

Article

SPFDNet: Water Extraction Method Based on Spatial Partition and Feature Decoupling

Xuejun Cheng ^{1,2}, Kuikui Han ³, Jian Xu ^{1,2}, Guozhong Li ^{1,2}, Xiao Xiao ^{1,2}, Wengang Zhao ⁴ and Xianjun Gao ^{3,*}

¹ Changjiang River Scientific Research Institute, Changjiang Water Resources Committee, Wuhan 430010, China; chengxj@mail.crsri.cn (X.C.); xujian@mail.crsri.cn (J.X.); liguozhong@mail.crsri.cn (G.L.); xiaoxiao@mail.crsri.cn (X.X.)

² Key Laboratory of River Regulation and Flood Control in Middle and Lower Reaches of Yangtze River of Ministry of Water Resources, Wuhan 430010, China

³ School of Geosciences, Yangtze University, Wuhan 430100, China; 2022720541@yangtzeu.edu.cn

⁴ Hunan Institute of Water Resources and Hydropower Research, Changsha 410007, China; ssky@slt.hunan.gov.cn

* Correspondence: junxgao@yangtzeu.edu.cn; Tel.: +86-13007165120

Abstract: Extracting water information from remote-sensing images is of great research significance for applications such as water resource protection and flood monitoring. Current water extraction methods aggregated richer multi-level features to enhance the output results. In fact, there is a difference in the requirements for the water body and the water boundary. Indiscriminate multi-feature fusion can lead to perturbation and competition of information between these two types of features during the optimization. Consequently, models cannot accurately locate the internal vacancies within the water body with the external boundary. Therefore, this paper proposes a water feature extraction network with spatial partitioning and feature decoupling. To ensure that the water features are extracted with deep semantic features and stable spatial information before decoupling, we first design a chunked multi-scale feature aggregation module (CMFAM) to construct a context path for obtaining deep semantic information. Then, an information interaction module (IIM) is designed to exchange information between two spatial paths with two fixed resolution intervals and the two paths through. During decoding, a feature decoupling module (FDM) is developed to utilize internal flow prediction to acquire the main body features, and erasing techniques are employed to obtain boundary features. Therefore, the deep features of the water body and the detailed boundary information are supplemented, strengthening the decoupled body and boundary features. Furthermore, the integrated expansion recoupling module (IERM) module is designed for the recoupling stage. The IERM expands the water body and boundary features using expansion and adaptively compensates the transition region between the water body and boundary through information guidance. Finally, multi-level constraints are combined to realize the supervision of the decoupled features. Thus, the water body and boundaries can be extracted more accurately. A comparative validation analysis is conducted on the public datasets, including the gaofen image dataset (GID) and the gaofen2020 challenge dataset (GF2020). By comparing with seven SOTAs, the results show that the proposed method achieves the best results, with IOUs of 91.22 and 78.93, especially in the localization of water bodies and boundaries. By applying the proposed method in different scenarios, the results show the stable capability of the proposed method for extracting water with various shapes and areas.



Citation: Cheng, X.; Han, K.; Xu, J.; Li, G.; Xiao, X.; Zhao, W.; Gao, X.

SPFDNet: Water Extraction Method Based on Spatial Partition and Feature Decoupling. *Remote Sens.* **2024**, *16*, 3959. <https://doi.org/10.3390/rs16213959>

Academic Editor: Konstantinos X. Soulis

Received: 19 August 2024

Revised: 18 October 2024

Accepted: 19 October 2024

Published: 24 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: water extraction; semantic segmentation; feature decoupling; multi-scale

1. Introduction

Accurate and efficient extraction of morphological features of water bodies from remote-sensing images is significant in surface resource change detection, flood disaster assessment and analysis, global water resources analysis, and other applications [1–3]. With

the rapid development of remote-sensing satellite technology [4–8], high-resolution remote-sensing images provide finer feature information for water body extraction. However, since the water shapes change and the features between the water body and water boundaries are different, water boundary extraction is not performed as well as that of the water body [9]. Therefore, exploring the automatic extraction algorithms of water bodies from high-resolution remote-sensing images has essential research value and significance.

The traditional water body extraction methods are mainly divided into threshold segmentation and machine learning-based methods. Among them, the threshold segmentation method primarily utilizes the property that water bodies have different reflectivity in different wavebands. It also suppresses other background features by designing the water index to highlight the water body features [10–12]. The machine learning-based approach mainly uses shallow modules to extract water body information by learning artificially designed features [13,14]. It mainly uses color, texture, and morphology characteristics to complete the extraction process by manual design features. However, these traditional methods cannot improve accuracy robustly in a changing environment. They cannot cope with massive, high-resolution remote-sensing images with complex information [15].

Recently, deep learning has been widely used in remote sensing image explanation because of its powerful autonomous learning and deep information extraction capabilities. Since water bodies in high-resolution remote-sensing images are characterized by sizeable inter-class variance, relevant algorithms proposed in computer vision for natural images cannot directly deal with the exhaustive feature information of remote-sensing images. As a result, inaccurate shoreline localization of water bodies often occurs [16]. Currently, the water feature extraction method based on deep learning mainly combines the characteristics of water bodies in high-resolution remote-sensing images. It is improved and designed based on the coding and decoding architecture [17–19]. The deep semantic features can be extracted by continuously expanding the sensory domain during encoding. The damaged water body boundary is constantly repaired during decoding by combining the spatial information from encoding. The final high-resolution probabilistic map of robust semantic features is obtained. Some studies focus on acquiring more detailed feature descriptions, extracting the multi-scale features for aggregation through convolution at different scales and rates [20–22], or aggregating multi-stage features through a pyramid structure during decoding [23,24]. Some scholars have strengthened features by improving or introducing self-attention mechanisms [25–28]. Alternatively, a multi-path architecture can be introduced to enhance the perception of multi-global information in the network [29]. Some scholars have improved the effective use of information by introducing multimodal data [30–33]. Some studies have focused on obtaining stable spatial patterns with finer boundaries [34,35]. For example, the finer boundary features can be obtained by reconstructing damaged spatial patterns and aggregating detailed information from the coding stage [36]. Some studies advocate directing the network to retain richer information in decoding through multiple output constraints [37,38]. Meanwhile, some methods are designed to achieve more accurate boundary localization by refining the output noise information [39].

The above studies achieve good results in water feature extraction, but they all essentially strengthen the feature extraction results by aggregating richer multi-level features. However, a vital issue is overlooked in this process since the model appears to have different requirements when learning internal versus boundary features of the water body. The internal features of water bodies need semantic solid information to ensure their integrity. Water boundary features need stable spatial information to ensure their positioning accuracy. Coupling water bodies and boundary features with different needs for derivation and optimization will inevitably lead to mutual interference and information competition [40]. This contradiction between the water body and boundaries can interfere with the consistency of the model, resulting in a model that cannot accurately identify these two different needs. Consequently, internal vacancies within the water body and inaccurate location of the water boundaries often occur at the same time.

To address the above problems, this paper adopts the decoupling approach to separate the external features of the water boundaries from the internal features of the water body. Thus, the information interference between the two opposite requirements is reduced when enhancing and supervising the features. Then, the water body features can be extracted more effectively. First, a multi-scale feature extraction architecture with spatial partitioning is designed. Two spatial branch blocks with rich spatial information and deep semantic features are obtained through interaction between fixed-resolution features. In decoding, the spatial branching features of the partition are decoupled utilizing internal flow field prediction so that they are divided into internal body and external boundary. Different feature layers strengthen and remove the semantic features of the internal body and the background noise around the boundary. Then, an integrated expansion recoupling module (IERM) is designed to strengthen the connection between the body and the boundary so that both information can be effectively integrated. It expands the information scope through continuous expansion and adapts the body and boundary features to be aggregated through the information guide map. Finally, the water body features are output by jointly supervising the decoupled boundary and body.

Specifically, the contributions of the proposed approach are illustrated as follows:

- (1) A water body extraction network with spatial partitioning and feature decoupling (SPFDNet) is designed. Feature decoupling is designed to separate the external boundaries of the water body from the internal body features. Thus, we can reduce the interference between the information when enhancing and supervising the features and thus extract the water body features more effectively. Numerous experiments on GID and GF2020 datasets demonstrate that SPFDNet exhibits low complexity and outperforms the other seven methods, particularly in its effectiveness in water boundary positioning.
- (2) A chunked multi-scale feature aggregation module (CMFAM) and an information interaction module (IIM) are designed to form the basic framework for water feature extraction. The framework generates two fixed-resolution spatial paths by spatial partitioning. The context path obtains the context information by continuously expanding the receptive domain. The two paths exchange information through IIM to ensure the stability of spatial information and the richness of deep features. The network can acquire rich multi-scale features at each stage for extracting fine tributaries and large-scale lakes.
- (3) A feature decoupling module (FDM) is explored to separate the features of these two opposing requirements and reinforce the separated features with different levels of features. Specifically, in decoding, the coupled water features are divided into internal body features and external boundary features. Then, deep features under context branching strengthened the decoupled internal body features. The boundary features are strengthened through deep guidance and information supplementation to reduce the information noise.
- (4) An integrated expansion recoupling module (IERM) is designed to repair the transition region between body and boundary features better. The decoupled body and boundary features are intensely supervised through joint supervision to ensure the accuracy of water body boundary localization.

The rest of the paper is organized as follows. In Section 2, some related works are introduced. In Section 3, we describe the methodological framework of this paper in detail. Section 4 introduces the experimental information, including datasets, experimental metrics, and the comparison analysis of different datasets. In addition, the effectiveness of each module is demonstrated through ablation experiments, module visualization, and scene detail illustration. To better illustrate the effectiveness of this proposed method, at the end of this section, we validate it through some representative scenarios with BIou metrics. In Section 5, the conclusion of our work is summarized.

2. Related Work

2.1. Encoder–Decoder Architecture

Deep learning-based water detection methods have given birth to many excellent algorithms, which are mainly based on codec architectures that need improvement. Some of them directly follow the U-Net architecture [41–43]. Some use ResNet as the coding structure [44] and convolution with different dilation rates to replace the downsampling layer to obtain deep semantic information [16]. Some studies have redesigned the feature extractor to build the encoder structure by considering the morphological diversity of the water body [45]. Some works are designed to increase the diversity of features through multiple paths [46].

With the application of the transformer in semantic segmentation, some scholars have introduced part of its structures to obtain more global features of water bodies. Combined with the local feature extraction capability of the convolutional neural network (CNN), the designed new architecture is to achieve finer water body extraction results [1]. To more accurately localize water boundaries, spatial information capability is critical. Many approaches attempt to improve the HRNet structure to retain more detailed information at each resolution of detail [3,47]. However, the multi-path design ensures the stability of spatial data but increases the computational resources to some extent. How to balance the accuracy and efficiency while determining the diversity of retained features and the stability of spatial information is the main problem faced by water body feature extraction.

2.2. Feature Aggregation

Feature aggregation is a commonly used structure in water body segmentation algorithms to improve the expressive power of features by fusing different layers of features. Ref. [48] reduced information redundancy at different scales after aggregation by erasing attention modules. Ref. [24] used channel attention to measure its channel weights and perform a concatenation operation on the weighted features to assign spatial dimension weights adaptively. This adaptive weight fusion suppresses noise and background information. Ref. [34] enhanced the communication capability of the two channels by bootstrapping at different scales. It enhanced the interconnectivity between the two types of feature representations by capturing features at various scales to realize the fusion of the two kinds of feature representations. Ref. [27] proposed a non-local neural network combined with global pooling to extract spatial location information of the image and embed the location information into the channel information. This approach enhanced the global attention module by enabling the network to mine further information along the horizontal and vertical directions of the feature map and captured the long-distance dependencies of the input feature map along one direction. This mitigated fine-grained spatial details at different scales from being corrupted by deeper features during aggregation. Ref. [36] proposed and designed a continuous attention fusion module. Ref. [49] proposed a gated-channel transform (GCT) module for skipping connections to fuse the shallow features of the encoder with the deep features of the decoder. Compared with the traditional skipped connections, the inclusion of the GCT module can adaptively adjust the weights of each channel to improve the performance of the whole network. TBiSeg improved the transformer block using two-layer routing attention to enhance understanding of water body structure [50]. SPNet guides the fusion of low and high-level semantic information through a multi-channel deep feature extraction module and combines it with a goal-directed attention mechanism to deal with noise. In contrast, better results were achieved in water body detection [51]. Ref. [52] integrated a histogram extraction layer designed for SAR content into a deep segmentation neural network to achieve better performance while being lightweight. MADFNet focuses on feature regions critical for recognizing water edges by fusing attentional weights in channel and spatial dimensions [53].

The above feature aggregation methods aggregate features at different levels to enhance the sampled features and compensate for the lost information to some extent. How-

ever, the lack of ignoring the coupling and redundancy of feature representations makes the aggregated features such that there are always opposite combinations of features.

2.3. Feature Decoupling

Feature decoupling decomposes input features into several mutually independent and non-interfering parts according to a specific method. Ref. [54] observed that the pixels inside the object are similar, and the pixels at the boundary often show differences. In this regard, they are divided into low- and high-frequency information by decoupling, and the obtained body features and residual edge features are further optimized under decoupling supervision. Ref. [55] decoupled all landmark features into single landmark features. In this process, dynamic multicore convolution is designed as the decoupling fundamental component to improve the feature robustness of inconspicuous landmarks. Ref. [56] decoupled the input features into body and edge features, and the two optimized features are combined into a comprehensive predictive representation using a feature fusion operator.

The above methods achieve better results in their respective applications, especially for boundary localization. The boundary of a water body has more complex characteristics than the interior, and simple decoupling is not sufficient to cope with the complex background region. Moreover, direct fusion does not connect the connecting area between the transition body and the boundary in recoupling. All these can cause errors in feature representation for the localization of water body boundaries.

3. Methodology

3.1. General Model Architecture

The overframe of the water extraction network based on spatial partitioning and feature decoupling network (SPFDNet) is shown in Figure 1. SPFDNet mainly consists of a chunked multi-scale feature aggregation module (CMFAM), an information interaction module (IIM), a feature decoupling module (FDM), and an integrated expansion recoupling module (IERM).

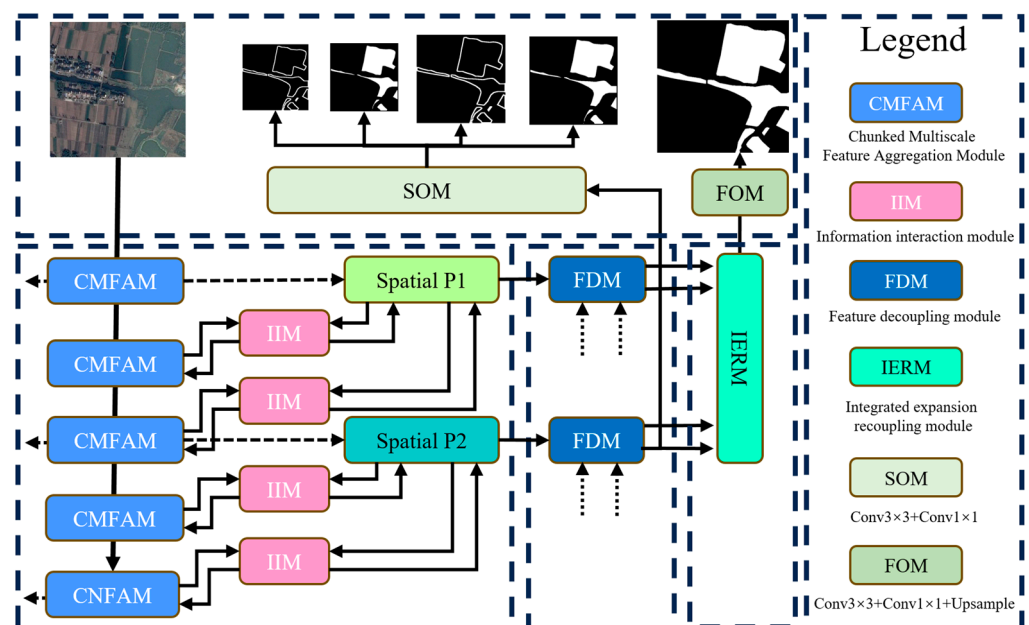


Figure 1. The general structure of the SPFDNet. SOM refers to the supervisory output module. FOM refers to the feature output module. The solid and dashed lines represent the transfer of features.

First, CMFAM is designed to extract richer features. Then, the image is input into two paths, including the contextual path and the spatial path. The contextual path is continu-

ously down-sampled to expand the perceptual domain. The spatial path retains the spatial information at $1/2$ (Spatial P1) and $1/8$ (Spatial P2) of the spatial resolution, respectively. The two path branches continuously exchange semantic features and detailed information through the IIM. In the decoding process, the features of the two spatial partitions are decoupled by IIM, and the decoupled volume and boundary features are enhanced by deep semantic information with the exact resolution details to remove redundant noise. Then, the decoupled features by IERM are recoupled to correct the connectivity region between the body and the boundary. The decoupled body and boundary features in two spatial partitions are remotely supervised to ensure that the output decoupled features are effectively supervised. The accurate positioning of the output water boundary is ensured by means of joint constraints.

3.2. Interactive Framework for Spatial Partition Extraction

The interactive framework for spatial partition extraction mainly relies on the CMFAM and IIM. The context path is composed of five layers. Each feature layer is extracted by a CMFAM. The maximum down-sampling multiplier is set to 32. The features of the first and third layers are kept as the initial spatial region. The context path continuously downsamples to obtain deep semantic features. The spatial path retains stable spatial information through fixed resolution. The two paths exchange the extracted semantic features with the retained detail information through IIM. Therefore, the network maintains stable detailed information while acquiring deep semantic information.

As shown in Figure 2, CMFAM mainly consists of multiple convolutional blocks, considering the balance between accuracy and efficiency. The dimension of the input features is adjusted by the first 3×3 convolution. Then, the output features are divided into four identical dimensions. The first and second dimensions are extracted by 3×3 convolution of features and fused with the next dimension. Considering that the water bodies are fine tributaries and large-scale lakes, the features are extracted in the last two dimensions with convolution kernel sizes of $1 \times 7, 7 \times 1$ versus $1 \times 11, 11 \times 1$, respectively. Finally, the partitioned features are cascaded.

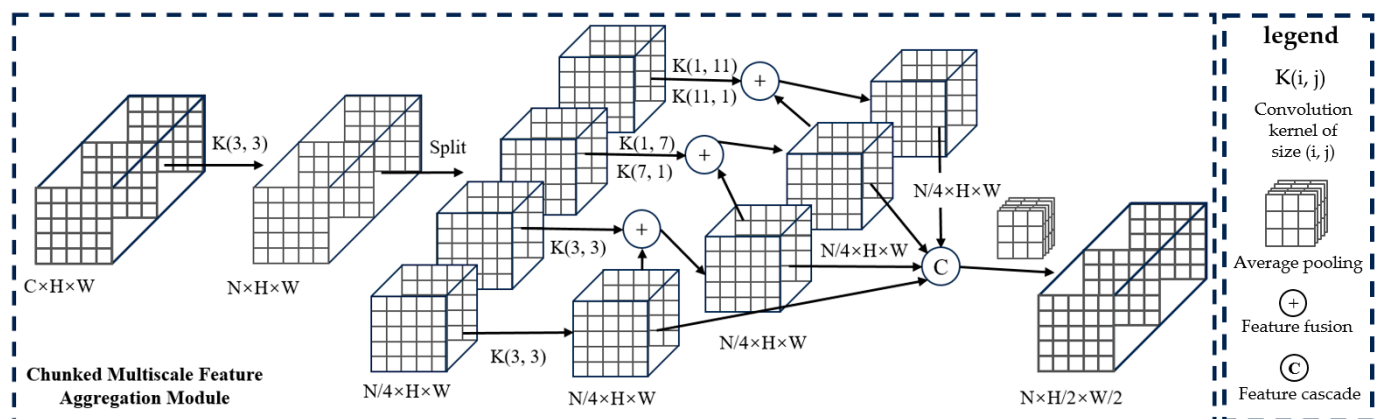


Figure 2. The structure of the chunked multi-scale feature aggregation module (CMFAM).

As shown in Figure 3, the IIM consists of semantic and spatial interactions. Semantic interaction generates a semantic probabilistic bootstrap map by refining features with up-sampling output and then fusing with spatial branches. Spatial interaction generates spatial information bootstrap maps by down-sampling refined features and combining them with contextual branches. Through the continuous interaction and fusion of CMFAM and IIM, they guide each other and finally constitute the base model framework. The specific parameters of the framework are shown in Table 1.

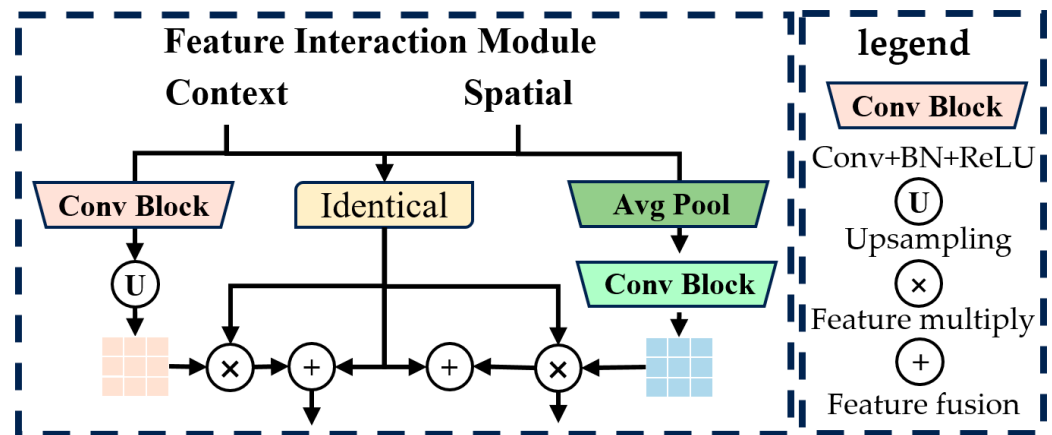


Figure 3. The structure of information interaction module (IIM).

Table 1. The parameter list of the base network with [module name, number of output channels], where 1×1 indicates a convolution kernel of 1 convolution.

	Stage	Input	Context Branch	Spatial Branch
Encoder	Stage 1	(3, 512, 512)	[CMFAM, 64]	
	Stage 2	(64, 256, 256)	[CMFAM, 128]	[IIM]
	Stage 3	(128, 128, 128)	[CMFAM, 256]	[IIM]
	Stage 4	(256, 64, 64)	[CMFAM, 512]	[IIM]
	Stage 5	(512, 32, 32)	[CMFAM, 512]	[IIM]
Decoder	Up1	(512, 16, 16)	[CMFAM, 256]	[Upsample]
	Up2	(256, 128, 128)	[CMFAM, 64]	[Upsample]
	Out	(64, 512, 512)	[1 × 1, 1]	

3.3. Feature Decoupling Module

To solve this coupled and redundant decoding method, which leads to redundancy of information and mislearning, we designed the feature decoupling module (FDM) as shown in Figure 4. The body features are separated by predicting the flow field inside the features and using feature erasure to obtain the boundary features [54].

$$F_1 = K_1^1(F_{cat}(F_{up}(F_{low}), F_{high})) \quad (1)$$

where $F_{low} \in \mathbb{R}^{C_1 \times H \times W}$ is a high-level feature at low resolution. $F_{high} \in \mathbb{R}^{C_2 \times H \times W}$ is a lowlevel feature at high resolution. F_{cat} is the splicing of features along the channel direction. F_{up} represents up-sampling using bilinear interpolation. K_k^d denotes features obtained by up-sampling using a convolutional kernel of size k and a dilation rate of size d . The characteristics after output enhancement are $F_1 \in \mathbb{R}^{C_2 \times H \times W}$.

$$F_{body} = \eta(K_3^1(F_1), F_1), F_{edge} = F_1 - F_{body} \quad (2)$$

where η denotes the resampling of F_1 by squares using the predicted internal flow, and the resulting internal body is characterized by $F_{body} \in \mathbb{R}^{C_2 \times H \times W}$. Use the body feature to erase the original feature to obtain the boundary feature $F_{edge} \in \mathbb{R}^{C_2 \times H \times W}$.

$$F_{Sbody} = K_1^1(F_{cat}(F_{body}, (F_{up}(F_{low}))) \quad (3)$$

Since the decoupled body features do not have strong semantic information, there is a lot of noise around the boundary features. Thus, deeper features are introduced

to supplement the strong semantic information of the internal body by cascading. The resulting enhanced body feature is $F_{Sbody} \in \mathbb{R}^{C_2 \times H \times W}$.

$$F_{Sedge} = F_{edge} + (K_1^1(F_{cat}(F_{edge}, F_{detail})) \otimes \sigma(K_7^1(F_{low}))) \quad (4)$$

The noise information around the boundary is weakened by employing semantic guidance, where $F_{detail} \in \mathbb{R}^{C_2 \times H \times W}$ represents the detailed information for high resolution. σ is expressed as a sigmoid function. $\otimes$ denotes the multiplication of features. High-level features F_{low} is used to obtain semantic information to guide the graph $\in \mathbb{R}^{1 \times H \times W}$. The noise around the boundary is suppressed by feature multiplication. Since the phase-reduced internal body features are used to obtain the boundary information, the features inside the boundary will not grow too much after the product. Finally, the original boundary features are fused with the original boundary features to output the strengthened boundary features $F_{Sedge} \in \mathbb{R}^{C_1 \times H \times W}$.

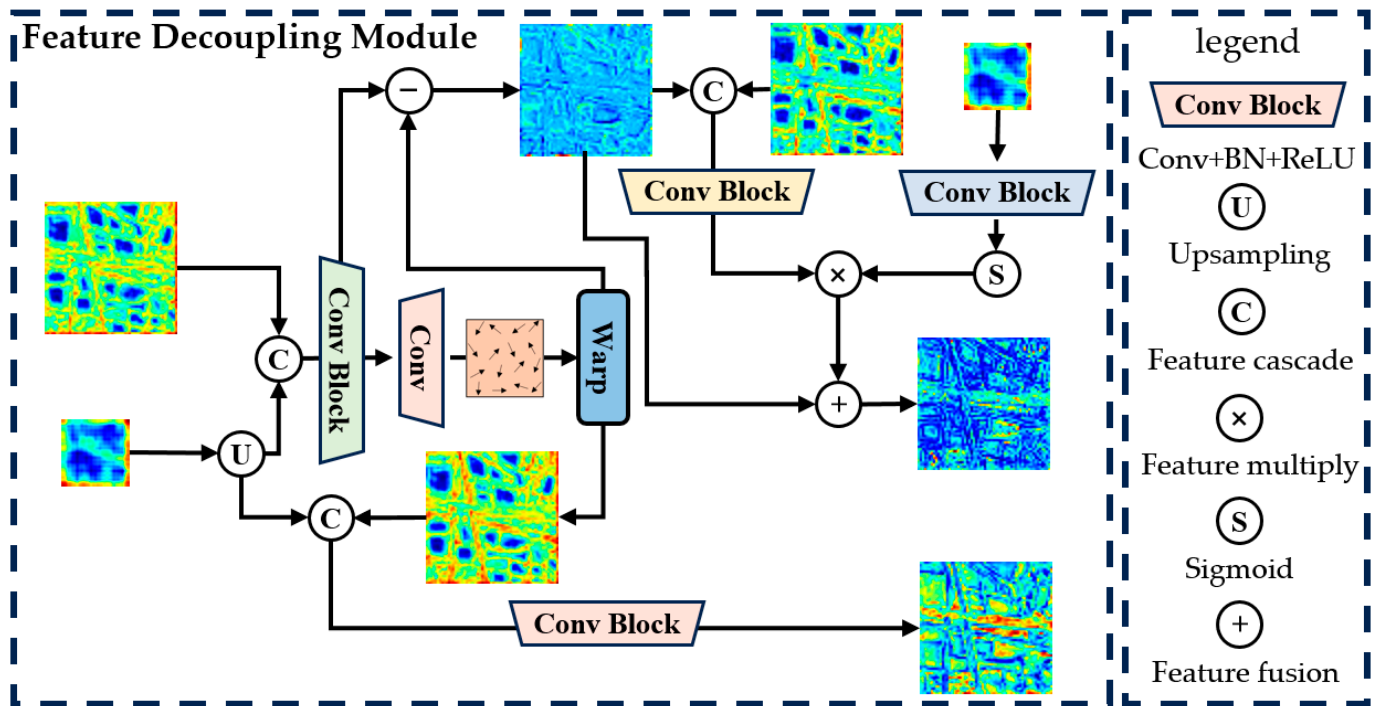


Figure 4. The structure of feature decoupling module (FDM).

3.4. Integration of Expansion Recoupling Modules

Directly fusing the decoupled features is prone to missing features in the connecting region of the body and boundary. Therefore, IERM is designed to strengthen the connection between body and boundary so that the information from both can be effectively integrated, as shown in Figure 5.

$$F_{edges} = F_{cat}(F_{up}(K_1^1(F_{edge1})), F_{edge2}) \quad (5)$$

Firstly, the body and boundary features under two spatial partitions were up-sampled to the same spatial resolution, where the output is $F_{edges} \in \mathbb{R}^{64 \times H \times W}$, and is the converged boundary feature. $F_{edge1}, F_{body1} \in \mathbb{R}^{64 \times H \times W}$ is the decoupled feature of the first spatial partition. $F_{edge2}, F_{body2} \in \mathbb{R}^{256 \times H/4 \times W/4}$ is the feature after decoupling the second spatial partition. The up-sampling of edge and body features under different spaces is adjusted to the same dimension at the same spatial resolution and spliced along the channel direction.

We use dimensionality reduction and expansion of the feeling domain to increase the scale range of the body and boundary.

$$F_{edges1} = F_{cat}(K_3^4(K_3^2(K_3^1(F_{edges})))) \quad (6)$$

Next, the body and boundary features are cascaded separately to be inflated by convolution with different dilation rates. Where $F_{edges1} \in \mathbb{R}^{64 \times H \times W}$ is the boundary feature after integrated expansion. K_k^d denotes the features obtained by up-sampling using a convolutional kernel of size k and a dilation rate of size d . Because, in Formulas (5) and (6), the boundary and the body are treated in the same way, only the processing of boundary features is illustrated as an example.

$$W_{edge} = \sigma(K_7^1(F_{edges1})), W_{body} = \sigma(K_7^1(F_{body1})) \quad (7)$$

$$F_{fuse} = K_1^1(W_1 \otimes F_{body1} + (1 - W_1) \otimes F_{edges1} + W_2 \otimes F_{edges1} + (1 - W_2) \otimes F_{body1}) \quad (8)$$

Then, the fused information at various scales is cascaded to generate an information guidance map through a large-scale convolution with an activation function. The boundary and body features are first compressed to generate a weighted bootstrap map $W_{edge} \in \mathbb{R}^{1 \times H \times W}$, $W_{body} \in \mathbb{R}^{1 \times H \times W}$. When coupling, the features are reinforced in the corresponding region by information guidance of the body and boundary features to compensate for the transition region between the body and the boundary. Raising the feature product strengthens the opposing body and boundary features, and finally, raising the one-dimensional convolution adjusts the feature information. $F_{fuse} \in \mathbb{R}^{64 \times H \times W}$ is the final output.

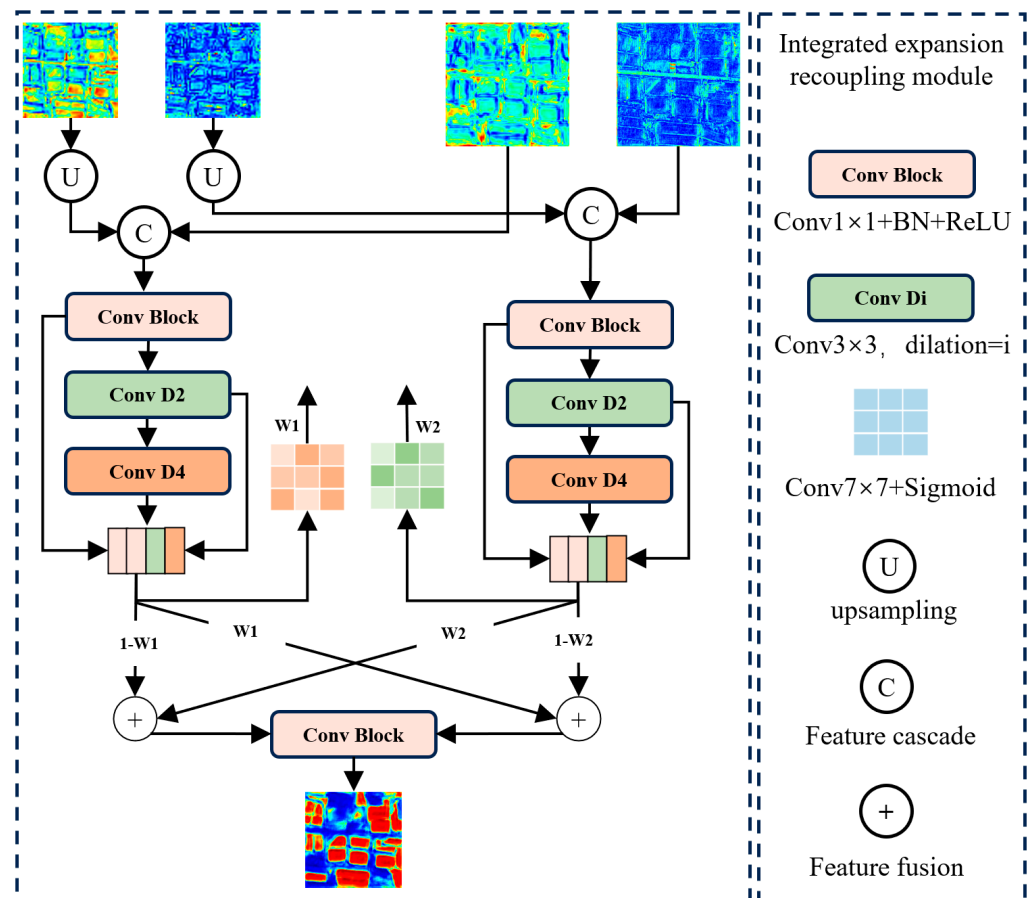


Figure 5. The structure of the integrated expansion recoupling module (IERM).

3.5. Multi-Level Constrained Loss Function

In order to better supervise the decoupled body and boundary, we decouple the body and boundary features for output supervision. The truth labels are processed by the canny operator to obtain the boundary information and then obtain the body features by phase reduction. Down-sampling matching is performed according to the resolution of the features to be supervised. Since the boundary pixels account for a relatively small number of pixels, the combination of binary cross entropy (BCE) [57] and dice loss (Dice) [58] is used to supervise the boundary constraints. The loss function is defined as follows:

$$L_{Dice} = 1 - \frac{2 \times \sum_{i=1}^N (g_i \times p_i)}{\sum_{i=1}^N g_i + \sum_{i=1}^N p_i} \quad (9)$$

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^N (g_i \times \log(p_i) + (1 - g_i) \cdot \log(1 - p_i)) \quad (10)$$

where L_{Dice} is the loss constraint for the last layer of output results. Where N is the total number of samples. g_i denotes whether a pixel in the label belongs to a water body or not, and i denotes the probability that the i th pixel belongs to a water body.

Since the boundary pixels account for a relatively small number of pixels, to balance the problem of positive and negative sample imbalance in the boundary output, we use BCE to constrain the boundary jointly with $Dice$. In our experiments, we set $\lambda_2 = 0.5$, $\mu_2 = 0.5$, $\lambda_1 = 1.0$ and $\mu_1 = 1.0$. They represent the constraint weights of the boundary to the body at 1/8 resolution and the boundary to the body at 1/2 resolution, respectively. The overall loss function of the model is $L_{SPDINet}$.

$$L_{SPDINet} = L_{Dice} + \lambda_2(L_{BCE}^1 + L_{Dice}^1) + \mu_2 L_{BCE}^2 + \lambda_1(L_{BCE}^3 + L_{Dice}^3) + \mu_1 L_{BCE}^4 \quad (11)$$

4. Experiment

4.1. Datasets

To better validate the effectiveness of the method in this paper, two publicly available datasets that contain water body features are used for testing, namely the GID [59] and the GF2020 [60].

4.1.1. GID Datasets

The WHU GID is a high-resolution land cover dataset acquired by the Gaofen-2 satellite. It consists of 150 images with a pixel size of 6800×7200 , and each image is composed of four bands: red, green, blue, and near-infrared. We composed the image in three bands: near-infrared, red, and green. Due to the computer memory limitation and loss of image edge features, the image is cropped to 512×512 as the window size is 64.

4.1.2. GF2020 Datasets

After cropping the overlap region, we remove a certain number of blank or full water body images and finally get 6978 images for the training set, 1744 for the validation set, and 2032 for the test set.

GF2020 is a public dataset provided in the automatic segmentation of water bodies in the 2020 high score challenge optical satellite imagery mission, which is collected by the GF-2 satellite and consists of 1000 high-resolution optical images with a pixel size of 492×492 . The datasets are divided into training, validation, and test sets in the ratio of 7:1:2.

4.2. Experimental Details

4.2.1. Experimental Environment

The experimental environment in this paper is built under a Windows system, equipped with a GeForce RTX 3090 graphics card with 24G video memory, and the deep learning

framework used is PyTorch. Adam is the optimizer, with the initial learning rate uniformly set to 0.0001. Considering the memory consumption of all models, the batch size is set to 8, and an automatic mixed-precision strategy is used to reduce the memory consumption.

4.2.2. Evaluation Metrics

To evaluate the performance of the model more comprehensively, we chose five commonly used semantic segmentation evaluation metrics, which are overall precision (OA), precision (P), recall (R), and F1 score (F1). In addition, in order to more accurately evaluate the precision of boundary localization, we introduce BIoU as an evaluation index. The specific equation for each index is as follows:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (12)$$

$$P = \frac{TP}{TP + FP} \times 100\% \quad (13)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (14)$$

$$F1 = \frac{2 \times P \times R}{P + R} \times 100\% \quad (15)$$

$$IOU = \frac{TP}{TP + FP + FN} \times 100\% \quad (16)$$

$$BIOU = \frac{|(G_d \cap G) \cap (P_d \cap P)|}{|(G_d \cap G) \cup (P_d \cap P)|} \times 100\% \quad (17)$$

where TP denotes the number of target pixels correctly identified, FP denotes the number of background pixels mistakenly detected as target pixels, TN denotes the number of target pixels recognized as background pixels, and FN denotes the number of negative samples judged as background pixels. In the $BIOU$ metric, G is the set of target pixels, P is the set of predicted target pixels, G_d denotes the set of pixels whose contour distance from the set of true target pixels is not greater than d , and P_d denotes the set of pixels whose contour distance from the set of predicted pixels is not greater than d . We set the size d by referring to the description in [61].

The structural similarity index (SSIM) [62] measures how similar two images are. It not only compares the pixel values but also integrates the brightness, contrast, and structural features of the images. Compared with traditional evaluation metrics such as mean square error (MSE) or peak signal-to-noise ratio (PSNR), SSIM pays more attention to the structural information of an image, thus better reflecting the human eye's perception of the image quality than just the difference between pixel values.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (18)$$

where μ_x, μ_y denotes the mean of image x and y , σ_x^2, σ_y^2 denotes the variance between image x and y , σ_{xy} denotes the covariance between image x and y , and C_1, C_2 denotes a stabilization constant with a zero denominator in m .

4.2.3. Comparison Methods

To verify the effectiveness of the proposed methods more comprehensively, seven methods are selected from the segmentation methods in the computer and remote sensing fields for comparison. UNet [63] is a classical coding and decoding structure. DeepLabV3+ [64] is a coding and decoding structure that includes multi-feature processing. ResUNet++ [65] is an improved extractor coding and decoding structure, which is similar to the direction of improve-

ment in this paper. In the field of remote sensing, MECNet [43], MFGF_UNet [49], MUNet [1], and MSResNet [44] are codec architectures for water body segmentation networks.

4.3. Comparative Experiments Analysis

4.3.1. GID Comparison Experiments Analysis

The test set results of all algorithms are compared, and some more representative scenes are selected for analysis. The visual results are shown in Figure 6. The fifth, seventh, and ninth subfigures are large-scale lake scenes, around which some depressions, rice fields, and other pseudo-targets are suspected to be lakes. Most algorithms are more prone to internal feature vacancies and false detections. Our approach has a more complete shape due to using CMFAM to build the base network, which retains rich multi-scale features at each stage. The third, fourth, and eighth subfigures are scenes with strong background interference and severe shoreline erosion, and it can be seen that all algorithms are unable to localize the shoreline boundary accurately. Since MECNet designs a multi-scale prediction fusion module, the boundary is more complete in the comparison methods. Our method used decoding of decoupling, and the boundary is strengthened and denoised to locate the boundary more accurately.

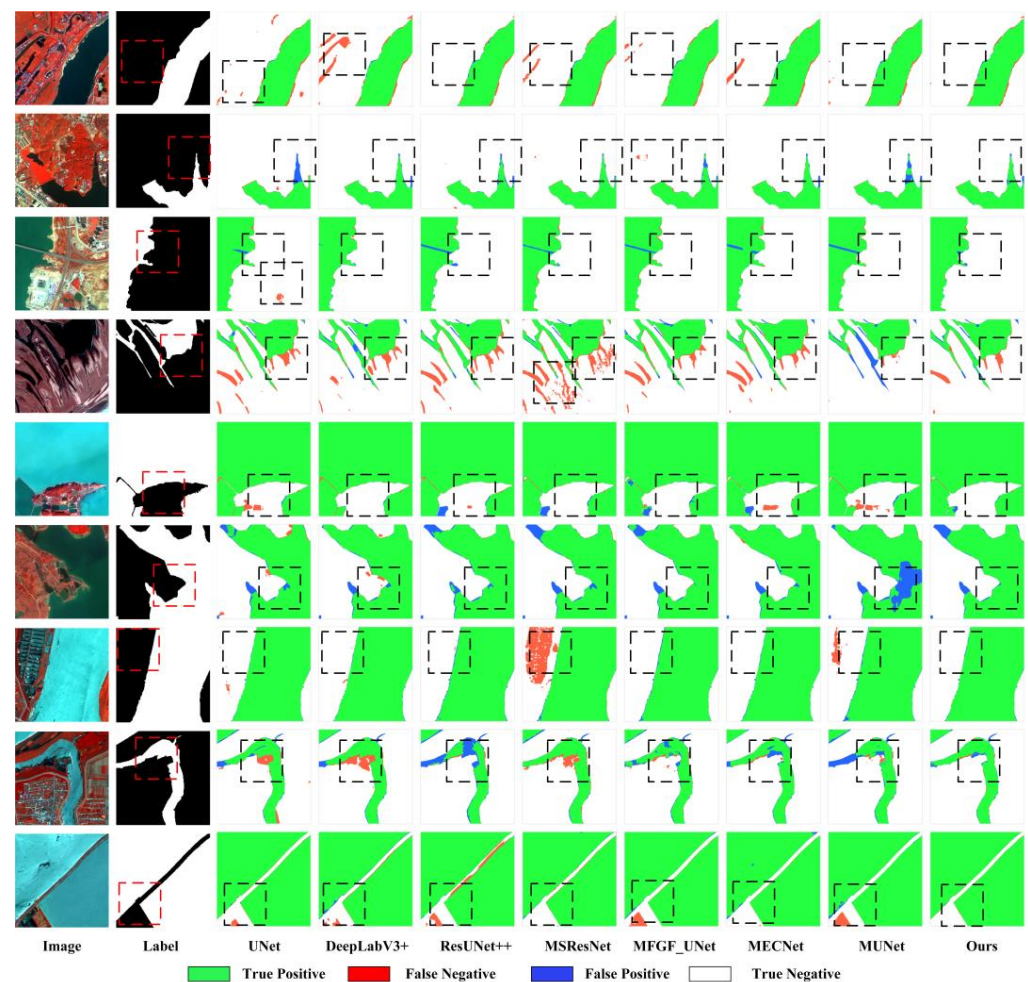


Figure 6. Comparison results on the GID.

The quantitative comparison results of the GID are shown in Table 2, which shows that our model achieves the best results in the three metrics of OA, IoU, and F1. Among them, IoU is higher than other algorithms, ranging from 0.43% to 2.03%. MUNet achieves the highest value in p due to the embedded MixFormer structure, which has better long-distance relationship modeling. Our method adopted the architectural strategy of two-path

interaction, which achieved good results in P and R. R improved by 1.39% compared to MUNet.

Table 2. Quantitative evaluation results of the comparative methods on the GID (black font underlined and bolded for optimal performance).

Methods	OA/%	P/%	R/%	IoU/%	F1/%
U-Net	96.54	93.31	95.29	89.19	94.29
DeepLabv3+	96.68	94.06	94.91	89.54	94.48
MSResNet	96.97	94.25	95.71	90.42	94.97
Resunet++	96.88	93.70	<u>96.03</u>	90.20	94.85
MFGF_UNet	96.91	94.94	94.71	90.17	94.83
MUNet	97.03	<u>95.46</u>	94.56	90.49	95.01
MECNet	97.09	94.78	95.56	90.79	95.17
ours	<u>97.23</u>	94.86	95.96	<u>91.22</u>	<u>95.41</u>

4.3.2. GF2020 Comparison Experiment Analysis

The visual comparison results are shown in Figure 7. The GF2020 has more fine rivers, and the lake information is more likely to be overwhelmed by the background. Therefore, it is suitable to test the ability to aggregate contextual and spatial detail information perception. As shown in Figure 6, a large number of data noise is the most common problem of the comparison methods. Because FDM is used to divide the features under the spatial path into subject and boundary, combined with multi-level features, the separated features are constrained and enhanced. The noise is well suppressed, and good results are obtained in the comparison. In addition, most methods exhibit lake shrinkage, which is an inaccurate boundary localization due to missing spatial information. Since our method performed spatial partitioning to save stable spatial information in different layer zones, it has better localization ability. It is worth mentioning that MECN et al., achieved good results in these scenarios thanks to the enhanced capture of the localized information.

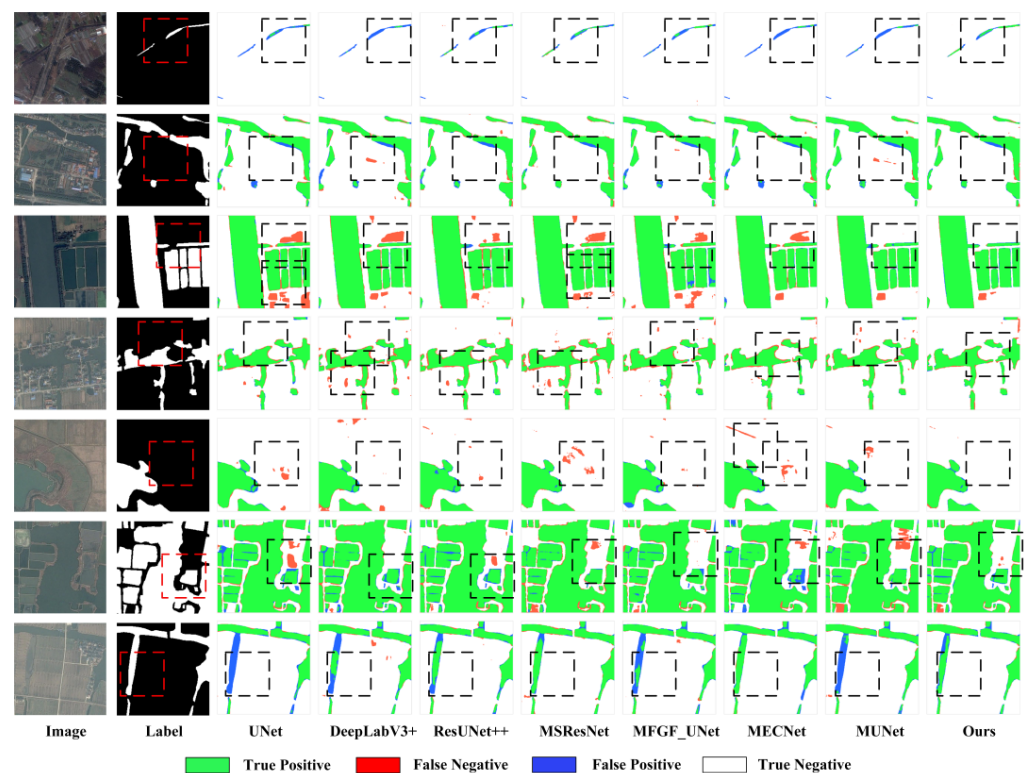


Figure 7. Comparison results of each method on GF2020.

The quantitative comparison results of GF2020 are shown in Table 3. Compared with GID, the image resolution of the GF2020 is lower, and the objects are primarily fine lakes. Thus, the whole water detection effect is worse. However, the method in this paper still achieves the best results in OA, IoU, and F1 indicators. The precision is 2.54% higher compared to DeepLabv3+. Benefiting from the multi-scale residual structure, the recall of MSResNet achieves the highest results, but other accuracy indexes are much lower than the proposed method.

Table 3. Quantitative evaluation results of the comparative methods on the GF2020 (black font underlined and bolded for optimal performance).

Methods	OA/%	P/%	R/%	IoU/%	F1/%
U-Net	94.53	84.59	86.58	74.78	85.57
DeepLabv3+	94.40	83.72	86.97	74.40	85.32
MSResNet	94.99	83.71	90.97	77.29	87.19
Resunet++	94.94	87.26	85.46	75.99	86.35
MFGF_UNet	94.97	86.54	86.63	76.34	86.58
MUNet	95.11	84.76	90.11	77.54	87.35
MECNet	95.03	85.92	87.86	76.80	86.88
ours	95.49	86.26	90.28	78.93	88.22

4.4. Ablation Experiment

We conducted ablation validation analyses in three directions to better validate the extent to which each module contributes to the model.

4.4.1. Quantitative Tests Analysis of Each Module

To comprehensively validate the effectiveness of each module, the model is partitioned into a trunk model and individual modules. The backbone model is encoded through the context path and decoded via the spatial path. Spatial paths are segregated into Spatial P1 and Spatial P2 with distinct resolutions for validation. Sequentially introducing FDM, IERM, and FOM to construct different networks, the results of model ablation experiments are shown in Table 4. Each module exhibits a positive correlation with enhancing model performance. IDM effectively addresses propriety and boundary requirements by decoupling and reinforcing features, resulting in the most significant improvement in network performance increase of 1.31% of IoU. Compared with Spatial P1, Spatial P2 achieves better effects since it obtains higher semantic features through IIM while retaining relatively complete spatial information. IERM enhances model accuracy by integrating expansion and adaptive guidance of information. SOM comprehensively improves the model through directional monitoring of decoupled volume and boundary features.

Table 4. Model ablation experiments.

Spatial P1	Spatial P2	FDM	IERM	SOM	FOM	OA	P	R	IoU	F1
✓					✓	94.83	88.07	83.74	75.21	85.85
	✓				✓	94.74	84.63	87.87	75.78	86.22
✓	✓				✓	94.94	86.32	86.73	76.25	86.52
✓	✓	✓			✓	95.15	85.23	89.60	77.56	87.36
✓	✓	✓	✓		✓	95.33	86.93	88.33	77.98	87.63
✓	✓	✓	✓	✓	✓	95.49	86.26	90.28	78.93	88.22

4.4.2. Visual Comparison Analysis

The results of the c-f network outputs are compared in the quantitative analysis, as shown in Figure 8. As indicated in the second, fourth, and sixth images, the output features are somewhat filled and strengthened after the addition of FDM. Because FDM decouples the body and boundary features, different levels of features enhance the internal body

features. Meanwhile, the boundary information is denoised. Therefore, the decoupled features are improved. In contrast with direct recoupling, the connectivity region between the body and boundary is compensated after adding IERM. As shown in the first and fourth images, with some strong semantic background features appearing in body features, the background information is weakened through the adaptive information guidance. SOM mainly supervises the decoupled body and boundary features. After supervision, the water body and boundary features get more guidance. Consequently, the effect of coupling is strengthened, leading to a more accurate ability to localize the output water body boundary.

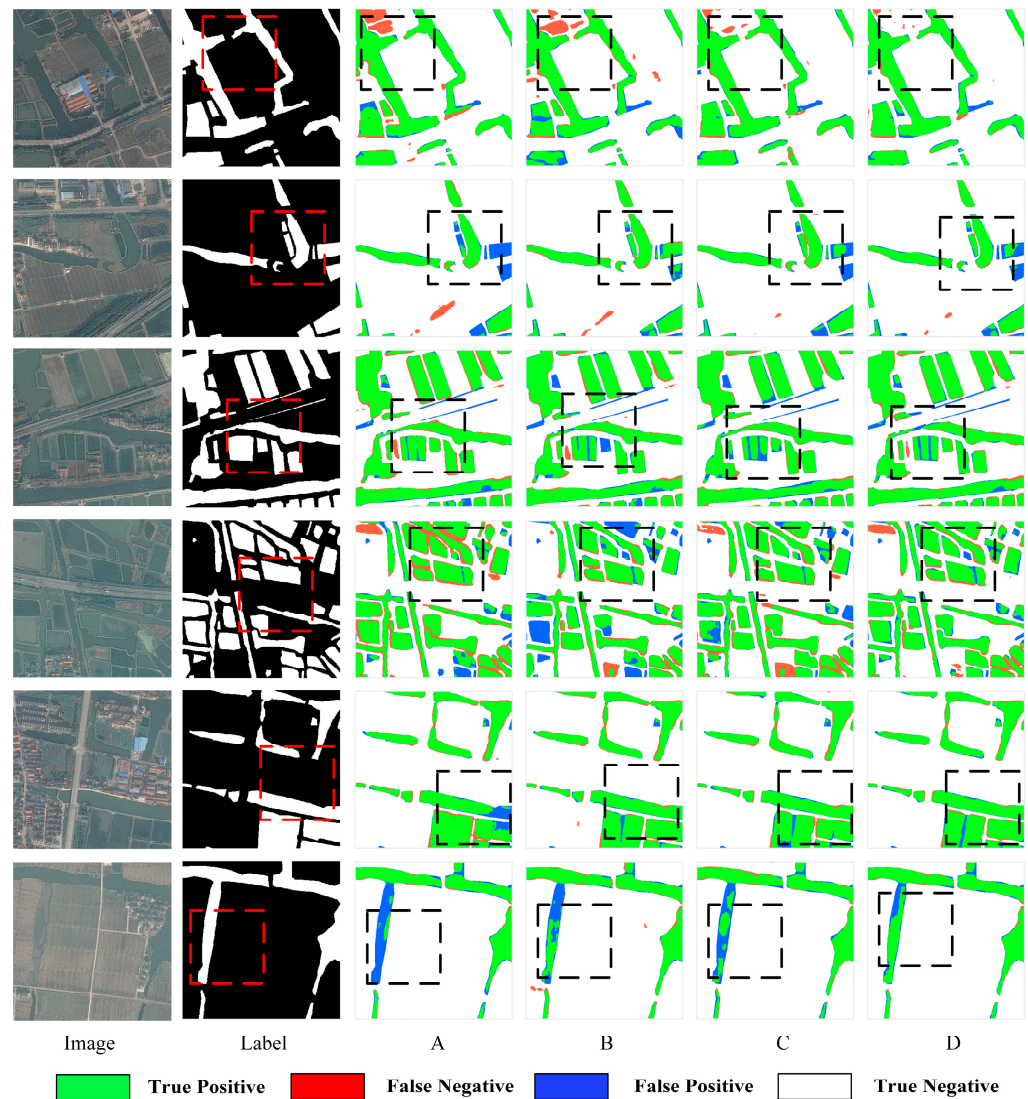


Figure 8. Comparison of the image effect of each module. (A) is the result after the output of the infrastructure containing context paths and spatial paths; (B) is the result after the output of adding FDM on A; (C) is the result after the output of adding IERM on B; (D) is the result after the output of adding SOM on C.

4.4.3. Visualization Comparison

To show more clearly how the internal features of the model change by introducing the modules, the resultant tensor in the form of the heat map results is shown in Figure 9. It can be seen that the spatial path under 1/8 resolution has richer semantic features, and the spatial path under 1/2 resolution has finer detail information. After decoupling the features under the two spatial branches, images C and D have more decisive semantic information, and the body and boundary features are more evident after decoupling supervision. Images

E and F show detailed information after decoupling more clearly as well. After coupling through the IERM, the background and body features are suppressed and strengthened.

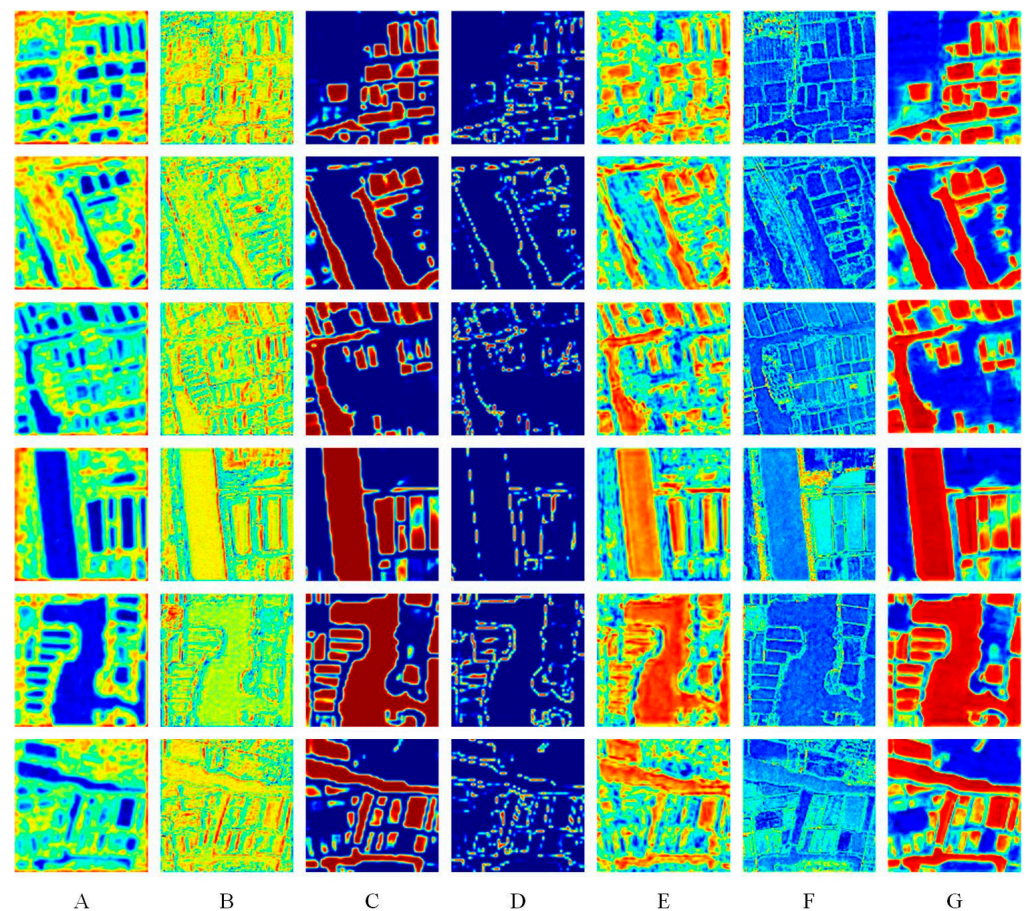


Figure 9. Visualization of the internal structure diagram of the model. The feature maps are uniformly sampled to the same resolution. (A) is the spatial path under 1/8 size; (B) is the spatial path under 1/2; (C) is the internal body feature output after decoupling supervision of subfigure label (A); (D) is the boundary feature output after decoupling supervision of subfigure label (A); (E) is the internal body feature output after decoupling supervision of subfigure label (B); (F) is the boundary feature output after decoupling supervision of subfigure label (B); (G) is the coupling feature after the output of the integrated expansion coupling module.

5. Discussion

5.1. Supervisory Weights Setting After Partition Decoupling

In the proposed method, features from different spatial partitions are decoupled to enhance or supervise the water body features according to the demand. First, each parameter is set to the same weight to estimate its approximate interval. The weight range can be set between 0.5 and 1.0 through groups A, B, and C sets. Secondly, to determine the weight relationship between different partitions and the influence change between the body and the boundary, groups D, E, F, and G experiments are set to validate.

Table 5 supervises the influence of weights. Where λ_i denotes the decoupled boundary feature, and μ_i is the body feature. When i is 1 or 2, it is denoted as Spatial P1 or Spatial P2, respectively.

From the results of groups A, B, and C, it can be seen that with the increase in the output weight information for each decoupled feature, the performance of each aspect is seen to grow. The maximum value is 0.5, and it can be judged that the optimal effect is concentrated between 0.5 and 1.0. From the experiments in groups D and E, it can be understood that shallow spatial partitioning retains richer spatial information and achieves

the best overall extraction effect compared with shallow spatial partitioning. In groups F and G, it can be seen that the boundary information has a greater degree of influence on the final effect. In summary, if more complete water body features are needed, the parameters of group D can be selected to supervise each decoupling feature, and its R-value is higher than that of other groups, ranging from 1.54% to 3.5%. If more accurate extraction of water body features is needed, group F parameters are selected, with the highest accuracy among all comparisons. If the overall segmentation effect is required, group E parameters are chosen for supervision, which achieves the optimal results in terms of overall accuracy, intersection and concurrency ratio, and F1 score.

Table 5. The influence of different weights.

Group ID	Parameter ($\lambda_2, \mu_2, \lambda_1, \mu_1$)	OA/%	P/%	R/%	IoU/%	F1/%
A	0.2, 0.2, 0.2, 0.2	95.30	85.70	89.89	78.17	87.75
B	0.5, 0.5, 0.5, 0.5	95.43	86.62	89.40	78.55	87.99
C	1.0, 1.0, 1.0, 1.0	95.37	86.42	89.31	78.32	87.84
D	1.0, 1.0, 0.5, 0.5	95.33	84.57	91.82	78.65	88.05
E	0.5, 0.5, 1.0, 1.0	95.49	86.26	90.28	78.93	88.22
F	1.0, 0.5, 1.0, 0.5	95.41	87.32	88.32	78.29	87.82
G	0.5, 1.0, 0.5, 1.0	95.34	86.49	89.05	78.17	87.75

5.2. Effectiveness of the Proposed Modules

Previous studies based on water body segmentation mainly focused on the fusion of multiple features. This fusion ignores the different requirements of the body and boundary on features, resulting in the coupling and redundancy of features. In the reinforcement and supervision of features, the expected effect cannot be achieved. For clearer comparison with previous multi-feature fusion methods and other decoupling methods. Based on the spatial partition interactive feature extraction framework proposed in this paper, we conducted comparison experiments with different multi-feature processing methods as variables, and the results are shown in Table 6.

Table 6. Comparison of effects of different fusion modules. The weight setting of SOM in different layers is based on the best results in Section 5.1.

Baseline	SOM	Methods	OA/%	P/%	R/%	IoU/%	F1/%
✓			94.94	86.32	86.73	76.25	86.52
✓	✓		94.85	83.53	90.33	76.67	86.80
✓		FDM	95.15	85.23	89.60	77.56	87.36
✓	✓	FFM [36]	95.06	84.45	90.24	77.38	87.25
✓	✓	MSFM [27]	95.16	85.36	89.47	77.57	87.37
✓	✓	IDM [56]	95.27	85.91	89.41	77.98	87.63
✓	✓	FDM	95.46	87.18	88.81	78.55	87.99

As can be seen from Table 6, SOM multi-feature constraints can improve the segmentation effect to a certain extent. To this end, we add other comparison modules in addition to SOM. Feature fusion module (FFM) [36] fuses features in a multi-feature cascade and uses attention mechanisms to reinforce features. Multi-scale fusion module (MSFM) [27] supplements the damaged information through the interaction of multiple layers of features and then fuses it. The two above multi-feature fusion methods, although there are differences in fusion timing and strengthening methods, essentially ignore the difference in features. This means the feature supervision cannot be effectively targeted, and the improvement effect is limited. The information decoupling module (IDM) [56] has a similar decoupling idea with this method through the Laplace pyramid decomposition feature and is superior to other contrast fusion methods in effect. However, IDM lacks information enhancement and background noise removal, which makes the result accuracy lower than 1.27% of the

proposed method. The FDM proposed in this paper can solve the above problems to a certain extent, and the effect is 0.57–1.17% higher than other fusion methods. In addition, through the comparison between SOM supervision directly and SOM supervision after FDM, it can be seen that the supervision effect of the decoupled body and boundary is more obvious. The method in this paper will produce more constraints on the feature decoupling supervision. With the increase in the number of decoupled feature layers, too many constraints will increase the training difficulty to some extent. However, this paper simplifies the number of feature layers by means of spatial partitioning, which alleviates the problem to some extent.

5.3. Analysis of Extraction Ability in Complex Scenarios

The completeness of lake morphology extraction and the accuracy of boundary localization are important indicators for assessing the generalization ability of the model. In this regard, this paper adds SSIM and BIoU as quantitative criteria. Some representative scenarios are selected from the dataset for discussion. The visualization results are shown in Figure 10.

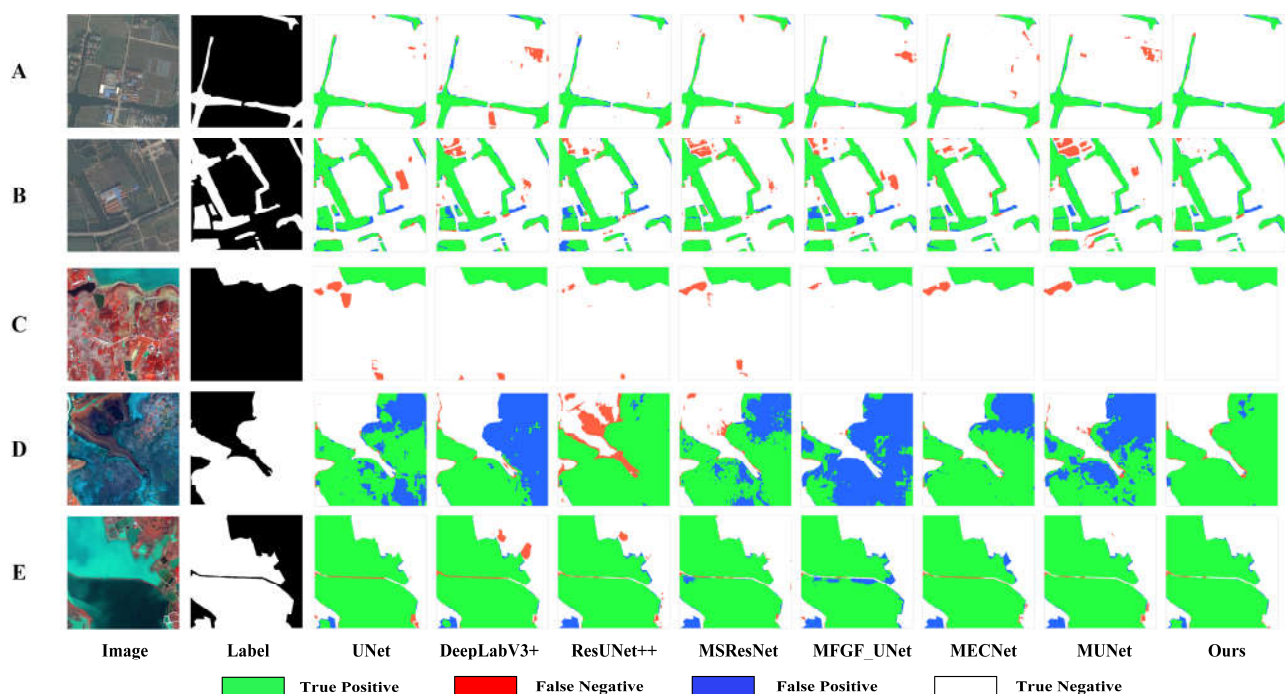


Figure 10. Boundary extraction effect of each method in complex scenes.

Scenarios A, C, and E consist of large-scale lakes or fine streams, which primarily test the algorithm's ability of multi-scale feature extraction and global context perception. Gated multi-filter inception in MFGF_UNet and multi-scale dilated convolution module in MSResNet, which consist of different scales of bar or cavity convolution, have similar effects with CMFAM in this paper, so they have a strong ability to capture fine tributaries. However, this paper retains more stable information through spatial partitioning while preserving multi-scale features, making it more effective in connecting some fine streams.

In scenarios B and D, inaccurate boundary localization effects are often produced due to shoreline erosion, spectral effects, background noise, and other disturbances. In this paper, we design FDM to separate the body from the boundary, weaken the boundary noise by semantic enhancement, and supervise the decoupled features by multi-level association. Compared with other decoding methods that aggregate multi-level features for output supervision. In the methodology of this paper, the strengthening of information

and the supervision do not produce a conflict of information, which strengthens the anti-information interference ability.

Scenarios B and E are the complex boundary regions. Although all methods can detect broader contours, none can produce better natural cuts. Since the spatial division retains detailed information and the IERM recouples the features, our method detects the water body boundary closer to the true value.

As shown in Table 7, the IoU, BIoU, and SSIM of this proposed method achieve the best results on both datasets. Among them, BIoU is 6.40% and 6.75% higher than UNet on both datasets, respectively. In order to better predict the fine contours of water bodies, MECNet adopts a multi-scale prediction fusion module to fuse the prediction results of different layers, which makes it achieve the second-best results in GID only compared with the method in this paper. MUNet combines CNN and MixFormer to model the local spatial detail information and global context information of the image to improve the network's ability to capture semantic features of water bodies. It is higher than other comparative methods in terms of internal feature capture. Unlike the multi-layer fusion approach mentioned above, this proposed approach considers the output differences of different features, decouples the features for supervision, and designs the IERM to adaptively couple the multi-layer features and fuse the body and boundary information using integrated expansion. Thus, our approach achieves the best results in IoU, BIoU, and SSIM metrics.

Table 7. Comprehensive model evaluation results.

Methods	GID			GF2020		
	IoU	BIoU	SSIM	IoU	BIoU	SSIM
U-Net	89.19	64.75	93.89	74.78	54.00	88.32
DeepLabv3+	89.54	67.52	94.27	74.40	52.15	87.96
MSResNet	90.42	67.75	94.14	77.29	58.73	88.34
Resunet++	90.20	68.28	94.36	75.99	54.63	88.65
MFGF_UNet	90.17	67.24	94.28	76.34	56.40	88.53
MUNet	90.49	68.87	94.82	77.54	59.23	89.06
MECNet	90.79	69.70	94.86	76.80	56.49	88.82
ours	91.22	71.15	95.07	78.93	60.75	89.69

5.4. Discussion of Confusion Matrix Relations

TP denotes the total number of pixels correctly predicted as water bodies, and TN denotes the total number of pixels correctly identified as background. FP denotes the total number of pixels incorrectly identifying the background as a water body. FN denotes the total number of pixels incorrectly identifying the water body as the background. The difference in the values is large because the two datasets differ in the number of samples and the percentage of water bodies, as can be seen in Figure 11. ResUNet++ and MSResNet show a strong contrast in the prediction results of the two datasets. They correctly recognize water bodies, while too many of the background elements are judged as water bodies. UNet and DeepLabV3+ are traditional deep learning algorithms, which are difficult to adapt to the environment of water body recognition, so the results are a little inferior. MECNet and the proposed method show strong stability in both data sets. Benefiting from the reinforcement and supervision of decoupled features, this paper can better distinguish the boundary and background information while extracting water body features.



Figure 11. The confusion matrix. The blue tint is the confusion matrix comparison of each algorithm on the GF2020 dataset, and the green tint is the GID. 1 to 8 are the rankings in a similar comparison. TP and TN are positively correlated, so the maximum value is 1. FP and FN are negatively correlated, so the minimum value is 1. The value in the rectangle is the total number of pixels in thousands for the corresponding situation in the prediction dataset. For better comparison, the font color of the top three is changed to white in this paper.

5.5. Analysis of Model Complexity

As shown in Table 8, the number of parameters and the calculation amount of the model are also compared to judge the model complexity. The results show that the computation amount of this paper's algorithm is only 49.24, second only to MFGF. The number of parameters and computation amount is only about 1/3 to 1/2 of MUNet, which has better comprehensive performance, indicating that our method maintains a certain degree of lightweightness while achieving excellent performance.

Table 8. Model parameters and calculations.

Methods	GID IoU	GF2020 IoU	Params (M)	GFlops (B)
U-Net	89.19	74.78	28.94	192.96
DeepLabv3+	89.54	74.40	59.34	88.81
MSResNet	90.42	77.29	69.55	121.49
Resunet++	90.20	75.99	4.06	63.20
MFGF_UNet	90.17	76.34	2.17	16.95
MUNet	90.49	77.54	23.29	172.75
MECNet	90.79	76.80	30.06	184.98
ours	91.22	78.93	9.61	49.24

5.6. Overall Analysis of the Experiment

Decoding by aggregating multiple feature layers seems to be the dominant form of water body segmentation networks. In this paper, we find that this widely adopted approach ignores the demand differences between features, which makes it impossible to effectively supervise the features at each level according to the demand and makes the differing information compete with each other in learning. In this regard, this paper proposes SPFDNet. The experimental results in Section 4.3 show that this paper achieves the best segmentation results compared to similar methods. To further verify the contribution of each module of this paper's method, a comparison experiment is conducted in Section 4.4. First of all, in the selection of the basic framework, this paper abandons the commonly used UNet framework and designs CMFAM with the multi-scale characteristics of the water body. In order to minimize the loss of the boundary information of the water body, the coding architecture is designed using the fixed-resolution interaction method. The results show that compared with UNet, the proposed infrastructure in this paper has superior performance and is lightweight. In decoding, the FDM adopts the optical flow method to separate the different features, which makes the reinforcement method more targeted and largely contributes to the model performance improvement. To set the supervisory weights according to the demand scenarios, this paper discusses the variation of supervisory weights in 5.1 to determine the best segmentation effect. It is worth mentioning that the FDM with joint supervision can be integrated into the water body segmentation network, which is based on the codec structure and can improve the reinforcement and constraints on the features. In Section 5.2, we compare with similar methods and analyze the advantages and disadvantages of the proposed method. In Section 5.3, we discuss those structures in this paper and comparative algorithms that are helpful in improving the ability to cope with complex scenes. Two metrics, SSIM and BIoU, are cited to evaluate feature completeness and boundary accuracy. By separating the different features in this paper, it can be seen that joint supervision is effective and achieves the best results. Then, we verify the stability of this paper's algorithm by comparing the content of the confusion matrices of all the algorithms on the two datasets in Section 5.4. To test the effectiveness of the algorithms in practice, we evaluate the methods lightly in Section 5.5. The results show that the methods in this paper can better maintain the balance between accuracy and efficiency.

The effect achieved in this paper depends on the data quality to a certain extent, and the lack of clouds and snow obscuring or covering the water body in the current water body dataset will reduce the effect of this paper to a certain extent when facing this scenario. Moreover, the deep learning-based water body segmentation method relies on the spatial distribution of the data. Thus, it is challenging to meet the practical requirements when dealing with images in unknown domains, especially with constantly updated massive datasets. Therefore, lightweight also limits the application scope of the model to some extent.

6. Conclusions

Deep learning-based water body extraction methods tend to ignore the differences between the interior and boundary requirements of the water body and treat them as coupled features for decoding. Thus, a SPFDNet is proposed to solve this problem. To enable the model to extract richer multi-scale features and stable spatial information, a contextual coding path is constructed based on CMFAM, and the underlying framework is composed of fixed-resolution interaction. When decoding, the different features are separated by FDM, and semantic guidance and multi-level supervision are provided to strengthen and guide the information transfer. Then, IERM compensates the transition region between the water body and the boundary to generate the output. Through comparative algorithm validation, scenario analysis, and ablation experiments, the results show that SPFDNet can make the internal features of the water body more complete and the boundary localization more accurate by monitoring and reinforcing the separated body and boundary features.

However, the results in this paper are based on the premise of having labeled samples. When confronted with unlabeled scenes across domains, the problem of insufficient model generalization ability often arises. In the future, we will explore unsupervised scenarios for shoreline extraction in water bodies to expand the application scope of shoreline detection.

Author Contributions: Conceptualization, X.C. and K.H.; methodology, X.C. and X.G.; validation, K.H. and X.X.; investigation, X.G. and W.Z.; data curation, J.X. and K.H.; writing—original draft preparation, X.C. and K.H.; writing—review and editing, X.C. and X.G.; visualization, J.X. and G.L.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Major water conservancy science and technology projects in Hunan Province, grant number XSKJ2022068-12; the Key Program of the National Natural Science Foundation of China, grant number U2240224; the National Basic Research Program of China, grant number 2023YFC3209503, 2023YFC3209502; The Special Fund of Chinese Central Government for Basic Scientific Research Operations in Commonwealth Research Institute, grant number CKSF2023313/KJ; the Key Project of the Scientific Research Plan of Hubei Provincial Department of Education, grant number D20231304; the Open Fund of National Engineering Laboratory for Digital Construction and Evaluation Technology of Urban Rail Transit, grant number 2023ZH01; the Open Fund of Key Laboratory of Mine Environmental Monitoring and Improving around Poyang Lake, Ministry of Natural Resources, grant number MEMI-2021-2022-08; the Tianjin Science and Technology Plan Project, grant number 23YFYSHZ00190, 23YFZCSN00280; the Hunan Natural Science Foundation Project Department Union Fund, grant number 2024JJ8327; and the Jiangxi Provincial Natural Science Foundation, grant number 20232ACB204032.

Data Availability Statement: Data are contained within the article.

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

References

1. Lv, Z.; Zhang, M.; Sun, W.; Benediktsson, J.A.; Lei, T.; Falco, N. Spatial-Contextual Information Utilization Framework for Land Cover Change Detection with Hyperspectral Remote Sensed Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4411911. [\[CrossRef\]](#)
2. Zhang, X.; Yu, W.; Pun, M.-O.; Shi, W. Cross-domain landslide mapping from large-scale remote sensing images using prototype-guided domain-aware progressive representation learning. *ISPRS J. Photogramm. Remote Sens.* **2023**, *197*, 1–17. [\[CrossRef\]](#)
3. Zhang, G.; Gao, X.; Yang, J.; Yang, Y.; Tan, M.; Xu, J.; Wang, Y. A multi-task driven and reconfigurable network for cloud detection in cloud-snow coexistence regions from very-high-resolution remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *114*, 103070. [\[CrossRef\]](#)
4. Song, H.; Yang, Y.; Gao, X.; Zhang, M.; Li, S.; Liu, B.; Wang, Y.; Kou, Y. Joint Classification of Hyperspectral and LiDAR Data Using Binary-Tree Transformer Network. *Remote Sens.* **2023**, *15*, 2706. [\[CrossRef\]](#)
5. Chang, J.; Gao, X.; Yang, Y.; Wang, N. Object-Oriented Building Contour Optimization Methodology for Image Classification Results via Generalized Gradient Vector Flow Snake Model. *Remote Sens.* **2021**, *13*, 2406. [\[CrossRef\]](#)
6. Lv, Z.; Liu, J.; Sun, W.; Lei, T.; Benediktsson, J.A.; Jia, X. Hierarchical Attention Feature Fusion-Based Network for Land Cover Change Detection with Homogeneous and Heterogeneous Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4411115. [\[CrossRef\]](#)
7. Yang, D.; Gao, X.; Yang, Y.; Jiang, M.; Guo, K.; Liu, B.; Li, S.; Yu, S. CSA-Net: Complex Scenarios Adaptive Network for Building Extraction for Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 938–953. [\[CrossRef\]](#)
8. Gao, X.; Zhang, G.; Yang, Y.; Kuang, J.; Han, K.; Jiang, M.; Yang, J.; Tan, M.; Liu, B. Two-Stage Domain Adaptation Based on Image and Feature Levels for Cloud Detection in Cross-Spatiotemporal Domain. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5610517. [\[CrossRef\]](#)
9. Wieland, M.; Fichtner, F.; Martinis, S.; Groth, S.; Krullikowski, C.; Plank, S.; Motagh, M. S1S2-Water: A Global Dataset for Semantic Segmentation of Water Bodies from Sentinel-1 and Sentinel-2 Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 1084–1099. [\[CrossRef\]](#)
10. Tan, J.; Tang, Y.; Liu, B.; Zhao, G.; Mu, Y.; Sun, M.; Wang, B. A Self-Adaptive Thresholding Approach for Automatic Water Extraction Using Sentinel-1 SAR Imagery Based on OTSU Algorithm and Distance Block. *Remote Sens.* **2023**, *15*, 2690. [\[CrossRef\]](#)
11. Liu, Q.; Tian, Y.; Zhang, L.; Chen, B. Urban Surface Water Mapping from VHR Images Based on Superpixel Segmentation and Target Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 5339–5356. [\[CrossRef\]](#)
12. Chen, J.; Wang, Y.; Wang, J.; Zhang, Y.; Xu, Y.; Yang, O.; Zhang, R.; Wang, J.; Wang, Z.; Lu, F.; et al. The Performance of Landsat-8 and Landsat-9 Data for Water Body Extraction Based on Various Water Indices: A Comparative Analysis. *Remote Sens.* **2024**, *16*, 1984. [\[CrossRef\]](#)

13. Hertel, V.; Chow, C.; Wani, O.; Wieland, M.; Martinis, S. Probabilistic SAR-based water segmentation with adapted Bayesian convolutional neural network. *Remote Sens. Environ.* **2023**, *285*, 113388. [\[CrossRef\]](#)
14. Li, K.; Wang, J.; Yao, J. Effectiveness of machine learning methods for water segmentation with ROI as the label: A case study of the Tuul River in Mongolia. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *103*, 102497. [\[CrossRef\]](#)
15. Isikdogan, F.; Bovik, A.C.; Passalacqua, P. Surface Water Mapping by Deep Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4909–4918. [\[CrossRef\]](#)
16. Weng, L.; Xu, Y.; Xia, M.; Zhang, Y.; Liu, J.; Xu, Y. Water Areas Segmentation from Remote Sensing Images Using a Separable Residual SegNet Network. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 256. [\[CrossRef\]](#)
17. Peña, F.J.; Hübinger, C.; Payberah, A.H.; Jaramillo, F. DeepAqua: Semantic segmentation of wetland water surfaces with SAR imagery using deep neural networks without manually annotated data. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *126*, 103624. [\[CrossRef\]](#)
18. Pires de Lima, R.; Karimzadeh, M. Model Ensemble with Dropout for Uncertainty Estimation in Sea Ice Segmentation Using Sentinel-1 SAR. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4303215. [\[CrossRef\]](#)
19. Qi, H.; Kong, X.; Cheng, L.; Hu, J.; Gu, J. Addressing Fine-Grained Lake Water Body Extraction: A Hybrid Approach Combining Vision Transformer and Geodesic Active Contour. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4204614. [\[CrossRef\]](#)
20. Wang, J.; Wang, S.; Wang, F.; Zhou, Y.; Wang, Z.; Ji, J.; Xiong, Y.; Zhao, Q. FWENet: A deep convolutional neural network for flood water body extraction based on SAR images. *Int. J. Digital Earth* **2022**, *15*, 345–361. [\[CrossRef\]](#)
21. Zhong, H.-F.; Sun, H.-M.; Han, D.-N.; Li, Z.-H.; Jia, R.-S. Lake water body extraction of optical remote sensing images based on semantic segmentation. *Appl. Intell.* **2022**, *52*, 17974–17989. [\[CrossRef\]](#)
22. Chang, J.-Y.; Xu, Z.-X. Enhanced Water Puddle Segmentation and Detection Using DCU-Net. *IEEE Internet Things J.* **2024**, *1*. [\[CrossRef\]](#)
23. Liu, W.; Chen, X.; Ran, J.; Liu, L.; Wang, Q.; Xin, L.; Li, G. LaeNet: A Novel Lightweight Multitask CNN for Automatically Extracting Lake Area and Shoreline from Remote Sensing Images. *Remote Sens.* **2020**, *13*, 56. [\[CrossRef\]](#)
24. Wang, Z.; Gao, X.; Zhang, Y. HA-Net: A Lake Water Body Extraction Network Based on Hybrid-Scale Attention and Transfer Learning. *Remote Sens.* **2021**, *13*, 4121. [\[CrossRef\]](#)
25. Zhang, X.; Li, J.; Hua, Z. MRSE-Net: Multiscale Residuals and SE-Attention Network for Water Body Segmentation From Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 5049–5064. [\[CrossRef\]](#)
26. Zhang, Z.; Liu, F.; Liu, C.; Tian, Q.; Qu, H. ACTNet: A Dual-Attention Adapter with a CNN-Transformer Network for the Semantic Segmentation of Remote Sensing Imagery. *Remote Sens.* **2023**, *15*, 2363. [\[CrossRef\]](#)
27. Dai, X.; Xia, M.; Weng, L.; Hu, K.; Lin, H.; Qian, M. Multiscale Location Attention Network for Building and Water Segmentation of Remote Sensing Image. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5609519. [\[CrossRef\]](#)
28. Xu, J.; Li, J.; Zhao, X.; Luan, K.; Yi, C.; Wang, Z. DANet-SMIW: An Improved Model for Island Waterline Segmentation Based on DANet. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 884–893. [\[CrossRef\]](#)
29. Chen, C.; Wang, Y.; Yang, S.; Ji, X.; Wang, G. A K-Net-based hybrid semantic segmentation method for extracting lake water bodies. *Eng. Appl. Artif. Intell.* **2023**, *126*, 106904. [\[CrossRef\]](#)
30. Yuan, K.; Zhuang, X.; Schaefer, G.; Feng, J.; Guan, L.; Fang, H. Deep-Learning-Based Multispectral Satellite Image Segmentation for Water Body Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7422–7434. [\[CrossRef\]](#)
31. Zhang, S.; Li, W.; Wang, R.; Liang, C.; Feng, X.; Hu, Y. DaliWS: A High-Resolution Dataset with Precise Annotations for Water Segmentation in Synthetic Aperture Radar Images. *Remote Sens.* **2024**, *16*, 720. [\[CrossRef\]](#)
32. Zhang, Y.; Yang, R.; Dai, Q.; Zhao, Y.; Xu, W.; Wang, J.; Wang, L. Boosting Semantic Segmentation of Remote Sensing Images by Introducing Edge Extraction Network and Spectral Indices. *Remote Sens.* **2023**, *15*, 5148. [\[CrossRef\]](#)
33. Ji, Y.; Wu, W.; Nie, S.; Wang, J.; Liu, S. Sea–Land Segmentation of Remote-Sensing Images with Prompt Mask-Attention. *Remote Sens.* **2024**, *16*, 3432. [\[CrossRef\]](#)
34. Hu, K.; Li, M.; Xia, M.; Lin, H. Multi-Scale Feature Aggregation Network for Water Area Segmentation. *Remote Sens.* **2022**, *14*, 206. [\[CrossRef\]](#)
35. Scala, P.; Manno, G.; Ciraolo, G. Semantic segmentation of coastal aerial/satellite images using deep learning techniques: An application to coastline detection. *Comput. Geosci.* **2024**, *192*, 105704. [\[CrossRef\]](#)
36. Lyu, X.; Jiang, W.; Li, X.; Fang, Y.; Xu, Z.; Wang, X. MSAFNet: Multiscale Successive Attention Fusion Network for Water Body Extraction of Remote Sensing Images. *Remote Sens.* **2023**, *15*, 3121. [\[CrossRef\]](#)
37. 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. [\[CrossRef\]](#)
38. Yang, R.; Zheng, C.; Wang, L.; Zhao, Y.; Fu, Z.; Dai, Q. MAE-BG: Dual-stream boundary optimization for remote sensing image semantic segmentation. *Geocarto Int.* **2023**, *38*, 2190622. [\[CrossRef\]](#)
39. Zhong, H.-F.; Sun, Q.; Sun, H.-M.; Jia, R.-S. NT-Net: A Semantic Segmentation Network for Extracting Lake Water Bodies from Optical Remote Sensing Images Based on Transformer. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5627513. [\[CrossRef\]](#)
40. Cheng, B.; Wei, Y.; Feris, R.; Xiong, J.; Hwu, W.M.; Huang, T.; Shi, H. Decoupled classification refinement: Hard false positive suppression for object detection. *arXiv* **2020**, arXiv:1810.04002.
41. 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. [\[CrossRef\]](#)

42. Weng, Y.; Li, Z.; Tang, G.; Wang, Y. OCNet-Based Water Body Extraction from Remote Sensing Images. *Water* **2023**, *15*, 3557. [\[CrossRef\]](#)
43. 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. [\[CrossRef\]](#)
44. Dang, B.; Li, Y. MSResNet: Multiscale Residual Network via Self-Supervised Learning for Water-Body Detection in Remote Sensing Imagery. *Remote Sens.* **2021**, *13*, 3122. [\[CrossRef\]](#)
45. Lyu, X.; Fang, Y.; Tong, B.; Li, X.; Zeng, T. Multiscale Normalization Attention Network for Water Body Extraction from Remote Sensing Imagery. *Remote Sens.* **2022**, *14*, 4983. [\[CrossRef\]](#)
46. Ding, L.; Tang, H.; Bruzzone, L. LANet: Local Attention Embedding to Improve the Semantic Segmentation of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 426–435. [\[CrossRef\]](#)
47. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; et al. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 3349–3364. [\[CrossRef\]](#)
48. 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. [\[CrossRef\]](#)
49. Wang, W.; Su, C. Semi-supervised learning for efficient water leakage segmentation in tunnel infrastructure. *Struct. Health Monit.* **2024**. [\[CrossRef\]](#)
50. Fu, C.; Li, M.; Zhang, B.; Wang, H. TBiSeg: A transformer-based network with bi-level routing attention for inland waterway segmentation. *Ocean Eng.* **2024**, *311*, 119011. [\[CrossRef\]](#)
51. Zhao, W.; Xia, M.; Weng, L.; Hu, K.; Lin, H.; Zhang, Y.; Liu, Z. SPNet: Dual-Branch Network with Spatial Supplementary Information for Building and Water Segmentation of Remote Sensing Images. *Remote Sens.* **2024**, *16*, 3161. [\[CrossRef\]](#)
52. Turkmenli, I.; Aptoula, E.; Kayabol, K. HistSegNet: Histogram Layered Segmentation Network for SAR Image-Based Flood Segmentation. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 4014705. [\[CrossRef\]](#)
53. Wang, J.; Jia, D.; Xue, J.; Wu, Z.; Song, W. Automatic Water Body Extraction from SAR Images Based on MADF-Net. *Remote Sens.* **2024**, *16*, 3419. [\[CrossRef\]](#)
54. Li, X.; Li, X.; Zhang, L.; Cheng, G.; Shi, J.; Lin, Z.; Tan, S.; Tong, Y. Improving Semantic Segmentation via Decoupled Body and Edge Supervision. In *Computer Vision—ECCV 2020; Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2020; pp. 435–452.
55. Li, X.; Lv, S.; Zhang, J.; Li, M.; Rodriguez-Andina, J.J.; Qin, Y.; Yin, S.; Luo, H. FDGR-Net: Feature Decouple and Gated Recalibration Network for medical image landmark detection. *Expert Syst. Appl.* **2024**, *238*, 121746. [\[CrossRef\]](#)
56. Su, Y.; Cheng, J.; Zhong, C.; Zhang, Y.; Ye, J.; He, J.; Liu, J. FeDNet: Feature Decoupled Network for polyp segmentation from endoscopy images. *Biomed. Signal Process. Control* **2023**, *83*, 104699. [\[CrossRef\]](#)
57. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
58. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
59. Tong, X.-Y.; Xia, G.-S.; Lu, Q.; Shen, H.; Li, S.; You, S.; Zhang, L. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sens. Environ.* **2020**, *237*, 111322. [\[CrossRef\]](#)
60. Sun, X.; Wang, P.; Yan, Z.; Diao, W.; Lu, X.; Yang, Z.; Zhang, Y.; Xiang, D.; Yan, C.; Guo, J.; et al. Automated High-Resolution Earth Observation Image Interpretation: Outcome of the 2020 Gaofen Challenge. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 8922–8940. [\[CrossRef\]](#)
61. Cheng, B.; Girshick, R.; Dollar, P.; Berg, A.C.; Kirillov, A. Boundary IoU: Improving Object-Centric Image Segmentation Evaluation. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 15329–15337.
62. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [\[CrossRef\]](#)
63. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015; Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2015; pp. 234–241.
64. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Computer Vision—ECCV 2018; Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2018; pp. 833–851.
65. 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. *arXiv* **2019**, arXiv:1911.07067.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.