SADA-Net: A Shape Feature Optimization and Multiscale Context Information-Based Water Body Extraction Method for High-Resolution Remote Sensing Images

Bin Wang, Zhanlong Chen, Liang Wu, Xiaohong Yang, and Yuan Zhou

***Abstract*—Convolutional neural networks (CNNs) have significance in remote sensing image mapping, and pixel-level representation allows refined results. Due to inconsistencies within a class and different scales of water bodies, the water body mapping has challenges, such as insufficient integrity and rough shape segmentation. To resolve these issues, we proposed an intelligent water bodies extraction method (named SADA-Net) for high-resolution remote sensing images. This method considers multiscale information, context dependence, and shape features. The network framework integrates three critical components: shape feature optimization (SFO), atrous spatial pyramid pooling, and dual attention modules. SADA-Net can accurately extract an extensive range of water bodies in complex scenarios. SADA-Net has certain advantages regarding small and dense water bodies extraction, as the SFO module effectively solves the defects of the unified processing of low-level features in the encoder stage of CNNs, which highlights the shape information of a water body. Two data types (red, green, and blue bands and multispectral images) are employed to verify the performance of the proposed network. The best result achieved an evaluation index F1-Score of 96.14% in large-scale image segmentation, and the structural similarity index measure reached 94.70%. Overall, the proposed method achieves the purpose of maximizing the integrity and optimizing the shape of a water body. Additionally, the SADA-Net proposed in this article has a specific reference value for high-resolution remote sensing image water bodies mapping.**

***Index Terms*—Atrous spatial pyramid pooling, dual attention, multispectrum, shape feature optimization, small water bodies.**

Manuscript received October 21, 2021; revised December 5, 2021 and January 2, 2022; accepted January 20, 2022. Date of publication January 27, 2022; date of current version February 18, 2022. This work was support in part by the National Natural Science Foundation of China under Grant 41871305, Grant 41871311, Grant 42001308, in part by the National Key R&D Program of China under Grant 2017YFC0602204, in part by the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) under Grant CUGQY1945, in part by the Opening Fund of Key Laboratory of Geological Survey and Evaluation of Ministry of Education, and in part by the Fundamental Research Funds for the Central Universities under Grant GLAB2019ZR02. *(Corresponding author: Zhanlong Chen.)*

Bin Wang, Zhanlong Chen, and Liang Wu are with the School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China (e-mail: wangbin@cug.edu.cn; chenzl@cug.edu.cn; wuliang@cug.edu.cn).

Xiaohong Yang is with the School of Computer Science, China University of Geosciences, Wuhan 430074, China (e-mail: yangxiaohong@cug.edu.cn).

Yuan Zhou is with the National Engineering Research Center of Geographic Information System, Wuhan 430074, China (e-mail: zhouyuan@cug.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3146275

I. Introduction

WATER is crucial for human survival. Water provides a resource that guarantees people's lives, industrial, and agricultural production; on the other hand, water is the habitat of various aquatic plants and animals. Therefore, reasonable water utilization and protection are vital to the stability of the ecosystem [1]. With recent advancements in earth observation technology, obtaining remote sensing images with high spatial, and time resolution has become more convenient, which greatly simplifies data collection for field surveyors. Therefore, the technology for using remote sensing images to analyse and infer relevant information about ground objects has received widespread attention. In the field of geospatial information, efficient and accurate water body mapping results using remote sensing images have significance for surveying water resources, natural disaster assessment, flood disaster monitoring, and water conservation planning [2]–[5].

Early manual interpretation methods have high accuracy, but they are gradually being replaced by less manual work and more efficient machine learning methods. Early water bodies extraction from remote sensing images mainly relied on spectral information, such as water index method [6]–[8]. However, the water index method does not pay attention to spatial information and has poor universal applicability [9]. In addition, since traditional machine learning algorithms need to manually construct many features for specific water bodies extraction tasks, this method is not suitable for different regions and images [10]–[12].

Currently, with the continuous development of deep learning technology and the existence of massive remote sensing data, the application of deep learning technology in remote sensing image interpretation has become a research hotspot. However, due to the complexity of water body distribution scenarios, inconsistencies of each water body, and differences between water bodies and other similar ground objects, the water body extraction is still challenging [13], [14]. The convolutional neural network (CNN), as a powerful tool of end-to-end image feature learning, is widely employed in water bodies extraction research of remote sensing images [15]–[17]. Among them, the fully convolutional network (FCN) is the first batch of networks utilized for image segmentation [18]. Certain scholars applied

the improved FCN network to realize water bodies extraction and verified the performance from data enhancement, input of different features and transfer learning [19]. However, the continuous downsampling of the FCN leads to a reduction in spatial resolution, which is not conducive to a refined water body extraction. To compensate for the shortcomings of the FCN, certain networks with encoder-decoder structures, such as U-Net [20] and SegNet [21], have been proposed. This type of network realizes the combination of high-level semantic information and low-level features in the decoder structure, enhancing the features' expression. However, compared with natural images, remote sensing images have higher complexity and multiscale characteristics, so these characteristics must be considered for water bodies extraction. Based on an improved U-Net network, Feng *et al.* [22] employed superpixel and conditional random field modules as constraints, greatly preserved the boundary information of a water body, and suppressed the generated salt and pepper noise. Analogously, Li *et al.* [23] increased the number of connections in the middle layer of the U-Net network, and fusing a wavelet transform and grey-level co-occurrence matrix, then achieved rapid and accurate identification of water bodies based on ultrahigh-resolution unmanned aerial vehicle images. Additionally, a method combining Google Earth Engine (GEE) and a multiscale CNN (MSCNN) was proposed to extract water bodies from Landsat images; the MSCNN was applied to train the model, and then the model was input into GEE for water bodies extraction; this method has good performance and is suitable for large-scale and long-term water body change detection in urban backgrounds [24]. However, these methods do not pay more attention to the low-level features of water bodies when encoding images, such as shape features. The methods still encounter the problem of insufficient integrity for large-scale water bodies extraction.

To solve the abovementioned problems, an intelligent water body mapping network (SADA-Net) is proposed in this article, aiming to optimize the shape feature of water bodies while improving the integrity of water bodies, to achieve the purpose of fine water bodies extraction on a large scale. The proposed network integrates shape feature optimization (SFO) [25], atrous spatial pyramid pooling (ASPP) [26], and dual attention (DA) [27] modules. The SFO module enhances the shape information of water bodies and is parallelized with the backbone network in the encoder structure. The ASPP module is utilized to capture multiscale context information using different receptive fields. The DA module includes a position attention module and channel attention module. Water bodies are greatly affected by illumination and location; the DA module can effectively aggregate spatial and channel dimension features, strengthen the context-dependence relationship, and thus, further improve the accuracy of water body extraction.

In summary, the main contributions of this article are listed as follows.

1) A water bodies extraction method (SADA-Net) was proposed to achieve the purpose of fine water bodies extraction on a large scale.
2) The combination of multiple important modules realizes the extraction of small and dense water bodies.
3) The loss function is designed accordingly, which is more conducive to stable convergence.
4) A water body dataset was constructed using Gaofen 2 remote sensing images, and the proposed network achieves excellent performance.

II. Related Works

Different kinds of remote sensing images have various numbers of bands and contain different information, such as optical remote sensing images [28], [29], radar remote sensing images [30], [31], and images generated from the fusion of multisource data [32]–[34]. Therefore, some research methods based on these images have emerged, such as the water index method, object-oriented analysis method, and pixel-level statistical identification.

The normalized difference water index (NDWI) was first proposed by McFeeters [35]. The normalized ratio of the green and the near infrared (NIR) band was applied to highlight water bodies. All positive NDWI numerical regions were considered water bodies, and the areas with negative values were recognized as nonwater areas. However, this method has certain limitations in areas with dense buildings. Xu [36] proposed the modified NDWI by using Landsat imagery to enhance the characteristics of a water body in areas with a large number of built-up areas. Furthermore, this method replaced the NIR band with the mid-infrared band, while the green band was unchanged. An automatic water extraction index was proposed based on Landsat-5 TM imagery [28]. This index contains two formulas that can be applied to unshaded and shadow-covered areas, and they can be associated to improve water bodies mapping. Sharma *et al.* [37] proposed a superfine water index (SWI) based on the Moderate Resolution Imaging Spectroradiometer (MODIS) multispectral imagery, conducting the robust global detection and identification of surface water bodies. To accomplish this task, the SWI replaces the green band in the NDWI with saturation (Sat) obtained from the Hue-Saturation-Value transformation of the red, green, and blue (RGB) bands of the MODIS data, while the NIR band is not changed, as follows:

$$\mathrm{SWI} = \frac{\mathrm{Sat}_{(\mathrm{RGB})} - 7{*}\mathrm{NIR}}{\mathrm{Sat}_{(\mathrm{RGB})} + 7{*}\mathrm{NIR}}. \tag{1}$$

The index method uses the difference in the reflectance of the multispectral band. The ratio method is then applied to amplify the differences between the bands to extract a water body. However, some optical remote sensing images have only RGB bands and lack the other bands required for the corresponding water index. Therefore, this method's application range is limited, and the method does not have generalization for different regions or images.

Object-oriented approaches include two steps: image segmentation and image classification, wherein image segmentation is a crucial procedure that directly determines the advantages and disadvantages of the results [38], [39]. Do *et al.* [40] used RapidEye and IKONOS imagery of the Lao Cai and Can Tho areas in Vietnam to study land cover classification and then achieved

good segmentation performance for water body mapping according to a selected optimal threshold. Object-oriented methods are generally employed in combination with other strategies [41], [42]. By combining the NDWI and object-oriented techniques, Kaplan and Avdan [43] applied Sentinel-2 imagery in the mountainous and urban areas of Macedonia to achieve an excellent mapping result, with kappa values (0.90 and 0.88) that were more than 0.5 higher than those obtained when using NDWI (0.39 and 0.35) alone. In addition, Zhao *et al.* [44] combined NIR spectroscopy and object-oriented methods to extract the central urban water bodies from IKONOS imagery, alleviating the difficulty of water body extraction caused by building shadows. In brief, object-oriented methods fully utilize the texture, shape, and spectrum characteristics of remote sensing images to produce excellent mapping performance. However, the scale segmentation process requires a significant amount of time. In the multiclassification task, certain categories were highlighted under the same scale conditions, while others performed poorly. Therefore, significant parameter adjustments are required when handling the segmentation of massive remote sensing data.

Traditional machine learning such as the support vector machine (SVM) and decision tree (DT) methods are generally pixel-level mapping methods. An SVM classification method is mainly employed to solve binary classification problems and obtain good results when the sample size is small [45]. An SVM is usually applied as a classifier, and the model's generalization and the reliability of the mapping results can reach a relatively good level [46]–[48]. In the processing of high-dimensional data, the SVM method has a significant advantage. Some researchers combined genetic algorithms with an SVM to the fusion of the Radarsat-2 SAR image and Thaichote multispectral imagery data, achieving the mapping of forests, agriculture, residential areas, and water bodies [49]. Additionally, multifeature fusion has an enhanced effect on water body extraction. The SVM method was utilized to handle the fusion result of geometric and attribute features, and the refined extraction of rivers, lakes, reservoirs, and ponds was achieved [50]. DT analysis methods provide high efficiency and descriptiveness, which is helpful for manual analysis. To accurately extract the categories of the regions of interest, McIver and Friedl [51] proposed a supervised DT classification algorithm that merged the prior probability, which was selected to segment the land cover, and a high water body extraction accuracy from coarse-resolution remote sensing images was possible. Acharya *et al.* [52] explored a J48 DT using four bands from Landsat-8 OLI imagery to extract water bodies in the northern part of the Han River basin in Gangwon-do, South Korea. The result showed that this method could obtain good results as the image resolution improved and as the image range gradually increased. Pixel-level water body mapping also includes other methods, such as the maximum likelihood function method [53], artificial neural networks, and random forests [54], which have good extraction effects in different application requirements.

CNN water body mapping is an end-to-end, pixel-level extraction method without manual construction of features. Owing to local connection and parameter sharing characteristics, CNN reduce the number of parameters, especially for the two-dimensional (2-D) matrix of images, to achieve excellent processing efficiency. Previously, many researchers have determined that CNN have a strong effect on water body mapping from remote sensing images [55]–[57]. Generally, coarse resolution is an important cause for poor extraction results, but data fusion, and multiscale feature fusion effectively alleviate such problems. Sun *et al.* [58] proposed a network model that combined multifilter and multiresolution data using radar data and high-resolution, optical remote sensing images as fusion input. The multiresolution data help segment the object boundary by reducing the impact of noise. Regarding multiscale feature fusion water body mapping, a rational network structure design can effectively suppress noise interference, such as building shadows and roads [59], [60]. For instance, a multiscale erasing-attention module was designed in a network to enhance the expression of features [61]. Similarly, Xia *et al.* [62] proposed a dense skip connection network with multiscale feature fusion and an attention mechanism for water body extraction. The shallow features and large-scale attention modules of the network are applied to locate the global information of the water body, whereas the depth feature and a small-scale attention module are used to refine the water body boundary. Li *et al.* [63] proposed an automatic water body extraction network that can be applied to different remote sensing images. The network embeds DenseNet in each layer to connect different feature maps and integrates spatial and spectral information through a dense-local-feature-compression module. The experiments show that the network has shown excellent performance on multiple datasets. Additionally, the selection of a loss function in CNN has a significant influence on the water body extraction. Dong *et al.* [64] enhanced the expression of water body features during the upsampling process of their proposed network and simultaneously improved the loss function to achieve strong generalization for optical remote sensing images. Moreover, a structure that compressed redundant layers in a deep CNN was proposed, and excellent water body segmentation results on Google imagery were achieved by constructing a new loss function to sharpen the boundary [65]. CNN water body mapping remains in continuous development because of the need for a large amount of sample data, the adjustment of the learning rate, and the appropriate choice of the loss function. Compared with traditional mapping methods that require expert experience, the CNN method is a better choice.

III. Data and Methodology

A. Dataset

In this article, Gaofen 2 satellite images were employed. The dataset, which is referred to as the Gaofen Image Dataset (GID), was produced and published by researchers at the State Key Laboratory of Wuhan University, China [66]. Gaofen 2 was the second high-definition earth observation system satellite that was launched by the China National Space Agency and the first civilian optical remote sensing satellite independently developed by China with a spatial resolution better than 1 m. The satellite is equipped with two high-resolution, panchromatic/multispectral cameras with a ground sampling distance of 1 m in panchromatic

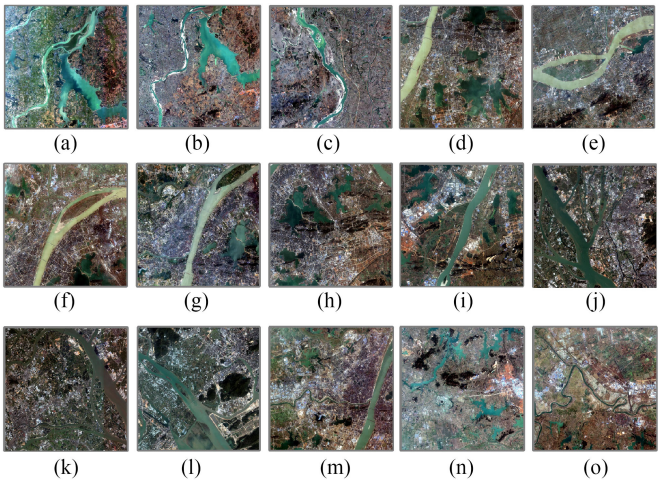

Fig. 1. Fifteen selected RGB, three-band images from the Gaofen Image Dataset (GID). (a)–(l) are the images selected for training, and (m)–(o) are selected for inference. Given space limitations, the multispectral images are not displayed here.

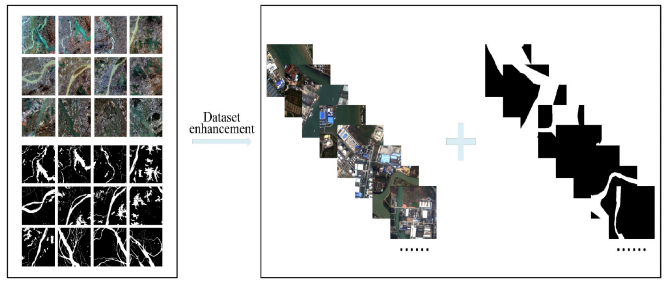

Fig. 2. Data enhancement and expansion. We selected 12 RGB, three-band images from the GID dataset employed for the training and verification sets. After data enhancement and expansion operations, such as translation, rotation, scaling, vertical mirroring, lighting adjustment, and adding pepper and salt noise points, the 256 × 256 pixel patch images were generated.

cameras and a ground sampling distance of 4 m in multispectral cameras. The subsatellite, point spatial resolution can reach 0.8 m, with submetre spatial resolution, and Gaofen 2 has high positioning accuracy and rapid attitude manoeuvrability. Since its launch in 2014, Gaofen 2 has been utilized for critical applications, such as land cover surveys, environmental monitoring, crop estimation, and construction planning. The GID dataset contains RGB, three-band images and multispectral, four-band images (RGB + NIR bands). In addition, the providers have produced a total of 150 labeled, complete land cover images that include five classes. The dimensions of each remote sensing image are 6800 pixels × 7200 pixels, and they cover an area of 506 km^2. The dataset includes nearly 60 cities in China and has a large coverage range. As this article focuses on water body mapping using remote sensing images, 15 three-band/four-band images with a wide range of water bodies were selected from the GID dataset. Only the RGB, three-band images are shown in Fig. 1; the multispectral bands' images are not displayed. The study areas contain numerous small water bodies, including small ponds, lakes, and rivers, which are generally composed of a small number of pixels that are densely distributed. Since the team that provided the dataset labeled the land types using 5 categories, we relabeled only the water bodies.

B. Data Preprocessing

A CNN needs to be trained with a large number of samples to obtain a well-weighted model. However, the dataset obtained for this study contains a small number of samples with only 15 images. As the learning outcome of the CNN is not affected when the dataset is rotated and shifted, the dataset in this article was enhanced and expanded. First, a remote sensing image and label data were randomly cropped into 256 × 256 pixel patches, which alleviates the problem of the graphics card capacity not being capable of handling large images. After cropping, the patched images underwent translation, rotation, scaling, vertical mirroring, lighting adjustment, and salt and pepper noise point additions to enhance sample diversity. The sample dataset after data enhancement and expansion is shown in Fig. 2.

C. SADA-Net Architecture

A CNN is generally composed of an input layer, hidden layers, and a fully connected layer. The hidden layers typically contain multiple convolutions and pooling operations [67]. After convolution and pooling operations, the features become more advanced, and the semantic information becomes more affluent. However, reducing the feature map dimensions leads to a severe loss of relevant information, such as spatial location, producing unsatisfactory segmentation results. Researchers have introduced the concept of atrous convolution [68], which reduces the downsampling rate while ensuring a sufficient receptive field, so the result produced is more refined. In addition, a CNN mostly uses an encoder–decoder structure, as indicated in current studies. This process uses gradual downsampling to extract features and upsampling to restore the feature map's size. In general, low-level features in the encoder structure are connected to the advanced semantic information in the decoder module, and both have the same size. The final output segmentation results are smoother and more accurate.

This article proposed a refined, multiscale water body mapping network that is based on water body shape characteristics and context information. The overall network has an encoder–decoder structure (see Fig. 3). The backbone network performs feature learning, while an SFO module is connected in parallel [25], which enriches the entire network's shape feature representation.

The encoder module of the network extracts different features. First, a backbone network was selected, and we chose the improved Xception used in DeeplabV3+ [26], [69] as the backbone network because of its strong feature-learning ability. However, there is a difference here. The original model output channel dimension is 2048, we set it in our experiments to 1024. After downsampling four times, the obtained feature map Layer-4 (16 × 16 pixels) was chosen as the input of two streams. The first stream is the DA module, which strengthens the spatial and channel context relationship. The second stream combines

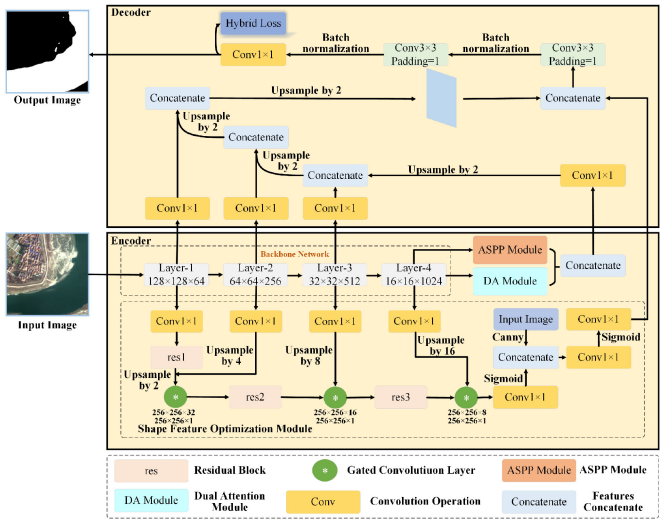

Fig. 3. Framework of the proposed SADA-Net. The backbone network is the improved Xception. The multiscale and context information is obtained via the ASPP and DA modules. In addition, the parallel SFO module strengthens the shape feature information of the water body.

the multiscale feature information of the water body through the ASPP module. The enhanced features of both streams were then connected. During the downsampling process, the parallel SFO module strengthens the expression of the shape features of the water body.

The network's decoder module fuses multiscale features, context information, and shape feature information. Simultaneously, the feature map is restored to the size of the input image. First, the feature map, which is the fusion of the enhanced features of both streams, is upsampled twice as the output. Second, the output is connected to the low-level features of Layer-3, and then the low-level features of Layer-2 and Layer-1 relate to the output of the previous fusion result. Third, after the low-level features of Layer-1 are connected to the output features of the previous unit, their output features are upsampled twice and fused with the shape features learned by the SFO module. Finally, an output image is obtained after multiple convolution and batch normalization operations.

D. ASPP Module

The ASPP module uses atrous convolution for the feature extraction of multiscale information. Atrous convolution mostly solves the low-resolution problem caused by the general convolution and pooling operation in downsampling. Atrous convolution is based on general convolution, with the addition of a dilation rate. In this article, we defined this rate as r (see Fig. 4). Fig. 4(a) shows the atrous convolution when r is 1, which is the general convolution. Fig. 4(b) and (c) shows the atrous convolution when r is greater than 1, representing atrous convolution when the dilation rates are 2 and 4, respectively. The input feature map samples every r–1 pixel in the feature map (red points in Fig. 4). Regarding the convolution kernel, atrous convolution expands the convolution kernel size, inserts r–1 zeros between adjacent points of the given convolution kernel, and then uses the

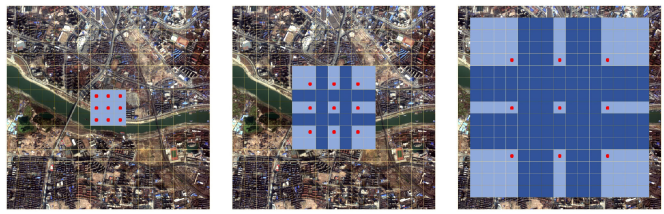

Fig. 4. Atrous convolution kernels at dilation rates of 1(a), 2(b), and 4(c).

input feature map and expanded convolution kernel to perform the convolution operation. Both of these processes increase the receptive field of the input image. In summary, the mathematical expression of the convolution kernel size of atrous convolution and output feature image size are provided in (2) and (3) as follows [70]:

$$K = k + (k-1) * (r-1) \tag{2}$$

where K is the size of the convolution kernel of atrous convolution, k is the size of the general convolution kernel, which is the number of rows and columns of sampled red points shown in Fig. 4, and r is the dilation rate

$$O = \text{Floor}\left(\frac{i + 2p - K}{s}\right) + 1 \tag{3}$$

where O represents the size of the output feature image, Floor produces a downward-rounded integer, i is the size of the input feature image, p is the padding, which represents the number of zeros at the edge of the matrix, and s is the stride.

Atrous convolution is employed to increase the size of the receptive field without changing the dimensions of the feature map. The receptive field calculation is provided in (4), and the blue grid area in Fig. 4 illustrates its representation. The receptive field is the size of the receptive range of the neurons at different positions within the network compared with the original feature image. The larger is the value, the wider the range of the original feature image processed by the neurons and the greater the amount of global information that is contained [71]

$$r_n = r_{n-1} + (k_n - 1) \prod_{i=1}^{n-1} s_i \tag{4}$$

where r_n is the size of the receptive field in the nth convolutional layer, r_{n-1} is the size of the receptive field in the $(n-1)$th convolutional layer, k_n is the size of the atrous convolution kernel in the nth convolutional layer, and s_i is the stride size of the ith convolutional layer, which is generally set to 1. It is common to set the receptive field size of the original image to 1.

The ASPP module uses atrous convolution with different dilation rates to extract the multiscale features of water bodies. To prevent the dilation rate from being the same size as the input feature map, which will cause the atrous convolution to be degraded into a 1×1 convolution kernel, the global information of the input feature map is obtained using global pooling. The SADA-Net applies the ASPP structure after downsampling the

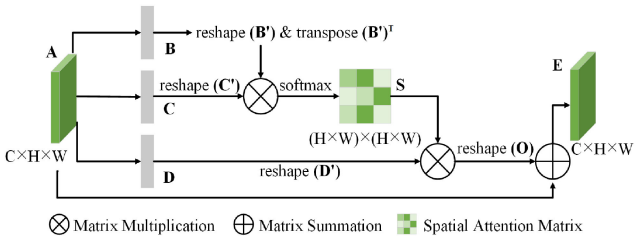

Fig. 5. Location attention module.

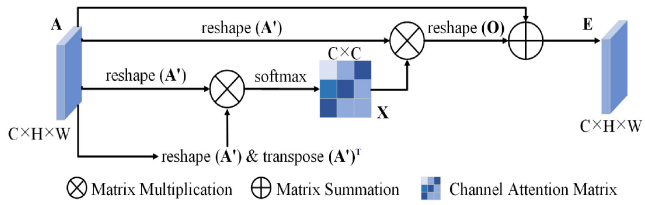

Fig. 6. Channel attention module.

input image four times, and the selected dilation rates are 1, 4, 8, and 12. As water bodies are unevenly distributed and the scale of each water body is different, local details are enriched via multiscale feature fusion.

E. DA Module

An attention mechanism [72], [73] imitates the human attention thinking mode and screens the extensive amount of information to obtain information that is valuable to the research subject. This information was applied to the field of computer vision early. In 2014, the Google Mind team added attention mechanisms to image classification and achieved good results in recurrent neural networks [74]. Thus, attention mechanisms became widely utilized in image and natural language processing. A DA mechanism is divided into location attention and channel attention modules. This research mainly uses a DA module to strengthen the global dependence of the spatial and channel dimensions of water bodies in remote sensing images.

1) Location Attention Module: A location attention module (see Fig. 5) is employed to obtain the weighted sum of the features of all the location units, that is, it aggregates the features of each location regardless of distance. Thus, the location attention can focus on the global spatial context information of a water body.

In this article, the feature map $A \in \mathbb{R}^{C\times H\times W}$ is the output of Layer-4 in the encoder structure. Features *B, C,* and *D* are then acquired using convolution operations, where $\{B, C, D\} \in \mathbb{R}^{C\times H\times W}$. Feature $B \in \mathbb{R}^{C\times H\times W}$ is dimensionally reshaped to $B' \in \mathbb{R}^{C\times(H\times W)}$, and $B' \in \mathbb{R}^{C\times(H\times W)}$ is transposed to obtain feature $(B')^T \in \mathbb{R}^{(H\times W)\times C}$. Subsequently, $C \in \mathbb{R}^{C\times H\times W}$ is reshaped to $C' \in \mathbb{R}^{C\times(H\times W)}$; matrix multiplication with $(B')^T \in \mathbb{R}^{(H\times W)\times C}$ is performed; and the output result is used to obtain the position attention map $S \in \mathbb{R}^{(H\times W)\times(H\times W)}$ of each relative position using the Softmax function. The calculation of each position element value of *S* is given as (5). $D \in \mathbb{R}^{C\times H\times W}$ is reshaped to $D' \in R^{C\times(H\times W)}$. Matrix multiplication between the transposed matrix values of $S \in \mathbb{R}^{(H\times W)\times(H\times W)}$ and $D' \in R^{C\times(H\times W)}$ is performed. The result is dimensionally reshaped to $O \in \mathbb{R}^{C\times H\times W}$ and added to the original feature map $A \in \mathbb{R}^{C\times H\times W}$ to obtain the final feature matrix $E \in \mathbb{R}^{C\times H\times W}$. The calculation for each positioning element in $E \in \mathbb{R}^{C\times H\times W}$ is given by (6)

$$s_{ji} = \frac{exp\,(B_i \cdot C_j)}{\sum_{i=1}^{N} exp\,(B_i \cdot C_j)} \tag{5}$$

where S_{ji} represents the influence of the *i*th position relative to the *j*th position, B_i and C_j represent the *i*th position value and *j*th position value, respectively, of features $(B')^T \in \mathbb{R}^{(H\times W)\times C}$ and $C' \in \mathbb{R}^{C\times(H\times W)}$, respectively, and *N* is the total number of dimensions

$$E_j = \alpha \sum_{i=1}^{N} (s_{ji} D_i) + A_j \tag{6}$$

where E_j is the element value in feature $E \in \mathbb{R}^{C\times H\times W}$, α is initialized to 0 and gradually adjusts to a higher weight, D_i represents the *i-th* position value of $D' \in \mathbb{R}^{C\times(H\times W)}$, and A_j is the *j*th position value of $A \in \mathbb{R}^{C\times H\times W}$.

2) Channel Attention Module: The purpose of a channel attention module is to create a correlation among the channels and enhance the response capability of the semantic information. The structure of the channel attention module is shown in Fig. 6.

The channel attention and location attention modules have similarities. First, the feature map $A \in \mathbb{R}^{C\times H\times W}$ is reshaped in dimensions to $A' \in \mathbb{R}^{C\times(H\times W)}$ to obtain the transposed $(A')^T \in \mathbb{R}^{(H\times W)\times C}$. Second, after completing the matrix multiplication operation of $A' \in \mathbb{R}^{C\times(H\times W)}$ and $(A')^T \in \mathbb{R}^{(H\times W)\times C}$, the channel attention matrix $X \in \mathbb{R}^{C\times C}$ is obtained following Softmax function normalization. The calculation of each position element value of $X \in \mathbb{R}^{C\times C}$ is given by (7). Third, matrix multiplication of the transposed $X \in \mathbb{R}^{C\times C}$ and $A' \in \mathbb{R}^{C\times(H\times W)}$ matrices are performed, and the results are reshaped into matrix $O \in \mathbb{R}^{C\times H\times W}$. Last, $O \in \mathbb{R}^{C\times H\times W}$ is added to the original feature map $A \in \mathbb{R}^{C\times H\times W}$ to obtain the output result $E \in \mathbb{R}^{C\times H\times W}$. The calculation for each positioning element in $E \in \mathbb{R}^{C\times H\times W}$ is given by (8)

$$x_{ji} = \frac{\exp(A_i' \cdot (A')_j^T)}{\sum_{i=1}^{C} \exp(A_i' \cdot (A')_j^T)} \tag{7}$$

where x_{ji} represents the influence of the *i*th channel relative to the *j*th channel, A_i' is the *i*th position value of $A' \in \mathbb{R}^{C\times(H\times W)}$, $(A')_j^T$ represents the value of the *j*th position of the transposed matrix $(A')^T \in \mathbb{R}^{(H\times W)\times C}$, and *C* represents the number of channels

$$E_j = \beta \sum_{i=1}^{C} (x_{ji} A'_i) + A_j. \tag{8}$$

Generally, the output feature matrix $E \in \mathbb{R}^{C\times H\times W}$ obtained via the channel attention module and output feature matrix $E \in \mathbb{R}^{C\times H\times W}$ obtained via the location attention module are added.

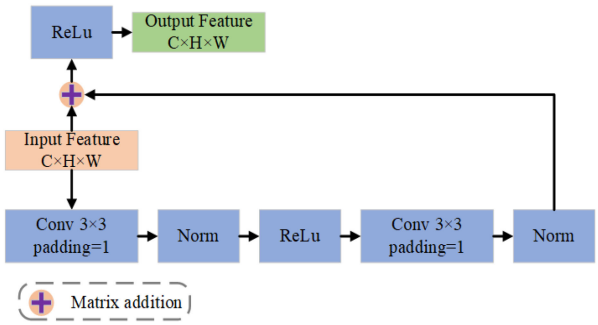

Fig. 7. Details of the residual block.

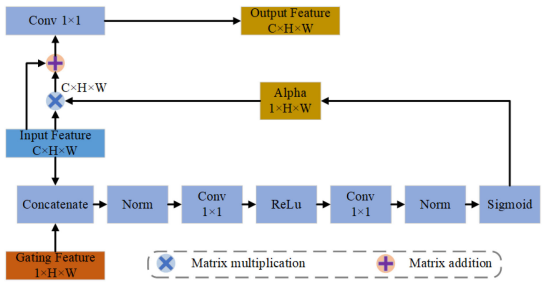

Fig. 8. Details of the GCL.

Thus, this article applied fused features to enhance the context information of the water bodies.

F. SFO Module

The SADA-Net introduced an SFO module in the backbone network of the encoder structure; its primary function is to optimize the shape information of water bodies. Shape information and other semantic features are connected by a gated convolutional layer (GCL), thus enriching the expression of semantic information in the feature fusion stage. The structure of the SFO module is composed of residual blocks and GCLs.

The detailed structure of the GCL is shown in Fig. 8, and the output feature size is $\mathbb{R}^{C\times H\times W}$. The Input Features $\in \mathbb{R}^{C\times H\times W}$ and Gating Feature $\in \mathbb{R}^{1\times H\times W}$ originate from two adjacent feature maps in the backbone network. For instance, Layer-1 successively undergoes a convolution operation (kernel size is 1×1) and residual block (see Fig. 7); After sampling up 2 times, a feature map of size $\mathbb{R}^{32\times H\times W}$ (Input Features) is obtained, which is one of the inputs to the GCL. In addition, the second feature map is obtained by applying a convolution operation (kernel size is 1×1) to Layer-2 and upsampling by a factor of 4; the result of this operation is Gating Feature $\in \mathbb{R}^{1\times H\times W}$. Similarly, Layer-3 and Layer-4 perform the same task as Layer-1.

As shown in Fig. 8, the Input Feature to the GCL undergoes a series of convolution and batch normalization operations after feature fusion and using a sigmoid function to generate a weight value, which is referred to as the attention graph $\alpha_t \in \mathbb{R}^{1\times H\times W}$ and defined as follows:

$$\alpha_t = \sigma(C_{1\times1}(S_t||r_t)) \tag{9}$$

where σ represents a sigmoid function, $C_{1\times1}$ represents the ordinary 1×1 convolution kernel, $||$ is the concatenate operation, $r_t \in \mathbb{R}^{C\times(H/m)\times(W/m)}$ and $s_t \in \mathbb{R}^{C\times H\times W}$ represent the middle layers of the backbone network and the SFO module, respectively, and m is the stride of the backbone network. In this article, m is 16.

The weight $\alpha_t \in \mathbb{R}^{1\times H\times W}$ is then multiplied by $s_t, \in \mathbb{R}^{C\times H\times W}$, which is the middle layer of the SFO module, and the result is summed with the Input Feature. After a 2-D convolution layer, the output result is expressed as the Output Feature shown in Fig. 8. The calculation is given as follows:

$$\begin{aligned}\hat{s}_t^{(i,j)} &= \left(s_t \,(\textcircled{\scriptsize o})\, w_t\right)_{(i,j)} \\ &= \left(\left(s_{t_{(i,j)}} \odot \alpha_{t_{(i,j)}}\right) + s_{t_{(i,j)}}\right)^T w_t\end{aligned} \tag{10}$$

where $\textcircled{\scriptsize o}$ symbolizes the gated convolution operation, ω_t is the weight of the channel-level convolution kernel, $\odot$ is the matrix multiplication, and $s_{t_{(i,j)}}$ and $\alpha_{t_{(i,j)}}$ represent the value of $s_t \in \mathbb{R}^{C\times H\times W}$ and $\alpha_t \in \mathbb{R}^{1\times H\times W}$, respectively.

In this article, the SFO module includes the features obtained after three gate operations and the features obtained using Canny edge detection [75]. The two types of features are fused.

G. Loss Function

Extracting water bodies from remote sensing images is a binary classification problem. Herein, we determined that the proportions of water bodies and background information after data enhancement are balanced. Therefore, in this article, the weighted sum of the binary cross-entropy function BCELoss, Dice Loss, and Boundary Loss (BD Loss) was adopted as the loss function [76]–[78] as it can make the loss function converge stably during the training and validation processes. Dice Loss is a regional loss metric that alleviates data imbalances. Boundary Loss focuses on the accuracy of boundaries and acts to refine boundaries. BCELoss has a role in stabilizing the hybrid loss function value. The calculations of BCELoss and Dice Loss are provided by (11) and (12), respectively, as follows:

$$\begin{aligned}&\mathrm{BCE}_{\mathrm{loss}} \\ &= -\left(y_{gt} \times \log\left(\delta\left(y_{pd}\right)\right) + \left(1 - y_{gt}\right) \times \log(1 - \delta(y_{pd}))\right)\end{aligned} \tag{11}$$

where δ is a sigmoid function, y_{pd} is the result of the sample predicted to be the water body, and y_{gt} is the label of the sample

$$Dice_{\mathrm{loss}} = 1 - \frac{2\left|y_{gt} \cap y_{pd}\right|}{\left|y_{gt}\right| + \left|y_{pd}\right|} \tag{12}$$

where $|y_{gt} \cap y_{pd}|$ represents the intersection between y_{gt} and y_{pd}.

The Boundary Loss consists of the following calculations:

$$\begin{aligned}y_{gt}^b &= \mathrm{pool}\left(1 - y_{gt}, \theta_0\right) - \left(1 - y_{gt}\right), y_{pd}^b \\ &= \mathrm{pool}\left(1 - y_{pd}, \theta_0\right) - \left(1 - y_{pd}\right)\end{aligned} \tag{13}$$

$$y_{gt}^{b,ext} = \mathrm{pool}\left(y_{gt}^b, \theta\right), \; y_{pd}^{b,ext} = \mathrm{pool}\left(y_{pd}^b, \theta\right) \tag{14}$$

$$P^c = \frac{\text{sum}\left(y_{pd}^b \odot y_{gt}^{b,ext}\right)}{\text{sum}\left(y_{pd}^b\right)},\ R^c = \frac{\text{sum}\left(y_{gt}^b \odot y_{pd}^{b,ext}\right)}{\text{sum}\left(y_{gt}^b\right)} \tag{15}$$

$$BF^c = \frac{2P^cR^c}{P^c + R^c},\ \text{BD}_{\text{loss}} = 1 - BF^c \tag{16}$$

where pool $(\cdot,\cdot)$ is the max-pooling operation, and θ_0 is the sliding window size, i.e., the convolution kernel size, which is usually set to 3. $y_{gt}^{b,ext}$ and $y_{pd}^{b,ext}$ represent the extended boundaries of the true map and predicted map, respectively, which are obtained by maximum pooling, where θ represents the size of the padding operation, which is set to 5 in this article; P^c and R^c represent precision and recall, respectively; sum $(\cdot,\cdot)$ is the sum operation, and $\odot$ is the matrix multiplication. Thus, the Boundary Loss (BD_{loss}) is obtained by calculating BF^c, which is defined by P^c and R^c.

The final form of the loss function ($\text{Loss}_{\text{hybrid}}$) is given by (17), where μ and ε are weighting coefficients between 0 and 1. In this article, when ω, μ and ε were set to 1.0, 0.5, and 0.5, respectively, the loss function converges stably

$$\text{Loss}_{\text{hybrid}} = \omega * Dice_{\text{loss}} + \mu * \text{BCE}_{\text{loss}} + \varepsilon * \text{BD}_{\text{loss}}. \tag{17}$$

H. Evaluation Metrics

The performance of the proposed method was evaluated using five evaluation indicators: precision (P), recall (R), F1-Score (F1), overall accuracy (OA), and a structural similarity index measure (SSIM) [79], which are defined as follows:

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{18}$$

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{19}$$

$$\text{F1} = 2 \times \frac{P \times R}{P + R} \tag{20}$$

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{21}$$

$$S\ (X, Y) = \frac{(2\mu_x\mu_y + C_1)\,(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{22}$$

where TP represents the true positives, the number of correct extraction pixels in this article, FN indicates the false negatives, the number of water body pixels not extracted in this article, FP denotes the false positives, the number of incorrect pixel extractions, and TN represents the true negatives, the number of nonwater body pixels that were extracted.

Additionally, *X* and *Y* denote the ground truth and prediction result map, respectively; μ_x and μ_y represent the average and standard deviation, respectively, of *X* and *Y*, respectively; σ_{xy} is the covariance of *X* and *Y*; and C_1 and C_2 are constants. Generally, the larger is the value, the more similar the structure of the two images, which means that the shape is better optimized.

TABLE I
EXPERIMENTAL RESULTS WITH DIFFERENT STRUCTURE COMBINATIONS

Methods	P	R	F1	OA	SSIM
Original network (Dice + BCE Loss)	0.8798	0.8826	0.8812	0.9544	0.8088
Original network	0.8925	0.8834	0.8879	0.9585	0.8137
With ASPP	0.9117	0.8926	0.9020	0.9629	0.8262
ASPP + DA	0.9104	0.9114	0.9109	0.9625	0.8325
SADA-Net	0.9345	0.9229	0.9287	0.9678	0.8460

IV. EXPERIMENTS AND ANALYSIS

A. Experimental Setup

The operating system utilized in the experiment was Linux; the experimental framework was Pytorch 1.4; the training process was accelerated on an NVIDIA GeForce RTX 2080Ti GPU, the video memory is 11 GB; and Python 3.7 was employed.

In this article, each of the 15 selected GID images occupied a large dimension; the water bodies were scattered; and there were many small water bodies, including ponds and lakes. In total, 12 images [see Fig. 1(a)—(l)] were applied as training data, and the remaining 3 images [see Fig. 1(m)—(o)] were utilized as predictions. Each of the 12 images was randomly divided into 2000 patches, wherein each patch size was 256 × 256 pixels, and the data were enhanced to obtain the input data. A total of 24000 patch images were obtained, of which 80% were selected as the training set, and 20% were selected as the validation set. The epoch of the training set was 50; the batch size was 6; the initial learning rate adopted a larger setting of 0.001; and the adaptive learning rate optimizer was Adam.

B. Ablation Experiments

The SADA-Net proposed herein integrates three important structures: the ASPP, DA, and SFO modules, which enrich multiscale context information, global dependence, and the expression of water body shape characteristics. First, this section discusses the influence of the loss function on the experimental results. Second, as each module has an important influence on the proposed network, the influence of the different structural modules on the quality of the experimental results are discussed. As shown in Fig. 9, the predicted input images had a size of 1024 × 1024 pixels, and they were cropped from Fig. 1(m)–(o). Table I summarizes the quantitative results.

As shown in Fig. 9(b) and (c), the prediction results of these two columns are the results of the original network under the combination of Dice Loss and BCE Loss, and the original network under the combination of Dice Loss, BCE Loss, and BD Loss. The combination of Dice Loss and BCE Loss is applied to make the loss convergence more stable of the training process. This combination is also commonly employed in various remote sensing image feature extraction studies [80]–[82]. Therefore, this article will not discuss these two loss functions. It can be seen from the results that the overall difference between the results of Fig. 9(b) and those of Fig. 9(c) is not particularly obvious, but it can be seen from the blue and red rectangles that after adding the BD Loss, it has certain advantages in extracting smaller dense water bodies, increasing the separation among

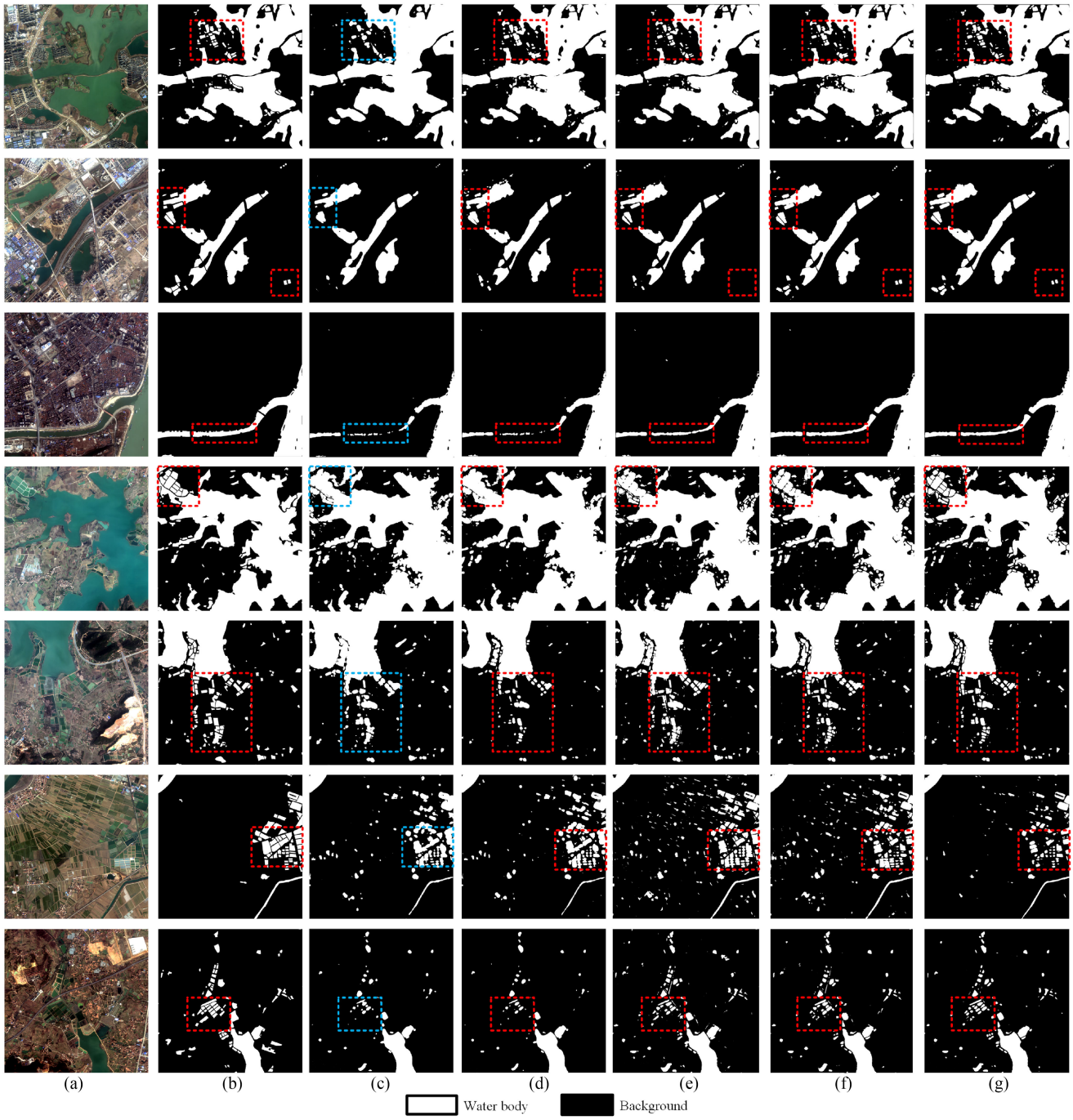

Fig. 9. Experimental results with different structure combinations. Among them, the blue rectangle marks show the experimental results of the original network (Dice + BCE Loss), while the red rectangle marks show the results under Dice + BCE + BD Loss. (a) Input image. (b) Ground truth. (c) Original network (Dice + BCE Loss). (d) Original network. (e) With atrous spatial pyramid pooling module. (f) With ASPP + dual attention (DA) module. (g) Proposed SADA-Net.

small water bodies. Specifically, the small water body in this article mainly refers to artificial ponds. The distribution of these water bodies is mainly scattered and dense, and the transition area between the background and the water body is small and difficult to divide. In brief, other experiments in this article are based on the combination of three loss functions to better promote the boundary separation among water bodies.

The original network extracted image features with Xception in the encoder structure and connected low-level features in the decoder stage, to achieve water bodies mapping. However, missing and incorrect extractions occurred (as shown in Fig. 9(d), marked with red rectangles). Thus, different modules were incorporated into the original network to make feature enhancements.

The ASPP module mainly expands the range of the receptive field, integrating the water body's context information at multiple scales. In the GID, the water bodies were different in terms of their shapes and sizes; certain water bodies account for

a large proportion of an image, and certain water bodies are too small. Therefore, the ASPP module is conducive to correlating context information when the water body size is different. As shown in Fig. 9(e), based on the precision value (0.9117) and recall value (0.8926), the result of adding the ASPP module can strengthen the separation of transitional areas between the water body and the background compared with the original network result. However, there are still some missing water bodies, as shown in the third and seventh prediction results of Fig. 9(e). It is speculated that the atrous rate setting is not dense enough. However, a dense atrous rate setting will cause a large increase in memory, so this article does not set excessively dense atrous rates.

To mitigate the shortcomings of the ASPP module, the DA module is integrated. The location and channel attention modules are beneficial for improving the integrity of water body extraction [as shown in Fig. 9(f)]. Therefore, with the feature fusion of the ASPP and DA modules, the water body extraction integrity was significantly improved. As a result, the recall reached 91.14%, which was 1.88% higher than when only the ASPP module was applied. Furthermore, the F1-Score was 91.09%, and the overall accuracy was 96.25%. The SSIM also increased to 83.25% due to an increase in the integrity of water bodies.

Although the enhanced features of the ASPP and DA modules are conducive to multiscale water body mapping and improvement in integrity, other ground objects may be mistaken as water bodies as the GID contains noise features similar to water bodies. Hence, the precision of the combination (ASPP + DA) is slightly lower than only the ASPP module. To resolve this issue, the SFO module was integrated since it can help enhance the water body shape feature; thus, it improves the status of incorrect extraction. This scenario achieved a precision of 93.45%, recall of 92.29%, and F1-Score and overall accuracy values of 92.87% and 96.78%, respectively. Furthermore, as shown in Fig. 9(g), marked with red rectangles, the SFO module enabling the SADA-Net to reasonably address the relationship between the water body and the background, which is beneficial to the extraction of small and dense water bodies. Hence, the SSIM is better, with a value of 84.60%. Therefore, we selected the network structure combining the three modules in this article as the optimal choice.

C. Results and Analysis

To verify the performance of the proposed SADA-Net, it was compared with SegNet [21], LinkNet [83], DeeplabV3+ [26], Attention U-Net [84], MECNet [85], and MSResNet [86]. The inferential input images are applied with a size of 3072 × 3072 pixels, originating from Fig. 1(n)–(o). The experimental results (RGB bands) are shown in Fig. 10. Given that the Gaofen 2 image datasets are multispectral images (RGB+NIR), this article also utilized the multispectral data of the GID to perform water body mapping and to verify the performance of the proposed network on multispectral images, as shown in Fig. 11. Quantitative calculations were performed simultaneously, and the results are summarized in Tables II and III.

TABLE II
EXPERIMENTAL RESULTS OF SADA-NET COMPARED WITH OTHER METHODS USING RGB DATA

Methods	P	R	F1	OA	SSIM
SegNet	0.8283	0.7880	0.8076	0.9348	0.7697
LinkNet	0.8608	0.7707	0.8133	0.9385	0.7730
DeeplabV3+	0.8743	0.8170	0.8447	0.9478	0.8094
Attention U-Net	0.8628	0.8381	0.8503	0.9487	0.8183
MECNet	0.8554	0.8386	0.8470	0.9474	0.8131
MSResNet	0.8805	0.8529	0.8665	0.9543	0.8386
SADA-Net	0.9203	0.8536	0.8857	0.9618	0.8579

TABLE III
EXPERIMENTAL RESULTS OF SADA-NET COMPARED WITH OTHER METHODS USING MULTISPECTRAL DATA

Methods	P	R	F1	OA	SSIM
SegNet	0.8735	0.8529	0.8631	0.9519	0.8276
LinkNet	0.8841	0.8546	0.8691	0.9553	0.8390
DeeplabV3+	0.8712	0.8639	0.8675	0.9542	0.8332
Attention U-Net	0.8887	0.8840	0.8863	0.9606	0.8606
MECNet	0.8904	0.8701	0.8801	0.9588	0.8517
MSResNet	0.8924	0.8738	0.8830	0.9598	0.8538
SADA-Net	0.9257	0.9473	0.9614	0.9840	0.9470

As shown in Fig. 10, the proposed SADA-Net and compared methods have good water body extraction performance. For the larger water bodies, each method showed high precision and F1-Score; all are greater than 80%. However, other methods are less effective at extracting small and dense water bodies (as shown in Fig. 10, marked with red ellipses).

SegNet is a classic encoder–decoder structure, which improves the shortcomings of the FCN. The fusion of high-level and low-level semantic features makes the output result have a high resolution, enabling it to recover to the original input image size through stepwise upsampling in the decoder module. On the GID RGB bands, SegNet achieved good results with a high precision of 82.83%. However, as it generates sparse feature maps in the encoder structure and only decodes the feature maps of the last layer, it is not conducive to describing the segmentation details. Hence, SegNet failed to show good performance for the integrity of the water body; its recall is 78.80%.

LinkNet is famous for its small network structure parameters, fast training speed, and real-time performance. LinkNet directly connects the encoder and decoder to improve the accuracy of the output result, which greatly reduces the network training time. Although LinkNet is aimed at real-time mapping, it compensates for lost details at different levels of the encoder layer by directly connecting the encoder and decoder, enabling high water body extraction precision. In this article, LinkNet trained a highly effective model within a short period. However, due to the lack of context information of complex scenes, there are some missing water body extraction cases, reaching a low recall of 77.07%, an F1-Score of 81.33%, and overall accuracy of 93.85%. However, LinkNet performs better than SegNet in terms of structural similarity, with an SSIM value of 77.30%.

The DeeplabV3+ network incorporates an ASPP structure for multiscale context information acquisition, which has significant

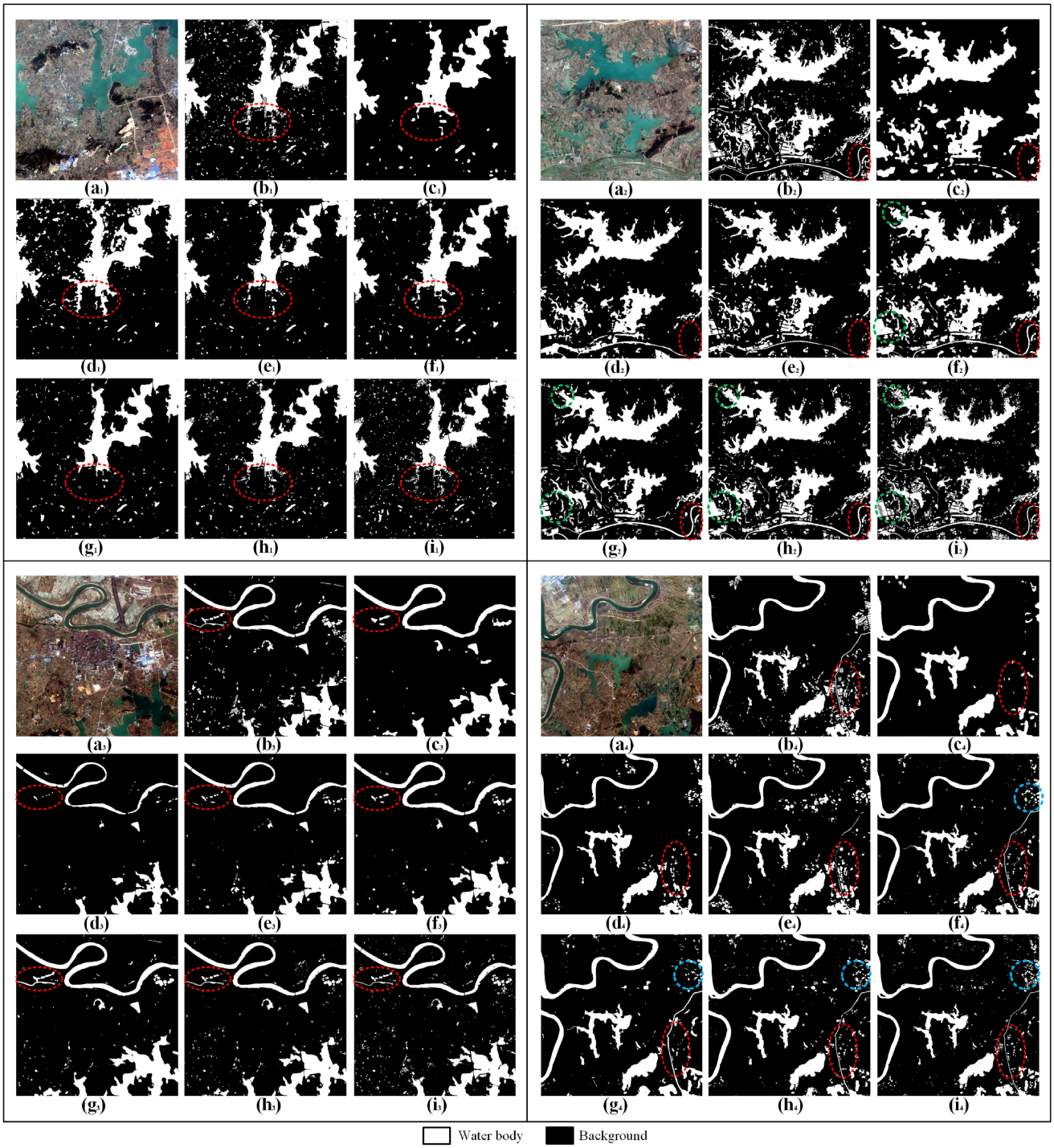

Fig. 10. Water body extraction results using the RGB bands of the GID with (a_1–a_4) Input images, (b_1–b_4) Ground truths, (c_1–c_4) SegNet, (d_1–d_4) LinkNet, (e_1–e_4) DeeplabV3+, (f_1–f_4) Attention U-Net, (g_1–g_4) MECNet, (h_1–h_4) MSResNet, and (i_1–i_4) Proposed SADA-Net.

advantages for large-scale water body extraction in complex scenarios. Simultaneously, the network combines high-level semantic features with low-level features in the decoder structure, making the output water body edge information smoother. However, the DeeplabV3+ network focuses more on the high-level semantic information of a water body and less on the low-level features; it only incorporates a low-level feature in the decoder stage. Therefore, the DeeplabV3+ network cannot describe the details for small and dense water bodies. DeeplabV3+ performed well on a wide range of water body extractions, achieving high precision, and reaching 87.43%, while it had a low recall because of the missing extraction of both small water bodies and thin water bodies. Thus, in general, the DeeplabV3+ network is suitable for large-scale, water body extraction.

With the inclusion of many low-level features, Attention U-Net well depicts boundaries; the SSIM value is 81.83%.

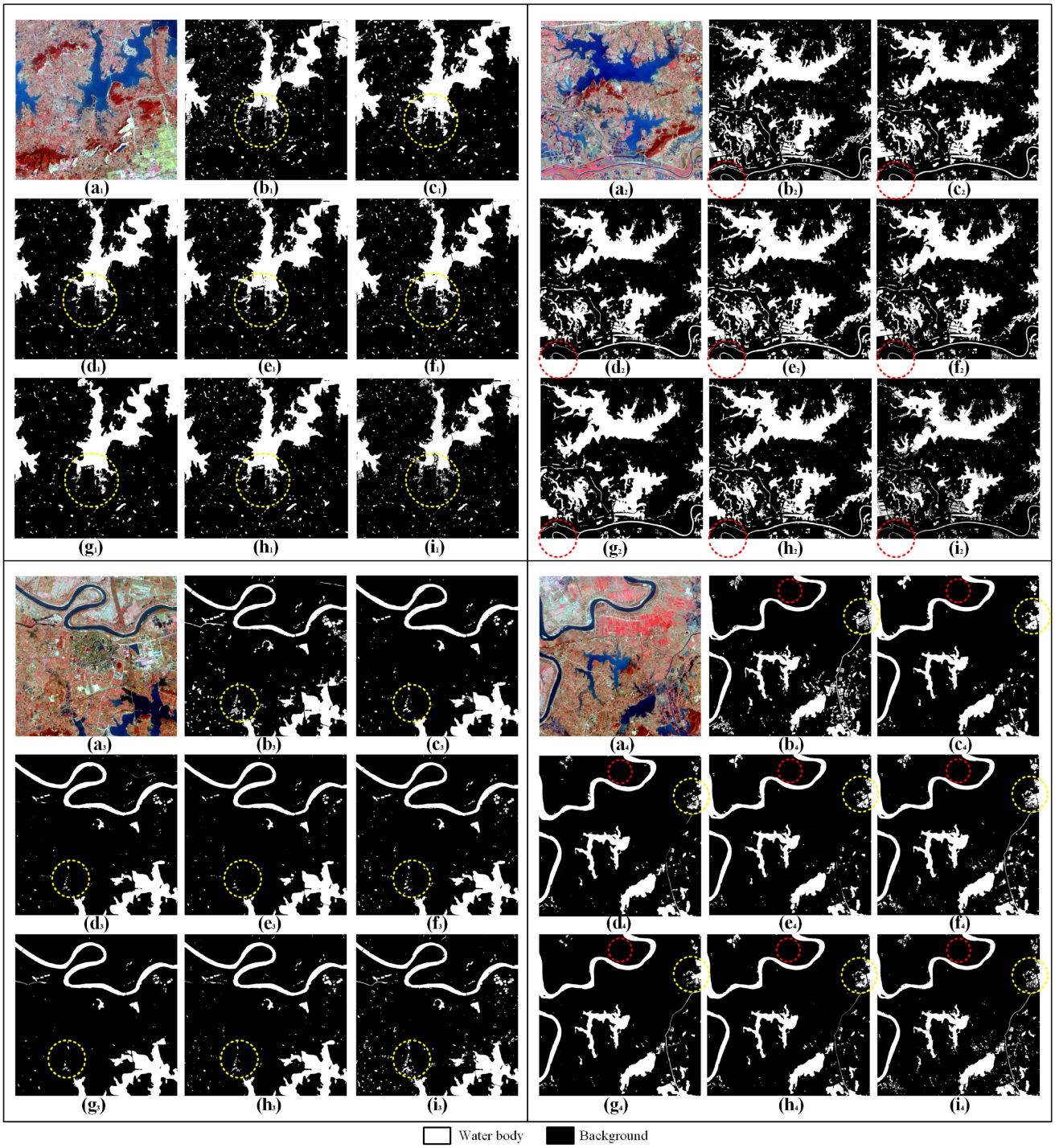

Fig. 11. Water body extraction results using the multispectral data (RGB + NIR) of the GID with (a_1–a_4) Input images, (b_1–b_4) Ground truths, (c_1–c_4) SegNet, (d_1–d_4) LinkNet, (e_1–e_4) DeeplabV3+, (f_1–f_4) Attention U-Net, (g_1–g_4) MECNet, (h_1–h_4) MSResNet, and (i_1–i_4) Proposed SADA-Net.

However, in this article, as the GID contains complex scenes, the Attention U-Net cannot help in separating certain small, intensive water bodies (artificial ponds). As shown in Fig. 10(f_2) and (f_4), marked with green and blue circles, a small number of details are not adequately portrayed, and certain small water bodies are missing. In general, Attention U-Net has quite good performance for multiscale, water body mapping in terms of the high F1-Score (85.03%).

MECNet and MSResNet are the most advanced models for water body extraction, but their performance is not optimal when applied to the GID RGB bands. MECNet is composed of multifeature extraction and combination modules, which can better predict the fine contour of water bodies. MSResNet is a state-of-the-art water bodies extraction method that can distinguish water bodies of different scale sizes and preserve detailed boundaries of water bodies. It can be seen from the experiment,

although good experimental results have been achieved in the large-scale water bodies mapping of GID RGB bands, the F1-Score of MECNet and MSResNet reached a high level, 84.70% and 86.65%, respectively. However, for densely distributed small water bodies, the recognition effect is poor, and the boundaries are not well separated, as shown in Fig. 10(g_2) and (h_2), marked with green circles.

The proposed SADA-Net is a blend of multistage, low-level features, wherein multiscale context information, global context information, and water body shape characteristics are considered, enabling it to produce an excellent performance for water body mapping, with a precision of 92.03%, recall of 85.36%, F1-Score of 88.57%, and an increased overall accuracy compared with Attention U-Net, reaching 96.18%. In brief, the SADA-Net proposed herein more accurately separates the transition zone and water bodies. In addition, from the green and blue circles marked in Fig. 10, the SADA-Net water body mapping results are more refined with fewer misclassifications and missed classifications, which has higher precision and recall, reaching 92.03% and 85.36%. Additionally, based on the SSIM value of 85.79%, SADA-Net dominates in shape.

Moreover, as the study area in the GID is highly complex, some ground objects, such as paddy fields, are misclassified into water bodies. Multispectrum remote sensing images store a large amount of information, which is very useful for identifying different features, especially using the NIR band. As an important influencing factor in water body mapping, multispectrum images have a large role. Therefore, for the proposed network that still has some misclassification problems on the GID RGB bands, experiments are carried out using the multispectral data of the GID. As shown in Table III, all experimental quantitative results are better than the RGB band results. The SADA-Net F1-Score is 7.57% higher than the result of the experiment using only the RGB, three-band dataset. Thus, multispectral data are helpful for solving the problem of misclassification and improving the accuracy and integrity of water body mapping. The yellow circles in Fig. 11 shows that the detailed description of a water body is more prominent. The red circles show that the extraction result of the multispectral band has less misclassification than that of the RGB band.

V. Discussion

This article proposed a SADA-Net that considers multiscale, water body context information, global dependence, and water body shape information. This SADA-Net ensured the complete extraction results of small and dense water bodies. There were fewer missing extractions of water bodies in a large-scale image. In both the vertical comparison experiments and horizontal comparison experiments, water body extraction performance was demonstrated well.

The selection of a loss function in this article mainly considers the imbalance between water bodies and the background, and highlights the refinement of the water body shape. After data enhancement, the water body and background proportion reached approximately 1:10, and the data were more diverse. As a large number of small water bodies, such as small artificial ponds and tiny rivers, exist in the GID study area, the Dice Loss fluctuates when encountering these small water bodies. The loss during training can be made more stable through the combination of Dice Loss, BCE Loss, and BD Loss, and it is beneficial for convergence. Fig. 12 shows the values of the loss function applied to the training and validation sets.

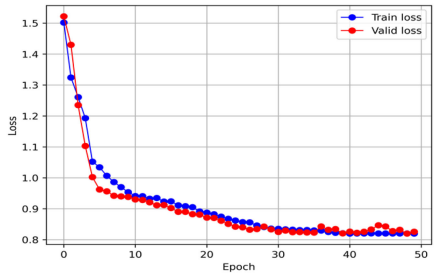

Fig. 12. Loss function values for the training and validation sets.

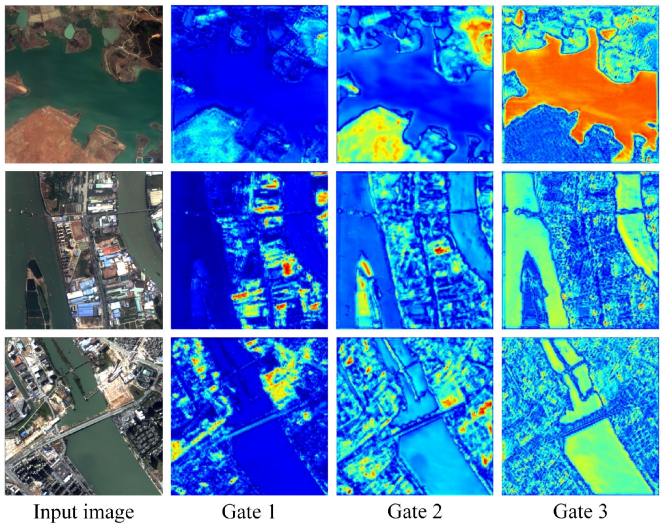

Fig. 13. Visualization of the intermediate layer from GCLs.

In this article, the SADA-Net pays attention to optimizing shape features, which is conducive to highlighting the gaps between nearby water bodies to obtain fine-mapping performance of water bodies. The GCL structure introduced in the SFO module was mainly employed to force the SFO module to process shape-related information only. The entire SFO module includes three gate structures. The first gate structure emphasizes low-level edge information, while the second and third gates emphasize object-level boundaries, effectively highlighting a water body shape feature. Fig. 13 shows the output results of the intermediate layer after the operation of the three gates, where Gate1 is the first gate with an output dimension of 32, Gate2 is the second gate with a dimension of 16, and Gate3 is the last gate with a dimension of 8. For all gate outputs, the maximum sequence values of the channel dimension are shown in Fig. 13. Through the operation of the three gates, the high-level semantic information was gradually enhanced. The output of Gate1 mainly displayed coarse texture and shape

information of the entire object. Gate2 obtained the low-level, water body boundary and shape features, and Gate3 obtained higher-level semantic information of the water body. Through the operation of the GCLs, a perfect shape map can be obtained, and a description of the shape of a water body was enhanced.

In conclusion, the SADA-Net has a strong generalization performance regarding the complex study areas and data types (RGB bands and multispectral images) in the GID dataset. Furthermore, the high F1-Score and SSIM indicate that the proposed network is excellent for water body mapping. In brief, the method proposed herein has a particular reference value when applied to high-resolution remote sensing image mapping.

VI. Conclusion

This article proposed a new intelligent convolutional neural network, SADA-Net, for water body mapping in high-resolution remote sensing images with complex scenes. The proposed network structure integrates ASPP, DA, and SFO modules and comprehensively considers multiscale context information, global dependence, shape information of water bodies. On the grounds of the comparison experiments, the performance of the SADA-Net (multispectral images) method is better, reaching a precision of 92.57% and recall of 94.73%. Moreover, the F1-Score, OA, and SSIM were improved by 7.51%, 2.34%, and 8.64%, respectively, compared with the Attention U-Net network, which is suitable for small water bodies extraction in complex scenes. Moreover, compared with MECNet and MSResNet, the most advanced water body mapping networks, the F1-Score, OA and SSIM have been greatly improved. As a result, these results suggest that the SADA-Net can extract large-scale water bodies and has good extraction performance for small and dense water bodies, such as small ponds, lakes, and narrow rivers. Furthermore, this method is robust in terms of data types and complexity of the study area. Future work will focus more on optimizing the network structure. Multisource data and multifeature fusion methods could be employed to improve the precision and integrity of water body mapping and to achieve larger-range fine-mapping.

References

[1] H. Hafizi and K. Kalkan, "Evaluation of object-based water body extraction approaches using landsat-8 imagery," *J. Aeronaut. Space Technol.*, vol. 13, no. 1, pp. 81–89, 2020.

[2] S. Huang, J. Li, and M. Xu, "Water surface variations monitoring and flood hazard analysis in Dongting Lake area using long-term Terra/MODIS data time series," *Natural Hazards*, vol. 62, no. 1, pp. 93–100, 2012.

[3] B. Brisco, A. Schmitt, K. Murnaghan, S. Kaya, and A. Roth, "SAR polarimetric change detection for flooded vegetation," *Int. J. Digit. Earth*, vol. 6, no. 2, pp. 103–114, 2013.

[4] Y. H. Dandawate and S. Kinlekar, "Rivers and coastlines detection in multispectral satellite images using level set method and modified Chan vese algorithm," in *Proc. 2nd Int. Conf. Adv. Comput. Netw. Secur.*, 2013, pp. 41–46.

[5] M. R. Rahman and P. K. Thakur, "Detecting, mapping and analysing of flood water propagation using synthetic aperture radar (SAR) satellite data and GIS: A case study from the Kendrapara district of Orissa state of India," *Egyptian J. Remote Sens. Space Sci.*, vol. 21, pp. S37–S41, 2018.

[6] W. Li *et al.*, "A comparison of land surface water mapping using the normalized difference water index from TM, ETM+ and ALI," *Remote Sens.*, vol. 5, no. 11, pp. 5530–5549, 2013.

[7] S. Wang *et al.*, "A simple enhanced water index (EWI) for percent surface water estimation using landsat data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 1, pp. 90–97, Jan. 2015.

[8] H. Su, Y. Peng, C. Xu, A. Feng, and T. Liu, "Using improved Deeplabv3+ network integrated with normalized difference water index to extract water bodies in Sentinel-2A urban remote sensing images," *J. Appl. Remote Sens.*, vol. 15, no. 1, 2021, Art. no. 018504.

[9] E. Özelkan, "Water body detection analysis using NDWI indices derived from landsat-8 OLI," *Polish J. Environ. Stud.*, vol. 29, no. 2, pp. 1759–1769, 2020.

[10] K. Rokni, A. Ahmad, A. Selamat, and S. Hazini, "Water feature extraction and change detection using multitemporal landsat imagery," *Remote Sens.*, vol. 6, no. 5, pp. 4173–4189, 2014.

[11] S. Yu, W. Yiquan, and D. Yimian, "Automatic river target detection from remote sensing images based on image decomposition and distance regularized CV model," *Comput. Elect. Eng.*, vol. 54, pp. 285–295, 2016.

[12] S. D. Jawak, K. Kulkarni, and A. J. Luis, "A review on extraction of lakes from remotely sensed optical satellite data with a special focus on cryospheric lakes," *Adv. Remote Sens.*, vol. 4, no. 3, 2015, Art. no. 196.

[13] F. Yang, J. Guo, H. Tan, and J. Wang, "Automated extraction of urban water bodies from ZY-3 multi-spectral imagery," *Water*, vol. 9, no. 2, 2017, Art. no. 144.

[14] B. Qi, Y. Zhuang, H. Chen, S. Dong, and L. Li, "Fusion feature multi-scale pooling for water body extraction from optical panchromatic images," *Remote Sens.*, vol. 11, no. 3, 2019, Art. no. 245.

[15] H. Guo, G. He, W. Jiang, R. Yin, L. Yan, and W. Leng, "A multi-scale water extraction convolutional neural network (MWEN) method for Gaofen-1 remote sensing images," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 4, 2020, Art. no. 189.

[16] J. Kang, H. Guan, D. Peng, and Z. Chen, "Multi-scale context extractor network for water-body extraction from high-resolution optical remotely sensed images," *Int. J. Appl. Earth Observ. Geoinformat.*, vol. 103, 2021, Art. no. 102499.

[17] P. Qin, Y. Cai, and X. Wang, "Small waterbody extraction with improved U-Net using Zhuhai-1 hyperspectral remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Jan. 2021, Art. no. 5502705, doi: 10.1109/LGRS.2020.3047918.

[18] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[19] L. Li, Z. Yan, Q. Shen, G. Cheng, L. Gao, and B. Zhang, "Water body extraction from very high spatial resolution remote sensing data based on fully convolutional networks," *Remote Sens.*, vol. 11, no. 10, 2019, Art. no. 1162.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.

[21] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Jan. 2017.

[22] W. Feng, H. Sui, W. Huang, C. Xu, and K. An, "Water body extraction from very high-resolution remote sensing imagery using deep U-Net and a superpixel-based conditional random field model," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 618–622, Dec. 2018.

[23] W. Li *et al.*, "Urban water extraction with UAV high-resolution remote sensing data based on an improved U-Net model," *Remote Sens.*, vol. 13, no. 16, 2021, Art. no. 3165.

[24] Y. Wang, Z. Li, C. Zeng, G. Xia, and H. Shen, "An urban water extraction method combining deep learning and Google Earth Engine," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 769–782, Feb. 2020.

[25] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-SCNN: Gated shape CNN for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 5229–5238.

[26] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.

[27] J. Fu *et al.*, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3146–3154.

[28] G. L. Feyisa, H. Meilby, R. Fensholt, and S. R. Proud, "Automated water extraction index: A new technique for surface water mapping using landsat imagery," *Remote Sens. Environ.*, vol. 140, pp. 23–35, 2014.

[29] P. S. Frazier and K. J. Page, "Water body detection and delineation with landsat TM data," *Photogrammetric Eng. Remote Sens.*, vol. 66, no. 12, pp. 1461–1468, 2000.

[30] W. Lv, Q. Yu, and W. Yu, "Water extraction in SAR images using GLCM and support vector machine," in *Proc. IEEE 10th Int. Conf. Signal Process.*, 2010, pp. 740–743.

[31] K. Irwin, D. Beaulne, A. Braun, and G. Fotopoulos, "Fusion of SAR, optical imagery and airborne LiDAR for surface water detection," *Remote Sens.*, vol. 9, no. 9, 2017, Art. no. 890.

[32] H. Wu, C. Liu, Y. Zhang, W. Sun, and W. Li, "Building a water feature extraction model by integrating aerial image and lidar point clouds," *Int. J. Remote Sens.*, vol. 34, no. 21, pp. 7691–7705, 2013.

[33] C. Huang, Y. Chen, S. Zhang, L. Li, K. Shi, and R. Liu, "Surface water mapping from Suomi NPP-VIIRS imagery at 30 m resolution via blending with landsat data," *Remote Sens.*, vol. 8, no. 8, 2016. Art. no. 631.

[34] X. Yang, Y. Chen, and J. Wang, "Combined use of Sentinel-2 and Landsat 8 to monitor water surface area dynamics using Google Earth Engine," *Remote Sens. Lett.*, vol. 11, no. 7, pp. 687–696, 2020.

[35] S. K. McFeeters, "The use of the normalized difference water index (NDWI) in the delineation of open water features," *Int. J. Remote Sens.*, vol. 17, no. 7, pp. 1425–1432, 1996.

[36] H. Xu, "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery," *Int. J. Remote Sens.*, vol. 27, no. 14, pp. 3025–3033, 2006.

[37] R. C. Sharma, R. Tateishi, K. Hara, and L. V. Nguyen, "Developing superfine water index (SWI) for global water cover mapping using MODIS data," *Remote Sens.*, vol. 7, no. 10, pp. 13807–13841, 2015.

[38] T. Blaschke, S. Lang, and G. Hay, *Object-Based Image Analysis: Spatial Concepts For Knowledge-Driven Remote Sensing Applications*. Berlin, Germany: Springer, 2008.

[39] D. C. Duro, S. E. Franklin, and M. G. Dubé, "A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery," *Remote Sens. Environ.*, vol. 118, pp. 259–272, 2012.

[40] H. T. Do, V. Raghavan, L. X. Truong, and G. Yonezawa, "Multi-scale object-based fuzzy classification for LULC mapping from optical satellite images," *Spatial Inf. Res.*, vol. 27, no. 2, pp. 247–257, 2019.

[41] I. Dronova, P. Gong, and L. Wang, "Object-based analysis and change detection of major wetland cover types and their classification uncertainty during the low water period at Poyang Lake, China," *Remote Sens. Environ.*, vol. 115, no. 12, pp. 3220–3236, 2011.

[42] X. Huang, C. Xie, X. Fang, and L. Zhang, "Combining pixel-and object-based machine learning for identification of water-body types from urban high-resolution remote-sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 5, pp. 2097–2110, Apr. 2015.

[43] G. Kaplan and U. Avdan, "Object-based water body extraction model using sentinel-2 satellite imagery," *Eur. J. Remote Sens.*, vol. 50, no. 1, pp. 137–143, 2017.

[44] L. Zhao, H. Yu, and L. Zhang, "Water body extraction in urban region from high resolution satellite imagery with near-infrared spectral analysis," in *Proc. Int. Symp. Photoelectronic Detection Imag., Adv. Infrared Imag. Appl.*, 2009, vol. 7383, Art. no. 73833I.

[45] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 988–999, Sep. 1999.

[46] X. Sun, L. Li, B. Zhang, D. Chen, and L. Gao, "Soft urban water cover extraction using mixed training samples and support vector machines," *Int. J. Remote Sens.*, vol. 36, no. 13, pp. 3331–3344, 2015.

[47] G. Sarp and M. Ozcelik, "Water body extraction and change detection using time series: A case study of Lake Burdur, Turkey," *J. Taibah Univ. Sci.*, vol. 11, no. 3, pp. 381–391, 2017.

[48] X. Li, X. Lyu, Y. Tong, S. Li, and D. Liu, "An object-based river extraction method via optimized transductive support vector machine for multi-spectral remote-sensing images," *IEEE Access*, vol. 7, pp. 46165–46175, Mar. 2019.

[49] C. Sukawattanavijit, J. Chen, and H. Zhang, "GA-SVM algorithm for improving land-cover classification using SAR and optical remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 284–288, Jan. 2017.

[50] X. D. Zhou, C. C. Yang, and N. N. Meng, "Method of remote sensing image fine classification based on geometric features and SVM," *Key Eng. Mater.*, vol. 500, pp. 562–568, 2012.

[51] D. K. McIver and M. A. Friedl, "Using prior probabilities in decision-tree classification of remotely sensed data," *Remote Sens. Environ.*, vol. 81, no. 2/3, pp. 253–261, 2002.

[52] T. D. Acharya, D. H. Lee, I. T. Yang, and J. K. Lee, "Identification of water bodies in a Landsat 8 OLI image using a J48 decision tree," *Sensors*, vol. 16, no. 7, 2016, Art. no. 1075.

[53] C. A. Rishikeshan and H. Ramesh, "An ANN supported mathematical morphology based algorithm for lakes extraction from satellite images," *ISH J. Hydraulic Eng.*, vol. 24, no. 2, pp. 222–229, 2018.

[54] B. C. Ko, H. H. Kim, and J. Y. Nam, "Classification of potential water bodies using landsat 8 OLI and a combination of two boosted random forest classifiers," *Sensors*, vol. 15, no. 6, pp. 13763–13777, 2015.

[55] Y. Chen, R. Fan, X. Yang, J. Wang, and A. Latif, "Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning," *Water*, vol. 10, no. 5, 2018, Art. no. 585.

[56] R. Wang, Y. Meng, W. Zhang, Z. Li, F. Hu, and L. Meng, "Remote sensing semantic segregation for water information extraction: Optimization of samples via training error performance," *IEEE Access*, vol. 7, pp. 13383–13395, Jan. 2019.

[57] L. Weng, Y. Xu, M. Xia, Y. Zhang, J. Liu, and Y. Xu, "Water areas segmentation from remote sensing images using a separable residual Segnet network," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 4, 2020, Art. no. 256.

[58] Y. Sun, X. Zhang, Q. Xin, and J. Huang, "Developing a multi-filter convolutional neural network for semantic segmentation using high-resolution aerial imagery and LiDAR data," *ISPRS J. Photogramm. Remote Sens.*, vol. 143, pp. 3–14, 2018.

[59] H. Guo, G. He, W. Jiang, R. Yin, L. Yan, and W. Leng, "A multi-scale water extraction convolutional neural network (MWEN) method for gaofen-1 remote sensing images," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 4, 2020, Art. no. 189.

[60] L. Li, Q. Wen, B. Wang, S. Fan, L. Li, and Q. Liu, "Water body extraction from High-resolution remote sensing images based on scaling efficient-nets," *J. Phys., Conf. Ser.*, vol. 1894, no. 1, 2021, Art. no. 012100.

[61] L. Duan and X. Hu, "Multiscale refinement network for water-body segmentation in high-resolution satellite imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 686–690, Jul. 2019.

[62] M. Xia, Y. Cui, Y. Zhang, Y. Xu, J. Liu, and Y. Xu, "DAU-Net: A novel water areas segmentation structure for remote sensing image," *Int. J. Remote Sens.*, vol. 42, no. 7, pp. 2594–2621, 2021.

[63] M. Li, P. Wu, B. Wang, H. Park, H. Yang, and Y. Wu, "A deep learning method of water body extraction from high resolution remote sensing images with multisensors," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3120–3132, Feb. 2021.

[64] S. Dong, L. Pang, Y. Zhuang, W. Liu, Z. Yang, and T. Long, "Optical remote sensing water-land segmentation representation based on proposed SNS-CNN network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 3895–3898.

[65] Z. Miao, K. Fu, H. Sun, X. Sun, and M. Yan, "Automatic water-body segmentation from high-resolution satellite images via deep networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 602–606, Feb. 2018.

[66] X.-Y. Tong *et al.*, "Land-cover classification with high-resolution remote sensing images using transferable deep models," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111322.

[67] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*, vol. 1. Cambridge, MA, USA: MIT Press, 2016.

[68] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. ICLR*, 2015.

[69] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1251–1258.

[70] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," 2016, *arXiv:1603.07285*.

[71] K. Lenc and A. Vedaldi, "Understanding image representations by measuring their equivariance and equivalence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 991–999.

[72] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. ICLR*, 2014.

[73] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. EMNLP*, 2014, pp. 1724–1734.

[74] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," in *Proc. NIPS*, 2014.

[75] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[76] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis.*, 2016, pp. 565–571.

[77] A. Bokhovkin and E. Burnaev, "Boundary loss for remote sensing imagery semantic segmentation," in *Proc. Int. Symp. Neural Netw.*, 2019, pp. 388–401.

[78] A. D. Desai, G. E. Gold, B. A. Hargreaves, and A. S. Chaudhari, "Technical considerations for semantic segmentation in MRI using convolutional neural networks," 2019, *arXiv:1902.01977*.
[79] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[80] H. Liu *et al.*, "DE-Net: Deep encoding network for building extraction from high-resolution remote sensing imagery," *Remote Sens.*, vol. 11, no. 20, 2019, Art. no. 2380.
[81] Y. Lin, D. Xu, N. Wang, Z. Shi, and Q. Chen, "Road extraction from very-high-resolution remote sensing images via a nested SE-deeplab model," *Remote Sens.*, vol. 12, no. 18, 2020, Art. no. 2985.
[82] J. Su *et al.*, "DV-Net: Accurate liver vessel segmentation via dense connection model with D-BCE loss function," *Knowl.-Based Syst.*, vol. 232, 2021, Art. no. 107471.
[83] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE Vis. Commun. Image Process.*, 2017, pp. 1–4.
[84] O. Oktay *et al.*, "Attention U-net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.
[85] Z. Zhang, M. Lu, S. Ji, H. Yu, and C. Nie, "Rich CNN features for water-body segmentation from very high resolution aerial and satellite imagery," *Remote Sens.*, vol. 13, no. 10, 2021, Art. no. 1912.
[86] B. Zhang and Y. Li, "MSResNet: Multiscale residual network via self-supervised learning for water-body detection in remote sensing imagery," *Remote Sens.*, vol. 13, no. 16, 2021, Art. no. 3122.

Bin Wang received the B.S. degree in geographic information science from the Guilin University of Technology, Guilin, China, in 2019. He is currently working toward the M.S. degree in cartology and geographic information engineering with the School of Geography and Information Engineering, China University of Geosciences, Wuhan, China.

His major research interests include intelligent interpretation of remote sensing images and geological text information extraction.

Zhanlong Chen received the B.S. and M.S. degrees in cartology and geographic information engineering from the Kunming University of Science and Technology, Kunming, China, in 2003 and 2006, respectively, and the Ph.D. degree in cartology and geographic information engineering from China University of Geosciences, Wuhan, China, in 2009.

He is currently a Full Professor with the School of Geography and Information Engineering, China University of Geosciences, Wuhan, China. His research interests include the intelligent mining of spatio-temporal data and high-performance spatial intelligence computation.

Liang Wu received the B.S. degree in computer science and application and the Ph.D. degree in cartology and geographic information engineering from China University of Geosciences, Wuhan, China, in 1998 and 2006, respectively.

He is currently a Full Professor with the School of Geography and Information Engineering, China University of Geosciences, Wuhan, China. His research interests include the geo-big data mining and high-performance spatial intelligence computation.

Xiaohong Yang received the B.S. degree in computer science and technology from the China University of Geosciences, Wuhan, China, in 2006, and the M.S. and Ph.D. degrees in spatial information science and technology from Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively.

She is currently a Lecturer with the School of Computer Science, China University of Geosciences, Wuhan, China. Her research interests include remote sensing application and water resource management.

Yuan Zhou received the B.S. degree in surveying and mapping engineering from Anhui Agricultural University, Anhui, China, in 2019. She is currently working toward the M.S. degree in surveying and mapping engineering with National Engineering Research Center of Geographic Information System, Wuhan, China.

Her major research interests include intelligent interpretation of remote sensing images and image processing.