water

MDPI

Article

A Deep Learning Method Coupling a Channel Attention Mechanism and Weighted Dice Loss Function for Water Extraction in the Yellow River Basin

Jichang Yang [1,2], Yuncong Lu [1,3], Zhiqiang Zhang [4], Jieru Wei [1,2], Jiandong Shang [1,2,*], Chong Wei [4,*], Wensheng Tang [1,2] and Junjie Chen [1,2]

[1] The School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450000, China; nsczzyjc@gs.zzu.edu.cn (J.Y.); luyuncong@rpc.edu.cn (Y.L.); weijr0117@gs.zzu.edu.cn (J.W.); tangwensheng@gs.zzu.edu.cn (W.T.); c17538273356@gs.zzu.edu.cn (J.C.)
[2] National Supercomputing Center in Zhengzhou, Zhengzhou University, Zhengzhou 450000, China
[3] Department of Image and Network Investigation, Zhengzhou Police University, Zhengzhou 450000, China
[4] College of Surveying and Geo-Informatics, North China University of Water Resources and Electric Power, Zhengzhou 450046, China; zhangzhiqiang@ncwu.edu.cn
* Correspondence: sjd@zzu.edu.cn (J.S.); weic85@ncwu.edu.cn (C.W.)

Abstract: The extraction of small water bodies in the Yellow River Basin has always been a key issue of concern in the fields of remote sensing technology application, water resource management, environmental science, and geographic information systems. Due to factors such as water bodies, human activities, and cloud cover, water body extraction becomes difficult. In addition, convolutional neural networks are prone to losing small water body feature information during the process of extracting local features, which can cause more imbalance between positive and negative samples of water bodies and non-water bodies. In response to these issues, this study focused on a specific research area—the middle and lower reaches of the Yellow River. We processed and analyzed high-resolution optical satellite images collected from the Yellow River Basin and other areas, with a particular emphasis on precise identification of small water bodies, and proposed a network structure, the SE-Attention-Residual-Unet (SE-ResUnet), for water extraction tasks.The main contributions of this article are threefold: (1) Introducing a channel attention mechanism with residual structure in the down-sampling process, and learning Unet's skipping structure for multi-scale feature extraction and compensation, thereby enhancing the feature extraction ability of small water bodies, including rivers, lakes, and reservoirs. (2) Introducing a weighted-Dice (W-Dice) loss function to balance positive and negative samples and enhance the generalization of the model. (3) In comparative experiments on improving the Unet model with semantic segmentation networks such as Unet, PSPNet, Deeplabv3+ on a self-built dataset and remote sensing interpretation public dataset, excellent performance and results were achieved on the mIoU, OA, and F1-score metrics. On the self-built dataset, compared with Unet, the mIoU, OA, and F1-score improved by 0.38%, 0.12%, and 0.08%, respectively. On the publicly available dataset, for remote sensing interpretation of water extraction, the mIoU, OA, and F1-score improved by 0.63%, 0.26%, and 0.25%, respectively. The experimental results demonstrate that a strategy combining an attention mechanism and a weighted loss function has a significant effect on the effectiveness of the collaborative improvement of neural network models in water extraction tasks.

Keywords: improved Unet; semantic segmentation; channel attention; water extraction; remote sensing images; weighted loss function

check for updates

Academic Editor: Achim A. Beylich

Received: 26 December 2024
Revised: 21 January 2025
Accepted: 25 January 2025
Published: 8 February 2025

Citation: Yang, J.; Lu, Y.; Zhang, Z.; Wei, J.; Shang, J.; Wei, C.; Tang, W.; Chen, J. A Deep Learning Method Coupling a Channel Attention Mechanism and Weighted Dice Loss Function for Water Extraction in the Yellow River Basin. *Water* **2025**, *17*, 478. https://doi.org/10.3390/w17040478

Copyright: © 2025 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/).

1. Introduction

The use of remote sensing technology, namely, optical remote sensing images, is becoming important in the extraction of water bodies due to its wide coverage, abundant information, and fast update rate. Remote sensing imageries provide efficient and accurate data on water bodies, offering significant assistance for related research and applications [1]. Water body extraction technology is founded on the capacity to identify, classify, and retrieve data pertaining to water bodies from remote sensing methods. This procedure utilizes technologies and knowledge from various domains, such as image processing, pattern recognition, geographic information systems, and others [2]. Water bodies typically exhibit distinct spectral and spatial properties, such as reflectivity, texture, shape, etc. By analyzing and using these attributes, we may efficiently gather information about water bodies. In recent years, the progress in computer and remote sensing technologies has resulted in notable improvements in water body extraction technology [3].

The water body index and image classification method are the two main conventional techniques employed for extracting water bodies from high-resolution optical remote sensing images. Several scholars have developed a variety of indicators for extracting information about water bodies. McFeeter proposed using the Normalized Differential Water Index (NDWI) for the green and near-infrared bands [4]. Nevertheless, the presence of buildings causes substantial interference with NDWI, hence posing difficulties in accurately identifying water bodies in such areas. In order to overcome the limitations of the NDWI, Xu et al. proposed the Modified Normalized Differential Water Index (MNDWI) [5], which replaces the near-infrared band with the short-wave infrared band in NDWI [5]. Yan et al. introduced the Enhanced Water Index (EWI) by merging NDWI and MNDWI [6]. Feyisa et al. developed the Automated Water Extraction Index (AWEI) [7] to reduce noise in mountainous and urban regions. Zhang et al. [8] combined a suspended particulate matter concentration and water index to reveal the multi-scale variation pattern of the surface water area in the Yellow River Basin. They used various bands (1, 2, 4, 5, and 7) of Landsat 5 TM. Other, less prevalent water indices encompass the shadow water index, pseudo-NDWI, Gaussian NDWI, MNDWI, and the new water index. However, the majority of water index techniques are unsuitable due to the limited number of spectral bands found in high-resolution images, often consisting of only four bands: blue, green, red, and near infrared.

In order to overcome the shortcomings of the water index method and make use of the spatial information of high-resolution optical remote sensing images, various image classification methods have been proposed. Image classification performs water extraction by combining spectral, shape, and texture features and using various classifiers in machine learning. Commonly used classifiers include decision tree [9], random forest [10], support vector machine [11–13], etc. Although image taxonomies can make use of a variety of features to achieve better results than water index methods, they require these features to be manually constructed for specific water extraction tasks. This will greatly reduce the efficiency and prolong the extraction time. In addition, the application scope of these artificially constructed features is limited, and it is difficult to extract water bodies in different regions [14].

Convolutional Neural Networks (CNNs) [15] have gained popularity in the fields of semantic segmentation, object identification, and scene categorization due to their ability to efficiently learn multiple layers of automated features. Water body extraction is a specific task in semantic segmentation that is commonly addressed by CNN models [16]. Chen et al. introduced an adaptive water extraction pooling layer [17] to reduce the loss of features during pooling. Most existing methods fail to meet the extraction requirements of water bodies of different sizes [18]. In addition, as features are extracted sequentially, the

size of the feature map will decrease, potentially leading to the omission of small water bodies with subtle characteristics, resulting in biased outcomes. Thus, it is necessary to employ multi-layer and multi-scale characteristics in order to address these challenges. Multi-layer features refer to features that have been extracted from different convolutional neural network (CNN) layers [19,20]. Several multi-scale feature extraction modules have been developed as a result of semantic segmentation, such as Spatial Pyramid Pooling (SPP), Pyramid Pooling Module (PPM) in PSPNet [21], and Atrous Spatial Pyramid Pooling (ASPP) in DeepLabv2. Unfortunately, data loss occurs due to the absence of multi-layer settings in these multi-scale feature extraction modules, which primarily execute pooling on feature maps. And Sun et al. [22] improved the Deeplabv3+ network to extract water bodies and attempted to combine DEM data with remote sensing image data to extract small water bodies [23]. Cao et al. [24] extracted water bodies from high-resolution remote sensing images by enhancing Unet networks and multi-scale information fusion. Yan et al. [25] used a novel water body extraction method to automate water body extraction from Landsat 8 OLI images. Cheng et al. [26] proposed a water body extraction method based on spatial partitioning and feature decoupling. Zhao et al. [27] proposed an unsupervised water body extraction method by combining the estimated probabilities of historical, neighboring, and neighborhood prior information using Bayesian model averaging. Chen et al. [28] refined the feature information by introducing dynamic semantic kernels to achieve high-precision extraction of lake water bodies. Mishra et al. [29] used principal component analysis to fuse panchromatic and infrared bands to classify the fused images into surface types. Wang et al. [30,31] incorporated a mixed-domain attention mechanism into the decoder structure to fully exploit the spatial and channel features of small water bodies in the image.

Traditional water extraction methods, such as the threshold method, spectral index method, object-oriented method, and machine learning classification method, mainly rely on manual statistical features and expert experience, which may not be effective in dealing with complex backgrounds or diverse water bodies. In addition, traditional methods often struggle to achieve a high degree of automation, which increases the complexity and time cost of operations. Interference factors such as shadows and buildings can also limit the accuracy of water extraction, and traditional methods may not be flexible and accurate enough in dealing with these interferences. Deep learning methods can automatically learn and extract useful features from input data without the need for manual feature selection and extraction, greatly improving the efficiency and accuracy of water extraction. Deep learning models can learn complex features and the intrinsic laws of water bodies by training on large amounts of data, thus having good generalization ability to unknown data. Furthermore, deep learning models have strong robustness to noise and interference, and can handle water extraction tasks in complex backgrounds. Deep learning models have transferability and can be applied to water extraction tasks in different regions and data sources. Moreover, with the continuous development of technology, deep learning models can be continuously optimized and expanded to meet more complex water extraction needs.

In light of this, the current study constructs a feature extraction network that utilizes the Unet's multi-scale feature architecture and jumping structure [32] to extract water body features at several layers and scales. The water body extraction task involves only two categories: background and water body. To ensure a balanced representation of positive and negative data during training, the W-Dice loss function is incorporated. To be more precise, the model minimizes unnecessary computation usage while accurately extracting the crucial information from the data by combining the optimization guidance of the loss function with the focusing ability of the attention mechanism [33]. This enhances the model's cognitive process and augments the precision of its predictions. The W-Dice loss function offers a distinct optimization objective for the model, while the SE channel

attention mechanism effectively captures the complex relationships among several channels in remote sensing images. To achieve the desired outcomes, this combination enables the model to consistently improve its performance and optimize the utilization of river features in remote sensing images [34].

The main contents of this paper are as follows:

(1) The down-sampling approach integrates the SE attention mechanism with a residual structure, and learns the jumping structure of the Unet for multi-scale feature extraction and compensation. This enhances the capability to extract the characteristics of small water bodies, including rivers, lakes, and reservoirs.

(2) In order to enhance the model's generalization and to ensure that both positive and negative samples are given equal importance, the W-Dice loss function is incorporated into the loss calculation during the training phase.

(3) The enhanced network model is compared and evaluated with semantic segmentation networks like PSPNet and Deeplabv3+ on both a self-built remote sensing image dataset and another public remote sensing interpretation dataset. The effectiveness of water body extraction is assessed using the mIoU as a crucial metric.

2. Materials and Methods

2.1. SE-ResUnet

The network configuration described in this paper, illustrated in Figure 1, is referred to as the SE-ResUnet. The coding layer is predominantly situated on the left side of the diagram, whereas the decoding layer is positioned on the right. The input image is reconstructed at the decoder stage after passing through the convolutional structure's feature extraction backbone network. To address the presence of features, the output feature images from several down-sampling layers are combined and overlaid using a hierarchical structure. An SE attention mechanism is incorporated into the third down-sampling layer of the ResnetBottleNeck [35] module within the residual structure. The primary distribution of the residual structure with the SE attention mechanism in the structure diagram of the SE-ResUnet is in the feature extraction backbone network, which is referred to as the "SRBlock".

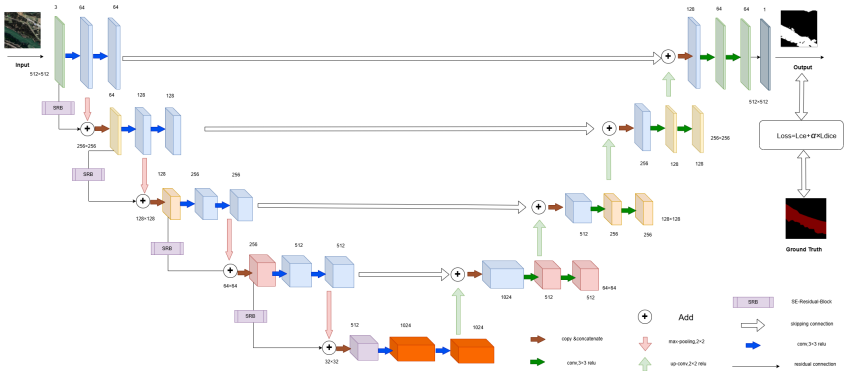


Figure 1. SE-ResUnet network structure.

Convolutional neural networks have advantages in extracting local feature information from remote sensing images, including remote sensing water body images. However, for small water bodies in the distribution of remote sensing images, they are easily lost during the down-sampling process, and the retained small water bodies pixels are more difficult to capture due to the small number of samples. Therefore, we introduce a residual structure with a channel attention mechanism on the basis of the Unet network to compensate for the lost small water bodies features during the down-sampling process. In addition, we

coupled the weighted Dice loss function to balance positive and negative samples, thereby reducing the impact of insufficient positive samples in small water bodies. Shown in Figure 2 is the improved flow chart of the SE-ResUnet model.

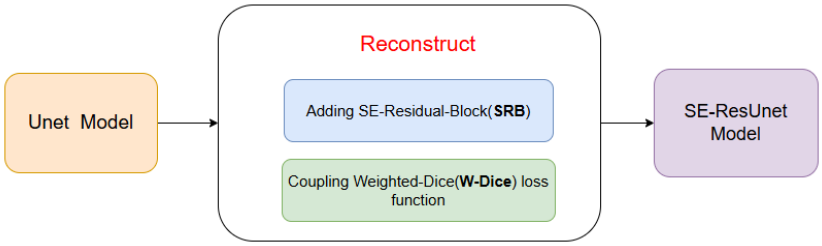


Figure 2. Flow chart of the SE-ResUnet model.

2.2. Squeeze-Excitation Residual Block Module

A unique architectural unit, called the Squeeze-Excitation Residual Block (SRB) module, is intended to improve the network representation capabilities by means of dynamic channel feature recalibration. The structure of the SRB module and ResBlock are comparable in that they both have residual connections, as seen in Figure 3. The SRB module's rightmost branch, in contrast to the typical ResBlock, further includes functions like global pooling, a fully connected layer, a ReLU activation function, and a sigmoid function. These functions, in turn, transform, compress, and excite images with input size $H \times W \times C$. It is easy to see how the image's width, height, and channel change during this procedure. In order to produce the attention mechanism effect, the final step involves performing feature fusion with the left branch after restoring the image to an $H \times W \times C$ tensor of the same size as the input through the zoom operation.

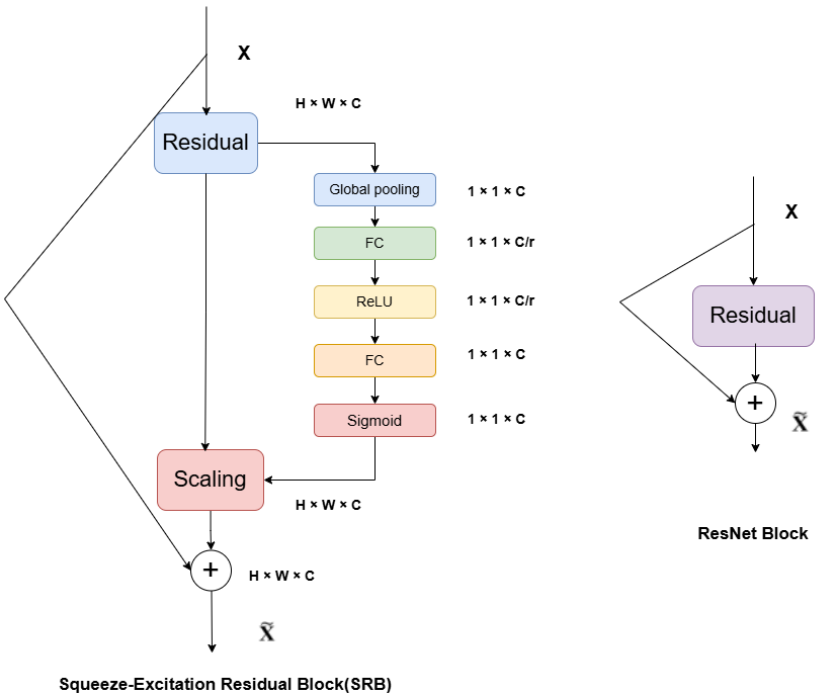


Figure 3. SRB structure.

We built the modified SE-ResUnet by first reconstructing the original Unet, adding the SE attention mechanism, and replacing the original loss function with the W-Dice loss function and the CE cross-entropy loss function. The modified SE-ResUnet is then applied to the remote sensing satellite images for water body extraction.

By explicitly modeling the interdependencies between convolutional feature channels, the Squeeze-and-Excitation technique [36] can enhance the network's representation capacity. This strategy enables the network to dynamically recalibrate features by selectively stressing desirable ones and suppressing undesirable ones, hence facilitating the learning of global information.

The SE module has three essential actions for recalibrating the input X feature channel, as depicted in Figure 4. After a number of convolutions and transformations, a feature with a specific number of feature channels, denoted as C2, is produced. This is based on the assumption that the input X has a certain number of feature channels, denoted as C1. F_{tr} represents the standard convolution operation, which transforms the feature map X into the feature map U, and F_{sq} represents the compression operation, specifically using global average pooling of channels to compress global spatial information into a channel descriptor. F_{ex} scales the channel dimension by a fixed coefficient to fully capture channel dependencies in order to utilize the information gathered during compression operations. F_{scale} weights the attention weights obtained earlier onto the features of each channel to obtain the final output of the module $\tilde{X}$.

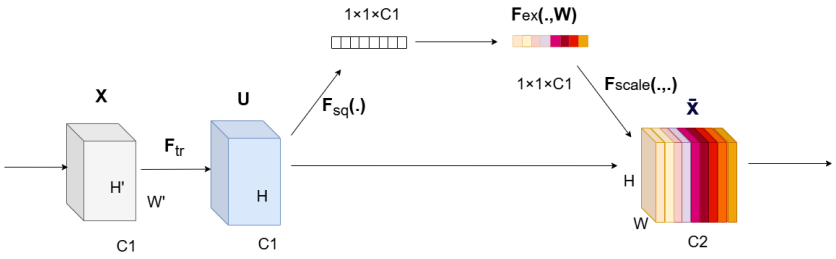


Figure 4. The Squeeze-Excitation module.

The first operation, known as squeezing, is responsible for feature compression in the spatial dimension. It turns each two-dimensional feature channel into a single real number, so creating a somewhat global receptive field. The output dimension is equal to the number of input feature channels, denoted as C1. It is highly beneficial in various activities as it represents the worldwide distribution of the response on the feature channel and allows the layers close to the input to also obtain the global receptive field. The second operation is excitation, which is analogous to the gating process seen in recurrent neural networks. The parameter w assigns weights that explicitly represent the relationship between the feature channels for each individual feature channel. The final step in re-calibrating the original features in the channel dimension is the reweighting operation. This operation utilizes the output weight of the excitation operation as a crucial indicator for each feature channel. It then applies this weight to the previous feature channels through multiplication, one channel at a time. The attention mechanism [37] is achieved through two crucial steps: the fully connected layer and feature multiplication fusion. Assuming that a fully connected layer and global pooling are employed, the input image with dimensions H $\times$ W $\times$ C1 is extended to a 1 $\times$ 1 $\times$ C1 size. The initial image is subsequently multiplied to allocate weights to each channel.

2.3. Loss Function

Cross-entropy loss is a metric for calculating the discrepancy between the actual and predicted maps. The cross-entropy loss function allows the network to have a faster gradient descent process during the first training stage, preventing gradient disappearance, when compared to other loss functions, like the mean square error and the L1 loss [38]. As shown in Formula (1), the divergence between the two probability distributions, j and k (j)

representing the actual distribution and k representing the predicted distribution) is known as the cross-entropy in information theory. One can obtain the cross-entropy loss LCE by:

$$L_{CE}(\mathrm{j},k) = -\sum_{i=1}^{n} j(x_i)\log(k(x_i)) \tag{1}$$

In the water body extraction task, j (xi) is the ground truth label and k (xi) is the predicted water body map, as shown in Formula (2). In the binary classification task (e.g., water body extraction), one can obtain the cross-entropy loss by:

$$L_{\mathrm{CE}}(\mathrm{j},k) = -[j\log k + (1-j)\log(1-k)] \tag{2}$$

To process the resulting feature map, the Softmax function and cross-entropy loss function are typically employed. The cross-entropy loss is computed after adding the predicted values of all the classes to one. The model's prediction improves with a decreasing cross-entropy value [39]. The dice coefficient, a commonly used set similarity measure function, is employed to assess the similarity between two samples. This coefficient ranges from 0 to 1 in Formula (3), as follows:

$$DiceCoefficient = \frac{2 \mid X \cap Y \mid}{\mid X \mid + \mid Y \mid} \tag{3}$$

where $\mid X \mid \cap \mid Y \mid$ represents the intersection of the $\mid X \mid$ and $\mid Y \mid$ sets, and $\mid X \mid$ and $\mid Y \mid$ represent the number of its elements, respectively. In the segmentation task, $\mid X \mid$ and $\mid Y \mid$ represent the segmented ground truth and the predict mask in Formula (4). Furthermore, the formula of the Dice loss is as follows:

$$DiceLoss = 1 - \frac{2 \mid X \cap Y \mid}{\mid X \mid + \mid Y \mid} \tag{4}$$

The enhanced model training incorporates the W-Dice loss, and the parameter is used to modify the Dice loss weight during backpropagation. The W-Dice loss and cross-entropy loss make up the two components of the overall model loss in Formula (5). In this experiment, we set the parameter value to 0.5, meaning that the optimal water body extraction effect is achieved when the ratio of the cross-entropy loss to the Dice loss is 2:1.

$$Loss = L_{CE} + \partial L_{Dice} \tag{5}$$

2.4. Evaluation Metrics

To assess our model's performance, we utilized three commonly-used evaluation metrics: the F1-score, Overall Accuracy (OA), and Mean Intersection over Union (mIoU). True positive (TP), false positive (FP), true negative (TN), and false negative (FN) are the elements of the confusion matrix from which the recall rate (Recall) and the precision indicator (Precision) can be obtained, as shown in Formulas (6)–(8). The F1-score can be obtained as:

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 - score = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{8}$$

The ratio of the results that the model properly predicts on all test sets to the total dataset is known as the overall precision in Formula (9). OA is best defined as:

$$OA = \frac{TP}{TP + FP + TN + FN} \tag{9}$$

The Intersection over Union (IoU) ratio between the prediction map and the ground truth map for each class can be calculated as follows:

$$IoU = \frac{TP}{TP + FP + FN} \tag{10}$$

In the water body extraction task, as shown in Formula (10), TP denotes that the number of pixels in the water body is the same as the ground truth value and the prediction result, while TP + FP + FN denotes the number of pixels in their union region [40]. Furthermore, the average IoU value obtained from all classes of the remote sensing images is referred to as the mIoU.

We use the weight file size obtained after model training as an indicator to measure the number of model parameters, and GFLOPS as an indicator to measure the speed of model training and inference. That is, the size of the weight file obtained after SE-ResUnet training is used to evaluate the number of model parameters, measured in MB. The smaller the space occupied, the lighter the model parameters, and GFLOPS are used to evaluate the efficiency of model training and inference. GFLOPS is the Giga Floating Point Operations Per Second, which means 1 billion floating-point operations per second. The larger the GFLOPS, the faster the model training and inference speed.

2.5. Dataset Description

The first self-built dataset used for the experiment consists of four main components of remote sensing images data. The UC-Merced Land-Use dataset is where the first component is taken from, including 99 images in total, each measuring 256 by 256 pixels and having a resolution of 0.3 m. The UC-Merced Land-Use dataset comprises 100 distinct image categories, providing a comprehensive collection of 21 levels of land use remote sensing data. The river images that were taken from the WHU-RS19 dataset make up the second component, including 56 images in total, each measuring 600 by 600 pixels and having a resolution of 0.5 m. The river images, extracted from the AID dataset, constitute the third component, including 410 images in total, each having a resolution of 0.5–8 m and measuring 600 by 600 pixels. The remote sensing image dataset AID was released by the Wuhan University and Huazhong University of Science and Technology in 2017, and contains a total of 30 categories of images, each containing 220–420 images and totaling 10,000 images. The remote sensing image data of the Yellow River Basin are mainly collected from the Gaofen-2 satellite, with a total of 1000 sample images. The image of one scene measures 6920 by 7300 pixels and the resolution is 3 m. In order to facilitate model training, the collected multi-scene images are segmented into small blocks, which measure 256 by 256 pixels. In the early stage of the experiment, the visual interpretation method and the 'labelme' tool were used to uniformly annotate and name these images, so as to produce images and labels for subsequent model training. The river semantic annotation format follows the specification of the VOC dataset, and is uniformly called Dataset1 in the study, as shown in Figure 5.

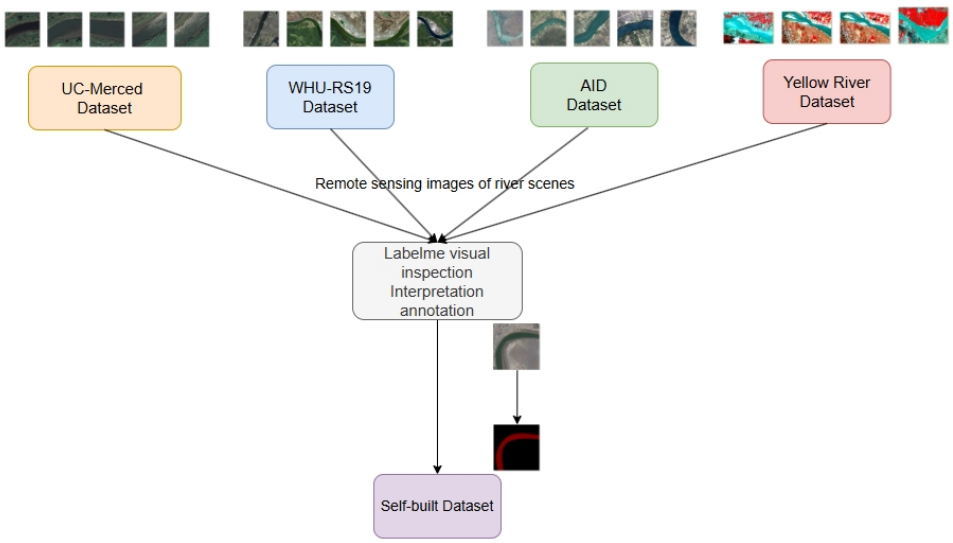


Figure 5. Dataset1 remote sensing image data collection and labeling procedure.

The Kaggle Remote Sensing Interpretation Public Dataset provides a second remote sensing image dataset for model training and generalization. The optical remote sensing image in question is part of a resource series and possesses a high resolution of 2 m. It is used to extract the water surface within the study area and is mostly directed towards domestic complicated water bodies. There are two categories in the dataset: background and water body. Each sample consists of two PNG files, Image and Label, that have identical file names. The Label file is in the single-channel 8-bit format, with a background pixel value of 0 and a water surface information value of 255, and the sample Image file is in 3-channel PNG format. The image size is 1024 × 1024 pixels, and the dataset contains a total of 2000 sets of data. All references to this dataset in the research are to Dataset2, as shown in Table 1.

Table 1. The spatial resolution, size, and quantity information of the datasets used in the experiment, Dataset2 denotes the Kaggle Remote Sensing Interpretation Public Dataset.

Dataset Name	Spatial Resolution	Size	Number
UC-Merced	0.3	256 × 256	99
WHU-RS19	0.5	600 × 600	56
AID	0.5–8	600 × 600	410
Yellow River	0.8	256 × 256	1000
Dataset2	2	1024 × 1024	2000

The design of the SE-ResUnet model takes into account the problem of adapting to the input of different image sizes. For the input of images of different sizes, the patch will be uniformly adjusted to the standard size of 512 × 512. If the input image size is larger than the standard size, the data preprocessing will slide-crop the image, and then train the cropped fragment in batches. If the input image is smaller than the standard size, it is filled to reach the standard patch size and then input to the network for training.

2.6. Data Enhancement

Figure 6 displays the samples and mask labels from Dataset1; Figure 7 displays the 2000 sets of samples and mask labels from Dataset2. We expanded Dataset1 by arbitrarily cropping, rotating, and mirroring images to fulfill the requirements of the deep learning model training. In the data preprocessing code section of the model, we use the Augmentor

library to perform data augmentation operations, such as cropping, rotating, and mirroring on the input source data, and expand the self-built dataset Dataset1 by controlling the number of remote sensing images used for training through parameter control. This study aimed to improve the model's generalizability by using geometric and color data enhancement approaches. Geometric data enhancement includes operations such as horizontal, vertical, and mixed flipping, while color data enhancement includes operations such as photometric distortion [41]. Photometric distortion adds noise to an image while adjusting its brightness, chroma, contrast, and saturation. The entire dataset covers a variety of water body types, such as lakes, rivers, canals, ponds and reservoirs, etc., as seen in Figures 6 and 7.

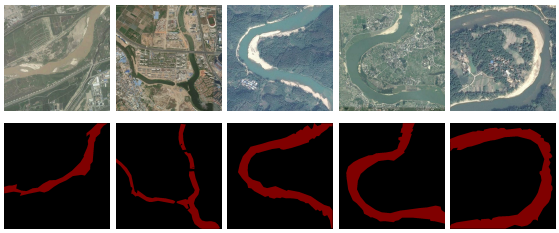


Figure 6. Original Dataset1 images and visual interpretation label images.

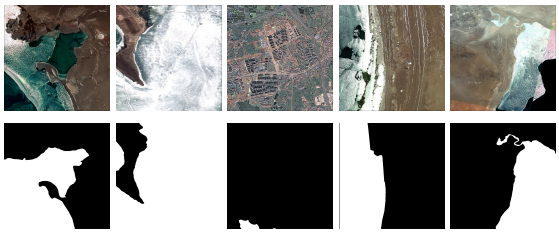


Figure 7. Original Dataset2 images and mask label images.

3. Results

3.1. Environment Settings

Our experiments were performed in a server environment (Intel Core w2223 quad core CPU and NVIDIA RTX 4070Ti GPU). The deep learning frameworks PyTorch 1.10.1 and Python 3.7 were used for all experiments, and CUDA 11.3 and CUDNN8.0 were used to speed up the training and inference process. The Adam optimizer was employed for training, with an initial learning rate of 1e-4, a momentum parameter of 0.9, and a learning rate reduction strategy that optimizes the objective function using the cosine annealing descent method. There are 1000 training iterations in total. In the first 300 iterations, the backbone network's training parameters were retrieved by freezing features, and in the final 700 iterations, the training parameters were unfrozen. After data enhancement, we obtained two water body datasets, namely, Dataset1 and Dataset2, which contained a total of 3000 sets of original images and mask samples. The dataset was divided into a training set, validation set, and test set at a ratio of 7:2:1. Subsequent contrast experiments with different models and ablations will be performed on these two datasets. The software and hardware environments used for specific experiments are given in Tables 2 and 3.

Table 2. Hardware environment.

	Hardware Environment
Operating System	Ubuntu20.04LTS
CPU	Intel-w2223 Quad Core 3.6 Ghz
GPU	NVIDIA GeForce RTX4070Ti (12,288 MB)
RAM	64 GB

Table 3. Software environment.

Software Environment	**Version**
PyTorch/torchvision	1.10.1/0.11.2
CUDA/CUDNN	11.3/8.0
python	3.7.13
Opencv-python	4.1.2.30
pillow	8.2.0
pyYAML	5.4.1
scipy	1.2.1
tqdm	4.60.0
numpy	1.17.0
matplotlib	3.1.2
setuptools	65.6.3
tensorboard	2.11.2

3.2. Ablation Study

We conducted a comparative analysis between the SE-ResUnet and the original Unet model in order to assess their performance in large-scale water body extraction tasks. Furthermore, a sequence of ablation tests were conducted to confirm the efficacy of including the SRB module and the weighted Dice loss function [15]. In particular, training was performed on the two previously introduced datasets using the original Unet feature extraction network as the baseline network, introducing the SRB module and the W-Dice loss in turn. The dataset is labeled and divided before being introduced into the model in batches for training. Tables 4 and 5 display the comparison results.

Table 4. Evaluation metrics from the ablation study on Dataset1 (SRB represents the residual structure with channel attention, and W-Dice represents the weighted Dice loss function).

Dataset1	**mIoU**	**OA**	**F1-Score**
Unet	94.44	98.41	97.82
+SRB	94.60	98.40	97.70
+W-Dice	94.17	98.34	97.62
+W-Dice + SRB	94.82	98.53	97.90

Table 5. Evaluation metrics from the ablation study on Dataset2 (SRB represents the residual structure with channel attention, and W-Dice represents the weighted Dice loss function).

Dataset2	**mIoU**	**OA**	**F1-Score**
Unet	96.67	98.31	98.31
+SRB	96.72	98.34	98.35
+W-Dice	96.22	98.09	98.10
+W-Dice + SRB	97.02	98.50	98.51

Tables 4 and 5 show the ablation experiment results on Dataset1 and Datset2. SRB represents the experimental results of introducing residual structures with attention. +W-

DICE represents the experimental results obtained by introducing the weighted Dice loss function. +W-DICE + SRB represents the experimental results of introducing both simultaneously. Upon careful examination of the results, it can be seen that introducing only one optimization strategy results in limited improvement on the model, even resulting in negative optimization when only the weighted Dice loss function is introduced. Introducing two optimizations simultaneously achieved the best water extraction effect, proving that the joint weighted Dice loss function and the residual structure with attention are effective in improving the model. The training phase statistics show that using just the original Unet network with Resnet as the backbone and adding the W-Dice loss function has a poor training effect. The tables show that the three metrics, mIoU, OA, and F1-score, dramatically decrease after the W-Dice loss is introduced to the Unet, whether on Dataset1 or Dataset2. There is a clear improvement in the training effect after SRB is added to the network. Compared to the original Unet, mIoU has a 0.1–0.2% improvement on both datasets. It is noteworthy that the training effect is greatly enhanced upon simultaneous introduction of the SRB module and the weighted Dice loss function. The model achieved the best mIoU on both datasets. This study confirms the efficacy of integrating SE attention and the weighted Dice loss function to enhance the accuracy of water body extraction in the model. Furthermore, it suggests that this combination is a promising approach for optimizing the model.

The inference results of some images from the ablation study on Dataset1 and Dataset2, GroundTruth, Unet, Unet + SRB, Unet + W-Dice and Unet + SRB + W-Dice are shown in Figures 8 and 9 from left to right, respectively. (a), (b), (c), (d), (e), and (f), respectively, represent Raw-Images, GroundTruth, Unet, Unet + SRB, Unet + W-Dice and Unet + SRB + W-Dice.



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**)

Figure 8. Ablation study on Dataset1. (**a**–**f**), respectively, represent Raw-Images, GroundTruth, Unet, Unet + SRB, Unet + W-Dice and Unet + SRB + W-Dice.

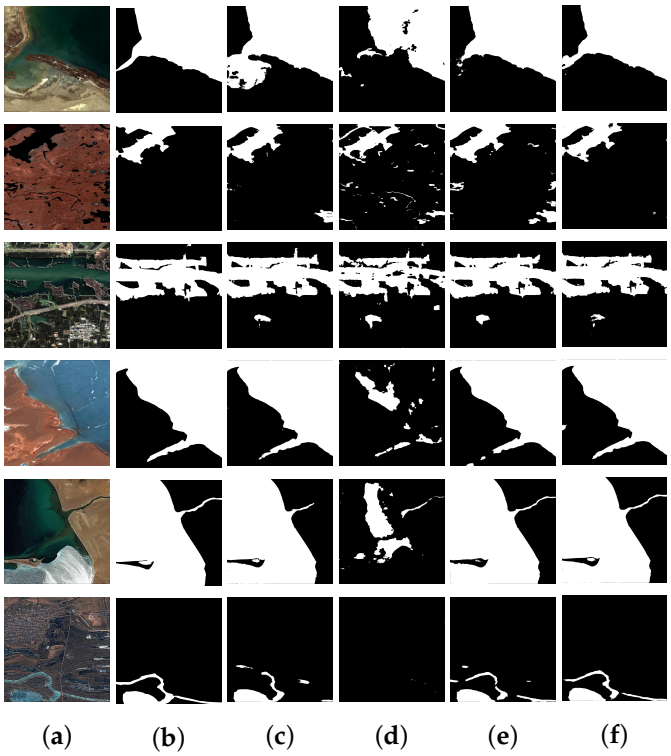


Figure 9. Ablation study on Dataset2. (**a**–**f**), respectively, represent Raw-Images, GroundTruth, Unet, Unet + SRB, Unet + W-Dice and Unet + SRB + W-Dice.

It is easy to see the results of the ablation experiment on Dataset 1 and Dataset 2. The introduction of the SRB module significantly improves the extraction effect of rivers compared to Unet. However, when only the W-Dice module is introduced, the water bodies at the end of the river branches cannot be extracted well. After introducing both the SRB module and the W-Dice mechanism, not only are the extracted river edges smoother, but the water bodies at the center of the river can also be identified well.

3.3. Contrast Experiment

To confirm that the model's performance had indeed improved when compared to other models, we trained SE-ResUnet and HRnet, Deeplabv3+, PSPnet, and the original Unet model [42] on two datasets and compared the results, as shown in Tables 6 and 7. The SE-ResUnet obtained the best evaluation metrics, indicating that it is more effective in water body extraction.

Table 6. Evaluation metrics from model performance comparison on Dataset1.

Dataset1	mIoU	Accuracy	F1-Score
HRnet	94.06	98.30	97.58
Deeplabv3+	94.05	98.30	97.58
Unet	94.44	98.41	97.82
PSPnet	93.39	98.10	97.30
SE-ResUnet	94.82	98.53	97.90

Table 5 displays the evaluation metrics of various models on Dataset 1 following 1000 training rounds. With a maximum mIoU of 94.82%, SE-ResUnet outperformed the original Unet semantic segmentation model by about 0.4%. Furthermore, SE-ResUnet's OA value of 98.53% was greater than the PSPNet model's OA value of 98.10%. At 97.9%, SE-ResUnet outperformed the Deeplabv3+ and HRNet models on the F1-score. These

findings demonstrate the effectiveness of the Unet improvement performed in this study, as well as the superior extraction effect on this water body dataset when compared to other semantic segmentation methods.

Figure 10 displays the outcomes of the model performance evaluation on Dataset1, including GroundTruth, Deeplabv3+, HRNet, PSPNet, and SE-ResUnet. (a), (b), (c), (d), (e), and (f), respectively, represent Raw-Images, GroundTruth, Deeplabv3+, HRNet, PSPNet, and SE-ResUnet.

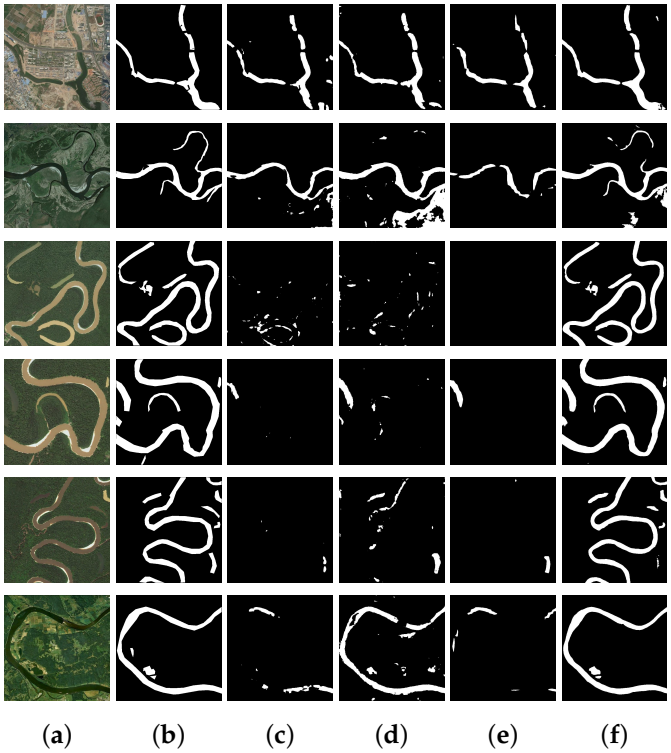


Figure 10. Contrast experiment on Dataset1. (**a**–**f**), respectively, represent Raw-Images, GroundTruth, Deeplabv3+, HRNet, PSPNet, and SE-ResUnet.

We can see from the comparative experiments on Dataset1 that compared with Ground Truth, Deeplabv3+, HRNet, and PSPNet, SE-ResUnet cannot extract the river center water body completely, and there is a different degree of missing extraction and over-extraction of the water body at the edge of the image. However, SE-ResUnet can extract the river edge more smoothly and the river center water body can be well recognized because it introduces the SRB module and W-Dice mechanism simultaneously, and the attention mechanism can pay more attention to the water body information between channels, as shown in Figure 9.

Table 7. Evaluation metrics from model performance comparison on Dataset2.

Dataset2	mIoU	Accuracy	F1-Score
HRnet	97.05	98.51	98.52
Deeplabv3+	96.64	98.30	98.31
Unet	96.67	98.31	98.31
PSPnet	94.94	97.42	97.44
SE-ResUnet	97.30	98.57	98.56

It was previously stated in the dataset description section that Dataset2 has a larger data volume than Dataset1. The training effect on Dataset2 is superior to that on Dataset1

from a variety of data indicator perspectives, and the three evaluation metrics (i.e., mIoU, OA, and F1-score) have increased numerically. To be more precise, SE-ResUnet scored 97.3%, 98.57%, and 98.56% for each of the mIoU, OA, and F1-score metrics. The mIoU, OA, and F1-score of the modified Unet increased by 0.63%, 0.26%, and 0.25%, respectively, compared to the original Unet. They grew by 2.36%, 1.15%, and 1.12%, respectively, compared to PSPNet; 0.25%, 0.06%, and 0.04% compared to HRNet; and 0.66%, 0.27%, and 0.25% compared to Deeplabv3+. It follows that the Unet network with the SRB module performs better in extracting the water body's semantic information, and the weighted Dice loss function speeds up the entire training process to the point of convergence.

The results of the model performance comparison on Dataset2, including GroundTruth, Deeplabv3+, HRNet, PSPNet, and SE-ResUnet are shown in Figure 11 in left to right order. (a), (b), (c), (d), (e), and (f), respectively, represent Raw-Images, GroundTruth, Deeplabv3+, HRNet, PSPNet, and SE-ResUnet.

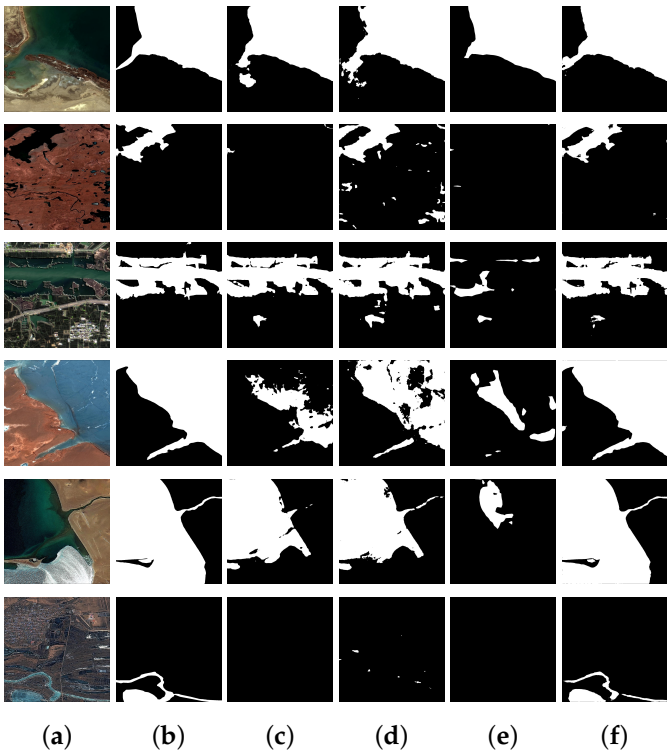


Figure 11. Contrast experiment on Dataset2. (**a**–**f**), respectively, represent Raw-Images, GroundTruth, Deeplabv3+, HRNet, PSPNet, and SE-ResUnet.

From the comparative experiments on Dataset2, we can see that compared with Ground Truth, PSPNet cannot extract the edge protruding type water body completely, and Deeplabv3+ and HRnet have a different degree of missing extraction and over-extraction of the water body at the edge of the image, and the main trunk of the extracted water body is also rough, as shown in Figure 11. However, SE-ResUnet can extract the river edge more smoothly and the river center water body can be well recognized because it introduces the SRB module and W-Dice mechanism simultaneously, and the weighted Dice loss function can pay more attention to the important water body information between channels. The river edge extracted is more smooth, and the water body at the edge protrusion and the river center can be well extracted.

Table 8 illustrates that after training, SE-ResUnet has the largest number of parameters, the largest weight file (177.5 MB), and the fastest inference speed (252.31 GFLOPS).

Table 8. Parameters and calculation speed of models involved in performance comparison.

	Param (MB)	GFLOPS (G)
HRNet	37.5	80.18
Deeplabv3+	22.4	52.87
Unet	167.9	200.57
PSPNet	25.5	55.6
SE-ResUnet	177.5	252.31

4. Discussion

This paper proposes the SE-ResUnet method for water extraction from satellite images. This method uses Unet's skipping structure for multi-scale feature extraction and compensation, the SE attention mechanism with residual structure for down-sampling, and the W-Dice loss function for loss calculation during training [35]. We conducted numerous performance comparison experiments and ablation studies on two datasets. The experiments demonstrated that, while adjusting the imbalance between water bodies and background distribution in some sample data, we can effectively acquire the semantic information of the river center and marginal water bodies in high-resolution remote sensing images.

The performance of SE-ResUnet was evaluated by comparing it with other semantic segmentation networks, such as Unet, PSPNet, HRNet, and Deeplabv3+. The evaluation was conducted using a self-built dataset and a publicly available remote sensing interpretation water body dataset. The comparison results demonstrated that SE-ResUnet shows superior water body extraction performance. The mIoU, OA, and F1-score of SE-ResUnet are improved by 0.38%, 0.12%, and 0.08%, respectively, on the generated dataset, while that on the public remote sensing dataset showed enhancements of 0.63%, 0.26%, and 0.25%, respectively, when compared with the original Unet. Based on the results displayed in Figures 10 and 11, SE-ResUnet was able to effectively extract target water bodies from those images containing spectral noise and other shadow interference, while fully extracting water bodies that are challenging to identify at the river's edge and center. The edge and ground truth also matched very well. The ablation study results, depicted in Figures 8 and 9, indicate a significant imbalance in the distribution of positive and negative samples between the water body and background in narrow rivers. Additionally, the SE-ResUnet model performs effectively on these specific test sets. The smoothness of the extracted water body's edge indicates that the SE channel attention mechanism with residual structure is more effective in capturing the relationship between different channels, especially the information at the image edge. However, solely using the W-Dice loss function is not enough to adequately supervise and focus on the features across channels. The comparison results and ablation study findings on the two datasets suggest that the weighted loss function can enhance the model's generalization, and help the SE-ResUnet model perform end-to-end water body extraction from multi-resolution remote sensing satellite images.

Our ablation experiment verified the feasibility of combining an attention mechanism and weighted loss function to optimize water extraction tasks, and achieved good experimental results on the two datasets used in the paper experiment. In the future, we will consider conducting experiments on more remote sensing image datasets and tasks to expand the application scope of this optimization strategy.

At the same time, it should be noted that there is still room for further experimentation with the Resnet backbone network selected by our proposed model. Additionally, the number and types of datasets are still limited, and there are not enough ways to enhance the data. This may result in poor generalization of our model on other datasets. However, based on our experimental results, our optimization strategy has shown good performance.

Despite achieving a good segmentation performance on Dataset1 and Dataset2, our suggested SE-ResUnet method still has significant limitations that will direct our future studies. First , there are a lot of parameters in the Unet model that are introduced into the SE attention mechanism, which means that training and deployment will be very expensive. We intend to use model pruning and distillation approaches to lower the model's complexity and processing requirements in order to increase efficiency in our follow-up research. The paper by Pandey et al. [43] includes a list of other performance metrics that we will attempt to use in other models to describe model performance and training effectiveness. Second, the left single CNN structure is not sensitive enough to fine features, especially when dealing with water body boundaries. The main reason is that this mode is an encoder-decoder network, which might result in less smooth water body boundary lines. Finally, a parameter is introduced to achieve a balance between the cross-entropy loss and Dice loss when computing the objective function using the W-Dice loss function. However, it is important to note that this parameter is not trained automatically. Due to the expense of training time, the experiment achieves its nearly optimal water extraction effect at a value of 0.5 after multiple manual adjustments. Subsequently, we can attempt to fine-tune it to the hyper-parameter of the network, and through the utilization of the training data, it will autonomously learn to reach a parameter threshold that is nearly optimal. Ozdemir et al. [44] used the automatic mask generator method of SAM to segment images and zero shot classification of fragments using CLIP. The proposed method accurately depicts water bodies under complex environmental conditions. In the future, combining SAM with massive remote sensing image data to complete segmentation and extraction tasks will be a research direction.

5. Conclusions

This paper proposes a new network structure, SE-ResUnet, which is used to extract small water bodies in the Yellow River Basin. Our experiments show that it has good generalization ability for water extraction tasks in large areas such as rivers, lakes, and reservoirs. We have derived the following conclusions:

(1) The distinguishing characteristic of SE-ResUnet is the incorporation of the SE attention mechanism with the residual structure during the down-sampling process. This allows for the learning of the Unet's jumping structure, which facilitates multi-scale feature extraction and compensation.
(2) SE-ResUnet enhances the ability to extract features from small water bodies in the Yellow River Basin, including large areas of water, such as rivers, lakes, and reservoirs. The W-Dice loss function was included in the training procedure.
(3) On a generated dataset and a public remote sensing interpretation water body dataset, SE-ResUnet was compared with semantic segmentation networks, including Unet, PSPNet, HRNet, and Deeplabv3+. The comparison results demonstrated that the SE-ResUnet shows superior water body extraction performance and generalization.
(4) SE-ResUnet's inference time cost is similar to the conventional Unet, demonstrating the high accuracy and high efficiency of our method. This suggests that the utilization of the channel attention mechanism in conjunction with the weighted Dice loss function is highly effective in enhancing the algorithm.

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

Funding: It should be noted that this research was funded by Science and Technology Innovation 2030 (2023ZD0120600), major science and technology project of Henan Province, China [221100210600], Higher Education Teaching Reform Research and Practice Project of NCWU (2024XJGXM061), Henan province water conservancy science and technology research projects (GG202404 GG202405, GG202338).

Data Availability Statement: The datasets that support the findings of this study are available on request from the corresponding author upon reasonable request.

Acknowledgments: We thank the laboratory for providing experimental equipment, including GPUs, and experimental datasets and for their support.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Zhang, Z.; Wang, W.; Zhang, X.; Zhang, H.; Yang, L.; Lv, X.; Xi, X. A Harmony-Based Approach for the Evaluation and Regulation of Water Security in the Yellow River Water-Receiving Area of Henan Province. *Water* **2024**, *16*, 2497. [CrossRef]
2. Duan, M.; Qiu, Z.; Li, R.; Li, K.; Yu, S.; Liu, D. Monitoring Suspended Sediment Transport in the Lower Yellow River using Landsat Observations. *Remote Sens.* **2024**, *16*, 229. [CrossRef]
3. Nagaraj, R.; Kumar, L.S. Extraction of Surface Water Bodies using Optical Remote Sensing Images: A Review. *Earth Sci. Inform.* **2024**, *17*, 893–956. [CrossRef]
4. McFeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [CrossRef]
5. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033. [CrossRef]
6. Yan, P.; Zhang, Y. A Study on Information Extraction of Water System in Semi-arid Regions with the Enhanced Water Index (EWI) and GIS Based Noise Remove Techniques. *Remote Sens. Inf.* **2007**, *6*, 62–67.
7. 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]
8. Zhang, Z.; Guo, X.; Cao, L.; Lv, X.; Zhang, X.; Yang, L.; Zhang, H.; Xi, X.; Fang, Y. Multi-Scale Variation in Surface Water Area in the Yellow River Basin (1991–2023) Based on Suspended Particulate Matter Concentration and Water Indexes. *Water* **2024**, *16*, 2704. [CrossRef]
9. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [CrossRef]
10. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
11. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
12. Huang, X.; Hu, T.; Li, J.; Wang, Q.; Benediktsson, J.A. Mapping Urban Areas in China Using Multisource Data with a Novel Ensemble SVM Method. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4258–4273. [CrossRef]
13. Koda, S.; Zeggada, A.; Melgani, F.; Nishii, R. Spatial and Structured SVM for Multilabel Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5948–5960. [CrossRef]
14. Kittler, J.; Hatef, M. On combining classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 226–239. [CrossRef]
15. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef]
16. Wu, P.; Fu, J.; Yi, X.; Wang, G.; Mo, L.; Maponde, B.T.; Liang, H.; Tao, C.; Ge, W.Y.; Jiang, T.T. Research on water extraction from high resolution remote sensing images based on deep learning. *Front. Remote Sens.* **2023**, *4*, 1283615. [CrossRef]
17. 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, Munich, Germany, 8–14 September 2018.
18. Walker, G.; Prowse, T.D.; Dibike, Y.B.; Bonsal, B.R. Climate Change Impacts on Precipitation Patterns and Water Storage in the High-latitude Dry Interior Climate of Northern Alberta, Canada. In Proceedings of the Agu Fall Meeting, San Francisco, CA, USA, 3–7 December 2012.
19. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Xiao, B. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3349–3364. [CrossRef]
20. Miao, Z.; Fu, K.; Sun, H.; Sun, X.; Yan, M. Automatic Water-Body Segmentation from High-Resolution Satellite Images via Deep Networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 602–606 [CrossRef]

21. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
22. Sun, D.; Gao, G.; Huang, L.; Liu, Y.; Liu, D. Extraction of water bodies from high-resolution remote sensing imagery based on a deep semantic segmentation network. *Sci. Rep.* **2024**, *14*, 14604. [CrossRef]
23. Sun, Q.; Li, J. A method for extracting small water bodies based on DEM and remote sensing images. *Sci. Rep.* **2024**, *14*, 760. [CrossRef] [PubMed]
24. Cao, H.; Tian, Y.; Liu, Y.; Wang, R. Water body extraction from high spatial resolution remote sensing images based on enhanced U-Net and multi-scale information fusion. *Sci. Rep.* **2024**, *14*, 16132. [CrossRef]
25. Yan, P.; Fang, Y.; Chen, J.; Wang, G.; Tang, Q. Automated Extraction for Water Bodies Using New Water Index from Landsat 8 OLI Images. *Acta Geod. Cartogr. Sin.* **2023**, *6*, 59–75.
26. 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. [CrossRef]
27. Zhao, B.; Wu, J.; Han, X.; Tian, F.; Liu, M.; Chen, M.; Lin, J. An improved surface water extraction method by integrating multi-type priori information from remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *124*, 103529. [CrossRef]
28. 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]
29. Mishra, V.K.; Chaudhary, P.K.; Pant, T. Image fusion based approach of water extraction from spectrally mixed water regions belonging to the sources of varying nature. *Multimed. Tools Appl.* **2023**, *82*, 39783–39795. [CrossRef]
30. Wang, Y.; Li, Y.; Wang, D. Extraction of small water body information based on Res2Net-Unet. In Proceedings of the 2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM), Seoul, Republic of Korea, 3–5 January 2023; pp. 1–5.
31. Shihao, A.; Xiaoping, R. A High-Precision Water Body Extraction Method Based on Improved Lightweight U-Net. *Remote Sens.* **2022**, *14*, 4127. [CrossRef]
32. 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 International Publishing: Cham, Switzerland, 2015; pp. 234–241.
33. Moknatian, M. Spatiotemporal Water Body Change Detection Using Multi-temporal Landsat Imagery: Case Studies of Lake Enriquillo and Lake Azuei. In Proceedings of the AGU Fall Meeting, San Francisco, CA, USA, 14–18 December 2015.
34. Tong, X.Y.; Xia, G.S.; Lu, Q.; Shen, H.; Zhang, L. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sens. Environ.* **2020**, *237*, 111322. [CrossRef]
35. 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.
36. Hu, J.; Shen, L.; Sun, G.; Albanie, S. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**.
37. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762.
38. Wang, L.; Lee, C.Y.; Tu, Z.; Lazebnik, S. Training Deeper Convolutional Networks with Deep Supervision. *arXiv* **2015**, arXiv:1505.02496.
39. Yuan, K.; Zhuang, X.; Schaefer, G.; Feng, J.; 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]
40. Yu, L.; Wang, Z.; Tian, S.; Ye, F.; Ding, J.; Kong, J. Convolutional Neural Networks for Water Body Extraction from Landsat Imagery. *Int. J. Comput. Intell. Appl.* **2017**, *16*, 1750001. [CrossRef]
41. Nie, C. Rich CNN Features for Water-Body Segmentation from Very High Resolution Aerial and Satellite Imagery. *Remote Sens.* **2021**, *13*, 1912. [CrossRef]
42. Duan, Y.; Zhang, W.; Huang, P.; He, G.; Guo, H. A New Lightweight Convolutional Neural Network for Multi-Scale Land Surface Water Extraction from GaoFen-1D Satellite Images. *Remote Sens.* **2021**, *13*, 4576. [CrossRef]
43. Pandey, M.; Arora, A.; Arabameri, A.; Shukla, U. Flood Susceptibility Modeling in a Subtropical Humid Low-Relief Alluvial Plain Environment: Application of Novel Ensemble Machine Learning Approach. *Front. Earth Sci.* **2021**, *9*, 659296. [CrossRef]
44. Ozdemir, S.; Akbulut, Z.; Karsli, F.; Kavzoglu, T. Extraction of Water Bodies from High-Resolution Aerial and Satellite Images Using Visual Foundation Models. *Sustainability* **2024**, *16*, 2995. [CrossRef]

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.