A Deep Learning Method of Water Body Extraction From High Resolution Remote Sensing Images With Multisensors

Mengya Li [ID], Penghai Wu [ID], *Member, IEEE*, Biao Wang [ID], Honglyun Park, Hui Yang [ID], and Yanlan Wu

Abstract—Water body extraction from remote sensing images is an important task. Deep learning has become a more popular method for extracting water bodies from remote sensing images. However, these methods are usually aimed at a specific sensor and are not applicable. Thus, we proposed a new network, called the dense-local-feature-compression (DLFC) network aiming at extracting water body from different remote sensing images automatic. In this network, each layer of the network can receive the feature maps of all layers before it by the densely connected module of DenseNet. The concatenate operation on the feature dimension is used when connecting across layers. It can realize the different levels of features reuse. The local-feature-compression module is introduced before concatenate operation. It can obtain the more abstract features further by the convolution operation. Through the DLFC, we can fuse the spatial and spectral information for the remote sensing images that can extract water body from different remote sensing images. Besides, we construct a new water body dataset based on GaoFen-2 (GF-2) remote sensing images. The proposed DLFC achieved excellent performance with GF-2, GaoFen-6, Sentinel-2, and ZY-3 remote sensing images. Compared with the traditional water body extraction method and contemporary networks, the DLFC exhibits noticeable improvement. The results indicate that the DLFC can realize water body extraction from multisource remote sensing images automatically and rapidly.

Index Terms—Deep learning, high resolution, multisource remote sensing images, water body.

I. Introduction

WATER is the source of life and an important component of the land ecological system, which plays an important role in the global water cycle [1]. Water significantly influences public health, the living environment, and economic development [2]. Therefore, investigating surface water and delineating its time-spatial distribution is of great importance [3]. Due to the features of a large scale, rapid update period, and dynamic monitoring, remote sensing has become a general method for surface water monitoring [4].

Water body extraction from remote sensing images mainly includes single-band density slicing [5], spectral water indexes [6]–[8], object-oriented approaches [9]–[12], and deep learning methods [13], [14]. The spectral water index-based method is a reliable method among all existing water body mapping methods. Since the 1990s, many different water indexes have been proposed [5], [15]. McFeeters proposed the normalized difference water index (NDWI) [16]. However, NDWI cannot perform well due to the effect of shadows in built-up areas. Because of the shortcomings of NDWI, Xu used the shortwave infrared (SWIR) band to replace the NIR band in NDWI, developing the modified normalized difference water index (MNDWI) [17]. The index has been used in experiments in which remote sensing images contain different water types, and most obtained better effects than those of NDWI, especially in urban areas. Du *et al*. mapped the water bodies of the Venice coastland, Italy, from Sentinel-2 imagery with MNDWI [8]. These methods improved the accuracy of water extraction to varying degrees. However, these traditional spectral methods always come down to the choice of the threshold value. The threshold values are different on different occasions. For each different image, we must choose a suitable threshold value that depends on the experience of the studies. Therefore, these methods are subjective and have a low level of automation [18]. Images with higher spatial resolution, such as SPOT6/7, IKONOS, Worldview and Quick-bird, GF-2, are used to extract water bodies. The higher the image resolution is, the more detailed the information the images have. However, it is impossible to use the MNDWI method with these fine-spatial-resolution images. However, MNDWI has good performance in water body extraction. These high-spatial-resolution images only have four bands (blue, green, red, and near infrared), lacking the SWIR band to compute MNDWI [3]. In addition, in order to realize the high accuracy in water body extraction

Mengya Li, Penghai Wu, and Biao Wang are with the School of Resources and Environmental Engineering, Anhui University, Hefei 230601, China (e-mail: x18301060@stu.ahu.edu.cn; wuph@ahu.edu.cn; wangbiao-rs@ahu.edu.cn).

Honglyun Park is with the Drone & Transportation Engineering, Youngsan University, Yangsan 288, South Korea (e-mail: hlpark@ysu.ac.kr).

Hui Yang is with the Information Materials and Intelligent Sensing Laboratory of Anhui Province, Institutes of Physical Science and Information Technology, Anhui University, Anhui 230601, China (e-mail: yanghui@ahu.edu).

Yanlan Wu is with the Information Materials and Intelligent Sensing Laboratory of Anhui Province Hefei, Anhui 230601, China, with the School of Resources and Environmental Engineering, Anhui University, Hefei 230601, China, and also with the Anhui Engineering Research Center for Geographical Information Intelligent Technology, Hefei 230601, China (e-mail: wuyanlan@ahu.edu.cn).

of the GF-2 images, Wu designed a new water index called the two-step urban water index (TSUWI) combining the urban water index and the urban shadow index to map the water surface [19]. The TSUWI can only extract water bodies from GF-2 images, which is not suitable for all types of remote sensing images. The spectral water index-based methods mentioned above all have some common problems such as the adaptability and automation of the index. What is more, the accuracy of these index methods is closely related to the operator's experience. Therefore, the index method has certain limitations.

With the extensive application of deep learning in remote sensing information extraction [20]–[30]. Li explored a novel network structure named DeepUNet for pixel-level sea-land segments on images from Google Earth [31]. The experiments showed that this new network achieves high-precision recall and F1-measure for both sea and land regions. To evaluate the 11 effective and efficient machine learning methods for the extraction of shorelines, Manaf tried to extract the water from the northwest coast of Peninsular Malaysia [32]. Among all the methods, Multilayer Perceptron Artificial Neural Network (MLP ANN) achieved the best performance. Yu extracted water bodies from Landsat imagery using convolutional neural networks (CNNs) [19]. In this study, CNN hierarchically extracted useful high-level features from input data and then used a logistic regression classifier to classify them. CNN has better performance than support vector machine (SVM) and artificial neural network (ANN). Yang introduced self-adaptive pooling into CNNs to extract urban water bodies from two categories of Chinese high-spatial-resolution remote sensing images (ZY-3 and GF-2 multispectral images) [3]. However, before the extraction, they needed to preprocess the data, such as color space transformation and adaptive simple linear iterative clustering. Li extracted urban water by combining deep learning and the Google Earth Engine [33]. They combined the Google Earth Engine (GEE) and Multiscale Convolutional Neural Network (MSCNN), which do not need to make many subjective decisions and maintain the advantage of high accuracy in urban water extraction using MSCNN. Deep learning has been increasingly popular in water body extraction from remote sensing images. However, as we can see, the aforementioned studies both used one deep learning network to extract the water bodies from only one remote sensing image, especially Landsat images. These methods were often difficult to be used to extract water body from other remote sensing images even the different images from the same sensor. Besides, for the methods of deep learning, each sensor needs to make its own training dataset, and the single-sensor sample training model cannot be effectively applied to other sensors.

To extract the water bodies from the high-resolution remote sensing images, we proposed a novel deep learning encoder–decoder framework, which is called the dense-local-feature-compression (DLFC) network. There is much complex and diverse information in the remote sensing images. The deep learning method needs to get a powerful ability of feature extraction. Densely connected module [34] has a strong ability to extract features in the encode part. In this part, DenseBlock can extract the different level features including the low-level features and the highly abstract features. Meanwhile combining the feature reuse mechanism of densely connected module, some low-level features will also be added to the decoder part. Through these two effective mechanisms, densely connected module takes advantage of the different level spectral features. Thus, we can make full use of the spectral information. To solve the data normalization problem and accelerate network convergence, we introduce group normalization [35] to replace traditional batch normalization [36]. Group normalization can be computed independently regardless of the batch size. Compared with batch normalization, group normalization has a stable accuracy in a wide range of batch sizes. In addition, to improve the adaptability of the method, we designed a local feature compression (LFC) module in the upsampling process. The LFC module can extract highly abstract feature that can further improve the adaptability of the model. Through the proposed network, we accomplished the target that extracts the water body from different images of one sensor and different sensors automatically. The extraction results are more objective than traditional index methods. It does not rely on human subjective experience. Besides compared with the remote sensing extraction algorithms, it suits more remote sensing images. The remote sensing extraction algorithms proposed before usually aim at one type of remote sensing data. We all know that most deep learning-based methods surely needs a large number of training samples. But there was not a water dataset that consists of remote sensing images. Thus, we constructed the dataset, which is made of GF-2 remote sensing images and the corresponding label. There are more than 16000 images in this dataset.

The rest of this article is organized as follows. A detailed description of the proposed method is given in Section II. Related data, data processing, and related experiments are presented in Section III. The results of the experiments are listed in Section IV. Finally, the conclusion is listed in Section V, and the discussion is in Section VI.

II. METHODOLOGY

A. DLFC Network

1) Architecture of the Proposed Network: Similar to other encoder–decoder architectures, we continued to use the classic encoder–decoder architecture in neural networks. In the encoder part, there is a convolution layer with a stride of two, convolution kernels of seven, a group normalization layer, and five DenseBlocks. Using this architecture, we can obtain a feature map that is highly abstracted. In the decoder, there are five DenseBlocks to extract features and five transpose layers to recover features. After the transpose layer, we introduced the DLFC module. The local-feature-compression module contains two convolution operations that can strengthen the high-level features. It combines the spatial information with the spectral information of the remote sensing images that ultimately realizes the information extraction. Then, we concatenated the high-level features after the upsampling and the low-level features from the encoder part by skip layers to improve the use of features and compensate for the loss of features. Through this operator, we ultimately improved the adaptability of the model. Finally, we used the convolution layer to recover the feature map as a binary
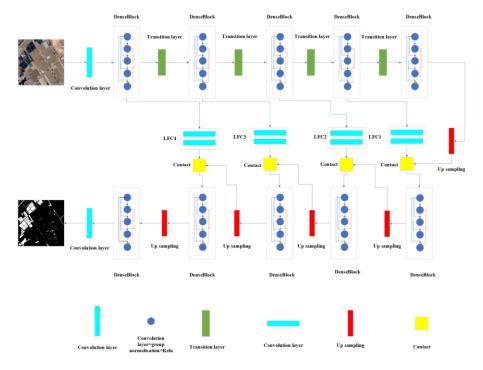
Fig. 1. Architecture of the DLFC network.

segmentation map that has the same size as the remote-sensing images. The architecture of the networks is shown in Fig. 1.

Batch normalization error increases rapidly when the batch size is smaller. Some computer vision tasks, such as detection and segmentation, usually require small batches because of the limitations created by memory consumption. Group normalization can be computed independently regardless of the batch size. Compared with batch normalization, group normalization has a stable accuracy in a wide range of batch sizes. We introduce group normalization to replace the traditional batch normalization in DenseBlocks, which changes the DenseBlock convolution template that consists of a batch normalization layer, a rectified linear unit (ReLU), and a 3×3 convolution layer into a group normalization layer, an ReLU, and a 3×3 convolution layer. Through this, we can make the network more reasonable.

2) Group Normalization: Therefore, GN was proposed to improve the shortage of BN. GN divides the channels into groups and calculates the mean and variance within each group for normalization. The dimension channel is reshaped into two dimensions (G and $C//G$), where C is the channel axis and G is the number of the groups. Then the mean and variance are calculated based on dimensions $C//G$, height (h) and width (w). Finally, the dimension is reshaped back to the original channel and returns the results of the linear transformation. The accuracy cannot be influenced regardless of the batch size. Because the calculation of GN is independent of the batch size, its accuracy is also stable in various batch sizes.

3) LFC Module: To better integrate spatial information and spectral information, we introduce a LFC module into the up-sampling process. This LFC contains two convolution layers. The first convolution layer has a 5×5 convolution operation with a 1 stride. The second convolution layer has a 7×7 convolution
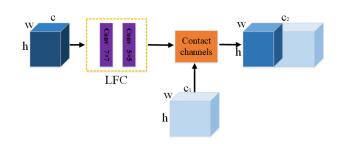

Fig. 2. Architecture of the LFC module, where h represents the height of the feature map, w represents the width of the feature map, and c_1, c_2, and c represent the channel of the feature map.

operation. After these two convolution layers, we can obtain higher level features that can help extract the important spatial and spectral features. It can improve the adaptability of the network. The architecture of the LFC module is shown in Fig. 2.

4) Implementation Details of the Network: The other important parameters in our proposed network were set properly to obtain fine extraction results. For example, every DenseBlock contained 3 layers and the growth rate was set as 32. In each transition layer, a 1×1 convolution operation and a dropout layer with a 0.2 rate were carried out, followed by a 2×2 average pooling operation. For the transpose layer, we used a 3×3 convolution operation with a 2 stride, and the activation function was often an ReLu. In the LFC module, there were 2 convolution operations: one was 5×5, and the other was 7×7 with a 1 stride.

In addition, the network proposed in this article used binary cross entropy as the loss function, the batch size was set to two, the number of rounds was 100, the number of iterations per round

was 5000. And the initial learning rate was set as $1e$-4. In order to better train the model, the learning rate will automatically adjust with the increase of the number of training rounds, that is, the learning rate will be reduced by 10 times every 20 rounds. The optimization used in this article was adam [37], which can also accelerate the convergence of the network.

B. Comparison Method

To validate the accuracy of the method proposed by us. We performed water extraction on selected images using other common methods, which are four deep learning methods, DeepLabV3+ [38], U-Net [39], DeepUNet[31], SegNet [40], and two remote sensing extraction algorithms, NDWI and SVM. DeepLab V3+ is the latest work of the DeepLab semantic segmentation network. It reached 89.0% of mIoU on Pascal VOC and 82.1% on Cityscape. U-net is a semantic segmentation network based on FCN, suitable for the segmentation of medical images. Basically, for all segmentation problems, it can try U-net to see the effect. The network won a big advantage in the 2015 ISBI Cell Tracking Challenge. In addition, U-net and DLFC are both the U-shaped structures. They both consist of two parts. U-net is consisted of a contracting path and an expansive path. Contraction path is consist of simple convolution layer and pooling layer, expansion path is made of a series of deconvolution layer. As for DLFC, it also has two parts, the encoder part and the decoder part. We introduced a series of dense block into these two parts. The dense block consists of several convolution layers. In every dense block, they both perform dense connection that means all layers in the network are connected in pairs. The concatenate operation on the feature dimension is used when connecting across layers, which is not the add operation. For each layer, the feature maps of all the preceding layers will be used as the inputs, and its own feature maps are also used as the inputs of the subsequent layers. DeepUNet is an improved network based on U-Net. SegNet is a deep network proposed by Cambridge to solve image semantic segmentation of autonomous driving or intelligent robots. NDWI and SVM are the most common methods for water body extraction.

C. Evaluation Metrics

For semantic segmentation, there are some common indexes to evaluate the accuracy of segmentation, including overall accuracy (OA), F1-score, IoU, etc. In this article, water extraction is regarded as a dichotomy problem, and the results of the prediction are either water or nonwater. For the dichotomy problem, the examples can be divided into four cases: true positive (TP), false positive (FP), true negative (TN), and false negative (FN) depending on the combination of the real category and the learning prediction category. These values can be calculated using the pixel-based confusion matrix per patch. If TP, FP, TN, and FN represent the corresponding sample numbers, it is obvious that TP + FP + TN + FN is the total number of samples. The F1-score is based on the harmonic average of precision and recall, and IoU is the intersection ratio. In the problem of semantic segmentation, the intersection ratio of the real value and the predicted value is calculated. The evaluation

index formula is as follows:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$F1 = \frac{2TP}{2TP + FN + FP} \tag{2}$$

$$IoU = \frac{TP}{FN + TP + FP}. \tag{3}$$

III. Data and Experiment

In this section, the effectiveness of the proposed scheme for water body extraction in very high-resolution remote sensing images was investigated. All networks were trained and tested with TensorFlow on a GPU.

A. Dataset

The Gaofen-2 (GF-2) satellite is the first civilian optical remote sensing satellite independently developed in China with a spatial resolution better than 1 m. GF-2 imagery is a good candidate for applying previous methods to extract urban water bodies. In this article, the dataset we used contains GF-2 remote sensing images containing the Changjiang River in AnHui Province and images containing the sea from southern China. These images, which have blue, green, red, and near infrared channels, include the sea, rivers, lakes, ponds, and other different types of water. Lakes, rivers, streams, and paddy fields are common forms of surface water in remote sensing images. Lakes and reservoirs are clear in the shape of a planar water body. The grass and trees around the lake, and the floating vegetation can result in a complex and mixed water spectrum. There are many types of river water systems, and the water bodies are usually linear, especially slender and small, or converge into an irregular surface. Small water bodies below the image resolution exist in the form of mixed pixels. Due to the extremely strong fluidity of the water body, its spatial distribution, geometric morphology, and other characteristics are affected by various factors such as topography, landform, water level, human modification, etc. The spectral information of the same body of water is often different in remote sensing images from different sensors at different times. In addition, due to the difference in resolution, there are more mixed pixels in low-resolution remote sensing images, especially at the junction of water and other objects. The water body shows various morphological characteristics in different situations, which increases the difficulty of water information extraction from remote sensing images. The number of different waters is generally the same.

The water masks of the selected GF-2 images were manually drawn by referring to the original images. The mapping of the water masks followed the same standard. First, slender water bodies with widths of less than 2 pixels were drawn with the line feature. Second, water bodies with less than 2 pixels were not drawn. Third, the edges of water areas were identified as water if the water features are obvious, otherwise as nonwater. Finally, a manually labeled reference water mask was created by setting the pixel values of water and nonwater to 1 and 0, respectively.

TABLE I
RELATED INFORMATION OF GF-2 MULTISPECTRAL IMAGERY

Satellite Parameters	GF-2 Multispectral Imagery
Product Level	1A
Number of bands	4
Wavelength (nm)	Blue: 450-520 Green: 520-590 Red: 630-690 NIR: 770-890
Spatial resolution (m)	1

In our study, all GF-2 images are of product type level 1A. These images have high quality and contain enough information for radiometric correction and geometric correction. All the images we used are free of clouds. A description of the GF-2 multispectral images is presented in Table I.

The dataset consists of 23 GF-2 images divided into two parts: training images and test images. The 23 images are not the same size. Given the limited memory of the GPU and obtaining more training samples, all the images had been split into small patches of size 640 × 640. Thus, 16722 images for training the network can be obtained. However, for testing, we chose another 9 images, which do not belong to training images and test images with a size of 2240 × 2240 to evaluate the extraction performance.

B. Reference Data

To validate the adaptability of our model in different satellite images, we chose three other types of high-resolution images: Sentinel-2A, ZY-3, and GF-6. These three types of images have different spatial resolutions, where the resolution of Sentinel-2A is 10 m, ZY-3 is 6 m, and GF-6 is 10 m. The band number of Sentinel-2A is 12, but we only chose four bands—blue band, green band, red band, and nir red band among all bands which are band 2, band3, band4, and band8. The band number of ZY-3 and GF6 both are 4, which are blue, green, red, and nir red bands. The detailed information of these three types of images was listed in Table II.

IV. RESULTS AND ANALYSIS

A. Extraction Results

After training the DLFC, we evaluated 9 test images, and the sizes of these 9 test images are all 2240 × 2240. The extraction results obtained by our method were perfect and more complete than those obtained by other networks, which are closer to the ground truth images. To accurately evaluate the performance, the average of the three evaluation indexes, the OA, F1-score, and IoU of all the test images were listed in Table III, which are 98.44%, 95.39%, and 91.25%, respectively. The highest OA among the 9 images is 99.13%, the highest F1 is 97.41%, and the highest IoU is 94.94%. The lowest OA among the 9 images is 97.27%, the lowest F1 is 91.62% and the lowest IoU is 84.54%. At the same time, to intuitively understand the extraction effectiveness of the DLFC, we showed the best and worst results in the test pictures, as shown in Fig. 3. From Fig. 3, we can see that the DLFC almost extracted all the water in the images. The boundary of the water matches the ground truth more perfectly, which cannot identify vegetation belonging to the river band as water. Slender water bodies are more complete and continuous. The completeness of the water is not influenced by waves appearing on the water. For ponds, the completeness of the extraction results is perfect. Every pond in the images is extracted regardless of the pond size. The DLFC can accurately distinguish urban water and building shadows. Low albedo features that are similar to water in the spectrum, such as roads, cannot be recognized as water. The DLFC never gives a wrong classification. In addition, it can exclude ships from the water, although ships in the water are always small. Therefore, the extraction accuracy of the DLFC is excellent. However, there are some shortcomings of the DLFC. The DLFC cannot effectively extract small water bodies of less than two pixels; small water bodies are usually extracted incompletely.

B. Comparison With DLFC Without the LFC Module, DLFC (BN), and Other Water Body Extraction Methods

To validate the effect of the LFC module, we deleted the LFC module in the upsampling. We trained the network under the same conditions as those of the DLFC. Then, we evaluated the same 9 test images used in Section IV-B. The results were shown in line 4, Fig. 4. The OA, F1-score, and IoU of all the test images, which are 98.50%, 95.26%, and 91.06%, respectively, were listed in Table IV. By comparing with the results of Table IV, we can see that the F1-score and IoU of the DLFC are 0.13% and 0.19% higher than those of the network without the LFC module. The OA of the DLFC is slightly smaller than that of the network without the LFC module, which only has a 0.04% difference. The extraction results are shown in Fig. 4. The results are also good. However, compared with those of the DLFC, the results of the DLFC (NLFCM) have many problems. Some ponds are not completely extracted. The small water body that is less than two pixels cannot be extracted. Therefore, the LFC module is useful for improving the reuse and fusion of the different levels of the feature map.

In order to prove the group normalization plays an important role in the network, we change the group normalization into batch normalization [DLFC (BN)]. We trained the network under the same conditions as those of the DLFC. The results were shown in line 5, Fig. 4. The OA, F1-score, and IoU of all the test images, which are 91.97%, 73.63%, and 59.87%. They are all lower than the results of the network with group normalization. The extraction results are shown in Fig. 4. From Fig. 4, we can see that the DLFC (BN) cannot extract the complete water in the images. Besides, it has many mistaken classifications. That is because batch normalization depends on the batch size, its accuracy is also unstable in various batch sizes. The small batch size cannot meet the requirement. Therefore, the group normalization is significant for the network.

To assess the relative performance of the DLFC, we performed water extraction on selected images using two type common methods. One is the remote sensing extraction algorithms, such as NDWI and SVM. Another method is deep learning method, such as DeepLabV3+ [37], U-Net [38] and DeepUNet [31]. We

TABLE II
INFORMATION ON REFERENCE DATA

Sensor	Sentinel-2A	ZY-3	GF-6
Spatial resolution	10m	6m	2m
Image size(pixel)	1024*1024	2240*2240	2240*2240
Wavelength(μm)	Band2:0.45-0.52 Band3:0.54-0.57 Band4:0.65-0.68 Band8:0.78-0.90	Band1:0.45-0.52 Band2:0.52-0.59 Band3:0.63-0.69 Band4:0.77-0.89	Band1:0.45-0.52 Band2:0.52-0.60 Band3:0.63-0.69 Band4:0.76-0.90
The number of validate images	4	4	4

TABLE III
RESULTS OF OA, F1, AND IoU ON THE VALIDATION IMAGES

Test Datasets	OA	F1-score	IoU
All images	98.44%	95.39%	91.25%
Image 1	98.39%	96.39%	93.03%
Image 2	98.29%	93.79%	88.31%
Image 3	98.18%	91.62%	84.54%
Image 4	98.77%	94.16%	88.96%
Image 5	99.13%	94.56%	89.68%
Image 6	98.76%	96.97%	94.12%
Image 7	98.20%	96.56%	93.35%
Image 8	97.27%	97.06%	94.29%
Image 9	98.97%	97.41%	94.94%

TABLE IV
COMPARISONS BETWEEN THE PROPOSED DLFC AND SEVERAL TYPICAL METHODS

Methods	OA	F1-score	IoU
DLFC (Ours)	98.44%	95.39%	91.25%
DLFC (NLFCM)	98.50%	95.26%	91.06%
DLFC (BN)	91.97%	73.63%	59.87%
DeepLabV3+	97.61%	93.19%	87.35%
U-Net	77.09%	59.15%	43.96%
DeepUnet	81.38%	15.29%	8.39%
NDWI	86.57%	71.19%	56.63%
SVM	97.83%	93.61%	88.14%

compared the extraction results of all methods mentioned which is shown in Table IV. Overall, all methods could accurately extract evident and clear urban waters, such as large rivers, lakes, reservoirs, and ponds. However, compared with remote sensing extraction algorithms and other deep learning networks, the DLFC performed better in the presence of the complex urban surface. Especially in urban centers with dense buildings, the DLFC could more accurately distinguish water from other nonwater objects. As seen from the table, the results of DLFC are much more accurate. Water detection on GF-2 imagery of different models is shown in Fig. 4, in which it can be seen that the DLFC generates more accurate water areas than those of other deep learning models, such as U-Net, DeepLabV3+, and DeepUNet. In addition, the DLFC is also better than NDWI, SVM which are used most frequently in water extraction from remote sensing imagery. From Table IV, we can see that the DLFC achieves a mean F1-score water accuracy of 95.14%.

The IoU of the DLFC is 90.79%, which is higher than that of both DeepLabV3+, U-Net, and DeepUnet. The water extracted by U-Net has faults in local regions, which results in U-Net having a lower OA, DeepUNet loss major part water in the test images. While the water body extraction results of DeepLabV3+ are generally satisfactory but are less accurate than those of NDWI, SVM, and DLFC. The mean water body extraction OA, F1-score, and IoU of NDWI are 86.57%, 71.19%, and 56.63%, F1-score and IoU of SVM are 97.83%, 93.61%, and 88.14%, respectively, which are much higher than those of deep learning methods. However, compared with the DLFC, NDWI, and SVM, both have some common shortcomings: roads with low albedo are identified as water, as are the shadows of buildings. The slender water whose length is smaller than 3 pixels cannot be extracted.

The DLFC has a strong ability to abstract and reuse the features. It can extract different level features and reuse them
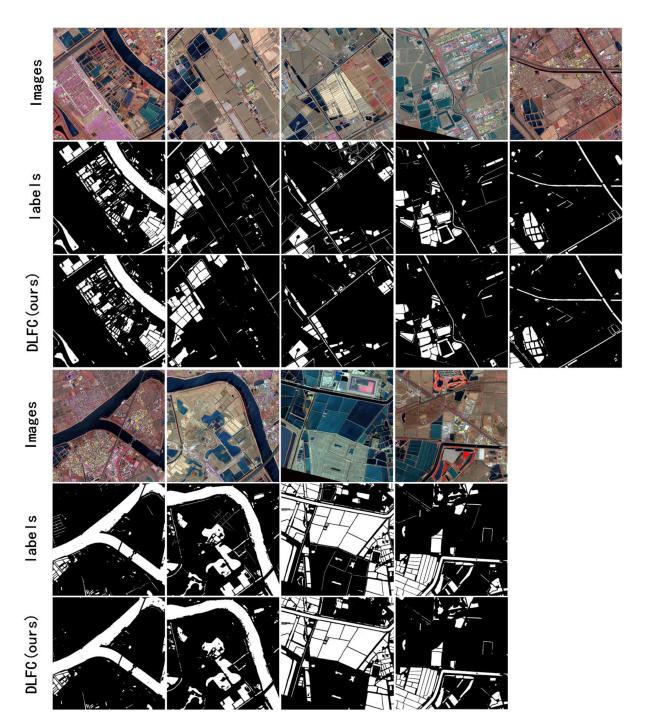
Fig. 3. Prediction results of nine validation images.

step by step. It can not only obtain highly abstract features but also reduce the loss of spatial information. Therefore, it can extract the small and slender water from the images. For other deep learning methods, they can't acquire advanced features and lose much information. As for the remote sensing extraction algorithm, some of them can acquire better results. However, they just extract the water body information by using their own spectral features. This can induce many problems, such as the condition of the same objects with different spectrum or same spectrum with different objects. Besides, the remote

sensing images, extraction algorithms generally depend on the experience of the operator.

Given the above, the DLFC is the best method of water body extraction from high-resolution remote sensing images.

C. Extraction Results of the Different Remote Sensing Images Using the DLFC Method

The pretrained DLFC can also be used to extract water bodies for other types of remote sensing images without any other
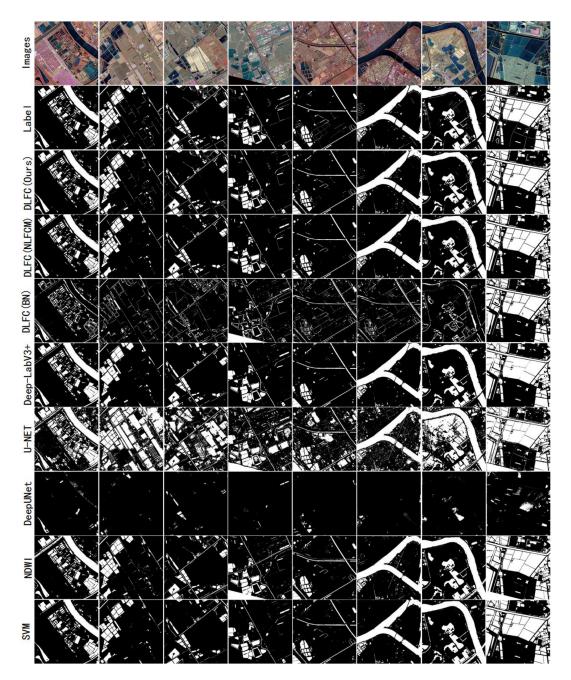
Fig. 4.   Prediction results of the DLFC (Ours), U-Net, DLFC (NLFCM), DLFC (BN), DeepLabV3+, DeepUNet, NDWI, and SVM.

training or parameter adjustment. We choose the other three types of satellite images, i.e., Sentinel-2A MSI (10 m), Ziyuan-3 MUX (6 m), and GF-6 (10 m). Before being put into the DLFC, some preprocessing steps need to be done, such as radiation calibration and FLAASH atmospheric correction. From Fig. 5, it can be seen that the DLFC generates visually satisfactory water body extraction for the multiple types of imagery. The results of water body extraction from Sentinel-2A are similar to the ground truth. The lake in the images is extracted completely. Almost all the small ponds in the image are extracted. Slender features in large areas of water have the correct recognition results. The water body in the ZY-3 images is also extracted perfectly, especially when ships floating on the water is called. The slender

water in the images, which hardly appears to be missing, is extracted well. The slender water continues and is complete. However, the extraction results of GF-6 are not great. Some water is missed. These missed waters are often slender bodies that are less than two pixels. However, the edge of the water body is accurate. Moreover, features such as the shadows of buildings and roads are not recognized as water. All three remote sensing images do not exhibit misclassification. The evaluation metrics results are listed in Table V. The average OA, F1, and IoU of GF-6 are 99.42%, 99.2%, and 98.42%, respectively. The average OA, F1, and IoU of Sentinel-2 are 98.93%, 98.98%, and 97.99%, respectively. The average OA, F1, and IoU of ZY-3 are 99.51%, 98.9%, and 97.83%, respectively. From Table V, we

TABLE V
RESULTS OF AVERAGE OA, F1, AND IOU ON THE OTHER REMOTE SENSING IMAGES

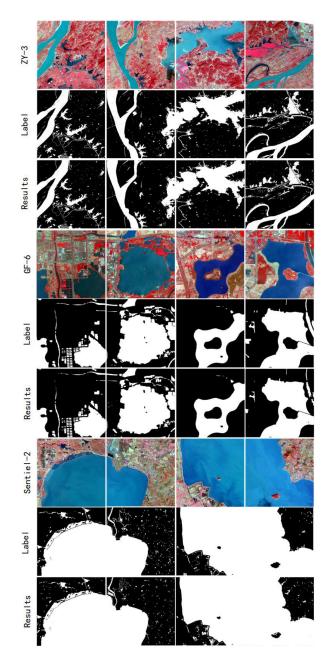| Model	OA	F1-score	IoU
GF-6	99.42%	99.2%	98.42%
Sentinel-2	98.93%	98.98%	97.99%
ZY-3	99.51%	98.9%	97.83%

Fig. 5.   Prediction results of ZY-3, GF-6, and Sentinel-2 using the DLFC (ours).

can see that the OA of the three different remote sensing images all exceeds 98% as well as the F1-score. The IoU of these three remote sensing images is more than 97%. The extraction results across multiple sensors are generally perfect. The accuracy of the extraction results satisfies the extraction requirements. In summary, this is the first method that can extract water bodies from so many types of satellite images with different resolutions that do not need any preprocessing.

D. Extraction Results of the Different Remote-Sensing Images Using the Other Deep Learning Method

We then chose three other deep learning networks, DLFC (NLFCM), DLFC (BN), DeepLabV3+, SegNet, U-Net, and DeepUNet, to validate the adaptability of different remote sensing images. These deep learning networks are often used for information extraction. To generate the objectivity of the study, we chose the remote sensing images used in the last chapter. The results are shown in Fig. 6. From Fig. 6, we can see that U-Net cannot extract water bodies from these remote sensing images. It is difficult for U-Net to extract water from remote sensing images without pretraining. Therefore, U-Net is the network with specificity. It does not have adaptability among different high-resolution remote sensing images. Although U-Net has brilliant performance in medical images. DeepLabV3+ can extract water from Sentinel-2A and ZY-3, especially the Sentinel-2A. However, the accuracy is poor. It also has incorrect recognition and the small area water cannot be extracted completely. However, when extracting water bodies from GF-6, it cannot extract anything from the images. With regard to SegNet, it has a good performance in GF-6 and Sentinel-2A. It almost completely extracts all the water. However, it also appears that it recognizes other things as water, such as roads and buildings. The rate of extraction error is high. As for ZY-3, the results are not satisfactory. The DLFC (NLFCM) also extracts almost all the water. It has good performance in Sentinel-2A. The average OA, F1-score, and IoU are 99.02%, 99.06%, and 98.13%, which are slightly higher than DLFC. However, for the small ponds and the slender water, it cannot extract completely. There is a discontinuous condition in slender river extraction. As for GF6 and ZY-3, it cannot extract the water in detail. The average OA, F1-score, and IoU of GF-6 are 99.06%, 98.59%, and 97.21%. The results of ZY-3 are 99.06%, 98.98%, and 97.21%. These all lower than DLFC. In addition, for ZY-3, there are some missing extraction. For the images of higher resolution, the accuracy of the extraction results is slightly worse than that of the DLFC. The DLFC (BN) cannot extract the water completely especially. It has a better performance in Sentinel-2A, but there was mistaken extraction. For GF-6 and ZY-3, it has worse performance compared with DLFC. For the big area water in GF-6 and ZY-3, DLFC (BN) can only extract water in the border. It cannot extract the whole water area. Therefore, the whole accuracy is lower than DLFC.

In our study, the GF-2 remote sensing images that have four channels (blue band, green band, red band, and NIR band)
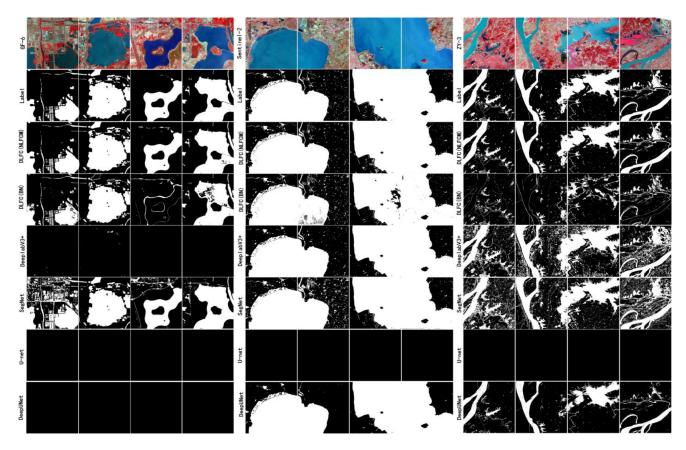
Fig. 6.   Prediction results of GF-6, Sentinel-2, and ZY-3 using different deep learning methods.

are imported into the deep learning network. Compared with the other three remote sensing images, we find that the blue band, green band, red band, and NIR band wavelength ranges of these three types of remote sensing images overlap greatly. The spectral and spatial features of the water in different source remote sensing images are similar. The method we proposed has two important parts, the densely connected module and the LFC module. Each layer of the network can receive the feature maps of all layers before it through the densely connected module. The concatenate operation on the feature dimension is used when connecting across layers. Therefore, it has a strong ability to extract different level water body features and reuse water body feature from the remote sensing images that have complex and diverse information. The highly abstract features are important for extracting water body information from different remote sensing images. The network extracts water body from different images through the highly abstract features. The LFC module can obtain highly abstract features. It can improve the adaptability of the network. Therefore, our method can not only extract the highly abstract spectral and spatial features from remote sensing images but also use the different level features of remote sensing images. Through these two modules, we can fuse the spatial and spectral information from different level features better.

As we all know, there is much information in remote sensing images including the water body information, building information, road information, vegetation information, etc. The information is complex and diverse. Some information may have the same features. Through simple convolution and pooling operation to extract features may lose most of the important information and cannot recover the feature as much as possible. Therefore, the wrong extraction often appears in the U-net, DeepUNet, SegNet, and DeepLabV3+. Besides, this method cannot extract highly abstract features because of the width of the network. They cannot realize water body extraction from different remote sensing images. The remote sensing images information extraction is a pixel-based binary classification problem. For the DeepLabV3+, it pays more attention to the encoder part. In this decoder part, the structure is so simple which cannot recover the feature well. So it cannot recover the feature of images so better that it has the bad extraction results.

In conclusion, the deep learning networks mentioned above cannot realize water body extraction from different remote sensing images without any preprocessing. The results of the evaluate metrics are listed in Table VI.

V. DISCUSSION

A. Advantages of the DLFC Method

In this article, we proposed the DLFC method for water extraction. Compared with the traditional water body extraction method, the DLFC improves the accuracy of water body extraction in various environmental backgrounds. In addition, it also provides a flexible and automatic way to extract water bodies from different high-resolution remote-sensing images without

TABLE VI
AVERAGE OA, F1-SCORE, AND IOU RESULTS OF SEVERAL TYPICAL MODELS

Methods	OA			F1-score			IoU		
	GF-6	Sentinel-2	ZY-3	GF-6	Sentinel-2	ZY-3	GF-6	Sentinel-2	ZY-3
DLFC (ours)	99.42%	98.93%	99.51%	99.2%	98.98%	98.9%	98.42%	97.99%	97.83%
DLFC (NLFCM)	99.06%	99.02%	98.98%	98.59%	99.06%	97.76%	97.21%	98.13%	95.62%
DLFC(BN)	92.58%	97.61%	86.99%	79.79%	97.61%	63.81%	75.12%	95.40%	47.64%
DeepLabV3+	63.64%	98.47%	84.81%	00.28%	98.44%	75.36%	00.14%	96.95%	60.82%
SegNet	91.74%	97.86%	85.55%	91.14%	97.66%	76.24%	83.90%	95.58%	62.60%
U-Net	63.60%	37.23%	76.08%	0	0	0	0	0	0
DeepUNet	63.60%	98.37%	94.36%	0	98.25%	86.56%	0	96.61%	76.32%

any processing steps, which accelerates the efficiency of water extraction.

The DLFC method can achieve higher precision requirements than those existing traditional water extraction methods, and it has stronger robustness for different types of water extraction in different urban environments. As a comparative experiment, NDWI can also achieve higher accuracy by adjusting the threshold. However, this requires a strong subjective consciousness, rich experience, and much time. It is not easy to extend this to large-scale automation applications. For the traditional deep learning method, the results of U-Net are not satisfactory. It classes other things such as land and buildings as water bodies. In contrast, the performance of DeepLabV3+ is not good. The DeepLabV3+ results are much more incomplete. Thin water bodies are not extracted by DeepLabV3+. The DLFC (NLFC) also has good performance in water body extraction. However, the accuracy may be lower than that of the DLFC.

Compared with the latest deep learning method proposed by other studies, the DLFC method can not only extract water bodies from GF-2 images but also extract water bodies from other remote sensing images. Above all, before extracting water from other remote sensing images, we do not need any other processing. This means that using only one dataset can realize water body extraction from different remote sensing images. These different remote sensing images do not need any other preprocessing, which largely saves time and experience.

In short, the DLFC method does not require many subjective decisions that affect the classification accuracy and it still maintains the advantage of high accuracy in water extraction from different types of remote sensing images. There is no need to worry about downloading and processing massive amounts of data. We can also apply the idea of the DLFC method to other land use classification studies.

B. Further Improvements

Although the DLFC method achieved satisfactory results in our study, there are still several problems. First, the water in the images with lengths of only two pixels often cannot be completely extracted. This performance may be created by convolution operations, which can be solved by improving the network. Second, some shadows of tall buildings are classified

as water due to the spectrum similarity between shadows and water. This problem may be solved by adding more accurate samples of pixels containing shadows and water to the training data. Finally, the accuracy of the extraction results from the different remote sensing images is not perfect. Therefore, the model parameters may be more suitable for the mapping of water in different remote sensing images.

VI. CONCLUSION

With the increase in image spatial resolution and the decrease in the available spectral channels, the traditional methods of water body extraction cannot meet the requirements. However, these methods have already performed well in medium and low-resolution imagery. The higher the resolution of the remote sensing images is, the more spectral information it has, which increases the difficulty of capturing features. In this condition, we use deep learning to realize the reuse of multiple features that can slow the degradation process.

The major contributions of our method were the proposed a novel deep learning encoder–decoder framework, which is called the DLFC network. At the same time, we introduce group normalization to replace traditional batch normalization and designed an LFC module in the upsampling process, which obviously improved the OA and F1 score and IOU (98.44%, 95.39%, and 91.25%). Through the proposed network, we accomplished the target that extracts the water body from different images of one sensor and different sensors automatically. Experiments were carried out on the multisensor dataset. Water bodies of different remote sensing images were extracted successfully with the proposed DLFC, and the results demonstrated the effectiveness and feasibility of the DLFC in improving the performance for water extraction. The proposed DLFC was compared with other traditional methods, such as NDWI, SVM, and other typical networks for semantic segmentation, such as the U-net, SegNet, and Deeplab-V3+ models. The experimental results showed that the proposed model performed better than other methods in extracting slender water bodies and distinguishing the effect of building shadows. In addition, the DLFC can be used to extract water bodies from other types of satellite imagery. The effectiveness of the DLFC means that it shows great promise for practical application with multiple types of satellite imagery.

REFERENCES

[1] N. Du, H. Ottens, and R. Sliuzas, "Spatial impact of urban expansion on surface water bodies—A case study of Wuhan, China," *Landsc. Urban Plan*, vol. 10, no. 2. pp. 175–185, 2010.

[2] X. Yang and Q. Qin, "Urban surface water body detection with suppressed built-up noise based on water indices from Sentinel-2 MSI imagery," *Remote Sens. Environ.*, vol. 9, no. 16, pp. 259–270, 2018

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

[4] J. Pekel *et al.*, "A near real-time water surface detection method based on HSV transformation of MODIS multi-spectral time series data," *Remote Sens. Environ.*, vol. 10, no. 8, pp. 704–716, 2014.

[5] E. Work and D. Gilmer, "Utilization of satellite data for inventorying prairie ponds and lakes," in *Proc. Photogramm. Eng. Remote Sens.*, 1976, pp. 685–694.

[6] W. Li *et al.*, "A comparison of land surface water mapping using the normalized difference water index from TM, ETM plus and ALI," *Remote Sens.*, vol. 5, pp. 5530–5549, 2013.

[7] Z. Du *et al.*, "Analysis of Landsat-OLI imagery for land surface water mapping," *Remote Sens. Lett.*, vol. 5, no. 7–9, pp. 672–681, 2014.

[8] Y. Du, Y. Zhang, F. Ling, Q. Wang, W. Li, and X. Li, "Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band," *Remote Sens.*, vol. 8, no. 354, pp. 1–19, Apr. 2016.

[9] Y. Zhou, G. Xie, and S. Wang, "Information extraction of thin rivers around built-up lands with false NDWI," *J. Geo-Inf. Sci.*, vol. 16, no. 1. pp. 102–107, 2014.

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

[11] G. Rolif and Fumera, "Support vector machines for remote sensing image classification," in *Proc. Vol. 4170, Image Signal Process. Remote Sens. VI*, 2001, pp. 160–166.

[12] E. Aung and T. Tint, "Ayeyarwady river regions detection and extraction system from Google Earth imagery," in *Proc. IEEE Int. Conf. Inf. Commun. Signal Process.*, 2018, pp. 74–78.

[13] Z. Liang, "Research on water information extraction method of multi-source remote sensing based on deep learning and its application," M.S. Dept. Ecol., AnHui Univ., AnHui, China, 2019.

[14] N. Wang *et al.*, "Application of U-Net model to water extraction with high resolution remote sensing data," *Remote Sens. Land Resour.*, vol. 32, no. 1. pp. 35–42, 2020.

[15] P. Frazier and K. Page, "Water body detection and delineation with Landsat TM data," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 12. pp. 1461–1467, 2000.

[16] S. 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.

[17] H. Xu, "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensedimagery," *Int. J. Remote Sens.*, vol. 27, no. 12/14, pp. 3025–3033, 2006.

[18] L. Yu *et al.*, "Convolutional neural networks for water body extraction from Landsat Imagery." *Int. J. Comput. Intell. Appl.*, vol. 16. no. 1, 2017, Art. no. 175001.

[19] W. Wu, Q. Li, Y. Zhang, X. Du, and H. Wang, "Two-step urban water index (TSUWI): A new technique for high-resolution mapping of urban surface water," *Remote Sens.*, vol. 10, no. 11, 2018, Art. no. 1704.

[20] X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[21] F. Isikdogan, A. Bovik, and P. Passalacqua, "Surface water mapping by deep learning," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 11, pp. 4909–4918, Nov. 2017.

[22] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. Johnson, "Deeplearning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 166–177, 2019.

[23] W. Fang *et al.*, "Recognizing global reservoirs from Landsat 8 images: A deep learning approach," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3168–3177, Sep. 2019.

[24] X. Yao *et al.*, "Land use classification of the deep convolutional neural network method reducing the loss of spatial features," *Sensors*, vol. 19, no. 12, 2019, Art. no. 2792.

[25] S. J. H. Li and S. Y. Fuchun, "Aerial scene classification with convolutional neural networks," in *Proc. Int. Symp. Neural Netw.*, 2015, pp. 258–265.

[26] H. Fan, G. Song, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14680–14707, 2015.

[27] G. Fu, C. Liu, R. Zhou, T. Sun, and Q. L. Zhang, "Classification for high resolution remote sensing imagery using a fully convolutional network," *Remote Sens.*, vol. 9, no. 5, p. 498, 2017, doi: 10.3390/rs9050498.

[28] C. Gong, Y. Ceyuan, Y. Xiwen, G. Lei, and H. Junwei, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, May 2018.

[29] S. Wang, H. Yang, Q. Wu, Z. Zheng, Y. Wu, and J. Li, "An improved method for road extraction from high-resolution remote sensing images that enhances boundary information," *Sensor*, vol. 20, no. 7, 2020, Art. no. 2064.

[30] H. Shen, Y. Jiang, T. Li, Q. Cheng, C. Zeng, and L. Zhang, "Deep learning-based air temperature mapping by fusing remote sensing, station, simulation and socioeconomic data," *Remote Sens. Environ.*, vol. 240, 2020, doi: 10.1016/j.rse.2020.111692.

[31] R. Li *et al.*, "DeepUnet: A deep fully convolutional network for pixel-level sea-land segmentation," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 3954–3962. Nov. 2018.

[32] S. Manaf, N. Md, N. Sulaiman, N. Husin, and M. Hamid, "Artificial neural networks for satellite image classification of shoreline extraction for land and water classes of the North West coast of peninsular Malaysia," *Adv. Sci. Lett.*, vol. 24, no. 2, pp. 53–62, Feb. 2018.

[33] Y. Wang, Z. Li, C. Zeng, G. Xia, and H. Shen, "Extracting urban water by combining deep learning and Google Earth Engine," *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* vol. 13, pp. 768–781, 2020.

[34] G. Huang, Z. Liu, V. D. M. Laurens, and K. Q. Weinberger "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2261–2269.

[35] Y. Wu and K. He, "Group Normalization," *Int. J. Comput. Vis.* 128, pp. 742–755, 2020, doi: 10.1007/s11263-019-01198-w.

[36] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Computer Science*, 2014, *arXiv:1412.6980*.

[38] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

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

[40] 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, Dec. 2017.

Mengya Li received the Graduate degree, in 2017, from Geographic Information Science, Anhui University, Hefei, China, where she is currently working toward the Postgraduate degree at the School of Resources and Environmental Engineering.

Her research interests include deep learning, remote sensing image data processing, and remote sensing image information extraction.

Penghai Wu (Member, IEEE) received the Ph.D. degree in cartography and geographical information engineering from Wuhan University, Wuhan, China, in 2014.

He is currently an Associate Professor with the School of Resources and Environmental Engineering, Anhui University, Hefei, China. His research interests include land surface temperature reconstruction and fusion, deep learning, and regional eco-environmental change.

Biao Wang received the Ph.D. degree from Chungbuk National University, Cheongju, South Korea, in 2015.

He is a Lecturer with the School of Resources and Environmental engineering, Anhui University, Hefei, China. His research interests include photogrammetry and remote sensing image processing, especially on change detection and object detection.

Hui Yang received the B.S. degree in geography from Anqing Normal University, Anqing, China, in 2010. He received the M.S. degree in surveying and the Ph.D. degree in cartography and geographical information engineering from Wuhan University, Wuhan, China, in 2012 and 2019, respectively.

He is currently a Lecturer with Information Materials and Intelligent Sensing Laboratory of Anhui Province, Institutes of Physical Science and Information Technology, Anhui University, Anhui, China. His research interests include remote sensing image semantic segmentation, artificial intelligence, and geographic information mining.

Honglyun Park received the B.S., M.S., and Ph.D. degrees from the Chungbuk National University, Cheongju, South Korea, in 2012, 2014, and 2019, respectively.

Since 2019, he has been an Assistant Professor with Youngsan University, Yangsan, South Korea. His research interests include remote sensing and GIS.

Yanlan Wu received the Ph.D. degree in cartography and geographic information system from Wuhan University, Wuhan, China, in 2004.

She is currently a Professor with the School of Resources and Environmental Engineering, Anhui University, Hefei, China. Her research interests include deep learning, remote sensing image data processing, and remote sensing image information extraction.