MESA-Net has 12.47 M parameters and requires 7.18 G FLOPs. While maintaining relatively low complexity, it is still capable of achieving strong segmentation performance, demonstrating high potential for practical deployment.

4. Discussion

From the analysis in Section 3.1 on Water-CD, LoveDA, and the karst water body dataset LJ-Water constructed in this study, it can be observed that significant differences exist among these datasets in terms of water body scale distribution, background complexity, and types of interference. Specifically, water boundaries in Water-CD are relatively clear, whereas LoveDA and LJ-Water contain a large number of small, elongated water bodies affected by shadows, bare rocks, and vegetation interference. Combined with the experimental results in Section 3.3, it can be found that the proposed method not only outperforms advanced general deep learning segmentation networks (DL) on all three datasets, but also consistently surpasses water-specific segmentation methods (WS). This demonstrates that the method possesses strong robustness and generalization capabilities in complex backgrounds and across diverse scenarios, such as karst landscapes. Further analysis based on comparative results shows that, as illustrated in Figure 8, CNN-based methods exhibit advantages in local edge and texture extraction; however, their limited receptive fields restrict effective modeling of slender water bodies and long-range continuous structures. Hybrid CNN–Transformer methods enhance global contextual representation but still suffer from insufficient preservation of small-scale water bodies and high-frequency boundary details. As shown in Figure 10, CNN–Mamba hybrid methods perform well in long-range dependency modeling and computational efficiency, yet they remain insufficient in capturing boundary information and weak target features under complex backgrounds. From the perspective of dedicated water segmentation networks, although existing methods are specifically designed for water body extraction, they still struggle to simultaneously address multi-scale perception, boundary refinement, and structural continuity when dealing with karst environments, fragmented small-scale water bodies, and high-resolution complex surfaces. Moreover, as observed from the visualization results in Figures 8–10, the proposed method demonstrates more stable performance in maintaining the continuity of slender water bodies and suppressing complex background interference. This can be mainly attributed to the DGHEF module, which enhances boundary representation through gradient priors while suppressing pseudo-edge responses, and the OGLMB module, which establishes effective collaboration between global and local features. This enables the model to maintain overall consistency while strengthening fine-grained detail representation. The complementary interactions among different modules further improve the model's adaptability across multiple scales and complex scenarios.

Although the proposed method generally achieves good segmentation results in complex remote sensing scenarios, minor local misclassifications or omissions may still occur in certain challenging cases. Figure 14 shows typical examples from the LJ-Water, Water-CD, and LoveDA datasets. As shown in Figure 14a, although the proposed method successfully preserves the overall structure of mountainous water bodies, a small number of mountain shadows with spectral characteristics similar to those of water bodies are incorrectly identified as water. In Figure 14b, the model accurately extracts the main river structure in the confluence area; however, due to sediment accumulation and spectral discontinuities, a very narrow segment of the water body is partially omitted, resulting in minor discontinuities in the local area. As shown in Figure 14c, under vegetation interference, the proposed method still maintains accurate segmentation for most water bodies, but minor omissions occur near the boundaries of shadows, as shadows alter the spectral response of the water surface. Overall, these results indicate that the proposed method maintains good robustness and structural continuity under complex background conditions; however, it still faces challenges in extremely difficult scenarios involving minute water bodies, severe spectral confusion, and highly irregular boundary details. Future work will focus on enhancing the ability to characterize fine-scale structures and improving robustness against complex illumination and shadow interference, thereby further improving segmentation accuracy in challenging remote sensing environments.

In addition, the current model is based on single-temporal remote sensing imagery and therefore lacks the ability to capture the temporal dynamics of water bodies. The generalization performance may also degrade under cross-sensor or multi-source data scenarios due to differences in spatial resolution, imaging conditions, and spectral characteristics across datasets. To address these issues, future work will further explore multi-temporal modeling, multi-source information fusion, and domain adaptation strategies to enhance the robustness and generalization capability of the model. In particular, unsupervised domain adaptation and domain generalization strategies could be introduced to alleviate distribution discrepancies caused by differences in spatial resolution, imaging conditions, and spectral characteristics across datasets. In addition, feature alignment and adversarial adaptation methods may help improve the transferability of the proposed model under cross-region and cross-sensor scenarios.

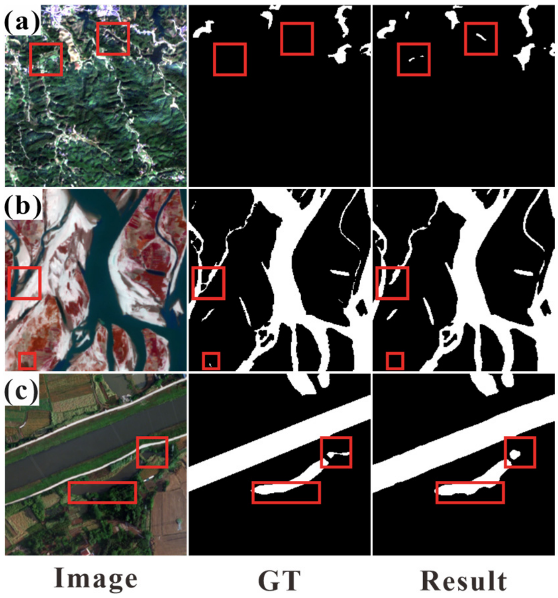

Figure 14. Failure case analysis under challenging scenarios; (**a**–**c**) represent the failure cases for the LJ-Water, Water-CD, and LoveDA datasets, respectively.

5. Conclusions

To address the challenges of object segmentation in karst regions, such as significant variations in scale, ambiguous boundaries, and severe background interference, this paper proposes a multi-directional edge-aware network with scale adaptation suitable for complex scenarios, named MESA-Net. By integrating the Cross-scale Adaptive Feature Fusion module, the Directional Gradient Histogram Edge-guided Fusion module, and the Omni-directional Global–Local Mamba Block, the proposed method effectively improves segmentation accuracy and stability in complex environments.

In terms of network architecture, ResNet-18 is adopted as the encoder to construct an encoder–decoder framework. In the decoding stage, a Cross-scale Adaptive Feature Fusion module (CAFF) is introduced, which performs multi-scale feature fusion and dynamic weight allocation across different encoder layers. This enables effective integration of multi-scale contextual information and enhances the model's ability to perceive both large water bodies and small-scale targets in karst regions. To address the issue of boundary interference in complex backgrounds, a Directional Gradient Histogram Edge-guided Fusion module (DGHEF) is designed. By incorporating explicit gradient histogram representations and boundary-guided mechanisms, it significantly improves fine-grained boundary delineation while suppressing pseudo-edge responses. Furthermore, an Omni-directional Global–Local Mamba Block (OGLMB) is proposed, which leverages omni-directional state space modeling and cross-attention mechanisms to achieve effective collaboration between global semantic information and local detail features. This enhances the model's capability in preserving structural continuity, particularly in karst scenes featuring elongated water bodies and multiple objects. In addition, the LJ-Water dataset constructed in this study focuses on the characteristics of karst water bodies, such as fragmentation, elongation, and high confusion, providing a challenging benchmark for water segmentation in complex scenarios and enabling a more comprehensive evaluation of model performance in real-world applications.

Extensive experiments were conducted on the self-constructed LJ-Water dataset as well as public datasets Water-CD and LoveDA. The results demonstrate that MESA-Net outperforms a variety of advanced general-purpose deep learning segmentation networks and dedicated water segmentation methods in terms of IoU and F1 scores. Specifically, it achieves 84.59% IoU and 91.65% F1 on LJ-Water, 92.15% IoU and 95.91% F1 on Water-CD, and 69.83% IoU and 82.24% F1 on LoveDA. These results not only validate the effectiveness of each module in multi-scale feature fusion, boundary enhancement, and global dependency modeling but also demonstrate the robustness and generalization capability of the proposed method under complex backgrounds and cross-region scenarios. In terms of model complexity and inference efficiency, MESA-Net has only 12.47 M parameters and a computational complexity of 7.18 GFLOPs, indicating that it strikes a good balance between lightweight design and feature representation capabilities, and thus holds significant potential for practical deployment.

Overall, MESA-Net exhibits superior performance in water body segmentation tasks under complex remote sensing environments, showing clear advantages in maintaining the continuity of slender water bodies, detecting small water targets, and suppressing background interference. Future work will further explore multi-temporal remote sensing data and multi-source information fusion to enhance the model's adaptability and practical value in dynamic environments.

Author Contributions: Conceptualization, B.S., Z.Z. and Z.C. (Zhili Chen); methodology, B.S. and Z.Z.; validation, B.S. and Z.Z.; formal analysis, B.S.; investigation, B.S. and T.Y.; resources, B.L., Y.C. and X.Z.; data curation, B.S., J.J. and Q.W.; writing—original draft preparation, B.S. and Z.Z.; writing—review and editing, Z.C. (Zhen Cao), B.L. and Z.Z.; funding acquisition, Z.C. (Zhili Chen) and B.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Guangxi Youth Talent Program, and partly by the Guangxi Science and Technology Project (grant No. GuikeFN2600640051 and GuikeAB25069093).

Data Availability Statement: The Water-CD dataset used in this study was published by Li et al. (2024) [46] and is available at: https://github.com/ChogoriPeak/Water-CD (accessed on 17 April 2026). The LoveDA dataset used in this study was published by Wang et al. (2021) [47] and is available at: https://github.com/Junjue-Wang/LoveDA (accessed on 17 April 2026) The LJ-Water dataset constructed in this paper is available at: https://github.com/zY-Yiz/LJ-Water (accessed on 17 April 2026).

Acknowledgments: The authors are deeply grateful for the comprehensive support provided by three research platforms of Guilin University of Technology: the Guangxi Guilin Observation and Research Station of Agricultural Water, Soil Resources and Environment, the Collaborative Innovation Center for Water Pollution Control and Water Safety in Karst Area, and the Guilin Lijiang River Ecology and Environment Observation and Research Station of Guangxi. Their outstanding laboratory facilities, long-term monitoring data, and on-site technical assistance were indispensable to the successful completion of this study. The experiments in this study utilized the Water-CD dataset published by Li et al. (2024) [46] and the LoveDA dataset published by Wang et al. (2021) [47]. The authors would like to thank the creators of these datasets for providing high-quality data for water body segmentation research. Furthermore, this work has benefited from previous research and open-source resources in the fields of remote sensing and deep learning.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, Z.; Chen, X.; Zhou, S.; Yu, H.; Guo, J.; Liu, Y. DUPnet: Water Body Segmentation with Dense Block and Multi-Scale Spatial Pyramid Pooling for Remote Sensing Images. *Remote Sens.* **2022**, *14*, 5567. [CrossRef]
2. Li, Z.; Huang, M.; Wu, H.; Lv, Z.; Shi, W.; Tao, T. TA-TransUNet: An Improved Deep Learning Network Model for Water Body Extraction from Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2026**, *64*, 5605415. [CrossRef]
3. Chen, S.; Ding, C.; Li, M.; Lyu, X.; Li, X.; Xu, Z.; Fang, Y.; Li, H. PAFNet: A Parallel Attention Fusion Network for Water Body Extraction of Remote Sensing Images. *Remote Sens.* **2026**, *18*, 153. [CrossRef]
4. Xu, P.; Dai, J.; Zhang, T.; Wu, Y. CWBSNet: A segmentation network for complex water bodies in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4206514. [CrossRef]

5. Lei, P.; Zhang, J.; Yi, J. Balancing Accuracy and Efficiency: HWBENet for Water Body Extraction in Complex Rural Landscapes. *Remote Sens.* **2025**, *17*, 3711. [CrossRef]
6. Zhang, T.; Ji, W.; Li, W.; Qin, C.; Wang, T.; Ren, Y.; Fang, Y.; Han, Z.; Jiao, L. EDWNet: A novel encoder–decoder architecture network for water body extraction from optical images. *Remote Sens.* **2024**, *16*, 4275. [CrossRef]
7. Bijeesh, T.V.; Narasimhamurthy, K.N. Surface water detection and delineation using remote sensing images: A review of methods and algorithms. *Sustain. Water Resour. Manag.* **2020**, *6*, 68. [CrossRef]
8. Rokni, K.; Ahmad, A.; Selamat, A.; Hazini, S. Water feature extraction and change detection using multitemporal Landsat imagery. *Remote Sens.* **2014**, *6*, 4173–4189. [CrossRef]
9. Koponen, S.; Pulliainen, J.; Kallio, K.; Haiilkainen, M. Lake water quality classification with airborne hyperspectral spectrometer and simulated MERIS data. *Remote Sens. Environ.* **2002**, *79*, 51–59. [CrossRef]
10. Feyisa, G.L.; Meilby, H.; Fensholt, R.; Proud, S.R. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sens. Environ.* **2014**, *140*, 23–35. [CrossRef]
11. McFeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [CrossRef]
12. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033. [CrossRef]
13. Kaplan, G.; Avdan, U. Object-based water body extraction model using Sentinel-2 satellite imagery. *Eur. J. Remote Sens.* **2017**, *50*, 137–143. [CrossRef]
14. Nguyen, D.D. Water body extraction from multi spectral image by spectral pattern analysis. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *39*, 181–186. [CrossRef]
15. Weng, Y.; Li, Z.; Tang, G.; Wang, Y. OCNet-based water body extraction from remote sensing images. *Water* **2023**, *15*, 3557. [CrossRef]
16. Farooq, B.; Manocha, A. Small water body extraction in remote sensing with enhanced CNN architecture. *Appl. Soft Comput.* **2025**, *169*, 112544. [CrossRef]
17. Lu, M.; Fang, L.; Li, M.; Zhang, B.; Zhang, Y.; Ghamisi, P. NFANet: A novel method for weakly supervised water extraction from high-resolution remote-sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5617114. [CrossRef]
18. Wang, Y.; Li, S.; Lin, Y.; Wang, M. Lightweight deep neural network method for water body extraction from high-resolution remote sensing images with multisensors. *Sensors* **2021**, *21*, 7397. [CrossRef] [PubMed]
19. Duan, L.; Hu, X. Multiscale Refinement Network for Water-Body Segmentation in High-Resolution Satellite Imagery. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 686–690. [CrossRef]
20. Ye, F.; Zhang, R.; Xu, X.; Wu, K.; Zheng, P.; Li, D. Water Body Segmentation of SAR Images Based on SAR Image Reconstruction and an Improved UNet. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 4010005. [CrossRef]
21. Li, R.; Liu, W.; Yang, L.; Sun, S.; Hu, W.; Zhang, F. DeepUNet: A deep fully convolutional network for pixel-level sea-land segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3954–3962. [CrossRef]
22. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015*; Springer: Cham, Switzerland, 2015; pp. 234–241.
23. Wang, S.; Chen, Y.; Yuan, Y.; Chen, X.; Tian, J.; Tian, X.; Cheng, H. TSAE-UNet: A novel network for multi-scene and multi-temporal water body detection based on spatiotemporal feature extraction. *Remote Sens.* **2024**, *16*, 3829. [CrossRef]
24. Wang, Z.; Gao, X.; Zhang, Y.; Zhao, G. MSLWENet: A novel deep learning network for lake water body extraction of Google remote sensing images. *Remote Sens.* **2020**, *12*, 4140. [CrossRef]
25. Zhang, Z.; Lu, M.; Ji, S.; Yu, H.; Nie, C. Rich CNN features for water-body segmentation from very high resolution aerial and satellite imagery. *Remote Sens.* **2021**, *13*, 1912. [CrossRef]
26. Billson, J.; Islam, M.D.S.; Sun, X.; Cheng, I. Water body extraction from Sentinel-2 imagery with deep convolutional networks and pixelwise category transplantation. *Remote Sens.* **2023**, *15*, 1253. [CrossRef]
27. Li, N.; Xu, X.; Huang, S.; Sun, Y.; Ma, J.; Zhu, H.; Hu, M.C. CRAUnet++: A new convolutional neural network for land surface water extraction from Sentinel-2 imagery by combining RWI with improved Unet++. *Remote Sens.* **2024**, *16*, 3391. [CrossRef]
28. Wang, J.; Jia, D.; Xue, J.; Wu, Z.; Song, W.Y. Automatic water body extraction from SAR images based on MADF-Net. *Remote Sens.* **2024**, *16*, 3419. [CrossRef]
29. Cheng, X.; Zhu, Q.; Song, Y.; Yang, J.; Wang, T.; Zhao, B.; Shen, Z.F. Precise city-scale urban water body semantic segmentation and open-source sampleset construction based on very high-resolution remote sensing: A case study in Chengdu. *Remote Sens.* **2024**, *16*, 3873. [CrossRef]

30. Dai, Z.; Liu, H.; Le, Q.V.; Tan, M. CoAtNet: Marrying convolution and attention for all data sizes. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 3965–3977.
31. Li, K.; Wang, Y.; Zhang, J.; Gao, P.; Song, G.L.; Liu, Y. UniFormer: Unifying convolution and self-attention for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 12581–12600. [CrossRef]
32. Yin, L.; Wang, L.; Lu, S.; Wang, R.; Yang, Y.; Yang, B.; Liu, S.; AlSanad, A.; AlQahtani, S.; Yin, Z.; et al. Convolution-Transformer for image feature extraction. *Comput. Model. Eng. Sci.* **2024**, *141*, 87–106. [CrossRef]
33. Kang, J.; Guan, H.; Ma, L.; Wang, L.; Xu, Z.; Li, J. WaterFormer: A coupled transformer and CNN network for waterbody detection in optical remotely sensed imagery. *ISPRS J. Photogramm. Remote Sens.* **2023**, *206*, 222–241. [CrossRef]
34. Wang, S.; Wei, B.; Shi, B.; Wang, N.; Zhang, Y.; Zhu, Y. MHNet: A masked hybrid network for robust water body segmentation from aerial images. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4208015. [CrossRef]
35. Tian, Y.; Cao, H.; Liu, Y.; Tian, C.; Wang, R. WB-Former: A hybrid model of CNN and Transformer for water body extraction in complex scenes. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4210918. [CrossRef]
36. Wang, L.; Li, D.; Dong, S.; Meng, X.; Zhang, X.; Hong, D. PyramidMamba: Rethinking pyramid feature fusion with selective space state model for semantic segmentation of remote sensing imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *144*, 104884. [CrossRef]
37. Li, F.; Wang, X.; Wang, H.; Karimian, H.; Shi, J.; Zha, G. LMVMamba: A hybrid U-shape Mamba for remote sensing segmentation with adaptation fine-tuning. *Remote Sens.* **2025**, *17*, 3367. [CrossRef]
38. Zhang, Q.; Geng, G.; Zhou, P.; Yan, L.; Zhou, M.; Li, Z.; Liu, Y. A Mamba-based vision transformer for fine-grained image segmentation of mural figures. *npj Herit. Sci.* **2025**, *13*, 204. [CrossRef]
39. Gu, A.; Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv* **2023**, arXiv:2312.00752.
40. Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; Liu, Y. VMamba: Visual state space model. *arXiv* **2024**, arXiv:2401.10166.
41. Cai, J.; Tao, L.; Li, Y. CM-UNet++: A multi-level information optimized network for urban water body extraction from high-resolution remote sensing imagery. *Remote Sens.* **2025**, *17*, 980. [CrossRef]
42. Li, Y.; Zhou, P.; Wang, Y.; Li, X.; Zhang, Y.; Li, X. Deep Learning Small Water Body Mapping by Transfer Learning from Sentinel-2 to PlanetScope. *Remote Sens.* **2025**, *17*, 2738. [CrossRef]
43. Liu, D.; Deng, C.; Olofsson, P.; Yan, S.; Wang, J.; Hu, I. SWD-Net: A Benchmark Dataset for Small Water Body Detection From Very High-Resolution Imagery. *IEEE Geosci. Remote Sens. Lett.* **2026**, *23*, 6010304. [CrossRef]
44. Hong, W.; Huang, Z.; Yang, Y.; You, L.; Cai, J.; Zhou, J. SI-Mamba: High-Resolution Sea Ice Recognition via RB-NDI Guided State-Space Model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *19*, 2783–2795. [CrossRef]
45. Liu, H.; Mu, G.; Li, K.; Zhang, H.; Sun, Y.; Sun, H.; Li, S. HyMambaNet: Efficient remote sensing water extraction method combining state space modeling and multi features. *Sensors* **2025**, *25*, 7414. [CrossRef]
46. Li, J.; Shen, Y.; Ye, W. Interactive segmentation method utilizing spectral and contour features: A case study on a dual-task remote sensing water body dataset. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 1503705. [CrossRef]
47. Wang, J.; Zheng, Z.; Ma, A.; Lu, X.; Zhong, Y. LoveDA: A remote sensing land-cover dataset for domain adaptive semantic segmentation. *arXiv* **2021**, arXiv:2110.08733.
48. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
49. Yang, Y.; Yuan, G.; Li, J. SFFNet: A wavelet-based spatial and frequency domain fusion network for remote sensing segmentation. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 3000617. [CrossRef]
50. Jha, D.; Smedsrud, P.H.; Riegler, M.A.; Johansen, D.; de Lange, T.; Halvorsen, P.; Johansen, H.D. ResUNet++: An advanced architecture for medical image segmentation. In Proceedings of the International Symposium on Multimedia, San Diego, CA, USA, 9–11 December 2019.
51. Nie, P.; Cheng, X.; Song, Z.; Mao, M.; Wang, T.; Meng, L. Rethinking BiSeNet: A lightweight network for urban water extraction. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4203910. [CrossRef]
52. Zhang, T.; Li, W.; Feng, X.; Ren, Y.; Qin, C.; Ji, W. Super-resolution water body extraction based on MF-SegFormer. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Athens, Greece, 7–13 July 2024.
53. Lu, C.; Zhang, X.; Du, K.; Xu, H.; Liu, G. CTCFNet: CNN-Transformer complementary and fusion network for high-resolution remote sensing image semantic segmentation. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5408717. [CrossRef]
54. Chen, H.; Luo, H.; Wang, C. AfaMamba: Adaptive feature aggregation with visual state space model for remote sensing images semantic segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 8965–8983. [CrossRef]

55. Xiao, P.; Dong, Y.; Zhao, J.; Peng, T.; Geiß, C.; Zhong, Y. MF-Mamba: Multiscale convolution and Mamba fusion model for semantic segmentation of remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5405916. [CrossRef]
56. Shen, L.; Li, C. Water body extraction from Landsat ETM+ imagery using adaboost algorithm. In Proceedings of the 18th International Conference on Geoinformatics, Beijing, China, 1–4 June 2010.