



Article

SPT-UNet: A Superpixel-Level Feature Fusion Network for Water Extraction from SAR Imagery

Teng Zhao ^{1,2,3} , Xiaoping Du ^{1,2,*} , Chen Xu ^{1,2} , Hongdeng Jian ^{1,2} , Zhipeng Pei ⁴, Junjie Zhu ^{1,2}, Zhenzhen Yan ^{1,2} and Xiangtao Fan ^{1,2}

- ¹ International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China; zhaoteng22@mails.ucas.ac.cn (T.Z.); xuchen@aircas.ac.cn (C.X.); jianhd@radi.ac.cn (H.J.); zhujunjie@aircas.ac.cn (J.Z.); yanzz202057@aircas.ac.cn (Z.Y.); fanxt@aircas.ac.cn (X.F.)
- ² Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
- ³ University of Chinese Academy of Sciences, Beijing 100049, China
- ⁴ School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541000, China; 22032303081@mails.guet.edu.cn
- * Correspondence: duxp@aircas.ac.cn

Abstract: Extracting water bodies from synthetic aperture radar (SAR) images plays a crucial role in the management of water resources, flood monitoring, and other applications. Recently, transformer-based models have been extensively utilized in the remote sensing domain. However, due to regular patch-partition and weak inductive bias, transformer-based models face challenges such as edge serration and high data dependency when used for water body extraction from SAR images. To address these challenges, we introduce a new model, the Superpixel-based Transformer (SPT), based on the adaptive characteristic of superpixels and knowledge constraints of the adjacency matrix. (1) To mitigate edge serration, the SPT replaces regular patch partition with superpixel segmentation to fully utilize the internal homogeneity of superpixels. (2) To reduce data dependency, the SPT incorporates a normalized adjacency matrix between superpixels into the Multi-Layer Perceptron (MLP) to impose knowledge constraints. (3) Additionally, to integrate superpixel-level learning from the SPT with pixel-level learning from the CNN, we combine these two deep networks to form SPT-UNet for water body extraction. The results show that our SPT-UNet is competitive compared with other state-of-the-art extraction models, both in terms of quantitative metrics and visual effects.

Keywords: SPT-UNet; CNN; transformer; superpixel; water exaction



Citation: Zhao, T.; Du, X.; Xu, C.; Jian, H.; Pei, Z.; Zhu, J.; Yan, Z.; Fan, X. SPT-UNet: A Superpixel-Level Feature Fusion Network for Water Extraction from SAR Imagery. *Remote Sens.* **2024**, *16*, 2636. <https://doi.org/10.3390/rs16142636>

Academic Editor: Massimo Menenti

Received: 13 June 2024

Revised: 6 July 2024

Accepted: 17 July 2024

Published: 18 July 2024



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

1. Introduction

Water is an indispensable and invaluable resource for human survival and social development. Water bodies extracted from remote sensing images have gained substantial popularity and usage in diverse fields such as surface water management and flood monitoring [1–3]. Optical satellites typically provide high spatial resolution images for water extraction, but they are often affected by thick clouds and rainfall, resulting in poor image quality. Recent studies [4,5] have highlighted the potential of GNSS-R for mapping inland water bodies. However, this technology is still in its early stages of development. In contrast, synthetic aperture radar (SAR) images are an ideal data source for water extraction due to their all-day, all-weather imaging capabilities and established technology [6–9]. However, the explosive growth of SAR images has presented challenges such as complex surface environments (e.g., speckle noise, shadow, and water-like surfaces). Consequently, in the era of big data [10,11], building a robust and accurate SAR image water extraction network has substantial practical importance.

In the previous decade, threshold-based methods have been widely used for water body extraction because water has weak scattering intensity in SAR images [12–15].

The underlying assumption of threshold-based methods is that the histogram of an SAR image comprises water and non-water pixels, represented by two partially overlapped distributions. The point of intersection between these distributions indicates the optimal threshold for minimizing errors. However, for SAR images with small water areas, the bimodal assumption is not satisfied, which makes it challenging to select the threshold. To alleviate this issue, local thresholding approaches have been introduced [16,17]. Despite the substantial enhancements in the precision of water body extraction achieved through the aforementioned threshold-based approaches, it is still challenging to find a threshold that can effectively distinguish water bodies from shadow interfered areas and water-like surfaces in SAR images [18]. Additionally, machine learning-based methods are also limited in extracting water bodies due to the reliance on manually selected features and empirical parameters [19,20].

Deep learning methods have been utilized for many applications of remote sensing imagery [21–24], owing to their ability for end-to-end learning of shallow detailed features and deep semantic features, as well as their characteristics of local connectivity and weight sharing. Many well-established deep-learning networks have been successfully utilized for water extraction from SAR imagery, yielding promising results. For instance, Dai et al. [25] designed an enhanced water extraction model based on a bilateral segmentation network for water body extraction. Similarly, Zhang et al. [26] introduced a fully convolutional up-sampling pyramid network to address the resolution loss caused by high stride convolution in conventional convolutional neural network (CNN) while utilizing a fully convolutional conditional random field to tackle the issue of water boundary blurring. Despite the substantial advancements afforded by these networks in SAR image water extraction, their capability to capture target features remains limited to local ranges, as illustrated in Figure 1a. Consequently, they fail to ensure the integrity and continuity of the targets.

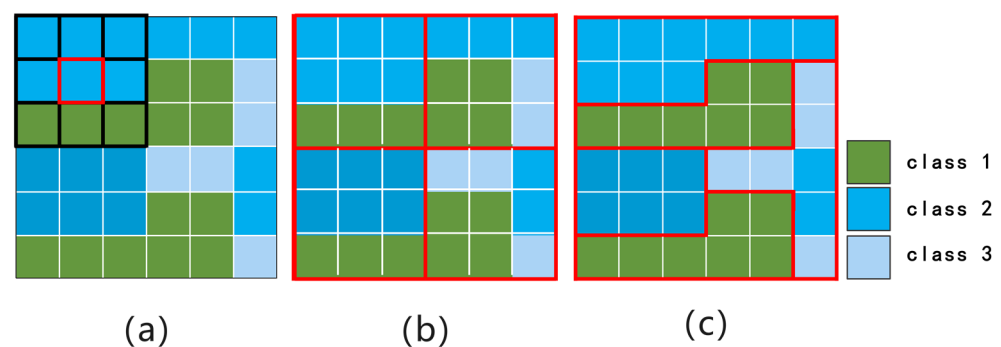


Figure 1. Illustrations of the three network processing modes. (a) The feature extraction of water bodies at the pixel level using CNN, where a black solid line box depicts the 3×3 Conv. (b) The feature extraction of water bodies at the patch level using the transformer-based models in the computer vision domain, where regular red solid line frames indicate patches (i.e., 3×3). (c) The feature extraction of water bodies at the superpixel level using SPT, where irregular red solid line frames represent superpixels.

Recently, the Transformer [27] has found extensive application in natural language processing (NLP) due to its capacity to capture global sequence semantic features. Inspired by the successes of NLP, many works have attempted to apply the transformer architecture in computer vision (CV) to capture the global contextual information of images [28–31]. For instance, the Vision Transformer (ViT) [28] has achieved state-of-the-art results in image classification tasks by applying the transformer architecture to images via regular patch partition. Recently, transformer-based models in the computer vision domain have also been utilized to extract features from the global range of SAR images and have obtained impressive classification performance [32,33]. However, when transformer-based models in the computer vision domain are used to extract water bodies from SAR images, there are two main challenges to be considered:

- (1) Predefined patch sizes (e.g., 16×16) are used for image partition, resulting in the issue of edge serration in water bodies extraction due to the heterogeneity within each patch;
- (2) The transformer-based models lack specific inductive biases, such as translation invariance and local correlation, which makes the model heavily rely on the dataset for accurate extraction of water bodies.

Therefore, we propose a Superpixel-based Transformer (SPT) to alleviate the above issues. Notably, superpixels are locally homogeneous regions composed of a collection of pixels with similar characteristics. Firstly, to overcome the problem of edge serration, the SPT substitutes the regular patch partition of the transformer-based models with irregular superpixel segmentation, as shown in Figure 1c. Additionally, to reduce the need for a large-scale water training dataset, the SPT implements a constrained optimization of the Multi-Layer Perceptron (MLP), putting forward a constrained MLP (C-MLP). Furthermore, inspired by heterogeneous networks [34,35], we incorporate a CNN and SPT into a unified network, termed SPT-UNet, where the CNN branch acquires pixel-level semantic features in the image to complement the features at the superpixel-level of the SPT branch.

Specially, the SPT-UNet mainly comprises four modules: the Superpixel Segmentation Module (SPSM), the Superpixel-level Transformer Sub-network (SPTsN), the Pixel-level Convolution Sub-network (PCsN), and the Feature Fusion Module (FFM). The SPSM mainly performs superpixel segmentation on SAR images. Subsequently, the SPTsN models the global spatial structure in images and forms semantic features of water bodies at the superpixel level. Meanwhile, the PCsN is utilized to extract pixel-level semantic features. Finally, the FFM incorporates both pixel- and superpixel-level features to enhance the performance of water extraction.

The primary contributions of the study are as follows:

- (1) To mitigate the issue of edge serration in water body extraction based on transformer-based models, we propose to use adaptive superpixels instead of regular patches;
- (2) To alleviate the dependence of transformer-based models on data, we introduce the C-MLP module, which imposes prior constraints on the SPT by utilizing a normalized adjacency matrix between superpixels;
- (3) Combining a CNN and SPT, we propose the SPT-UNet, which facilitates simultaneous learning on both the pixel and superpixel-level semantic features of water bodies. The experiments on the public dataset demonstrate that SPT-UNet performs competitively compared to other state-of-the-art extraction networks, which can partially alleviate the interference caused by terrain-induced shadows and water-like surfaces.

The rest of the paper is organized as follows. Section 2 introduces the SPT-UNet and provides a comprehensive overview of each part. Section 3 introduces the public water dataset involved in the experiments and the relevant experimental settings. In Section 4, we conducted a series of experiments and discussions. Finally, Section 5 provides the conclusions.

2. Proposed Method

The extraction task of water bodies is designed to identify whether each pixel is water. We define the SAR image as $I \in \mathbb{R}^{3 \times H \times W}$, where 3 corresponds to the bands (VV, VH, and SDWI) and H, W represents the height and width of the image, respectively. Notably, the VV and VH symbolize the polarization modes of the SAR image, and the SDWI symbolizes the Sentinel-1 Dual-Polarized Water Index to strengthen the features of water bodies [36]. The main aim of this study is to extract water bodies from I by training an extraction network of water bodies. To achieve this goal, we propose the SPT-UNet, as depicted in Figure 2, which comprises four modules: (1) the SPSM that generates the normalized adjacency matrix $\tilde{A}$ and the mapping matrix M ; (2) the SPTsN that performs global modeling of superpixels in the SAR image and forms water body features at the superpixel level; (3) the PCsN that performs the local modeling of pixels in the SAR image and forms water body features at

the pixel level; (4) the FFM that integrates the pixel- and superpixel-level semantic features of water bodies. The training process of SPT-UNet is shown in Algorithm 1. The four modules are detailed in the following subsections. Additionally, the hybrid weighted loss is also presented.

Algorithm 1: Training procedure for the SPT-UNet framework

Input: SAR dataset $Z_s = \{I^i\}_{i=1}^N$, DEM dataset $Z_d = \{D^i\}_{i=1}^N$, JRC-gsw dataset $Z_j = \{J^i\}_{i=1}^N$,
 Ground Label $L = \{L^i\}_{i=1}^N$, The number of Superpixels Z , the hyperparameter a .
 Output: Water bodies extraction results $E = \{E^i\}_{i=1}^N$.
 for $i = 1; i < N; i++$ do
 Randomly batch sample I^i from Z_s , Batch sample D^i consistent with I^i from Z_d , Batch sample J^i
 consistent with I^i from Z_j , Batch sample L^i consistent with I^i from L .
 // Stage 1 SPSM
 $\tilde{A}, M = \text{SPSM}(I^i)$
 // Stage 2 SPTsN
 $V^* = \text{SPTsN}(I^i, D^i, J^i, \tilde{A}, M)$
 // Stage 3 PCsN
 $F = \text{PCsN}(I^i)$
 // Stage 4 FFM
 $E^i = \text{FFM}(F, V^*)$
 // Calculate the segmentation loss
 $\text{Loss}_{all} = a \cdot \text{BCELoss}(E^i, L^i) + (1 - a) \cdot \text{DiceLoss}(E^i, L^i)$
 end for

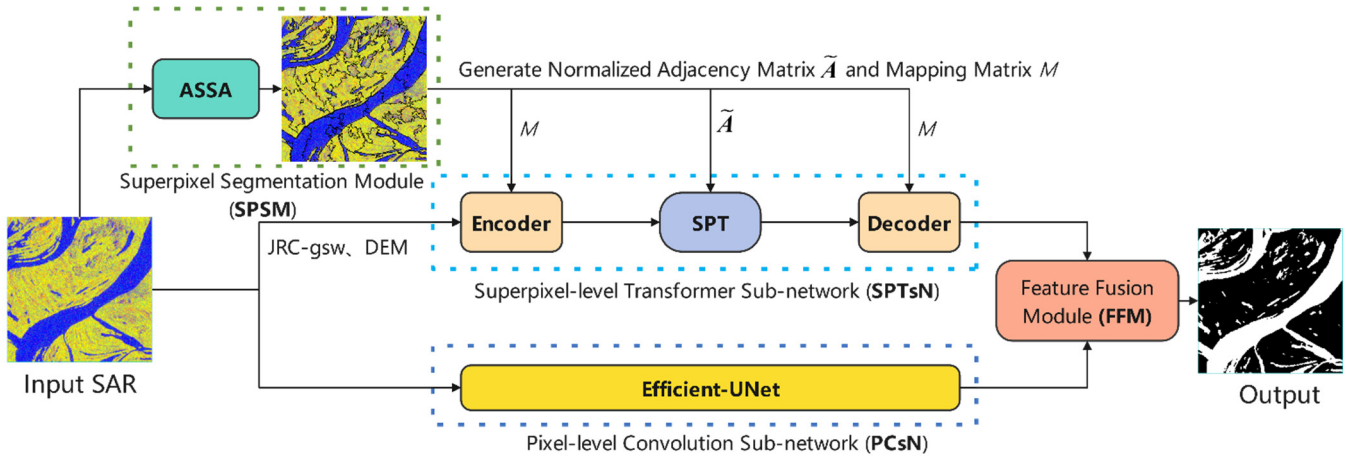


Figure 2. The architecture of SPT-UNet.

2.1. SPSM

In the SPT, the internal homogeneity of superpixels significantly impacts the subsequent water extraction performance. Therefore, we incorporate the Adaptive Superpixel Segmentation Algorithm (ASSA) [37] into the SPSM. The ASSA consists of three steps: (1) The initialization stage sets the required algorithm parameters. (2) The priority queue traversal stage adjusts or generates seed points based on the homogeneity formula defined in (1). (3) Finally, the post-processing stage merges superpixels smaller than a predefined threshold to alleviate the interference from speckle noise.

$$|G(I(p_i)) - IM_j| < 2 \cdot \min(\sigma_m, m = 1, 2, \dots, M) \quad (1)$$

where $G(\cdot)$ represents the standard Gaussian kernel function, $|\cdot|$ represents absolute value, $\min(\cdot)$ represents the minimum standard deviation σ of all Gaussian distributions in the Gaussian Mixture Model (GMM) [38], $p_i = (x, y)$ represents the coordinates of the i th pixel in the SAR image, and IM_j represents the average intensity of the j th superpixel.

Integrating a CNN and SPT poses a significant challenge, primarily because of the inherent differences in data representation structures between pixel- and superpixel-level features. Pixel-level features are traditionally extracted utilizing uniform grids, while superpixel-level features are derived based on irregular, internally homogeneous superpixels. The SPSM transforms features between the pixel- and superpixel-level to address the above issue by utilizing the M [39]. This transformation achieves effective feature interaction and fusion between the pixel- and superpixel-level.

The primary process of the SPSM is as follows: Firstly, the ASSA is applied to segment the I into superpixels, obtaining a superpixel set $S = [S_i]_{i=1}^Z$, where $S_i = [I(p_i^j)]_{j=1}^{N_i}$ denotes the i th superpixel, and p_i^j denotes the j th pixel within the i th superpixel. Z and N_i denote the number of superpixel objects and the number of pixels within the i th superpixel object, respectively.

Then, to achieve the conversion between pixel- and superpixel-level features, the Encoder and Decoder are defined [40], which can be written as follows:

$$V = \text{Encoder}(I; M) = \hat{M}^T \text{Flatten}(I) \quad (2)$$

$$I^* = \text{Decoder}(V^*; M) = \text{Reshape}(MV^*) \quad (3)$$

where $V \in \mathbb{R}^{Z \times 5}$ denotes the initial superpixel-level features, V^* denotes the superpixel-level water semantic features after SPTsN, I^* denotes the pixel-level features after mapping from superpixel-level features, $\text{Reshape}(\cdot)$ indicates the restoration of spatial dimensions of the image, and $\hat{M}$ denotes the column-normalized $M \in \mathbb{R}^{HW \times Z}$ which is defined as follows:

$$M_{i,j} = \begin{cases} 1, & \text{if } \hat{I}_i \in S_j \\ 0, & \text{otherwise} \end{cases} \quad \hat{I} = \text{Flatten}(I) \quad (4)$$

where $\text{Flatten}(\cdot)$ denotes flattening SAR images by the spatial dimension, $\hat{I}_i$ denotes the i th pixel of $\hat{I}$, and S_j denotes the j th superpixel.

Furthermore, to enhance the generalization ability of the SPTsN, we define the $\tilde{A} \in \mathbb{R}^{Z \times Z}$ [41] for use within the SPTsN. Its definition is as follows:

$$\tilde{A} = \tilde{D}^{-\frac{1}{2}} \hat{A} \tilde{D}^{-\frac{1}{2}} \quad (5)$$

$$\tilde{D}_{ii} = \sum_j \hat{A}_{ij} \quad (6)$$

$$\hat{A} = A + I_N \quad (7)$$

$$A_{i,j} = \begin{cases} 1, & \text{if } S_i \text{ and } S_j \text{ are adjacent} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where I_N denotes the identity matrix, $\hat{A}$ denotes the adjacency matrix considering the superpixel itself, $\tilde{D}$ denotes the degree matrix of $\hat{A}$, and A denotes the adjacency matrix.

In summary, leveraging the ASSA, we construct superpixel-level initial features. Additionally, we define the Encoder and Decoder to promote the transformation between pixel- and superpixel-level features. Further, the $\tilde{A}$ is defined for the subsequent training of the SPTsN.

2.2. SPTsN

The transformer-based models are widely used in the CV domain due to their powerful global modeling capability. However, directly applying it to water body extraction from SAR images faces two main challenges: the issue of edge serration caused by regular patch partition and the dependence on data caused by weak generalization ability. We define the SPT in SPTsN to alleviate the above issues, as depicted in Figure 3.

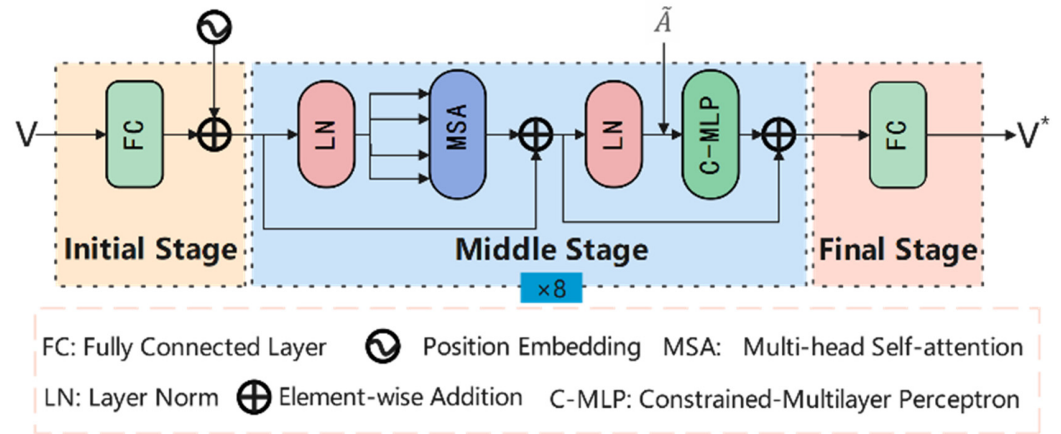


Figure 3. The Flowchart of the SPT.

In the SPT, we first input the V and $\tilde{A}$, which are obtained from the SPSM. Here, the V incorporates JRC Global Surface Water (JRC-gsw) [42] and DEM [43] to reduce the probability of water misclassification.

In the initial stage of the SPT, a fully connected layer (FC) is used to map V to a D -dimensional space. To capture spatial positional information among the superpixels, we introduce a learnable one-dimensional position encoding denoted as $P \in \mathbb{R}^{Z \times D}$. The formula for feature mapping and incorporating the position encoding into $V_0 \in \mathbb{R}^{Z \times D}$ is as follows:

$$V_0 = VW + P \quad (9)$$

where W denotes the parameters of the FC.

In the middle stage of the SPT, we perform 8 layers to learn the superpixel-level semantic features of water bodies in SAR images. As shown in Figure 3, the middle stage mainly consists of the multi-head self-attention (MSA) and the C-MLP. The MSA is used to model the spatial relationships between superpixels globally. Then, the C-MLP ensures that all superpixels in the image propagate and learn features solely through themselves and their neighbors. The formula of C-MLP is as follows:

$$C - \text{MLP}(V) = \text{Dropout}(\tilde{A} \text{Relu}(\tilde{A}VW^{(0)})W^{(1)}) \quad (10)$$

where $W^{(0)}$ and $W^{(1)}$ denote the parameters of the FC, $\text{Relu}(\cdot)$ denotes the activation layer, and $\text{Dropout}(\cdot)$ denotes the dropout layer that randomly discards neurons with a certain probability. $\tilde{A}$ essentially reflects the prior correlation weights among any superpixels in the image. Incorporating $\tilde{A}$ into the MLP can be seen as introducing an inductive bias capability into the SPT. Consequently, it enhances the ability of the SPT to generalize in small-scale datasets. The formula for the output in the middle stage during the m th layer is as follows:

$$V'_m = \text{MSA}(\text{LN}(V_{m-1})) + V_{m-1} \quad (m = 1, \dots, M) \quad (11)$$

$$V_m = C - \text{MLP}(\text{LN}(V'_m)) + V'_m \quad (m = 1, \dots, M) \quad (12)$$

where V_m denotes the output of the m th middle stage and V'_m denotes the output of the MSA module in the m th middle stage.

In the final stage of the SPT, the FC is used to compress the superpixel-level features, which capture abundant semantic information related to water bodies. These features are mapped to a D' -dimensional space, resulting in the output superpixel-level features denoted as $V^* \in \mathbb{R}^{Z \times D'}$.

The SPT within the SPTsN is designed to capture spatial relationships between different superpixels at the global scale of SAR images and establish superpixel-level features. We represent the SPT as a function denoted by $V^* = SPT(V; \tilde{A})$.

2.3. PCsN

To enable the model to possess pixel-level local feature learning capability, we propose Efficient-UNet in the PCsN, as illustrated in Figure 4.

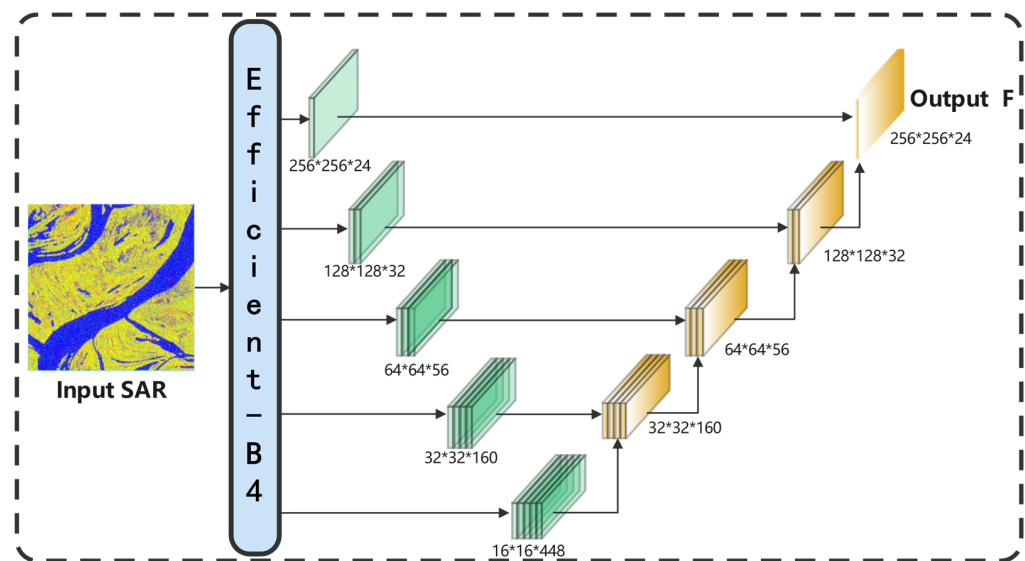


Figure 4. The Flowchart of the Efficient-UNet.

The Efficient-UNet follows a standard encoder–decoder architecture. EfficientNet [44], a well-known convolutional neural network model introduced by Google, balances parameter efficiency and accuracy by uniformly scaling the width, depth, and image resolution of the network using a new compound scaling method. For the encoder part, we select the Efficient-B4 network to extract water bodies' semantic features from the I , taking into account the model's parameter count. The upsampling operation is used in the decoder part to gradually restore the low-resolution features obtained from the encoder to the original image size. Additionally, a skip connection is utilized to concatenate the low-level features from the encoder with the high-level features from the decoder, thus recovering the detailed characteristics of the image.

Generally, the Efficient-UNet within the PCsN captures spatial relationships between pixels at the local scale of SAR images and establishes pixel-level features. We represent the *Efficient-UNet* as a function, denoted by $F = \text{Efficient-UNet}(I)$.

2.4. FFM

To fully integrate the pixel- and superpixel-level semantic features of water bodies, we construct the FFM, as shown in Figure 5.

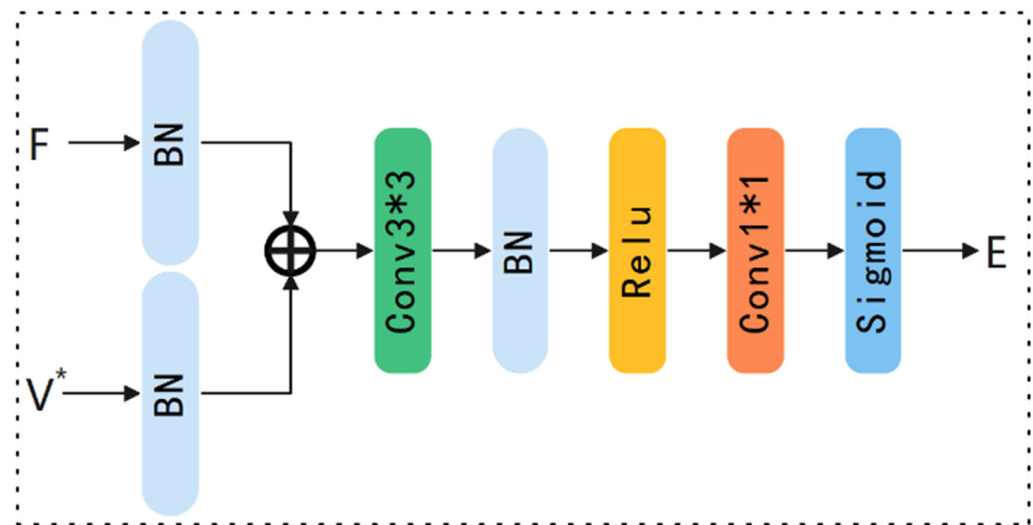


Figure 5. The Flowchart of the FFM.

In the FFM, the pixel-level feature F and the superpixel-level feature V^* are first inputted. Then, the BN [45] is applied to both F and V^* to address the issue of inconsistent scales between the pixel- and superpixel-level features and reduce the influence of internal covariate shifts for subsequent steps of the FFM. Next, the Add layer is used to sum up the pixel- and the superpixel-level features, obtaining the fused features. Afterward, the fused feature is further processed through a 3×3 convolution layer, a BN layer, and an activation layer to extract spatial and semantic features about the water bodies. Finally, the 1×1 convolutional and Sigmoid layers process the refined features and output the water body extraction result.

Overall, the FFM combines the pixel- and superpixel-level semantic features of the water bodies, leveraging their respective advantages.

2.5. Hybrid Weighted Loss

The SPT-UNet, which has been proposed in this study, is an end-to-end pixel-wise classification model for which we use a hybrid weighted loss, denoted as $Loss_{all}$. It combines the binary cross-entropy loss (BCELoss) and the Dice loss (DiceLoss). This combination leverages the stable training process of the BCELoss and the advantage of the DiceLoss in mitigating the imbalance in the dataset distribution. The formula is presented below:

$$BCELoss(E, L) = -\frac{1}{N} \sum_{i=1}^N [L(p_i) \log(E(p_i)) + (1 - L(p_i)) \log(1 - E(p_i))] \quad (13)$$

$$DiceLoss(E, L) = 1 - \frac{2 |E \cap L|}{|E| + |L|} \quad (14)$$

$$Loss_{all} = a \cdot BCELoss + (1 - a) \cdot DiceLoss \quad (15)$$

where $E(p_i)$ denotes the probability of the i th pixel, which is water in the predicted image, and a denotes the weight factor.

3. Dataset and Experimental Setup

3.1. Data Description

Our experiments used the C2S-MS dataset [46], a benchmark dataset for water extraction in SAR images. The C2S-MS dataset was obtained globally through sampling from different complex surface environments (e.g., the interference of the shadow and water-like surfaces). Consequently, evaluating the superiority and effectiveness of the model in extracting water bodies from this dataset presents a more significant challenge.

The C2S-MS dataset consists of 8100 images from Sentinel-1 with a size of 256×256 . Four typical samples are shown in Figure 6. The images in the dataset have a spatial resolution of 10 m. To facilitate more effective extraction of water features and assess model generalization ability, the training, validation, and testing data sets were divided in a ratio of 7:1:1, as shown in Table 1.

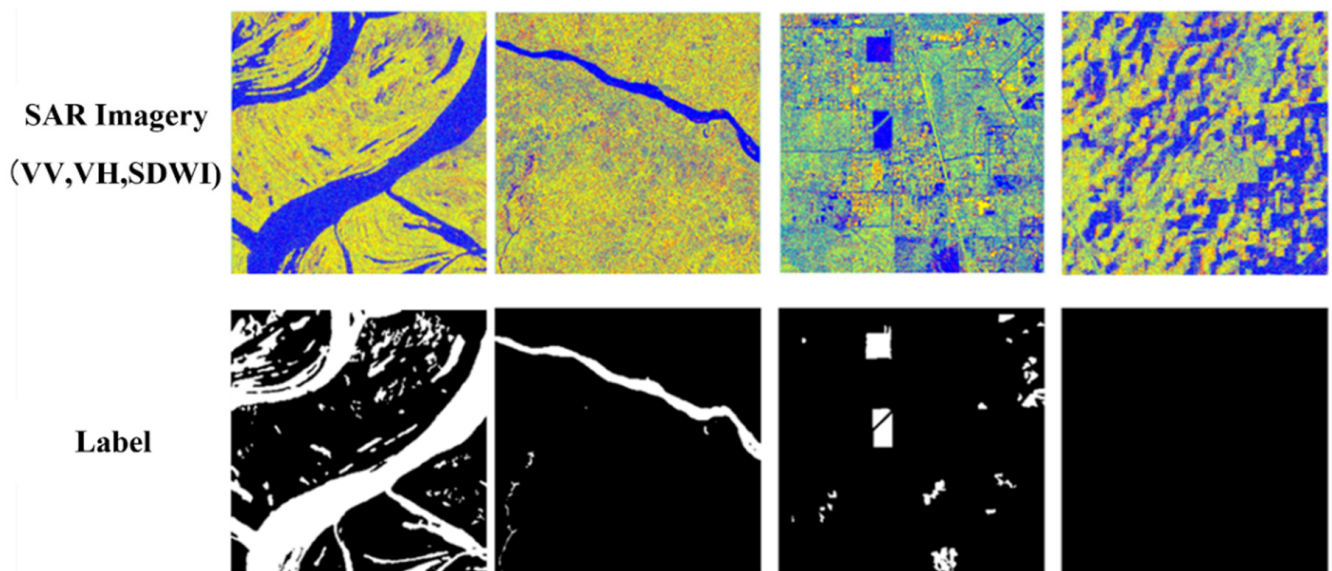


Figure 6. The dataset samples.

Table 1. The dataset parameters.

C2S-MS	
band	VV, VH
Train.	6300
Val.	900
Test.	900

3.2. Experimental Settings

The proposed network is implemented using PyTorch and trained on an NVIDIA TESLA P100 GPU platform. Adam was chosen to speed up the training process. The training process is conducted for 300 epochs, with an initial learning rate of 3×10^{-4} . A cosine annealing optimization method is utilized to adjust the learning rate dynamically to enhance model performance.

3.3. Evaluation Indices

To measure the water extraction performance of the model, we chose the intersection over union (IoU), Precision, Recall, and F1 score as evaluation indices:

$$\text{IoU} = \frac{\text{TP}}{\text{TN} + \text{TP} + \text{FN}} \quad (16)$$

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}} \quad (17)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (18)$$

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (19)$$

where TP denotes the total number of correctly classified water pixels, TN denotes the total number of correctly classified background pixels, FP denotes the total number of background pixels wrongly classified as water pixels, and FN denotes the total number of water pixels wrongly classified as background pixels. The values of these four indices range from 0 to 1, and when the values are close to 1, it indicates higher accuracy in water segmentation and better algorithm robustness.

4. Experiment Results

4.1. Comparison of Extraction Performance

To evaluate the superiority and effectiveness of the proposed SPT-UNet, we compared it with various well-known segmentation algorithms, including U-Net [47], DeepLabv3+ [48], GCN [41], Swin-UNet [49], FWENet [50], and DWENet [51]. All baseline networks were performed under the same conditions to ensure a fair comparison, utilizing the hyperparameters specified in the original paper. The quantitative and visualized results of water body extraction with different networks are depicted in Table 2 and Figure 7.

Table 2. The accuracies of different networks. The best result is highlighted in bold.

Network	IoU	Recall	Precision	F1
U-Net	72.8	82.81	77.42	80.02
DeepLabv3+	72.48	82.11	84.12	83.10
FWENet	72.42	86.80	79.79	83.15
DWENet	73.00	78.94	89.03	83.68
GCN	66.22	74.87	82.87	78.67
Swin-UNet	66.35	78.77	78.63	78.70
SPT-UNet	80.10	86.20	91.07	88.57

Table 2 shows that the SPT-UNet displays the best value of IoU, Precision, and F1 score, outperforming other networks in extracting water bodies. Experimental results illustrate that our networks can capture the complementary pixel- and superpixel-level features of water bodies to reduce misjudgments and improve details of water bodies.

It is worth mentioning that the proposed SPT-UNet provides the PCsN branch to capture the pixel-level features of water bodies to supplement the SPTsN, thus surpassing the patch-based Swin_UNet and the superpixel-based GCN. Similarly, due to the provision of the SPTsN branch to capture superpixel-level water bodies features to supplement the PCsN, it surpasses pixel-based networks like U-Net and DeepLabv3+. Additionally, the mainstream extraction networks of water bodies, such as FWENet and DWENet, which consider the relevant characteristics of water bodies, are primarily pixel-level networks. As a result, they cannot model global structures, limiting their extraction accuracy.

Figure 7 shows that the water bodies extracted by the proposed SPT-UNet had more fitting boundaries. In contrast to the Swin-UNet, the proposed SPT-UNet significantly reduced edge serration, which validates the effectiveness of superpixel segmentation over patch partitioning. Additionally, the water bodies had fewer misjudgments. The above phenomena further prove the complementary effect of the PCsN and SPTsN in the proposed SPT-UNet.

4.2. Ablation Study

In this section, we performed a series of ablation experiments to assess the performance of various modules. Considering the essentiality of combining SPSM with SPTsN, we merge them by default. The performance of water body extraction under different module combinations is shown in Figure 8.

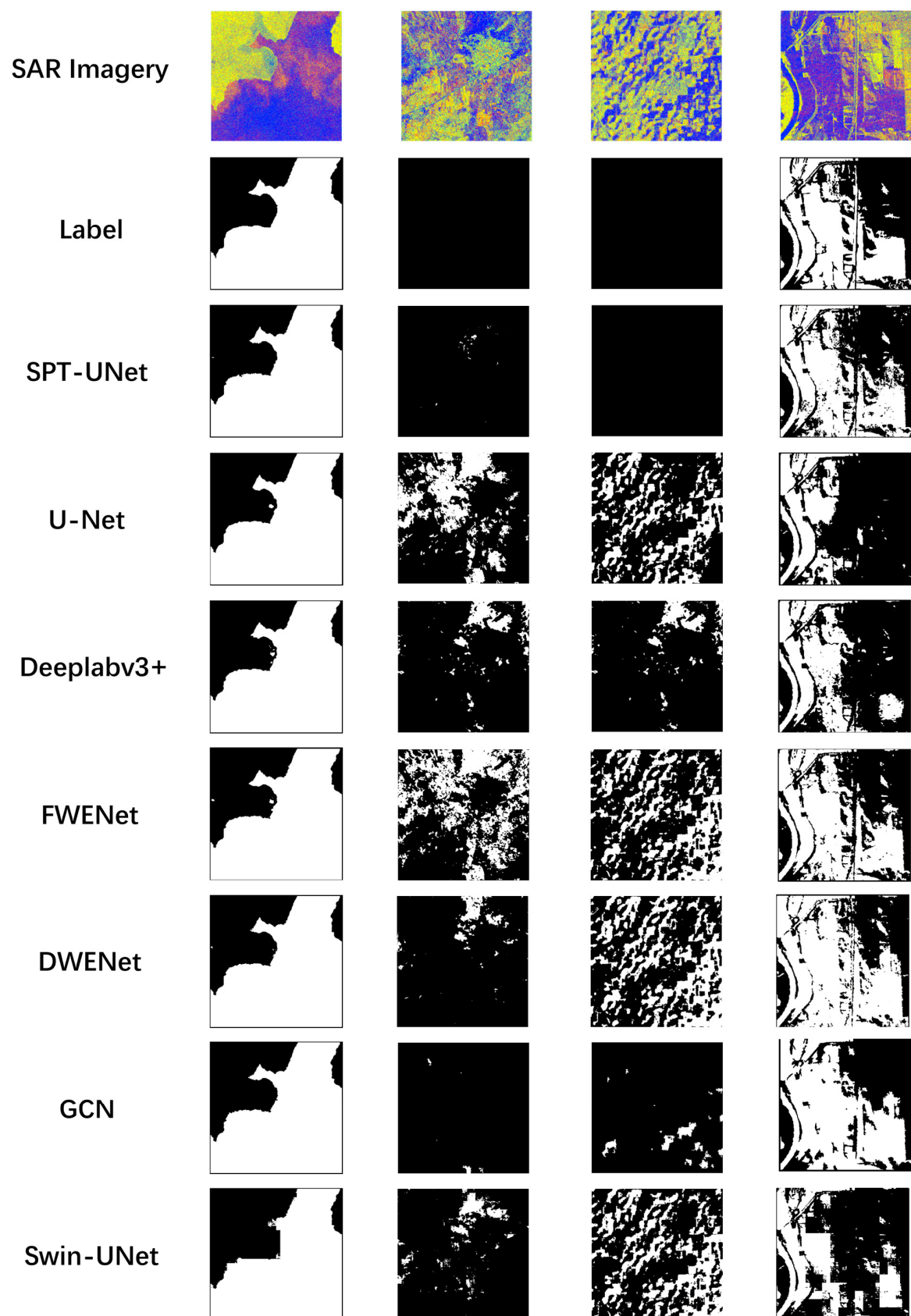


Figure 7. Visual comparison of water extraction results.

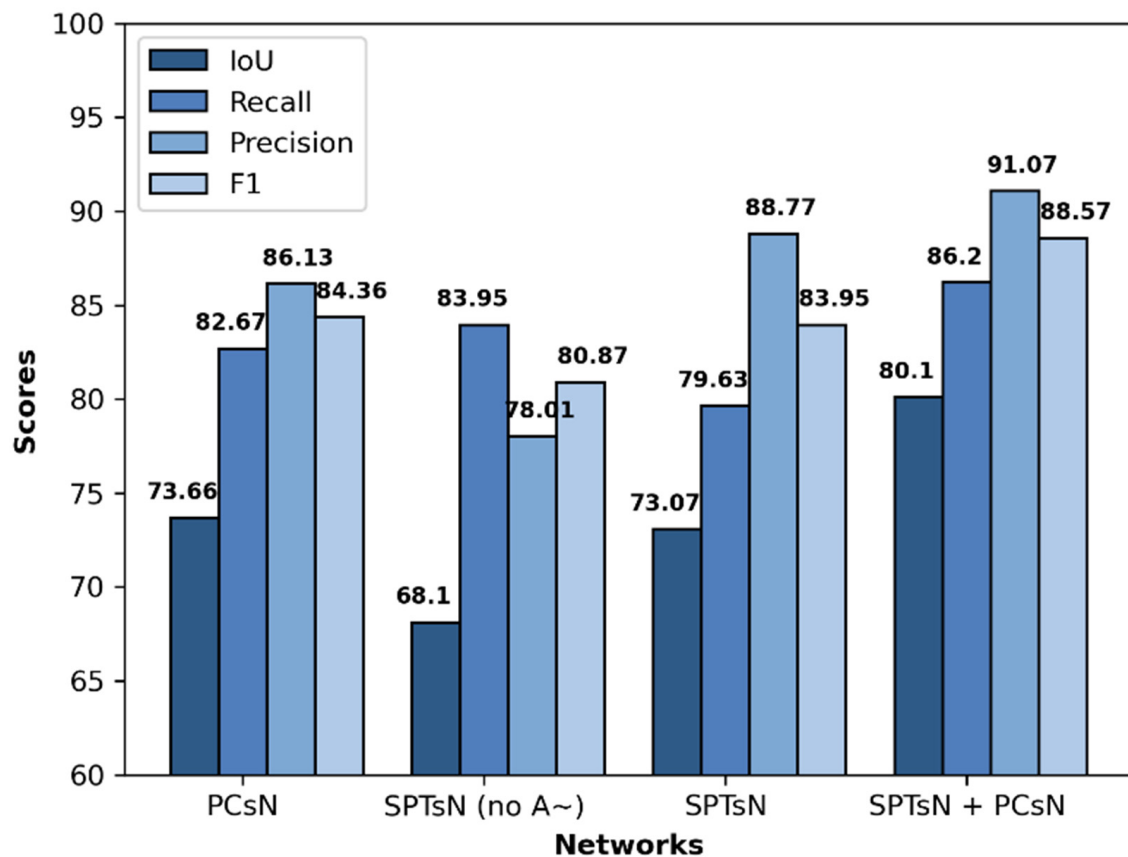


Figure 8. The accuracies of different modules.

Compared to using either SPTsN or PCsN modules individually, combining both achieved higher extraction performance in water body extraction, indicating the complementarity between the pixel- and superpixel-level features. Furthermore, the extraction performance of the SPTsN module, with the inclusion of the normalized adjacency matrix $\tilde{A}$, outperformed that of the SPTsN module without $\tilde{A}$. This finding indicates that using the $\tilde{A}$ to impose knowledge constraints on superpixels is beneficial for enhancing the generalization ability of the SPTsN module in small-scale datasets.

Furthermore, to analyze the impact of combining the pixel- and superpixel-level features, we employed the Grad-CAM [52] on different networks, a visualization method that utilizes gradients to calculate the spatial importance of positions within a layer. We compared the visualization results of the last convolution layer in Efficient-UNet based on pixels, SPT based on superpixels, and SPT-UNet, which combines pixel and superpixel features. The results of Figure 9 indicate that SPT, relying on superpixels, plays a crucial role in maintaining accurate water body boundaries and minimizing misclassifications. Conversely, Efficient-UNet, operating at the pixel level, excels in extracting the detailed features of water bodies. In essence, the pixel- and superpixel-level features within the SPT-UNet complement each other effectively.

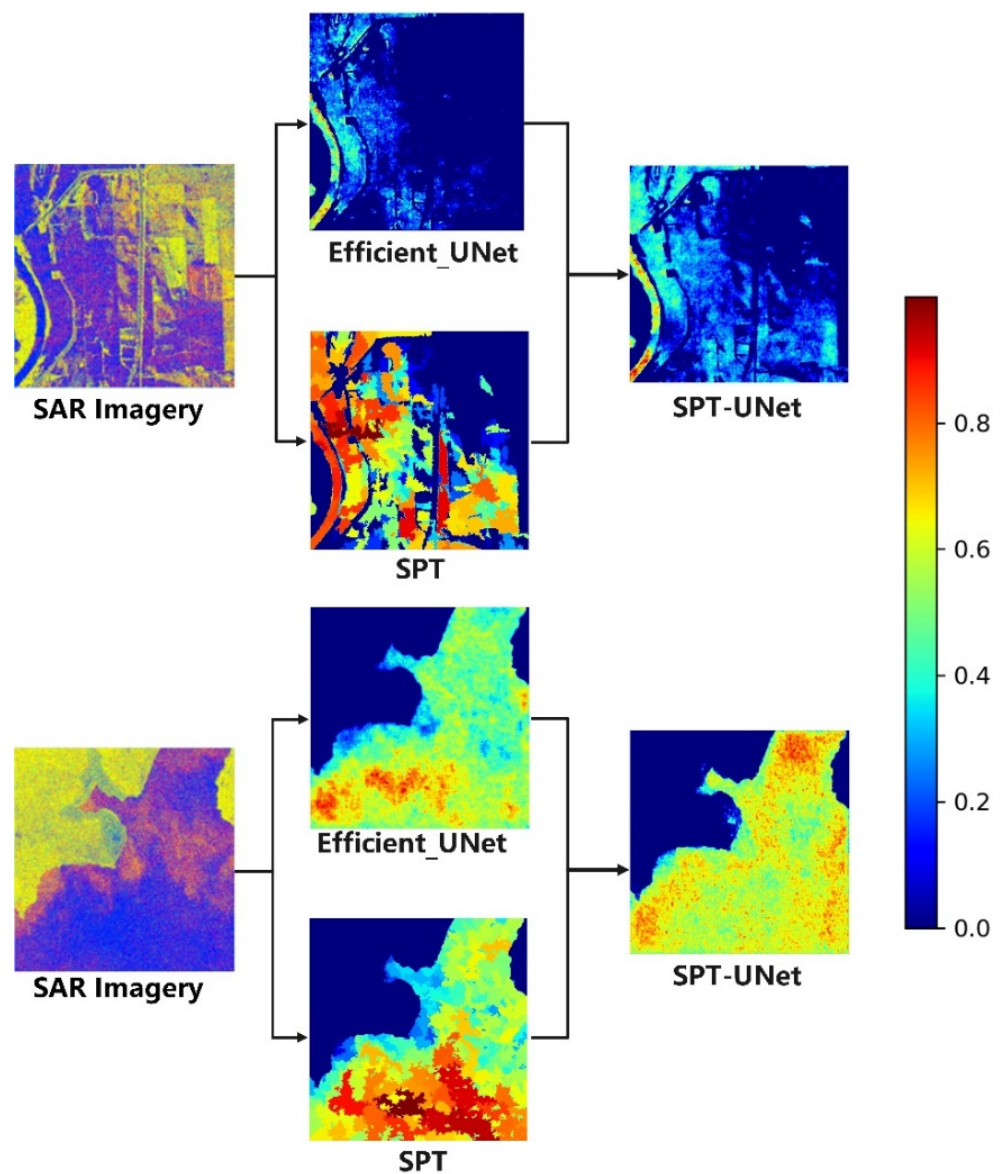


Figure 9. Visualization of the results. We compare the visualization results of different networks, including the pixel-based Efficient-UNet, superpixel-based SPT, and SPT-UNet combined pixel and superpixel. The heatmap reveals the probability of pixels belonging to water.

5. Discussion

5.1. Influence of Segmentation Scale

The number of superpixels, Z , could affect the performance of the SPT-UNet. To analyze its influence on our network, we set the number of superpixels, Z , to 256, 512, 800, 1024, and 1200, respectively, and tested their extraction accuracies, as shown in Table 3.

Table 3. The accuracies of different Z . The best result is highlighted in bold.

Z	IoU	Recall	Precision	F1
256	77.36	87.27	86.21	86.74
512	80.10	86.20	91.07	88.57
800	75.40	83.03	87.22	85.07
1024	79.80	87.20	89.57	88.37
1200	76.84	82.45	91.40	86.69

The segmentation scale, Z , decides the smallest size of the superpixels in the image. As can be seen, the values of Recall mainly decreased with the increase in scale. One probable explanation is that a larger scale leads to small superpixel units that may contain a relatively large number of noise points, making it challenging to learn features at the superpixel level. In contrast, the values of Precision primarily increased with the increase in scale. One probable explanation is that a smaller scale results in large superpixel units that can effectively resist the interference of noise points but lack the feature learning of small objects. Considering the balance between noise point interference and the extraction effect of small water bodies, we fixed Z to 512 in all experiments.

5.2. Impact of a

The hyperparameter a in $Loss_{all}$ is used to balance the benefits and drawbacks of BCELoss and DiceLoss. To validate the appropriateness of the set, we compared the extraction indices for different a values, as depicted in Table 4.

Table 4. The accuracies of different a . The best result is highlighted in bold.

a	IoU	Recall	Precision	F1
0.3	77.11	86.57	86.20	86.38
0.4	78.96	83.86	92.19	87.83
0.5	78.82	85.87	90.03	87.90
0.6	80.10	86.20	91.07	88.57
0.7	78.03	86.54	88.06	87.29

The hyperparameter a can affect the training process of the networks. The accuracies of water extraction mainly decreased with the too-small a . This phenomenon may be attributed to the instability in model training caused by the increased weight of the DiceLoss, affecting the degree to which the parameters of the SPT-UNet fit the samples. Similarly, the accuracies of water extraction also primarily decrease with the too-large a , which may be caused by the imbalance between water bodies and the background in the dataset. Therefore, considering all factors, we set the a value to 0.6.

5.3. Impact of Additional Data

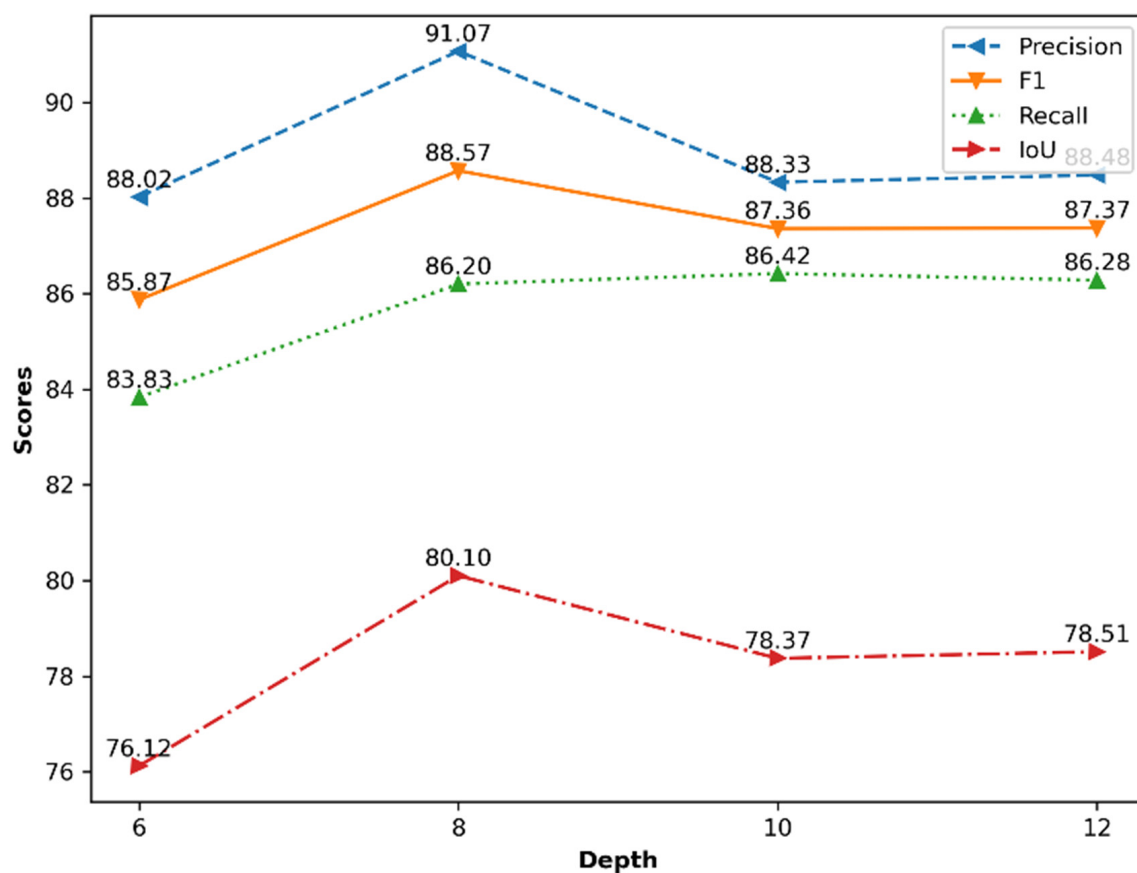
Additional data, such as SDWI image, JRC-gsw image, and DEM image, are introduced to enhance the performance of the SPT-UNet in water extraction. To investigate the effectiveness of the additional data, we carried out a series of experiments for different combination methods, the details of which can be found in Table 5. Compared to the SPT-UNet, which is composed only of VV and VH inputs, the water extraction indicators of the network generally improved after adding SDWI images. This phenomenon may be attributed to the enhancement of water features in the input section, making the network more inclined to learn such features. Adding DEM to the inputs of SPT-UNet composed of VV, VH, and SDWI resulted in improvements in the values of IoU, Recall, and F1 indicators. This phenomenon may be explained by alleviating the interference caused by terrain-induced shadow areas. Adding JRC-gsw to the inputs of SPT-UNet, which consists of VV, VH, and SDWI, leads to improvements in all four indicators. This may be caused by alleviating the impact of interfering objects in SAR images on water extraction. Simultaneously adding DEM and JRC-gsw to the inputs of SPT-UNet, which consists of VV, VH, and SDWI, achieves a balance between Recall and Precision, achieving optimal values in F1 and IoU.

Table 5. The accuracies of different combination methods. The best result is highlighted in bold.

Combination Method	IoU	Recall	Precision	F1
VV + VH	72.98	81.58	86.20	83.83
VV + VH + SDWI	74.55	81.20	88.09	84.50
VV + VH + SDWI + DEM	75.11	87.14	83.37	85.21
VV + VH + SDWI + JRC-gsw	77.66	82.40	92.26	87.05
VV + VH + SDWI + DEM + JRC-gsw	80.10	86.20	91.07	88.57

5.4. Impact of Depth in SPT

In the SPT middle stage, we performed eight layers to extract and learn the water features in SAR images at the superpixel level. To validate the rationality of the aforementioned configuration, we performed the following experiments: training the SPT-UNet with 6, 8, 10, and 12 layers, respectively. A comparison of extraction indicators is shown in Figure 10.

**Figure 10.** The accuracies of different depths.

The SPT-UNet achieved the highest segmentation IoU, Precision, and F1 score when the layer number of the SPT middle stage was eight. When depth sets were too small or large, the value of the four evaluation indicators all decreased. One probable explanation is that a small depth reduces the ability of SPT to model global semantic dependencies. In contrast, a large depth makes modeling difficult due to the small dataset scale, hindering network optimization. Thus, an eight-layer SPT middle stage was incorporated into the SPT-UNet, offering superior segmentation results.

6. Conclusions

Extraction of water bodies from SAR images has been crucial in many applications. In this study, we introduce an innovative network architecture, SPT. (1) The network leverages

the ASSA for irregular superpixel segmentation instead of regular patch partition, which can mitigate edge serration in water body extraction. (2) The network uses the C-MLP to alleviate the problem of the weak generalization ability in small-scale datasets. The C-MLP imposes knowledge constraints on superpixels, which limits the learning range of the superpixels. (3) Additionally, to fuse the superpixel-level features extracted by SPT with pixel-level features, we proposed the SPT-UNet—a unified network based on CNN and SPT. Experimental results of the SPT-UNet showed that the superpixel-level water bodies' semantic features complement pixel features, further improving the performance of water bodies extraction. Further work will improve the computational efficiency of the model using parallel strategies and other optimization methods and refine the SPT-UNet for the multi-scale features fusion at pixel- and superpixel-level in an adaptive manner to implement water bodies extraction.

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

Funding: This research was supported by the National Key Research and Development Program of China (2022YFC3800704), the Innovation Drive Development Special Project of Guangxi (GuikeAA20302022), China Postdoctoral Science Foundation (2023M743586), the National Natural Science Foundation of China (41974108), the Chinese Academy of Sciences, Project title: CASEarth (XDA19080100) and Postdoctoral Fellowship Program of CPSF (GZC20232758).

Data Availability Statement: The code and dataset will be available at <https://github.com/SuperPixelPioneer/SPT-UNet> (accessed on 10 July 2024). The JRC-gsw and DEM images were provided by the Microsoft (<https://planetarycomputer.microsoft.com/dataset/jrc-gsw> (accessed on 10 July 2023)) and <https://planetarycomputer.microsoft.com/dataset/cop-dem-glo-30> (accessed on 10 July 2023)).

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

References

1. Dong, Z.; Liang, Z.; Wang, G.; Amankwah, S.O.Y.; Feng, D.; Wei, X.; Duan, Z. Mapping inundation extents in Poyang Lake area using Sentinel-1 data and transformer-based change detection method. *J. Hydrol.* **2023**, *620*, 129455. [CrossRef]
2. Qiu, J.; Cao, B.; Park, E.; Yang, X.; Zhang, W.; Tarolli, P. Flood Monitoring in Rural Areas of the Pearl River Basin (China) Using Sentinel-1 SAR. *Remote Sens.* **2021**, *13*, 1384. [CrossRef]
3. Proud, S.R.; Fensholt, R.; Rasmussen, L.V.; Sandholt, I. Rapid response flood detection using the MSG geostationary satellite. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 536–544. [CrossRef]
4. 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*, 1–5. [CrossRef]
5. 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*, 1–14. [CrossRef]
6. Kitajima, N.; Seto, R.; Yamazaki, D.; Zhou, X.; Ma, W.; Kanae, S. Potential of a SAR Small-Satellite Constellation for Rapid Monitoring of Flood Extent. *Remote Sens.* **2021**, *13*, 1959. [CrossRef]
7. Giustarini, L.; Hostache, R.; Matgen, P.; Schumann, G.J.P.; Bates, P.D.; Mason, D.C. A Change Detection Approach to Flood Mapping in Urban Areas Using TerraSAR-X. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2417–2430. [CrossRef]
8. Martinez, J.-M.; Letoan, T.; Le Toan, T. Mapping of flood dynamics and spatial distribution of vegetation in the Amazon floodplain using multitemporal SAR data. *Remote Sens. Environ.* **2007**, *108*, 209–223. [CrossRef]
9. Matgen, P.; Schumann, G.; Henry, J.-B.; Hoffmann, L.; Pfister, L. Integration of SAR-derived river inundation areas, high-precision topographic data and a river flow model toward near real-time flood management. *Int. J. Appl. Earth Obs. Geoinform.* **2007**, *9*, 247–263. [CrossRef]
10. Xu, C.; Du, X.; Fan, X.; Yan, Z.; Kang, X.; Zhu, J.; Hu, Z. A modular remote sensing big data framework. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–11. [CrossRef]
11. Xu, C.; Du, X.; Fan, X.; Giuliani, G.; Hu, Z.; Wang, W.; Liu, J.; Wang, T.; Yan, Z.; Zhu, J.; et al. Cloud-based storage and computing for remote sensing big data: A technical review. *Int. J. Digit. Earth* **2022**, *15*, 1417–1445. [CrossRef]
12. Boni, G.; Ferraris, L.; Pulvirenti, L.; Squicciarino, G.; Pierdicca, N.; Candela, L.; Pisani, A.R.; Zoffoli, S.; Onori, R.; Proietti, C.; et al. A prototype system for flood monitoring based on flood forecast combined with COSMO-SkyMed and Sentinel-1 data. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2016**, *9*, 2794–2805. [CrossRef]

13. Greifeneder, F.; Wagner, W.; Sabel, D.; Naeimi, V. Suitability of SAR imagery for automatic flood mapping in the Lower Mekong Basin. In *Remote Sensing the Mekong*; Routledge: Oxfordshire, UK, 2018; pp. 111–128.
14. Manjusree, P.; Kumar, L.; Bhatt, C.M.; Rao, G.S.; Bhanumurthy, V. Optimization of threshold ranges for rapid flood inundation mapping by evaluating backscatter profiles of high incidence angle SAR images. *Int. J. Disaster Risk Sci.* **2012**, *3*, 113–122. [\[CrossRef\]](#)
15. Chen, L.; Liu, Z.; Zhang, H. SAR Image Water Extraction based on Scattering Characteristics. *Remote Sens. Technol. Appl.* **2014**, *29*, 963–969.
16. Chini, M.; Hostache, R.; Giustarini, L.; Matgen, P. A hierarchical split-based approach for parametric thresholding of SAR images: Flood inundation as a test case. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6975–6988. [\[CrossRef\]](#)
17. Liang, J.; Liu, D. A local thresholding approach to flood water delineation using Sentinel-1 SAR imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 53–62. [\[CrossRef\]](#)
18. Martinis, S.; Kersten, J.; Twele, A. A fully automated TerraSAR-X based flood service. *ISPRS J. Photogramm. Remote Sens.* **2015**, *104*, 203–212. [\[CrossRef\]](#)
19. Huang, Z.; Wu, W.; Liu, H.; Zhang, W.; Hu, J. Identifying Dynamic Changes in Water Surface Using Sentinel-1 Data Based on Genetic Algorithm and Machine Learning Techniques. *Remote Sens.* **2021**, *13*, 3745. [\[CrossRef\]](#)
20. Qiu, F.; Guo, Z.; Zhang, Z.; Wei, X.; Jing, M. Water Body Area Extraction from SAR Image based on Improved SVM Classification Method. *Geo Inf. Sci.* **2022**, *24*, 940–948.
21. He, Y.; Yao, S.; Yang, W.; Yan, H.; Zhang, L.; Wen, Z.; Zhang, Y.; Liu, T. An extraction method for glacial lakes based on Landsat-8 imagery using an improved U-Net network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6544–6558. [\[CrossRef\]](#)
22. Wei, S.; Zhang, H.; Wang, C.; Wang, Y.; Xu, L. Multi-temporal SAR data large-scale crop mapping based on U-Net model. *Remote Sens.* **2019**, *11*, 68. [\[CrossRef\]](#)
23. Qin, H.; Wang, J.; Mao, X.; Zhao, Z.; Gao, X.; Lu, W. An Improved Faster R-CNN Method for Landslide Detection in Remote Sensing Images. *J. Geovis. Spat. Anal.* **2024**, *8*, 2. [\[CrossRef\]](#)
24. Solórzano, J.V.; Mas, J.F.; Gao, Y.; Gallardo-Cruz, J.A. Land use land cover classification with U-net: Advantages of combining sentinel-1 and sentinel-2 imagery. *Remote Sens.* **2021**, *13*, 3600. [\[CrossRef\]](#)
25. Dai, M.; Leng, X.; Xiong, B.; Ji, K. An Efficient Water Segmentation Method for SAR Images. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 1129–1132.
26. Zhang, J.; Xing, M.; Sun, G.-C.; Chen, J.; Li, M.; Hu, Y.; Bao, Z. Water Body Detection in High-Resolution SAR Images with Cascaded Fully-Convolutional Network and Variable Focal Loss. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 316–332. [\[CrossRef\]](#)
27. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
28. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2021**, arXiv:2010.11929.
29. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021.
30. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2020; pp. 213–229.
31. Wang, H.; Zhu, Y.; Green, B.; Adam, H.; Yuille, A.; Chen, L.C. Axial-deeplab: Stand-alone axial-attention for panoptic segmentation. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2020; pp. 108–126.
32. Wang, H.; Xing, C.; Yin, J.; Yang, J. Land Cover Classification for Polarimetric SAR Images Based on Vision Transformer. *Remote Sens.* **2022**, *14*, 4656. [\[CrossRef\]](#)
33. Dong, H.; Zhang, L.; Zou, B. Exploring Vision Transformers for Polarimetric SAR Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [\[CrossRef\]](#)
34. Liu, X.; Wu, Y.; Liang, W.; Cao, Y.; Li, M. High Resolution SAR Image Classification Using Global-Local Network Structure Based on Vision Transformer and CNN. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [\[CrossRef\]](#)
35. Dong, Y.; Liu, Q.; Du, B.; Zhang, L. Weighted Feature Fusion of Convolutional Neural Network and Graph Attention Network for Hyperspectral Image Classification. *IEEE Trans. Image Process* **2022**, *31*, 1559–1572. [\[CrossRef\]](#)
36. Jia, S.; Xue, D.; Li, C.; Zheng, J.; Li, W. Study on New Method for Water Area Information Extraction Based on Sentinel-1 Data. *Yangtze River* **2019**, *50*, 213–217.
37. Zhao, T.; Du, X.; Yan, Z.; Zhu, J.; Xu, C.; Fan, X. Adaptive superpixel segmentation of SAR images using an adaptive adjustment strategy for seeds. *Natl. Remote Sens. Bull.* **2023**, *1*–12. [\[CrossRef\]](#)
38. Celik, T.; Tjahjadi, T. Automatic Image Equalization and Contrast Enhancement Using Gaussian Mixture Modeling. *IEEE Trans. Image Process* **2012**, *21*, 145–156. [\[CrossRef\]](#)
39. Jampani, V.; Sun, D.; Liu, M.Y.; Yang, M.H.; Kautz, J. Superpixel Sampling Networks. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 352–368.
40. Liu, Q.; Xiao, L.; Yang, J.; Wei, Z. CNN-Enhanced Graph Convolutional Network with Pixel- and Superpixel-Level Feature Fusion for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 8657–8671. [\[CrossRef\]](#)

41. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2017**, arXiv:1609.02907.
42. 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](#)] [[PubMed](#)]
43. Crippen, R.; Buckley, S.; Agram, P.; Belz, E.; Gurrola, E.; Hensley, S.; Kobrick, M.; Laval, M.; Martin, J.; Neumann, M.; et al. Nasadem Global Elevation Model: Methods and Progress. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*, 125–128. [[CrossRef](#)]
44. Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; Volume 97, pp. 6105–6114.
45. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7–9 July 2015; Volume 37, pp. 448–456.
46. Cloud to Street, Microsoft, Radiant Earth Foundation. A Global Flood Events and Cloud Cover Dataset. Radiant MLHub. 2022. Available online: <https://registry.opendata.aws/c2smsfloods/> (accessed on 20 June 2022).
47. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the 18th International Conference, Munich, Germany, 5–9 October 2015; pp. 234–241.
48. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 801–818.
49. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-unet: Unet-like pure transformer for medical image segmentation. *arXiv* **2021**, arXiv:2105.05537.
50. Wang, J.; Wang, S.; Wang, F.; Zhou, Y.; Wang, Z.; Ji, J.; Xiong, Y.; Zhao, Q. FWENet: A deep convolutional neural network for flood water body extraction based on SAR images. *Int. J. Digit. Earth* **2022**, *15*, 345–361. [[CrossRef](#)]
51. Zhao, B.; Sui, H.; Liu, J. Siam-DWENet: Flood inundation detection for SAR imagery using a cross-task transfer siamese network. *Int. J. Appl. Earth Obs. Geoinform.* **2023**, *116*, 103132. [[CrossRef](#)]
52. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2020**, *128*, 336–359. [[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.