remote sensing

MDPI

Article

MEF-TransUNet: A Newly Developed Remote Sensing Detection Model for Micro Water Body Targets

Yongkang Yu [1], Sijia Li [2], Xingming Zheng [2], Kai Li [1] and Jianhua Ren [1,*]

[1] College of Geographical Science, Harbin Normal University, Harbin 150025, China; 2024301062@stu.hrbnu.edu.cn (Y.Y.); likai@stu.hrbnu.edu.cn (K.L.)
[2] Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China; lisj983@nenu.edu.cn (S.L.); zhengxingming@iga.ac.cn (X.Z.)
* Correspondence: renjianhua86@hrbnu.edu.cn

Highlights

What are the main findings?

- A MEF-TransUNet model was developed for fine-scale extraction of micro water bodies, integrating multi-scale enhancement and transformer-based feature modeling.
- The proposed MEGA and HLFD modules effectively repair topological fractures and suppress shadow-induced noise, achieving 91.74% precision and a 90.07% F1-score on PlanetScope imagery.

What are the implications of the main findings?

- The method significantly improves boundary connectivity and structural integrity of micro water bodies compared with DeepLabV3+ and U-Net variants.
- This approach provides a robust solution for high-resolution mapping of small and fragmented water bodies in complex surface environments.

Abstract

Micro water bodies are essential to regional ecosystems but are difficult to extract from high-resolution remote sensing images due to fragmentation and building shadows. To address edge breakage and high false-alarm rates in existing semantic segmentation models, this study proposes MEF-TransUNet, an improved TransUNet-based model for fine micro water body extraction. The model integrates a multi-scale edge-guided attention module (MEGA), a high–low-frequency decomposition fusion module (HLFD), and a convolutional block attention module (CBAM). Specifically, MEGA extracts edge priors using a Laplacian pyramid to repair topological breaks in slender water bodies. HLFD uses frequency-domain decoupling to suppress high-frequency background noise and reduce confusion between water bodies and shadows. CBAM enhances channel and spatial feature attention. Experiments using PlanetScope images from the Songhuajiang River Basin in Daqing City of the Heilongjiang Province in China showed that MEF-TransUNet achieves 91.74% precision, a 90.07% F1-score, a recall of 90.22%, and a B-IoU of 43.88%. For the GID dataset, the model attains a precision of 91.85%, an F1-score of 91.48%, a recall of 92.01%, and a B-IoU of 55.42%. Its overall performance clearly outperforms DeepLabV3+, SegFormer, U-Net, AttenUNet, and UNet++, enabling accurate micro water body localization, high output purity, and reduced manual correction costs, thus supporting fine water resource management in complex surface environments.

Check for updates

Academic Editors: Nan Xu, Xin Li, Junfeng Xiong and Linyang Li

Received: 5 March 2026
Revised: 25 April 2026
Accepted: 14 May 2026
Published: 17 May 2026

Copyright: © 2026 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.

Keywords: micro water bodies; deep learning; edge guidance; frequency-domain decomposition; PlanetScope imagery

1. Introduction

Water bodies are vital to regional ecosystems and play key roles in hydrological regulation, biodiversity maintenance, wetland preservation, ecological security and socio-economic development [1]. Micro water bodies, such as ponds, small lakes, low-order streams, ditches, and springs, are closely linked to daily human life and influence living environments. Although micro water bodies are small in area compared with large lakes and rivers, they are numerous and widely distributed, providing irreplaceable functions in flood control, groundwater recharge, habitat provision, and ecological network maintenance [2]. In rapidly urbanizing and industrializing regions, land-use and environmental changes drive the spatiotemporal evolution of surface water, especially small water bodies. These changes are highly sensitive to surface runoff, ecological patterns, and water quality. Even minor expansions or losses can indicate human activity intensity and may negatively affect urban climate, or trigger flooding and waterlogging [3,4]. Therefore, high-precision identification and dynamic monitoring of micro water bodies are essential for regional ecological protection, fine water resource management, and urban planning decisions [5].

Optical remote sensing provides an effective data source for water body extraction, which has been widely applied in water body monitoring, flood assessment, wetland mapping, and lake change detection [6,7] based on the spectral index method and image classification method [8]. Specifically, the spectral index method does not require training samples; it mainly enhances the contrast between water and land by threshold processing of water body index images to extract water bodies [9]. Common spectral indices include the Normalized Difference Water Index (NDWI) [10], the Modified Normalized Difference Water Index (MNDWI) [11], and the Automated Water Extraction Index (AWEI) [12]. Although appropriate threshold selection can maximize the contrast between water and land, the threshold varies with circumstances and has high uncertainty [13,14]. Moreover, the index values of low albedo targets (such as shadows) and high albedo targets (such as clouds and densely built-up areas) are often similar to or higher than those of water bodies, which can lead to serious misclassification in complex urban scenes [15]. Due to environmental influences and the existence of mixed pixels, the index values of small water bodies are also usually lower than those of pure water bodies, making threshold optimization difficult [16–18].

Supervised image classification based on machine learning provides a direct and scalable technical approach for water body extraction. Under the premise of having a sufficient number of representative training samples, such methods can significantly improve the accuracy and robustness of water body recognition in complex surface scenes. DeepLabV3+ [19,20] uses Atrous Spatial Pyramid Pooling (ASPP) to aggregate multi-scale context, improving water-shadow discrimination and suppressing background noise. However, ASPP downsampling reduces spatial resolution, causing breaks in narrow rivers. U-Net [21] uses an encoder–decoder architecture with skip connections to fuse shallow spatial details and deep semantic features, effectively preserving boundaries and shape consistency of small water bodies. However, its local receptive field limits global context modeling, making it prone to misclassifying dark bare soil or vegetation shadows. To break through the inherent modeling limitations of CNNs, recent research has gradually shifted toward hybrid architectures that integrate convolution and self-attention mechanisms. For example, TransUNet [22] embeds a Transformer encoder into the U-Net decoding path to

capture long-range dependencies and global context, improving segmentation accuracy and preserving water network topology, yet it struggles with sub-pixel targets, high-frequency textures, and sharp edges. SegFormer [23], a pure Transformer architecture, uses a hierarchical encoder and lightweight pointwise decoder to enhance multi-scale feature learning and global semantic integration, achieving strong generalization in complex heterogeneous scenes. However, its decoder lacks explicit spatial constraints for fine-grained structures, causing edge blurring, detail loss, and topological breaks, especially for fragmented, tortuous, or sub-pixel water bodies. To address these challenges, several U-Net variants tailored for water body extraction have been proposed. Zhou et al. [24] introduced D-LinkNet, embedding hybrid dilated convolutions at the encoder–decoder bottleneck to enlarge the receptive field without sacrificing spatial detail. Fu et al. [25] proposed DANet, using dual position and channel attention to adaptively fuse global semantics, thereby improving spatial consistency and contextual coherence. Li et al. [26] designed DeepUNet with Plus and U-shaped connections forming dense blocks, strengthening fine-boundary modeling. To reduce shadow-dark-land spectral confusion, Li et al. [27] developed SU-Net, incorporating S-shaped cyclic connections for enhanced feature reuse and GLCM-derived texture statistics for better class discrimination. Liu et al. [28] proposed MSFENet, deploying dense dilated convolutions across encoder levels and integrating pixel-level contrastive learning to jointly optimize large and micro water body extraction under limited samples. Yet despite these advances, they remain sensitive to data distribution and scene configuration, falling short of a universal solution that simultaneously ensures high localization accuracy, boundary integrity, and topological connectivity. It should be noted that in addition to the wide application of traditional optical remote sensing data, microwave remote sensing technology has also increasingly attracted academic attention and has become an important complementary approach for water body mapping research. Among them, microwave data obtained through Global Navigation Satellite System Reflectometry (GNSS-R) technology, when integrated with improved deep learning methods such as GA-LinkNet [29], BARTs [30], and DiffWater [31], has achieved significant progress in water body monitoring under complex surface environments. However, due to the spatial resolution limitations of current GNSS-R data products, this technology still faces bottlenecks in fine identification and boundary demarcation of small water bodies such as ditches and small ponds.

To balance multiple feature advantages and address the difficulty of achieving both denoising and fidelity, this study develops a micro water body extraction model based on an optimized TransUNet to overcome background confusion and target omission. The technical framework is shown in Figure 1. A multi-dimensional collaborative feature enhancement strategy is constructed by integrating a multi-scale edge-guided attention module (MEGA) [32], a high–low-frequency decomposition fusion module (HLFD) [33], and a convolutional block attention module (CBAM) [34]. This strategy combines edge constraints, frequency-domain analysis, and attention mechanisms to suppress complex background noise such as building shadows while enhancing sensitivity to weak and fine water signals. Comparative experiments with different deep learning models [35,36] further verify the model's generalization ability and practical value in complex scenes. The proposed approach aims to improve the recognition accuracy of concealed water bodies in high-resolution remote sensing images and resolve edge fracture and shadow misclassification, demonstrating the collaborative effectiveness of feature enhancement modules and the advantages of fine-grained segmentation and providing a high-precision and low-cost solution for large-scale micro water body monitoring.

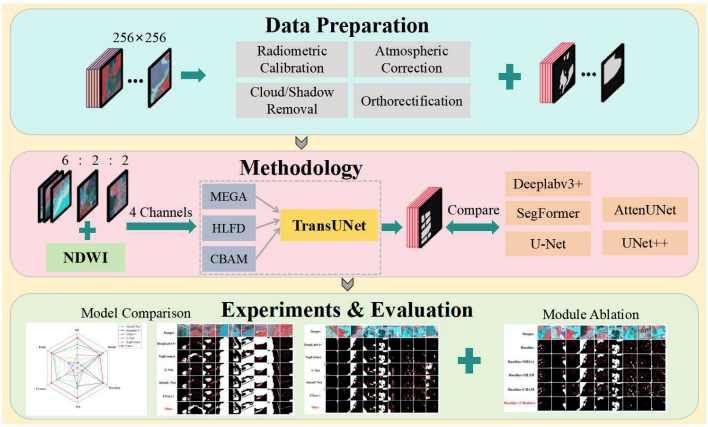


Figure 1. Technical flowchart used in this study.

2. Materials and Methods

2.1. The Self-Built Dataset

This self-built dataset covers the Songhua River Basin in Daqing City, Heilongjiang Province, China (Figure 2), which is a typical area in the Songnen Plain, characterized by flat terrain, a dense water network, concentrated lakes, swamps, and artificial water bodies. Its hydrological environment is highly complex: numerous small, scattered, morphologically diverse water bodies intermingle with wetland vegetation, irrigated farmland, and urban land, resulting in high spatial heterogeneity and ambiguous boundaries. Under the temperate continental monsoon climate, water bodies exhibit strong seasonality (expanding in summer and freezing in winter) and are further affected by high evaporation and human activities. This wetland-water composite landscape thus offers a representative benchmark for evaluating fine-scale micro water body extraction and robustness against background interference [37–41]. The dataset uses multispectral PlanetScope imagery acquired on 4 September 2021 (3 m resolution; blue, green, red, near-infrared bands) [42]. Preprocessing included radiometric calibration, atmospheric correction to surface reflectance, cloud/shadow removal, spatial cropping, multi-temporal co-registration, and band normalization [43–46]. Pixel-level labels were manually drawn using spectral-spatial features to precisely delineate water while excluding vegetation shadows and spectrally similar dark bare soil. Images were then cropped to 256 × 256 pixels, yielding a standardized binary classification dataset of 9232 samples, split into training, validation, and test sets at a 6:2:2 ratio.

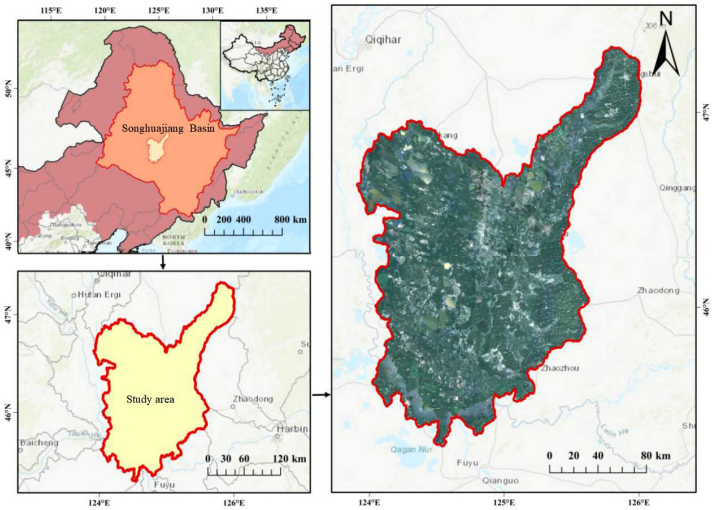


Figure 2. Study area.

2.2. The GID Public Dataset

To verify the generalization ability of the proposed model, an external public GID dataset was introduced to evaluate its cross-dataset adaptability [47]. This dataset is publicly available via Science Data Bank (https://doi.org/10.57760/sciencedb.18218 (accessed on 13 May 2026)). Specifically, the GID dataset consists of high-resolution satellite images with a spatial resolution of 0.8 m, covering diverse regions such as the North China Plain, Northeast Plain, the middle and lower reaches of the Yangtze River, and Southwestern China. Due to differences in image acquisition time and region, its data distribution is substantially different from the self-built dataset, making it suitable for evaluating cross-regional generalization ability. Since this study focuses on water body extraction, only the water body category in the GID dataset was retained and converted into a binary label, with water bodies regarded as positive samples and other categories as background. A dataset containing 27,300 512 × 512-pixel images was obtained and divided into training, validation, and test sets in a 6:2:2 ratio consistent with the self-built dataset.

2.3. Construction of MEF-TransUNet Model

Although TransUNet combines CNN local feature extraction with Vision Transformer (ViT) long-range dependency capture, it struggles with spectral recognition and high-frequency detail recovery for micro water bodies, causing small river fractures or shadow confusion. To address this, this study proposes MEF-TransUNet, an enhanced segmentation network based on TransUNet for fine micro water body extraction. Four-channel (Green, Blue, NIR, NDWI) inputs were used to enhance spectral saliency and improve target recognition. The overall structure is shown in Figure 3. Before the decoder outputs feature maps to the final segmentation head, three enhancement modules—MEGA, HLFD, and CBAM—are sequentially applied to improve boundary and detail capture. In the encoder, ResNet-50 [48] is used as the backbone, with the first convolution layer adjusted for four-channel input. Multi-level features are extracted, and deep C4 feature maps are flattened and sequenced before being sent to a 12-layer ViT module, where self-attention captures long-range dependencies. In the decoder, cascaded upsampling gradually restores spatial resolution. A selective feature fusion strategy concatenates upsampled deep semantic features only with C4, C3, and C2 encoder stages (1/16, 1/8, 1/4 resolution) and discards C1 shallow features (1/2 resolution) rich in high-frequency noise, enabling robust detail reconstruction while suppressing noise.

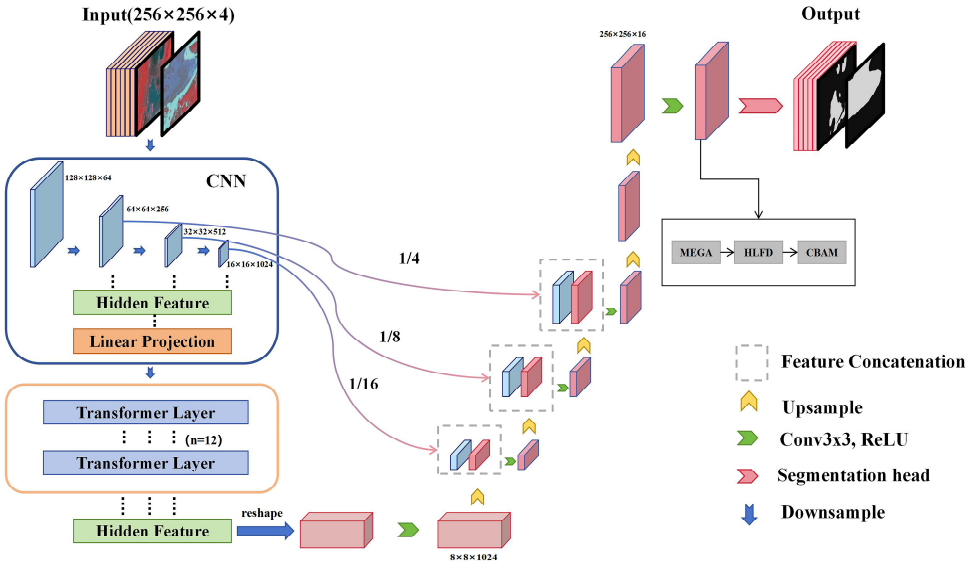


Figure 3. Structure diagram of MEF-TransUNet model.

2.4. Decoder Enhancement Module

In order to reconstruct the fine structure of micro water bodies without introducing shallow noise, three complementary enhancement modules were also connected at the decoder end.

2.4.1. Multi-Scale Edge-Guided Attention Module (MEGA)

To overcome the limitation of defining micro water body boundaries using only one feature source, the multi-scale edge-guided attention module (MEGA) with a three-way parallel attention mechanism was proposed (Figure 4). MEGA takes three inputs: the current feature map F_{in}, the Sigmoid-activated prediction map P, and the original image I. From I, it extracts high-frequency texture E_{org} from the first Laplacian pyramid layer [49] and upsamples it to recover detail lost during downsampling. From P, it generates a background mask M_{bg} (to suppress non-water areas) and a boundary mask M_{edge} (to enhance water contours) via reverse Laplacian operations and edge detection. These three signals are then separately weighted and fused with F_{in} to produce enhanced features, and the generation of feature components can be formalized as:

$$F_{bg} = F_{in} \odot (1 - \sigma(P)) \tag{1}$$

$$F_{edge} = F_{in} \odot L(\sigma(P)) \tag{2}$$

$$F_{org} = F_{in} \odot UP(E_{org}) \tag{3}$$

where σ(.) represents the Sigmoid activation function, F_{bg} represents background features, F_{edge} represents boundary features, F_{org} represents edge enhancement features, L(.) represents the Laplacian operator, ⊙ represents element-wise multiplication, and UP represents the upsampling operation. These three features are concatenated along the channel dimension and processed through convolutional layers to generate a unified attention map A. This attention map dynamically reweighs the features, which are then refined via residual addition with the original input and further enhanced by the CBAM (mentioned later). The calculation formula of F_{MEGA} is as follows:

$$F_{MEGA} = CBAM\left(F_{in} + A \odot Conv\left(\left[F_{bg}, F_{edge}, F_{org}\right]\right)\right) \tag{4}$$

where F_{MEGA} represents the final output of the module, and Conv represents the convolution operation.

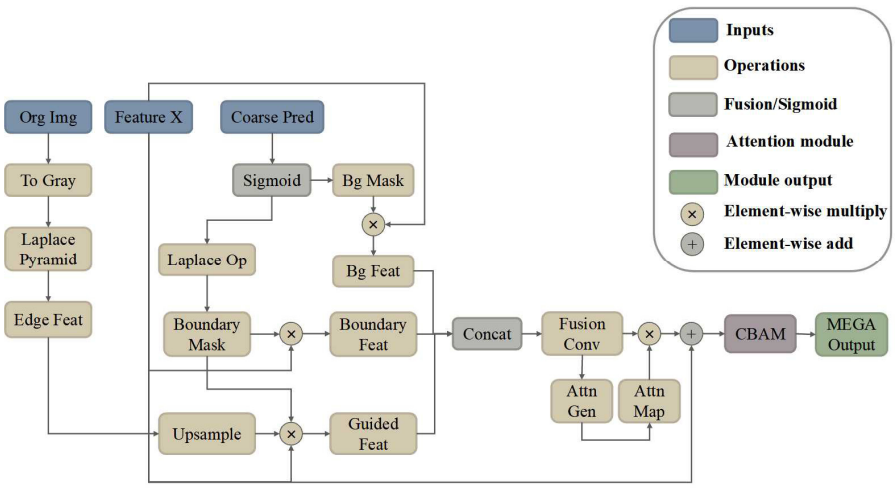


Figure 4. Structure diagram of MEGA module.

2.4.2. High–Low-Frequency Decomposition Fusion Module (HLFD)

Given that micro water bodies are prone to losing high-frequency details during deep network downsampling, HLFD introduces a frequency-domain analysis perspective and uses a decoupling independent enhancement adaptive fusion strategy to reconstruct features, as shown in Figure 5. The input features are first extracted as low-frequency overview components F_{low} through average pooling operation (AvgPool), and high-frequency detail components F_{high} containing edge textures are separated through interpolation and subtraction to achieve preliminary decoupling in the frequency domain. This decomposition process can be defined as:

$$F_{low} = AvgPool \odot (F_{in}) \quad (5)$$

$$F_{high} = F_{in} - UP(F_{low}) \quad (6)$$

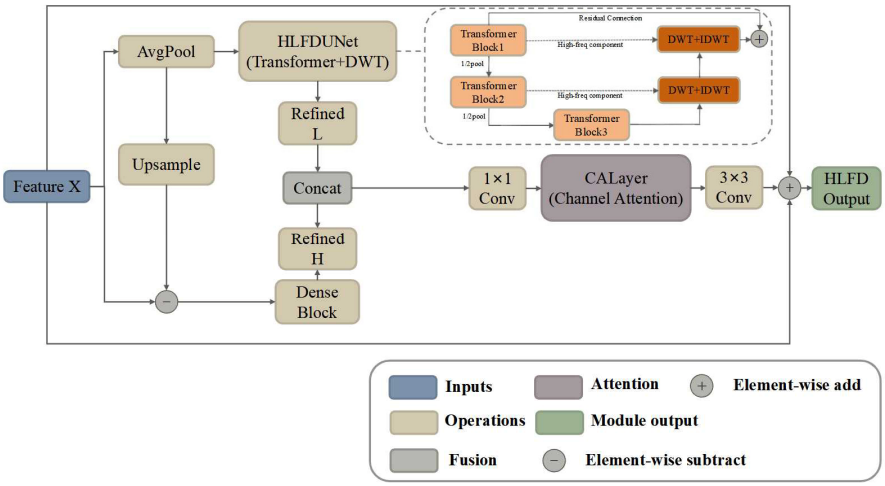


Figure 5. Structure diagram of HLFD module.

The low-frequency components are enhanced by a DWT–Transformer subnetwork that uses Transformer Blocks to establish long-range dependencies for precise localization of water target structures, and Discrete Wavelet Transform and Inverse Transform (DWT/IDWT) are combined in skip connections to achieve lossless detail compensation, yielding F'_{low}. The high-frequency components pass through dense convolutional blocks to improve gradient flow and feature reuse for fine textures, producing F'_{high}. These features are concatenated along the channel dimension, then processed by a 1×1 convolution and a channel attention layer (CALayer) [50], which adaptively weights channels to balance high- and low-frequency contributions in the final output. The reconstructed feature is denoted as F_{HLFD} from the following equation:

$$F_{HLFD} = F_{in} + f_{3\times3}\left(CALayer\left(f_{1\times1}\left(\left[F'_{high}, UP(F'_{low})\right]\right)\right)\right) \quad (7)$$

where F_{HLFD} represents the output features of the module, F'_{low} is the processed low-frequency features, F'_{high} represents the enhanced high-frequency features, and f refers to the convolution operation.

2.4.3. Convolutional Block Attention Module (CBAM)

To further improve feature map sensitivity to water targets and suppress background noise, CBAM was integrated at the end of the feature enhancement path (Figure 6). This module follows a serial attention design to adaptively recalibrate features. The channel attention submodule applies global average pooling (AvgPool) and global maximum pool-

ing (MaxPool) in parallel to the input feature map F, generating spatial context descriptors. These are fed into a shared MLP to model inter-channel dependencies, producing a channel weight vector M_c that recalibrates F along the channel dimension to yield F′:

$$M_C(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (8)$$

$$F' = M_c(F) \odot F \quad (9)$$

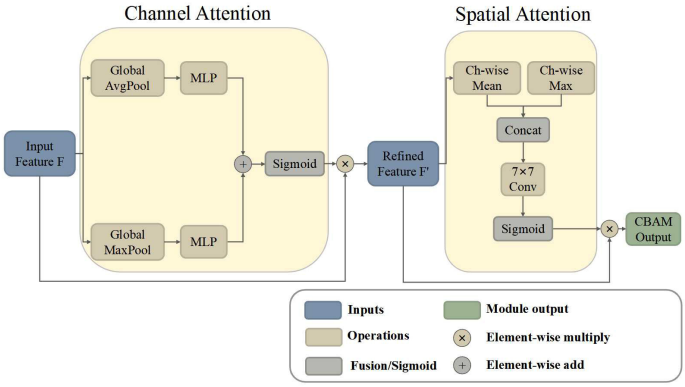


Figure 6. Structure diagram of CBAM.

Subsequently, the channel-calibrated feature maps F′ undergo average and max pooling along the channel dimension to generate 2D spatial descriptors. A large-kernel convolution then aggregates local spatial information to produce the spatial attention map M_s, which is applied to F′ to yield the final refined output F_{CBAM}:

$$M_S(F') = \sigma(f_{7\times7}([AvgPool(F'), MaxPOOL(F')])) \quad (10)$$

$$F_{CBAM} = M_s(F') \odot F' \quad (11)$$

Through these steps, the model achieves pixel-level localization of water targets while robustly suppressing irrelevant background interference.

2.5. Model Training, Testing, and Evaluation

The model is trained using supervised learning. To address the class imbalance between water and non-water pixels in remote sensing imagery, a hybrid loss function was adopted combining Binary Cross-Entropy (BCE) and Dice loss to balance pixel-level classification accuracy and overall region overlap [51–53]. The BCE loss is defined as:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \quad (12)$$

where y_i and p_i respectively represent the true label and predicted probability of the i-th pixel. Dice loss is used to measure the degree of overlap between the predicted area and the actual water body area, with its form derived from Equation (13):

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} p_i y_i + \varepsilon}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} y_i + \varepsilon} \quad (13)$$

where ε is a small constant preventing division by zero. BCE and Dice losses are jointly optimized via the following weighted summation:

$$L = \lambda L_{BCE} + (1 - \lambda)L_{Dice} \quad (14)$$

where λ is a balancing hyperparameter, typically set to 0.5.

The model was trained using the Adam optimizer [54] with an initial learning rate of 1×10^{-4} and a weight decay of 1×10^{-3}. The learning rate follows cosine annealing, dropping to a minimum of 1×10^{-5} to accelerate early convergence and stabilize late-stage optimization. In addition, batch size and training runs in this study were set at 8 and 80 epochs. To improve generalization and prevent overfitting, rigorous data augmentation was applied: geometric transformations (random 90° rotation, horizontal/vertical flipping) and color jitter (random hue, saturation, brightness, and contrast perturbations) to simulate lighting and sensor variability. Each epoch was evaluated using IoU, precision, and recall on the validation set. Model weights are saved at the epoch with the lowest validation loss; if the loss improves after epoch 60, those weights are retained as the final checkpoint, ensuring optimal performance without compromising stability. Note that the entire training process is conducted in a GPU parallel environment to improve the efficiency of large-scale high-resolution remote sensing image processing.

To evaluate generalization, the model was tested using the checkpoint with the lowest validation loss and set to evaluation mode, where BatchNorm layers use training-computed moving statistics to ensure deterministic inference. Test preprocessing mirrors training but omits random augmentations to preserve spatial integrity. Inference disables gradients for efficiency and a lower memory footprint. Outputs are passed through Sigmoid and thresholded at 0.65 to produce binary segmentation masks. Final metrics, including IoU, recall, precision, overall accuracy, F1-score, and B-IoU, are averaged over the test set to quantify accuracy and consistency in micro water body extraction.

2.6. Comparative Experiment

This study conducted comparative experiments on both the self-built and the GID public datasets. Comparison models, such as U-Net variants (U-Net, AttenUNet, UNet++), DeepLabV3+, and the Transformer-based SegFormer, were introduced to evaluate MEF-TransUNet for micro water body identification. This assesses both architectural improvements over the U-Net backbone and cross-architecture performance.

2.6.1. Accuracy Evaluation

Classification results were verified using quantitative evaluation and qualitative analysis. Quantitative metrics include overall accuracy (OA), precision, recall, F1-score, intersection over union (IoU), and the boundary intersection–union ratio (B-IoU). OA measures the overall pixel-wise classification accuracy; precision and recall indicate the model's accuracy and completeness in identifying water; F1-score, as their harmonic mean, reflects overall recognition performance; and IoU evaluates spatial consistency between predicted and actual water body extents, serving as a core metric in semantic segmentation. The calculation formulas for these indicators are as follows:

$$\mathrm{OA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{15}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{16}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{17}$$

$$\mathrm{F1} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{18}$$

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{19}$$

$$B - IoU = \frac{TP_d}{TP_d + FP_d + FN_d} \quad (20)$$

where TP is true water pixels correctly predicted, TN represents true non-water pixels correctly predicted, FP refers to non-water pixels falsely predicted as water, and FN indicates water pixels falsely predicted as non-water.

2.6.2. Model Complexity Evaluation

To comprehensively evaluate the feasibility of water body extraction models in practical remote sensing applications, a model complexity assessment system was constructed in this study, including model parameters (Params), floating-point operations (FLOPs), and inference speed (FPS). Specifically, Params refer to the total number of learnable variables in the network, measured in millions (M), reflecting the model's storage requirements and representational capacity; FLOPs denote the number of floating-point operations required for a single forward pass on a fixed-resolution image, measured in gigaflops (G), indicating the theoretical computational burden of the architecture; FPS refers to the number of image frames that the model can process per second under uniform hardware and input size, directly measuring deployment efficiency. A comprehensive analysis of these three aspects enables an objective judgment of whether the model can enhance the accuracy of water body refinement while balancing computational efficiency and practicality.

2.6.3. Visual Evaluation

To examine MEF-TransUNet's spatial applicability, typical areas were selected in this study for visual comparison of classification results. The model's ability to delineate micro water body boundaries, preserve morphology, and distinguish them from surrounding land, along with its detection and misjudgment performance, was analyzed. Furthermore, the stability of MEF-TransUNet under conditions of boundary clarity, noise, and complex backgrounds was also evaluated by comparing with other deep learning models mentioned above.

2.7. Ablation Experiment

This study conducted ablation experiments on both the self-built and the GID public datasets. To analyze MEF-TransUNet's improvement strategy and verify MEGA, HLFD, and CBAM effectiveness, ablation experiments were designed to isolate and quantify each module's performance. Starting from the TransUNet baseline, four progressive experimental schemes were constructed. Specifically, MEGA, HLFD, and CBAM were independently embedded to test edge guidance, frequency-domain denoising, and attention enhancement for addressing water fragmentation, background confusion, and weak signals. Afterward, the three modules were integrated to form the final MEF-TransUNet (ours) to evaluate overall fine-grained segmentation improvement via multi-dimensional feature collaboration. All ablation experiments were performed under identical training conditions and parameter settings to ensure fair and comparable results.

3. Results

3.1. Results on the Self-Built Dataset

3.1.1. Comparison of Accuracy Indicators

Table 1 summarizes the accuracy metrics of all models. The U-Net series models achieve high recall but lower precision in small water body extraction: U-Net (recall 89.99%, precision 88.82%, IoU 81.70%, B-IoU 28.09%) detects well but struggles with boundaries and false positives. AttenUNet and UNet++ improve across all metrics (recall 90.82%/91.02%, IoU 82.47%/82.79%, B-IoU 42.53%/42.50%) but still produce some false positives and

ambiguous boundaries. DeepLabV3+ attains the highest precision (91.71%) but suffers from lower recall (85.87%), IoU (80.86%), and B-IoU (31.53%), indicating missed small water bodies and poor boundary detail. SegFormer delivers balanced results (IoU 81.28%, recall 88.62%, precision 89.38%, OA 95.79%, F1 88.46%, B-IoU 39.14) but limited boundary refinement. In contrast, MEF-TransUNet achieves the best overall performance (precision 91.74%, recall 90.22%, IoU 83.48%, OA 96.42%, F1 90.07%, B-IoU 43.88), demonstrating superior segmentation accuracy, stronger boundary consistency, and enhanced detection of small water bodies.

Table 1. Evaluation of model accuracy indicators on self-built dataset.

Method	IoU (%)	Recall (%)	Precision (%)	OA (%)	F1-Score (%)	B-IoU (%)
DeepLabV3+	80.86	85.87	91.71	96.12	88.29	31.53
SegFormer	81.28	88.62	89.38	95.79	88.46	39.14
U-Net	81.70	89.99	88.82	95.85	88.87	28.09
AttenUNet	82.47	90.82	88.86	96.05	89.43	42.53
UNet++	82.79	91.02	89.00	96.19	89.61	42.50
Ours	83.48	90.22	91.74	96.42	90.07	43.88

Figure 7 shows the radar chart of the accuracy indicators in Table 1 after 0–1 standardization processing. From the figure, it can be seen that U-Net, AttenUNet, and UNet++ exhibit expansion along the recall axis and contraction along the precision axis, indicating strong detection capability but limited false-positive control. U-Net shows a pronounced contraction in B-IoU, reflecting insufficient boundary representation, while AttenUNet and UNet++ show moderate improvement in B-IoU, indicating better boundary accuracy. DeepLabV3+ differs from the U-Net series, showing higher precision but lower recall and IoU, with relatively weak B-IoU, indicating fewer false positives but more missed detections and limited boundary detail representation. SegFormer shows balanced but unremarkable performance, with only slight improvement in B-IoU. In contrast, MEF-TransUNet achieves significant gains in precision, IoU, and F1-score, while maintaining only a slight reduction in recall, and retains high B-IoU, showing coordinated improvement with IoU. This indicates that the model achieves higher segmentation accuracy and stronger boundary consistency and robustness, outperforming all comparison models.

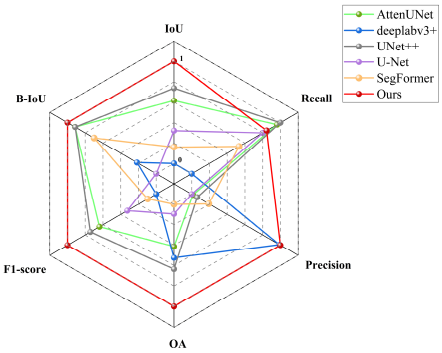


Figure 7. Accuracy index radar charts of different models on self-built dataset.

3.1.2. Comparison of Complexity and Efficiency

As shown in Table 2, the proposed model strikes an effective balance between computational efficiency and extraction accuracy. With 133.75 M parameters, it delivers strong feature learning capability; yet its computational cost is only 28.56 G FLOPs, substantially lower than U-Net++ and AttenUNet. These parameters are used effectively to enhance small water body detection and complex boundary segmentation without increasing over-

head. Moreover, the model maintains high segmentation accuracy while achieving 11.50 FPS inference speed, meeting requirements for remote sensing image processing.

Table 2. Comparison of model parameter count, computational cost, and inference speed.

Method	Params (M)	FLOPs (G)	FPS
DeepLabV3+	39.76	43.35	13.17
SegFormer	27.35	14.18	26.48
U-Net	31.39	55.98	20.53
AttenUNet	57.16	135.34	10.07
UNet++	36.63	138.66	8.5
Ours	133.75	28.56	11.50

3.1.3. Visual Comparison

To further evaluate generalization in complex real scenes, water extraction results of MEF-TransUNet and other segmentation networks were visually compared. Figure 8 shows differences in micro water body morphology integrity. The red dashed circles mark areas where the model's recognition performance is poor; the larger the circle, the more pronounced the recognition error. Specifically, DeepLabV3+ performs well on large water areas but poorly on fragmented, narrow, or micro water bodies, yielding the least ideal overall extraction. SegFormer, U-Net, AttenUNet, and UNet++ preserve more spatial details, improving micro water body detection, yet U-Net still shows edge roughness and partial breakage. MEF-TransUNet achieves the best performance, accurately capturing narrow rivers and fragmented ponds while maintaining connectivity and water structure integrity.

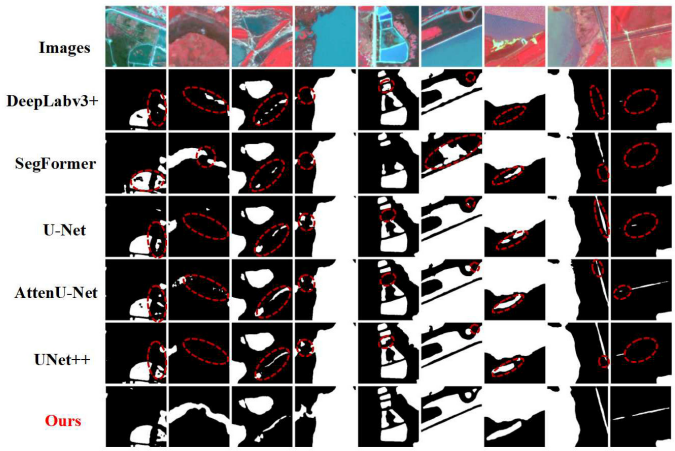


Figure 8. Visual missed detection results of different models on self-built dataset.

Figure 9 shows visual comparisons of background noise suppression in micro water bodies. It can be seen that models differ in misjudgment under strong interference from building shadows and dark roads. DeepLabV3+ is most robust against misjudgment but sacrifices recall for small targets, representing conservative high-precision predictions. SegFormer also performs well in false-positive control, ranking second to DeepLabV3+, but at the expense of reduced recall for small targets. U-Net, AttenUNet, and UNet++ preserve details but introduce noise, producing many spot-like false positives in areas with spectral features similar to water. MEF-TransUNet achieves a balanced anti-interference performance; its suppression of false positives in some difficult shadow areas is slightly lower than that of DeepLabV3+ but still excels at removing non-water background.

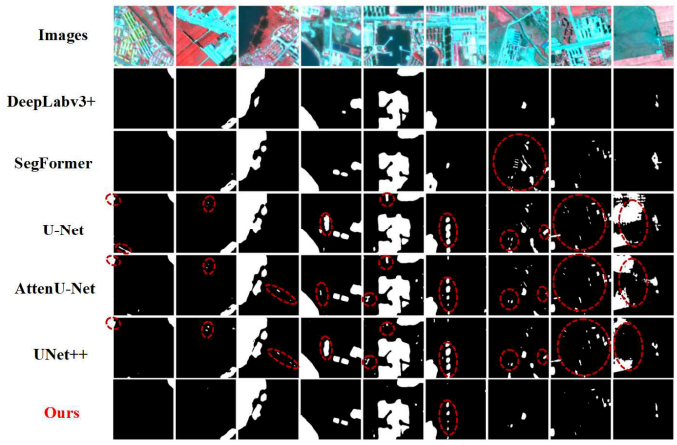


Figure 9. Visual misjudgment results of different models on self-built dataset.

3.1.4. Evaluation of Ablation Experiment

Table 3 shows the ablation experiment results. The TransUNet baseline performs moderately, with IoU 82.90% and precision 89.91%. The B-IoU is relatively low, reaching only 28.62%. The baseline adding MEGA increases precision to 90.27% but slightly reduces recall and B-IoU. The baseline adding HLFD raises precision to 91.04%, and B-IoU also shows a slight improvement. The baseline adding CBAM achieves optimal IoU, recall, and B-IoU among single-module improvements; B-IoU is further significantly improved, reaching 43.81%. The baseline integrating all three modules yields the best overall performance and robustness. Although IoU is slightly lower than CBAM alone, precision, OA, and B-IoU lead at 91.74%, 96.42%, and 43.88%, respectively.

Table 3. Evaluation of ablation experimental indicators on self-built dataset.

Method	IoU (%)	Recall (%)	Precision (%)	OA (%)	F1-Score (%)	B-IoU (%)
TransUNet	82.90	90.07	89.91	96.27	89.63	28.62
TransUNet + MEGA	82.86	89.81	90.27	96.23	89.67	28.56
TransUNet + HLFD	83.27	89.53	91.04	96.40	89.97	29.25
TransUNet + CBAM	83.56	90.56	90.35	96.38	90.16	43.81
TransUNet + three modules	83.48	90.22	91.74	96.42	90.07	43.88

Table 4 shows the increased computational cost for each module. As can be seen from the table, with each additional module, the number of parameters increases only slightly, while the computational load rises slowly and the inference speed decreases by half. Compared to the performance improvement of each module, the increased computational cost of this model improvement is within a manageable range.

Table 4. Comparison of model parameter count, computational cost, and inference speed.

Method	Params (M)	FLOPs (G)	FPS
TransUNet	133.69	26.83	27.67
TransUNet + MEGA	133.70	27.31	16.78
TransUNet + MEGA + HLFD	133.75	28.55	12.83
TransUNet + three modules	133.75	28.56	11.50

Figure 10 shows visual segmentation results for different module combinations. Red dashed circles indicate regions with poor recognition performance, and larger circles represent more significant recognition errors. The TransUNet baseline struggles in complex urban backgrounds, often misclassifying building shadows and dark surfaces as water and

showing adhesion or blur at micro water boundaries. Introducing MEGA or HLFD alone improves background noise suppression and reduces misjudgment. Adding CBAM alone increases sensitivity to micro water bodies, detecting more fine targets but also amplifying shadow noise, causing higher misjudgment and chaotic segmentation. In addition, the baseline integrating all modules yields the best overall performance, preserving water structure integrity while keeping the background relatively clean, effectively identifying fragmented and micro water bodies and suppressing non-water interference.

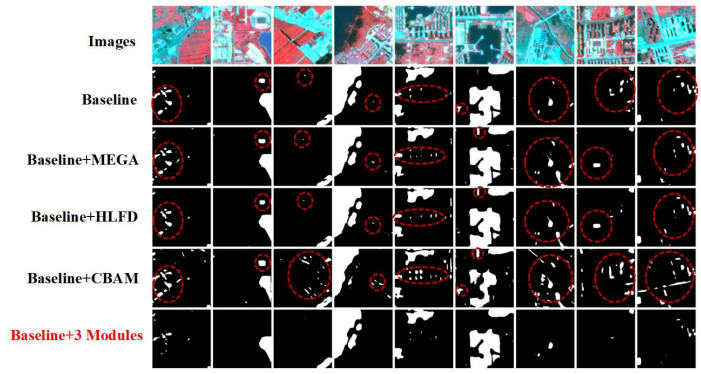


Figure 10. Visualization results of ablation experiments on self-built dataset.

3.2. Results on the Public Dataset

3.2.1. Comparison of Accuracy Indicators

Table 5 shows quantitative results on the GID public dataset. It can be seen that U-Net series models perform differently on this public dataset than on our self-built one. For small water body extraction, they all achieve high boundary IoU (B-IoU), recall, and precision, but U-Net (baseline) lags behind: though recall is strong (89.05%), its IoU (81.14%), precision (88.85%), and B-IoU (46.52%) are lowest among them, revealing weak boundary modeling. AttenUNet improves precision (90.18%) and B-IoU (54.31%) but only marginally boosts IoU and recall. UNet++ ranks second overall, achieving top B-IoU (57.07%) and strong scores across IoU, recall, precision, OA, and F1. In contrast, DeepLabV3+ improves on GID (precision 89.95%, recall 87.86%, IoU 81.32%, B-IoU 52.11), while SegFormer declines; its precision slightly exceeds U-Net's, but all other metrics fall short, suggesting poor cross-dataset generalization. The proposed MEF-TransUNet achieves the best performance across IoU (85.67%), precision (91.85%), recall (92.01%), F1-score (91.48%), and OA (97.60%), confirming superior segmentation accuracy and robust boundary representation.

Table 5. Evaluation of model accuracy indicators on GID dataset.

Method	IoU (%)	Recall (%)	Precision (%)	OA (%)	F1-Score (%)	B-IoU (%)
DeepLabV3+	81.32	87.86	89.95	96.97	88.49	52.11
SegFormer	79.37	85.97	89.20	96.41	86.63	43.42
U-Net	81.14	89.05	88.85	96.56	88.33	46.52
AttenUNet	81.11	87.81	90.18	96.44	88.32	54.31
UNet++	83.76	89.56	91.67	97.37	90.06	57.07
Ours	85.67	92.01	91.85	97.60	91.48	55.42

Figure 11 shows the radar chart of the accuracy indicators in Table 5 after 0–1 standardization processing. Specifically, U-Net exhibits contraction in precision and B-IoU, indicating false positives and poor boundary delineation. UNet++ achieves the most balanced performance, especially in boundary accuracy, with gains across all metrics. AttenUNet improves B-IoU over U-Net but contracts in OA, suggesting better boundaries yet more false negatives. DeepLabV3+ performs consistently but conservatively due to fewer false positives, more false negatives, and limited boundary detail. SegFormer shows

compact, unbiased performance, indicating it is stable but unremarkable. In contrast, the proposed model achieves the widest performance envelope: it has the highest precision, IoU, and F1-score; maintains strong B-IoU and recall synergy; and effectively balances false positives and false negatives, delivering superior boundary consistency and overall segmentation accuracy.

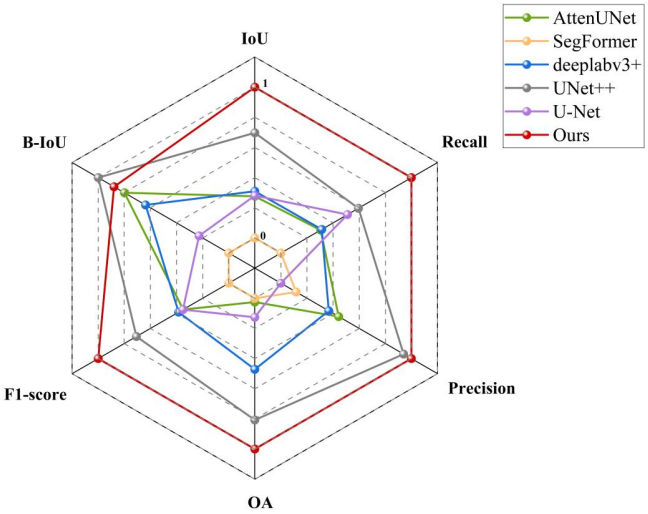


Figure 11. Accuracy index radar charts of different models on GID dataset.

3.2.2. Visual Comparison

Figure 12 shows the comparison results of different models in terms of small water body morphology preservation. The figure indicates that there are significant differences among the models in terms of the preservation of water body topology and boundary details. As a classic semantic segmentation model, DeepLabV3+ shows discontinuous segmentation of fragmented and elongated small water bodies, performing the worst overall. SegFormer has slightly better topological integrity but still fails to meet the demand for continuous extraction. U-Net, AttenUNet, and UNet++ preserve spatial details better, displaying better overall integrity of small water bodies, but still have omissions and local breaks. Among them, U-Net and AttenUNet have weaker boundary delineation, while UNet++ has more refined boundaries. In comparison, the proposed MEF-TransUNet performs the best, accurately capturing extremely narrow rivers, demonstrating the strongest connectivity, stability, and overall morphological preservation ability.

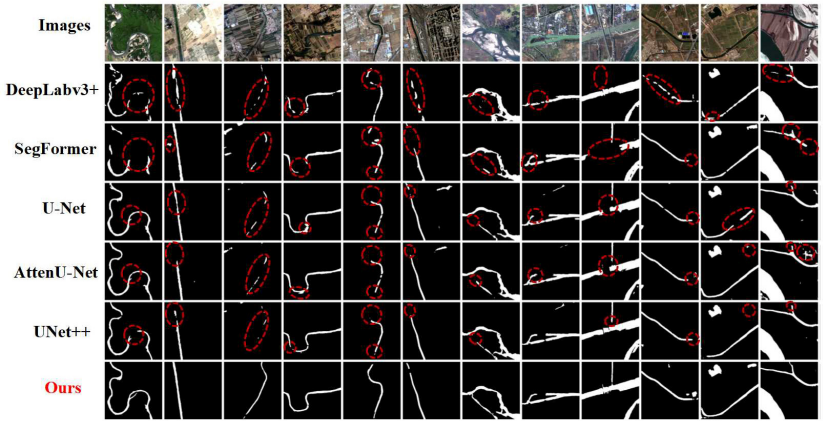


Figure 12. Comparison of visual results on GID dataset.

3.2.3. Evaluation of Ablation Experiment

Table 6 presents the ablation experiment results on the GID dataset. Compared with the self-built dataset, the contributions of each module on the public dataset are slightly different. When the MEGA module is introduced alone, except for a slight drop in accuracy to 91.83%, all other metrics improve. After adding the HLFD module, IoU, F1, and B-IoU reach 85.18%, 91.24%, and 53.76%, respectively, which is the best performance among single modules. In addition, the CBAM also brings continuous improvement to all metrics when used alone. After integrating the three modules into TransUNet, the overall performance and robustness of the model are the best with IoU, recall, OA, F1, and B-IoU reaching 85.67%, 92.01%, 97.60%, 91.48%, and 55.42%, respectively, comprehensively outperforming each single-module variant, highlighting the synergy of different modules.

Table 6. Evaluation of ablation experimental indicators on GID dataset.

Method	IoU (%)	Recall (%)	Precision (%)	OA (%)	F1-Score (%)	B-IoU (%)
TransUNet	84.45	90.22	92.27	97.21	90.68	52.42
TransUNet + MEGA	84.62	90.62	91.83	97.28	90.75	53.03
TransUNet + HLFD	85.18	90.67	92.58	97.26	91.24	53.76
TransUNet + CBAM	84.73	90.48	92.46	97.35	90.92	52.77
TransUNet + three modules	85.67	92.01	91.85	97.60	91.48	55.42

Figure 13 shows the ablation experiment results of the model on the GID public dataset. Red dashed circles indicate regions with poor recognition performance, where larger circles correspond to more significant errors. It can be observed that TransUNet, as the baseline model, has limitations in handling slender water bodies, often missing narrow rivers and ditches and misclassifying some background features as water bodies. When the MEGA or CBAM modules are introduced individually, the model shows improved ability to preserve the topological integrity of slender water bodies, although some misclassification and missed detection still remain. In contrast, the model with only the HLFD module demonstrates better performance, with significantly enhanced sensitivity to small water bodies, improved detection of fragmented targets, and better connectivity and integrity of narrow rivers, while reducing misclassification to some extent. When all three modules are integrated, the model achieves the best overall performance, with more complete preservation of slender river structures and further reduction in both misclassification and missed detection, enabling more accurate and continuous small water body extraction.

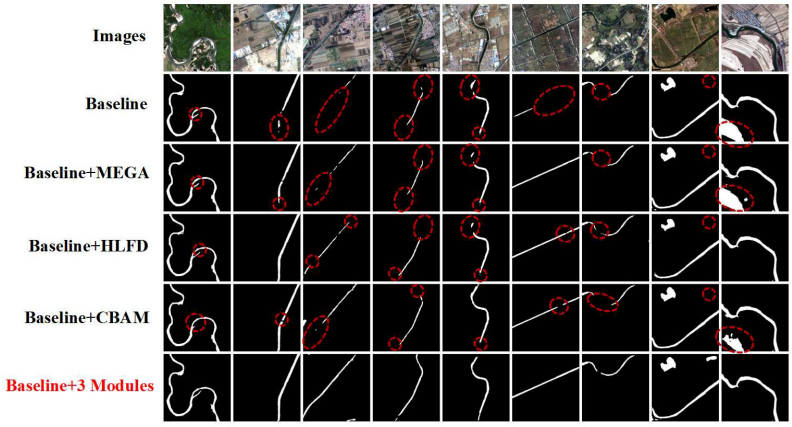


Figure 13. Visualization results of ablation experiments on GID dataset.

4. Discussion

This study proposes the MEF-TransUNet model, integrating MEGA, HLFD, and CBAM modules, to address fragmented morphology and strong background interference in high-resolution small water body extraction. Evaluated on a self-built dataset and the public Songhua River Basin GID dataset, the proposed model's mechanisms, applicability boundaries, and spatial generalization are analyzed against existing models, which exhibit distinct performance bottlenecks. Specifically, DeepLabV3+ demonstrates conservative, high-precision false-positive control on the self-built dataset but sacrifices small-target recall and boundary quality (B-IoU), exposing the tendency of traditional dilated convolutions to lose fine spatial details, a flaw remaining despite its robust overall GID performance. Similarly, the pure-Transformer SegFormer captures global contexts well but lacks targeted local detail extraction, leading to significant omissions, poor contiguous extraction, and bottom-ranking metrics across both datasets. Conversely, the U-Net family leverages encoder–decoder architectures to better preserve spatial details; however, on the complex self-built dataset, the base U-Net suffers from low precision and shadow-induced noise, AttenUNet improves recall but fails to eliminate false alarms, and UNet++ achieves detailed boundaries and high recall but remains trapped in a high-recall, low-precision bottleneck with local noise, despite its strong GID performance. Overcoming these limitations, MEF-TransUNet demonstrates superior adaptability and cross-regional generalization. Driven by multi-scale feature enhancement and progressive detail recovery, it achieves comprehensive leadership in IoU, precision, F1-score, and B-IoU on the highly disturbed self-built dataset, optimally balancing interference suppression with precise topological connectivity to capture narrow targets and eliminate shadow artifacts. Simultaneously, its consistently top-tier metrics on the representative GID dataset confirm its robust feature extraction and ability to avoid over-segmentation. Finally, despite an increased parameter count from global and multi-scale mechanisms, MEF-TransUNet maintains controlled forward computation and stable inference speeds, successfully balancing theoretical extraction accuracy with practical engineering deployment overhead.

A longitudinal ablation analysis on the self-built dataset and the publicly available GID dataset reveals the contributions and synergistic effects of each improved module in small water body extraction, as well as their different behaviors under varying data distributions. Small water bodies are characterized by narrow spatial extent, fragmented morphology, and weak spectral signals, which pose significant challenges to the baseline TransUNet. On the self-built dataset with complex urban backgrounds and strong land-cover interference, the baseline model is prone to misclassifying building shadows as water bodies due to spectral ambiguity, accompanied by obvious boundary adhesion. In contrast, on the GID dataset with broader scenes and relatively better image quality, its limitations are mainly reflected in fragmentation and missed detection of narrow rivers and streams. Conventional edge enhancement methods tend to indiscriminately amplify gradient responses, introducing false edges such as building outlines and increasing false positives, while traditional frequency decomposition may overly smooth high-frequency information, leading to semantic loss of weak boundary features of small water bodies [55]. Therefore, each improved module in this study is designed with clear and targeted objectives to address these specific challenges. The MEGA module primarily addresses boundary blurring and adhesion caused by spectral similarity between small water bodies and dark ground features, effectively suppressing non-water edge interference through more discriminative edge constraints. The HLFD module targets high-frequency shadow noise in complex scenes, using an optimized frequency-domain decomposition strategy to remove background noise in the self-built dataset, thereby improving precision on both datasets without sacrificing true boundary information. In addition, a CBAM attention mechanism

is introduced to mitigate feature loss and signal attenuation during downsampling of slender water bodies, enhancing weak feature capture and significantly improving recall. However, in the densely interfered self-built dataset with many fragmented targets, this increased sensitivity also amplifies background shadows. Notably, using CBAM alone yields relatively high IoU, even slightly higher than the final fused model, which can be attributed to the trade-off between metric sensitivity and feature activation: CBAM enhances pixel-level activation and increases detected water regions, improving IoU, but also introduces local false positives. When MEGA and HLFD are further integrated, their spatial and frequency-domain constraints suppress shadows and high-frequency noise but also conservatively remove some boundary transition pixels, leading to a slight reduction in predicted positive regions and a corresponding minor decrease in IoU. Ultimately, the full MEF-TransUNet achieves a better balance across datasets: on the complex self-built dataset, MEGA and HLFD effectively regulate CBAM activation, maintaining high recall while improving precision and significantly enhancing B-IoU, thereby suppressing shadow interference and improving boundary integrity. On the GID dataset with relatively better image quality, the fusion model effectively converts CBAM's sensitivity into improved continuity of narrow river extraction, demonstrating strong cross-scene adaptability and module synergy. Ablation experiments further show that the additional computational cost of MEGA and HLFD is modest, mainly causing a slight increase in complexity and minor FPS reduction, while CBAM introduces only a minimal computational overhead. Overall, the model achieves substantial performance gains with low computational cost.

MEF-TransUNet performs well for micro water bodies but has some limitations in complex scenes. First, preserving high-frequency details enhances sensitivity to small targets but increases false alarms from spectrally similar interferences, especially shadows. Second, deep network downsampling may still cause topological fractures for extremely narrow or boundary-blurred water bodies due to sub-pixel information loss, even with multi-level feature recovery. Furthermore, regarding the feature fusion strategy, this paper adopts a selective skip connection mechanism that retains only the C2–C4 stage features for decoding and reconstruction, without incorporating the shallow C1 feature. This design is primarily based on considerations of semantic differences across feature levels and the interference of high-frequency texture noise in remote sensing imagery, aiming to enhance the model's robustness in complex backgrounds. However, a comprehensive quantitative evaluation of different feature-level combination approaches through systematic ablation experiments has yet to be conducted, and the applicability of this strategy under different tasks or data scenarios remains to be further validated. Future improvements will focus on two aspects: introducing more effective feature selection by integrating multi-source data like MNDWI and DEM, using water spectral and low-lying terrain priors to suppress shadow interference, and increasing shadow sample weights to improve discrimination; and optimizing feature preservation for fine water bodies by maintaining high-resolution branches, and designing connectivity-oriented loss functions, such as topological or boundary-weighted loss, to enhance learning of structure integrity and pixel-level extraction accuracy [56–59]. Furthermore, comparative experiments on feature fusion strategies at different levels will be conducted, such as introducing attention mechanisms or adaptive weighting strategies to dynamically screen shallow features, thereby suppressing noise interference while preserving detailed information. This will enable the development of a more robust water body extraction model with stronger generalization ability.

5. Conclusions

To address morphological fragmentation, boundary blurring, and background interference in micro water body extraction from high-resolution remote sensing images, this

study proposes a newly developed MEF-TransUNet model, which integrates multi-scale edge guidance (MEGA), high–low-frequency decomposition (HLFD), and convolutional attention (CBAM). The model combines Transformer's long-range modeling with CNN's local feature extraction, overcoming the limitations of single architectures in fine-grained segmentation. Specifically, MEGA repairs slender target topologies, CBAM enhances sensitivity to weak broken signals, and HLFD removes spectral interference from objects like shadows. The synergistic integration of the three modules achieves high recall, improves accuracy and purity, demonstrates strong robustness, reduces manual correction, and provides reliable technical support for refined water surveys while laying the foundation for future multi-source and super-resolution monitoring research.

Author Contributions: Conceptualization, J.R.; methodology, Y.Y.; software, Y.Y.; validation, S.L.; formal analysis, K.L.; investigation, J.R.; resources, S.L.; data curation, X.Z.; writing—original draft preparation, Y.Y.; writing—review and editing, J.R.; visualization, X.Z.; supervision, S.L.; project administration, J.R.; funding acquisition, J.R.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Natural Science Foundation of Heilongjiang Province of China (No. PL2025D015), the National Natural Science Foundation of China (No. 42371381 and No. 42171333), the Natural Science Foundation of Jilin Province of China (No. YDZJ202501ZYTS466), and the Program for Young Talents of Basic Research in Universities of Heilongjiang Province (No. YQJH2024113).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Mullen, A.L.; Watts, J.D.; Rogers, B.M.; Carroll, M.L.; Elder, C.D.; Noomah, J.; Williams, Z.; Caraballo-Vega, J.A.; Bredder, A.; Rickenbaugh, E.; et al. Using High-Resolution Satellite Imagery and Deep Learning to Track Dynamic Seasonality in Small Water Bodies. *Geophys. Res. Lett.* **2023**, *50*, e2022GL102327. [CrossRef]
2. Biggs, J.; von Fumetti, S.; Kelly-Quinn, M. The importance of small waterbodies for biodiversity and ecosystem services: Implications for policy makers. *Hydrobiologia* **2016**, *793*, 3–39. [CrossRef]
3. Pekel, J.F.; Cottam, A.; Gorelick, N.; Belward, A.S. High-resolution mapping of global surface water and its long-term changes. *Nature* **2016**, *540*, 418–422. [CrossRef]
4. Taubenböck, H.; Wurm, M.; Netzband, M.; Zwenzner, H.; Roth, A.; Rahman, A.; Dech, S. Flood risks in urbanized areas—Multi-sensoral approaches using remotely sensed data for risk assessment. *Nat. Hazards Earth Syst. Sci.* **2011**, *11*, 431–444. [CrossRef]
5. Carstens, D.; Amer, R. Spatio-temporal analysis of urban changes and surface water quality. *J. Hydrol.* **2019**, *569*, 720–734. [CrossRef]
6. Gu, Z.; Zhang, Y.; Fan, H. Mapping inter- and intra-annual dynamics in water surface area of the Tonle Sap Lake with Landsat time-series and water level data. *J. Hydrol.* **2021**, *601*, 126644. [CrossRef]
7. Mao, D.; Wang, Z.; Du, B.; Li, L.; Tian, Y.; Jia, M.; Zeng, Y.; Song, K.; Jiang, M.; Wang, Y. National wetland mapping in China: A new product resulting from object-based and hierarchical classification of Landsat 8 OLI images. *ISPRS J. Photogramm. Remote Sens.* **2020**, *164*, 11–25. [CrossRef]
8. Chen, F.; Chen, X.; Van de Voorde, T.; Roberts, D.; Jiang, H.; Xu, W. Open water detection in urban environments using high spatial resolution remote sensing imagery. *Remote Sens. Environ.* **2020**, *242*, 111706. [CrossRef]
9. Li, L.; Su, H.; Du, Q.; Wu, T. A novel surface water index using local background information for long term and large-scale Landsat images. *ISPRS J. Photogramm. Remote Sens.* **2021**, *172*, 59–78. [CrossRef]
10. McFeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **2007**, *17*, 1425–1432. [CrossRef]
11. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **2007**, *27*, 3025–3033. [CrossRef]
12. 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]

13. Otsu, N. A threshold selection method from gray-level histograms. *Automatica* **1975**, *11*, 23–27. [CrossRef]
14. Campos, J.C.; Sillero, N.; Brito, J.C. Normalized difference water indexes have dissimilar performances in detecting seasonal and permanent water in the Sahara–Sahel transition zone. *J. Hydrol.* **2012**, *464–465*, 438–446. [CrossRef]
15. Fan, L.; Dong, Y.; Zhao, J.; Geiß, C.; Wang, L.; Taubenbock, H. Mapping of Small Water Bodies with Integrated Spatial Information for Time Series Images of Optical Remote Sensing. In *Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium*; IEEE: New York, NY, USA, 2022; pp. 6224–6227.
16. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
17. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
18. 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]
19. 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; pp. 801–818. [CrossRef]
20. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587. [CrossRef]
21. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer: Cham, Switzerland, 2015; pp. 234–241.
22. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv* **2021**, arXiv:2102.04306. [CrossRef]
23. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. In *Advances in Neural Information Processing Systems 34*; Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P.S., Vaughan, J.W., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2021; pp. 12077–12090. [CrossRef]
24. Zhou, L.; Zhang, C.; Wu, M. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 182–186. [CrossRef]
25. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual Attention Network for Scene Segmentation. In *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE Computer Society: Los Alamitos, CA, USA, 2019; pp. 3141–3149.
26. Li, R.; Liu, W.; Yang, L.; Sun, S.; Hu, W.; Zhang, F.; Li, W. 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]
27. Li, W.; Li, Y.; Gong, J.; Feng, Q.; Zhou, J.; Sun, J.; Shi, C.; Hu, W. Urban Water Extraction with UAV High-Resolution Remote Sensing Data Based on an Improved U-Net Model. *Remote Sens.* **2021**, *13*, 3165. [CrossRef]
28. Liu, B.; Du, S.; Bai, L.; Ouyang, S.; Wang, H.; Zhang, X. Water extraction from optical high-resolution remote sensing imagery: A multi-scale feature extraction network with contrastive learning. *GIScience Remote Sens.* **2023**, *60*, 2166396. [CrossRef]
29. Yan, Q.; Chen, Y.; Jin, S.; Liu, S.; Jia, Y.; Zhen, Y.; Chen, T.; Huang, W. Inland water mapping based on GA-LinkNet from CyGNSS data. *IEEE Geosci. Remote Sens. Lett.* **2022**, *20*, 1500305. [CrossRef]
30. Yan, Q.; Liu, S.; Chen, T.; Jin, S.; Xie, T.; Huang, W. Mapping surface water fraction over the pan-tropical region using CYGNSS data. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5800914. [CrossRef]
31. Chen, Y.; Yan, Q.; Jin, S.; Huang, W. DiffWater: A conditional diffusion model for estimating surface water fraction using CyGNSS data. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5801817. [CrossRef]
32. Ma, X.; Zhou, C.; Kong, X.; He, J.; Gui, L.; Neubig, G.; May, J.; Zettlemoyer, L. Mega: Moving average equipped gated attention. *arXiv* **2022**, arXiv:2209.10655. [CrossRef]
33. Ma, X.; Zhang, X.; Ding, X.; Pun, M.-O.; Ma, S. Decomposition-Based Unsupervised Domain Adaptation for Remote Sensing Image Semantic Segmentation. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5645118. [CrossRef]
34. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19. [CrossRef]
35. Oktay, O.; Schlemper, J.; Le Folgoc, L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Glocker, B.; Rueckert, D.; et al. Attention U-Net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999. [CrossRef]
36. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Cham, Switzerland, 2018; Volume 11045, pp. 3–11. [CrossRef]
37. Lu, Z.; Zhang, J.; Li, C.; Wang, X.; Lei, G.; Li, K. Modeling hydrological processes and analyzing water balance utilizing remote sensing data and physical hydrological models in the Songnen Plain, China. *PLoS ONE* **2025**, *20*, e0329816. [CrossRef]
38. Chen, H.; Zhang, W.; Gao, H.; Nie, N. Climate Change and Anthropogenic Impacts on Wetland and Agriculture in the Songnen and Sanjiang Plain, Northeast China. *Remote Sens.* **2018**, *10*, 356. [CrossRef]

39. Yu, X.; Ding, S.; Zou, Y.; Xue, Z.; Lyu, X.; Wang, G. Review of Rapid Transformation of Floodplain Wetlands in Northeast China: Roles of Human Development and Global Environmental Change. *Chin. Geogr. Sci.* **2018**, *28*, 654–664. [CrossRef]
40. Xiao, W.; Chen, W.; Yue, W.; Mu, J.; Xu, J. Waterbody loss due to urban expansion of large Chinese cities in last three decades. *Sci. Rep.* **2022**, *12*, 17498. [CrossRef]
41. Jiang, H.; Ji, L.; Yu, K.; Zhao, Y. Analysis of the Substantial Growth of Water Bodies during the Urbanization Process Using Landsat Imagery—A Case Study of the Lixiahe Region, China. *Remote Sens.* **2024**, *16*, 711. [CrossRef]
42. Roy, D.P.; Huang, H.; Houborg, R.; Martins, V.S. A global analysis of the temporal availability of PlanetScope high spatial resolution multi-spectral imagery. *Remote Sens. Environ.* **2021**, *264*, 112586. [CrossRef]
43. Frazier, A.E.; Hemingway, B.L. A Technical Review of Planet Smallsat Data: Practical Considerations for Processing and Using PlanetScope Imagery. *Remote Sens.* **2021**, *13*, 3930. [CrossRef]
44. Zhang, X.; Zhou, Y.; Qiao, P.; Lv, X.; Li, J.; Du, T.; Cai, Y. Image Registration Algorithm for Remote Sensing Images Based on Pixel Location Information. *Remote Sens.* **2023**, *15*, 436. [CrossRef]
45. Wang, T.; Shi, J.; Letu, H.; Ma, Y.; Li, X.; Zheng, Y. Detection and Removal of Clouds and Associated Shadows in Satellite Imagery Based on Simulated Radiance Fields. *J. Geophys. Res. Atmos.* **2019**, *124*, 7207–7225. [CrossRef]
46. Miller, L.; Pelletier, C.; Webb, G.I. Deep learning for satellite image time-series analysis: A review. *IEEE Geosci. Remote Sens. Mag.* **2024**, *12*, 81–124. [CrossRef]
47. 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]
48. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
49. Burt, P.J.; Adelson, E.H. The Laplacian pyramid as a compact image code. *IEEE Trans. Commun.* **1983**, *31*, 532–540. [CrossRef]
50. Xu, K.; Qin, M.; Sun, F.; Wang, Y.; Chen, Y.-K.; Ren, F. Learning in the frequency domain. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 1740–1749. [CrossRef]
51. 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)*; IEEE: New York, NY, USA, 2016; pp. 565–571.
52. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988. [CrossRef]
53. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
54. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980. [CrossRef]
55. Marmanis, D.; Schindler, K.; Wegner, J.D.; Galliani, S.; Datcu, M.; Stilla, U. Classification with an edge: Improving semantic image segmentation with boundary detection. *ISPRS J. Photogramm. Remote Sens.* **2018**, *135*, 158–172. [CrossRef]
56. Lyu, S.; Zhao, Q.; Zhou, Z.; Li, M.; Zhou, Y.; Yao, D.; Cheng, G.; Zhou, H.; Shi, Z. Deep learning based domain adaptation methods in remote sensing: A comprehensive survey. *arXiv* **2025**, arXiv:2510.15615. [CrossRef]
57. 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]
58. Bokhovkin, A.; Burnaev, E. Boundary loss for remote sensing imagery semantic segmentation. In *Advances in Neural Networks—ISNN 2019*; Lu, H., Li, H., Wang, D., Li, Z., Wang, X., Cohen, G., Eds.; Springer: Cham, Switzerland, 2019; pp. 388–401. [CrossRef]
59. Hu, X.; Li, F.; Samaras, D.; Chen, C. Topology-preserving deep image segmentation. In *Advances in Neural Information Processing Systems 32*; Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: Vancouver, BC, Canada, 2019; pp. 5657–5668.

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.