Classification of High-Resolution Remote Sensing Images in the Feilaixia Reservoir Based on a Fully Convolutional Network

PINGHAO WU1,2,5, KAIWEN ZHONG2,3, HONGDA HU2,4, JIANHUI XU2,4, YUNPENG WANG1, AND YI ZHAO1,2,5

1 Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, China
2 Key Laboratory of Guangdong for Utilization of Remote Sensing and Geographical Information System, Guangdong Open Laboratory of Geospatial Information Technology and Application, Guangzhou Institute of Geography, Guangzhou 510070, China
3 Guangdong Natural Resources Science and Technology Collaborative Innovation Center, Guangzhou 510070, China
4 Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, China
5 College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

Corresponding author: Kaiwen Zhong (zkw@gdas.ac.cn)

This work was supported in part by the Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) under Grant GML2019ZD0301; in part by the Science and Technology Planning Project of Guangdong Province under Grant 2018B020207002, Grant 2019B020208013, and Grant 2019B020208004; in part by the Provincial Agricultural Science and Technology Innovation and Promotion Project of Guangdong Province under Grant 2019KJ1147; in part by the GDAS’ Project of Science and Technology Development under Grant 2020GDASYL-20200104004; and in part by the Guangdong Innovative and Entrepreneurial Research Team Program under Grant 2016ZT06D336.

ABSTRACT In technologically underdeveloped areas, water pollution threatens the living environment of local residents, so remote sensing monitoring of the features around reservoirs is necessary. Fully convolutional networks (FCNs) offer great potential for extracting high-resolution features due to their unlimited input image size and higher accuracy compared to convolutional neural networks. Therefore, a proposal to classify WorldView-2 images is implemented with a sixty-eight thousand iterations of fine-tuning and fully trained combined training method based on a fully convolutional network (SEFCN). The chosen images depict the urban area of Yingde, which is located to the northeast of the Feilaixia Reservoir, Qingyuan, Guangdong Province, China. The SEFCN combines an FCN-32s and FCN-16s to better integrate the deep features and shallow features and effectively improve the classification accuracy. Additionally, the loss value fully converges with enough iterations, and the overfitting caused by the superposition of two full trainings is avoided. The SEFCN model achieves the highest accuracy among all compared classification models and obtains the best classified results on the WorldView-2 images, as demonstrated by attaining the highest F1 scores in each category. After the classified images are optimized using a conditional random field, the current status of the study area is analyzed, and several suggestions for land use in the urban area of Yingde are made. The experiments still have deficiencies in the application of high-resolution remote sensing image classification with different sensors, and the classification results can be optimized and improved in other aspects.

INDEX TERMS Feilaixia Reservoir, classification, fully convolutional network, SEFCN, WorldView-2.

I. INTRODUCTION

Reservoir projects have functions such as flood control, water storage irrigation, water supply regulation, and power generation. They play a crucial role in regulating local ecological environments and promoting local economic development.

Economic growth, industrialization, and urbanization have accelerated in recent years. The ecological spatial patterns of landscapes, such as cultivated land, forests, and settlements, have changed significantly. Various illegal activities, such as sand mining, breeding, and reclamation, have resulted in increased water pollution. This has threatened the living environment of local residents [1]–[4]. Due to technological underdevelopment, there are a large amount of areas...
that are unmonitored by traditional reservoir monitoring and management systems. This leads to a low land use efficiency around reservoirs and the disorganized management of features, highlighting the urgent need for dynamic remote sensing monitoring.

Deep convolutional nets have enabled recent breakthroughs in image processing [5]–[8], and convolutional neural networks (CNNs) have made significant progress in remote sensing image interpretation. Wu designed a shallow CNN with a 5-layer network layer and a complex CNN utilizing a 9-layer network layer and observed that the complex convolutional neural network had a higher accuracy [9]. Wang researched and designed an end-to-end deep convolutional neural network (DCNN) model allowing high-accuracy input of differently sized images [10]. To reduce the storage space required for DCNN models and to increase the portability of the network, Ding proposed a lightweight network method [11]. Tziolos et al. used a CNN to present a novel Earth observation (EO)-based soil monitoring approach leveraging open-access Copernicus Sentinel data and the Google Earth Engine platform [12].

Currently, CNNs have increasingly important applications in remote sensing image change detection and object detection, especially in monitoring building changes and performing ship identification. Maria et al. used a supervised classification program to obtain a large amount of training data. This allowed for the presentation of an automated building detection framework with DCNNs [13]. Shunta et al. proposed a CNN for extracting buildings in remote sensing images, utilizing a channelwise inhibited softmax to avoid overfitting. In this method, 8 shifted input maps were matched and merged with 8 output maps in a single model to simultaneously detect and extract buildings [14]. Zhang et al. proposed a novel approach for high-speed ship detection in synthetic aperture radar (SAR) images based on a grid convolutional neural network (G-CNN) [15]. Wu et al. proposed a new coarse-to-fine ship detection network (CF-SDN) that directly achieves end-to-end mapping from image pixels to bounding boxes with confidence [16].

A fully convolutional network (FCN) can perform semantic segmentation and is well suited for the pixelexcision classification of remote sensing data, which helps us obtain more details from images. This is due to the skip-connection operation in the model structure, which is significantly different from the operations in a typical CNN [17]–[20]. Tai proposed a classification method based on FCNs for high-resolution remote sensing imagery and designed an FCN-based SAR fusion image classification framework and method [21]. Mboga et al. developed a methodology based on FCNs that are trained in an end-to-end fashion using only aerial RGB images as input [22]. Li et al. presented a new end-to-end FCN-based deep network architecture for hyperspectral image (HSI) change detection [23].

Some modified algorithms have been proposed for fewer parameters than FCNs [24]–[26]. Most of these models are an encode-decode structure and have at least two upsampling layers. However, E-Net [27] denies the completely symmetrical structure of upsampling and downsampling in U-Net [28]. UNet++ [29] also proposes that skip connections combine the shallow features of the encoder with the deep features of the decoder, which would produce a semantic gap. The open question concerns how many upsampling layers would satisfy the needs of training a classified model. Furthermore, determining the length of the skip connection that would produce a semantic gap is yet unresolved. In an effort to explore these questions, we used FCNs, in which FCN-32s have no skip connection, while FCN-8s have a long skip connection, and FCN-16s have a short skip connection. We start our experiment with a pretrained visual geometry group (VGG) since fine-tuning has been proven to have better potential than training from scratch, as Tajbakhsh et al. [30] mentioned in the paper.

Many public datasets of urban areas can be acquired from previous works, but most are unsuitable for our research area. The study area contains the Beijiang River crossing, rich vegetation, dense buildings, especially some low buildings. We develop the sample database from our remote sensing data located in the study area and weaken the dependence of the dataset on manual work because expanding the number of samples by nonvisual interpretation is an important consideration. Most research on FCN model fine-tuning has allowed a reduction of the base learning rate and an adoption of one-third of the full number of training iterations, which saves a significant amount of time. However, there is currently no study that has determined the number of iterations with the same base learning rate for fine-tuning during training that would result in a higher accuracy for the subsequent full training.

The main goal of this study was to present an SEFCN model to classify the WorldView-2 images of the Yingde urban area, which is located to the northeast of the Feilaixia Reservoir. We also aimed to analyze the current status of the study area to improve the land use monitoring level in the area. To obtain sufficient sample data, we used mirror flipping to expand the dataset, which was acquired from a manually performed visual interpretation of WorldView-2 images. We trained several FCN models (FCN-8s, FCN-16s, and FCN-32s) of high-resolution remote sensing images based on a VGG16 network model. Then, we also trained a U-Net on the same images with the same initial weight and other training conditions. After our experiment using a trained FCN-16s with the initial weight of a fine-tuning FCN-32s model was performed, it was iterated more than 10 thousand times. Each group of 4000 iterations for fine-tuning the FCN-32s model provided the initial weight for a full-training FCN-16s model, and the seven models with the smallest loss values were applied to high-resolution remote sensing images to compare the classification accuracies of these models.

II. METHODOLOGY

The experimental process was divided into three sections, which are illustrated in Figure 1. The dataset included a
large array of images and labels that were separately used for training and validation. After adequate data preparation, the FCN-8s, FCN-16s, FCN-32s, U-Net and SEFCN were separately trained. In the training process, the fluctuation of the loss value reflected the model convergence, and the training parameters with nonconvergence were adjusted. Some parts of the data were used to validate the trained model, pixel accuracy, and mean accuracy, while other indices were evaluated in those classification models. Continuous adjustments were performed until the best performing model was found for WorldView-2 image classification. The boundary of the classification results for a given feature was optimized by a CRF.

**A. CONVOLUTIONAL NEURAL NETWORK**

After AlexNet, increasingly deeper CNN models were proposed [31]. In comparison to traditional algorithms, CNNs have a more effective range of interconnectivity, along with improved self-adaptability, self-organization, fault tolerance, and self-learning. The basic structure of a CNN includes five parts: input layer, convolution layer, pooling layer, fully connected layer, and output layer. Among all the layers, the convolution layer plays a vital role. The original input image is convolved with a set of trainable kernels that scan across the entire input image, resulting in a group of feature maps [32]. Each convolution kernel can extract different features, and there are multiple convolution kernels in each convolution layer that extract the features of the same picture [33]. The result of each convolution is the feature map delivered to the next layer. After the convolutional layer, the pooling layer is always used to condense the output of the feature maps from the convolutional layer. This operation is known as “downsampling”, and it reduces the dimensions of the feature map and the number of parameters to be learned by the network. This prevents overfitting, expands the receptive field, and achieves invariance (translation, rotation, and scale invariance). The hierarchy also allows the CNN network to learn data at multiple levels [34].

**B. FULLY CONVOLUTIONAL NETWORK**

A CNN becomes an FCN when the fully connected layers are replaced by convolutional layers. An FCN inherits the high-accuracy feature for image-label classification from the standard CNN. It maintains the 2-D spatial information of the input image while performing dense class prediction [35]. The FCN uses a skip-connection operation, in which the high-level abstract semantic information and the low-level fine semantic information are added directly according to the corresponding channels to form new features [36]. This kind of skip-connection operation is known as upsampling. FCNs use deconvolution to upsample the feature map of the last convolutional layer and restore it to the same size as the input image. This enables a prediction to be generated for each pixel while preserving the spatial information in the original input image. Finally, pixel-by-pixel classification can be performed on the upsampled feature map to achieve pixel-to-pixel mapping [37].

An FCN calculates the cross-entropy for each pixel and sums them across all the pixels and all the training samples in a training batch to obtain the cost, which can be represented by [38]:

$$C = \frac{1}{\sum_{i=1}^{n} r_i \times c_i} \sum_{i=1}^{n} \sum_{p=1}^{r_i} \sum_{q=1}^{c_i} \sum_{j=1}^{m} \left( \frac{1}{1 + \exp \left(-y_{p,q}^j\right)} \right) \times \ln \left( \frac{1}{1 + \exp \left(-y_{p,q}^j\right)} \right)$$

(1)
where \( I_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases} \) \( A \) is a given set, \( d_{p,q} = \frac{e^{\theta_{p,q}}}{\sum_{k=1}^{n} e^{\theta_{k,q}}} \), \( n \) is the total number of training samples in a given training batch, and \( m \) is the total number of feature types. If the location of the class \( j \) training sample \( i \) is the pixel in row \( p \) and column \( q \), the softmax output is \( \theta_{p,q} \). \( \theta_{p,q} \in (0, 1) \). If the ground-truth class ID for the pixel in row \( p \) and column \( q \) is \( j \), this output value is 1. Otherwise, the value is 0.

### C. SEFCN

According to the upsampling operations of the different pooling layers, there are three types of FCN models: FCN-8s, FCN-16s and FCN-32s. Figure 2 shows the formation of these models. The FCN-32s corresponds to an upsampling stride of 32 after the primary network, while the FCN-16s adds the feature information of pool 4 in the upsampling operation based on FCN-32s. Because the stride is 16, it is known as FCN-16s. In the same way, FCN-8s adds the data information of pool 3 based on FCN-16s.

![FIGURE 2. The overall process of FCN-8s, FCN-16s, and FCN-32s.](image)

An upsampling with a stride of 32 appearing after the primary network produces the FCN-32s. The final FCN-16s prediction is fused from the output of the primary network and the pool 4 layer, while the FCN-8s adds the pool 3 layer on this basis.

Traditional FCNs usually use VGG16 or another pretrained model as the initial weight input for training, but it might be difficult to obtain the best classification model. In this case, it is better to use fine-tuning based on our dataset to obtain the pretrained model as the initial weight for training.

First, we trained a fine-tuning FCN-32s model because the FCN-32s does not have a skip connection, and the feature map from pool 5 is directly upsampled to the final result. Then, the fine-tuning model is used as the initial weight and input for the FCN-16s full-time training. There is a short skip connection in the FCN-16s structure, from pool 4 to the twice upsampled pool 5, which was upsampled to match the size of the feature map for pool 4.

For commonly implemented fine-tuning operations, the step often has no significant effect on the final trained models. However, in SEFCN, the fine-tuning experiment of the FCN-32s was iterated sixty-eight times, which can significantly improve the full-training results of the FCN-16s.

For the FCN-8s model, it is commonly believed that the model adds the features of pool 3, pool 4 and pool 5 to make full use of the deep and shallow features in the samples, and the fully trained models can obtain the best classification performance among the FCNs. The long skip connection from pool 3 to the summation of pool 4 that is upsampled twice and pool 5 that is upsampled four times combines the shallow features and deep features. However, this approach also produces a semantic gap. The gap is manifested in the small convergence of the loss value in the training, and the final fully trained FCN-8s model is unable to well recognize all categories.

The SEFCN is a fine-tuning based and fully trained combined training method that combines the feature maps from pool 4 and pool 5 to maximize the utilization of the shallow features and deep features in the two pooling layers. The fine-tuning step neither prevents the loss value from sufficiently converging nor results in overfitting for the superposition of the two full trainings.

### D. OPTIMIZATION BY A CONDITIONAL RANDOM FIELD

A conditional random field (CRF) performs well in border optimization [39]. It couples adjacent nodes with spatial near-end pixels to facilitate the allocation of the same label. The CRF is connected to the fully convolutional neural network and extended to a deep CRF. It is then combined with the single-pixel prediction and shared structure by establishing a pair of semantics on all pixel pairs in the remote sensing image to restore the object boundary [40], [41]. The function of this model is:

\[
E(x) = \sum_{i} \theta_{i} (x_i) + \sum_{i,j} \theta_{i,j} (x_i, x_j),
\]

where \( x \) represents the label assignment of each pixel and \( \theta_{i} (x_i) \) represents the unary likelihood of the pixel method. \( \theta_{i,j} (x_i, x_j) \) is equivalent to \(- \log P(x_i)\), where \( P(x_i) \) is the label assignment probability of pixel \( i \).

Fully connected graphs that connect all pairs of image pixels \( i \) and \( j \) are used to establish a pair of potentials. From this, effective interference can be achieved. The paired edge potential can be defined as a linear combination of Gaussian kernels in the form [42]:

\[
\theta_{i,j} (x_i, x_j) = \mu (x_i, x_j) \sum_{m=1}^{K} w_m k_m (f_i, f_j),
\]

where \( \mu \) is the label of the compatibility function and \( k_m (f_i, f_j) \) is a Gaussian kernel. This kernel is a feature extracted from pixels \( i \) and \( j \) (defined by \( f \)) and is determined by the parameter \( w_m \). \( \mu (x_i, x_j) = 1 \) if \( x_i \neq x_j \). If it is 0, as in the Potts model, only the nodes with different labels will be impacted.

### III. EXPERIMENTAL DESIGN

#### A. STUDY AREA AND REMOTE SENSING DATA

The Feilaixia Water Conservancy Hub is located on the Bei-jiang River section approximately 40 kilometers northeast of Qingyuan city in the north-central part of Guangdong.
Province, China. It is currently the largest comprehensive water conservancy project in Guangdong Province [43], [44]. It is primarily used for flood prevention, and it is also used for power generation, shipping, water supply management, and improvement of the ecological environment, which is a crucial project for the comprehensive management of the Beijiang River Basin. The study area contains the urban area of Yingde City, which is located in the northeast section of the Feilaixia Reservoir [45].

Figure 3 shows the geographic location of the study area. The area covers both banks of the Beijiang River form the Yingde City district. This is the largest town in the Feilaixia Reservoir. The Yingde City flood basin and the large station flood basin in the study area are two major flood storage areas, which enable the Feilaixia Reservoir to withstand a 50-year flood and a 30-year flood, respectively. Overall, the Feilaixia Reservoir includes the most abundant feature types and the most concentrated feature samples in this area.

The WorldView-2 satellite was launched into space on October 6, 2009, and has accumulated sufficient data so far. Its primary advantages include flexible operation, a large capacity, fast return visits, accurate shooting, high-definition image capturing, and the availability of multiple color bands [46]. These advantages provide us with sufficient and reliable high-resolution remote sensing data sources. The parameters of WorldView-2 are shown in Table 1. WorldView-2 images with a spatial resolution of 0.5 meters, acquired on January 19 and February 13, 2018, were used for classification in the study. The images have three spectral bands: blue (440–510 nm), green (510–580 nm), and red (630–690).

**B. PREPROCESSING AND TRAINING**

The original WorldView-2 images were acquired and preprocessed according to the following steps: radiometric correction, image registration, mosaicing and cutting. Atmospheric correction was used to eliminate the errors caused by atmospheric scattering, absorption and reflection. Image registration and resampling were conducted to ensure that corresponding image points coincide in the same coordinate system. After mosaicing and cutting, the study area of WorldView-2 images was obtained.

To evaluate the distribution of features in the study area, we comprehensively considered the characteristics of water crossings, rich vegetation, and dense buildings in the study area, among other characteristics. The remote sensing images were classified into six types of features: bare land, buildings, roads, ships, vegetation, and water. Figure 4 shows the ground truth of one small piece of a WorldView-2 image. We divided a 14602 × 10346-pixel WorldView-2 image with a complete and concentrated distribution of features within the study area into smaller images and marked the original image to obtain the ground-truth image. Then, we used a 256 × 256 window to slide along the ground-truth image and trim it in 100 steps. After removing the images with the background value, we obtained 12,975 images. Among these images, ten percent were randomly selected and used for mirror flipping, and finally, we obtained a total of 14,272 images for network training and validation, which make up our dataset.

In the experiment, we divided the dataset images according to a ratio of eight to two, and they were used as training images and verification images, respectively. A total of 11,517 images was used for training, and 2855 images were used for validation. The VGG16 pretrained model was input as the initial weight in the full training of the FCN-8s, FCN-16s, FCN-32s, and U-Net. We input the fine-tuning
FCN-32s model, which was iterated sixty-eight thousand times, as the initial weight of the FCN-16s model for full training. The same base learning rate was used to acquire the trained SEFCN model. After these class models were trained, the WorldView-2 images of the study area were applied to these classifiers. We also employed 23 fine-tuning FCN-32s models with more than 10 thousand iterations as the initial weights of the fully trained FCN-16s, and six models with similar loss values to that of the SEFCN model were used in the WorldView-2 images classification. When the classifiers were running the class operation, the deploy file was modified according to the network structure of the classifier. The trained FCN-8s model, the trained FCN-32s and the trained U-Net model had relevant deploy files, respectively. The trained SEFCN model and the other fully trained FCN-16s models all used the deploy file of the FCN-16s model.

The experiments were implemented using Python 3.5 configured in the Caffe environment and were performed on the Windows 10 operating system installed on an NVIDIA GeForce RTX 2080 Ti graphic device with an 11 GB graphics card. The average loss and weight decay were set to 20 and 0.005, respectively. According to the training epochs, we adjusted the initial learning rate, which was set to 1e-10. In the U-Net training experiment, we set the batch size and the initial learning rate to 20 and 1e-8 to achieve the best final accuracy, and other parameters were consistent with the other experiments. The maximum number of iterations was 100 thousand.

To measure and compare the classification results of these trained models on WorldView-2 images, 5000 randomly distributed verification regions of interest (ROIs) were established in the study area. The classified images of the SEFCN and the trained U-Net model had relevant deploy files, respectively. The trained SEFCN model and the other fully trained FCN-16s models all used the deploy file of the FCN-16s model.

The assessment index values of the trained FCN-8s, FCN-16s, FCN-32s, U-Net and SEFCN models are shown in Table 2. Based on the statistics of the five class models, the SEFCN model demonstrated the highest pixel accuracy and mean accuracy. The MIoU and FWIoU were both higher for the FCN-16s model than in the other four models, and the pixel accuracy and mean accuracy were both above 0.90. The FCN-16s model was shown to be more effective than the FCN-32s model. Even though the mean accuracy of the FCN-16s model was slightly lower than that of the FCN-32s model, the OA, MIoU, and FWIoU of the FCN-16s model were all higher than those of the FCN-32s model. The FCN-8s and MIoU are standard measures of semantic segmentation. It effectively calculates the ratio of the ground truth to the predicted segmentation and is represented by:

\[ \text{MIoU} = \frac{1}{k + 1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}}. \]  

Where \( k \) represents the number of classes and \( p_{ij} \) represents the number of pixels that belong to class \( i \) but are predicted to be class \( j \). In other words, \( p_{ii} \) represents the true quantity, and \( p_{ij} \) and \( p_{ji} \) are interpreted as a false positive and false negative, respectively.

FWIoU is an improvement of the MIoU. This metric sets the weight for each class according to the frequency and is represented by:

\[ \text{FWIoU} = \frac{1}{\sum_{i=0}^{k} \sum_{j=0}^{k} p_{ij}} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}}. \]  

To compare the classification results of these trained models on WorldView-2 images, a confusion matrix was used to obtain the overall accuracy (OA), kappa coefficient, user accuracy (UA) of each category, and producer accuracy (PA) of each classification feature. The precision, recall, and F1 score were also used as metrics to quantitatively evaluate the performances of different classifiers [47]–[49]. The OA is the ratio of correctly classified pixels to the total number of pixels in the confusion matrix. The kappa coefficient measures the consistency of the reference data and the detailed map. A true positive (TP) indicates a pixel of a positive example that is classified correctly. A false negative (FN) indicates a pixel of a negative example that is misclassified. A false positive (FP) indicates a pixel that is a negative example but that the model misjudges as a positive example. The functions for calculating the precision and recall are:

\[ \text{Precision} = \frac{TP}{TP + FP}. \]  

\[ \text{Recall} = \frac{TP}{TP + FN}. \]  

The F1-score is the harmonic average of the precision and recall, and it measures the balance between the precision and recall. It is formulated as follows:

\[ F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \]  

To evaluate the performance of the model more intuitively, the precision-recall curve (PRC) is used to reflect the trend of the precision and recall. The PRC is generated based on the IoU, which represents the overlapping area between the detection results and the ground-truth object. A result is considered to be a TP if the IoU exceeds a set threshold; otherwise, it is considered to be a FP. In addition, if multiple detection results overlap with the same ground-truth bounding box, only one is considered a TP, and the others are marked as FPs.

IV. RESULTS

The assessment index values of the trained FCN-8s, FCN-16s, FCN-32s, U-Net and SEFCN models are shown in Table 2. Based on the statistics of the five class models, the SEFCN model demonstrated the highest pixel accuracy and mean accuracy. The MIoU and FWIoU were both higher in the SEFCN model than in the other four models, and the pixel accuracy and mean accuracy were both above 0.90. The FCN-16s model was shown to be more effective than the FCN-32s model. Even though the mean accuracy of the FCN-16s model was slightly lower than that of the FCN-32s model, the OA, MIoU, and FWIoU of the FCN-16s model were all higher than those of the FCN-32s model. The FCN-8s
and the U-Net exhibited the worst performance among all classification models, and the metrics of the U-Net were lower than those of the FCN-8s.

We employed the WorldView-2 images of the research area to verify the accuracy of these class models. After we obtained the image classification results from the trained models, we used the confusion matrix to evaluate these results. Table 3 indicates the values of the OA and kappa coefficient of each classification model. It also shows the calculated precision, recall, and F1 score for each classification model.

**TABLE 2.** Comparison of the trained FCN-8s, FCN-16s, FCN-32s, U-Net and SEFCN models.

| Model   | Pixel Accuracy | Mean Accuracy | mIoU   | FWIoU  |
|---------|----------------|---------------|--------|--------|
| FCN-8s  | 0.8740         | 0.8103        | 0.7140 | 0.7792 |
| FCN-16s | 0.8999         | 0.8727        | 0.7770 | 0.8204 |
| FCN-32s | 0.8808         | 0.8851        | 0.7432 | 0.7938 |
| U-Net   | 0.8279         | 0.6801        | 0.6624 | 0.7557 |
| SEFCN   | **0.9171**     | **0.9065**    | **0.8128** | **0.8491** |

The statistics indicated that the SEFCN classification model obtained the best performance among the models. The recall of the SEFCN was approximately 0.9299, with both a precision and an OA above 0.88. The F1 score of the SEFCN was approximately 0.9129, which was the highest among these classification models. The F1 scores of the FCN-16s and FCN-32s were both approximately 0.9, while the F1 scores of the FCN-8s and U-Net ranged from 0.8 to 0.9. It was proven that FCN-16s and FCN-32s achieved similar accuracy levels, and the classification accuracies of U-Net and FCN-8s were also similar, while that of U-Net was lower than that of FCN-8s.

Figure 5 compares the classifications of the five models for three original images. This allows a more intuitive visualization of the differences among the classification models. Images (a) to (g) are the original images, the results of the FCN-8s, the results of the FCN-16s, the results of the FCN-32s, the results of the U-Net, the results of the SEFCN, and the ground-truth image, respectively. The left image in Figure 5a is located near a park in an urban area and primarily includes bare land, buildings, roads, and a lake. The middle image is next to the water of the reservoir and

**TABLE 3.** Comparison of the classified results of the trained FCN-8s, FCN-16s, FCN-32s, U-Net and SEFCN models on the WorldView-2 images.

| Model   | Precision | Recall | F1   | OA    | Kappa |
|---------|-----------|--------|------|-------|-------|
| FCN-8s  | 0.8577    | 0.9395 | 0.8967 | 0.8689 | 0.8201 |
| FCN-16s | 0.8868    | 0.9266 | 0.9062 | 0.8316 | 0.8316 |
| FCN-32s | 0.8781    | 0.9286 | 0.9026 | 0.8743 | 0.8278 |
| U-Net   | 0.8087    | 0.8234 | 0.8160 | 0.7579 | 0.6776 |
| SEFCN   | **0.8889** | **0.9379** | **0.9128** | **0.8444** | **0.8872** |

**FIGURE 5.** The classification results of three original images. (a) The original remote sensing image. (b) The classification result of the trained FCN-8s model. (c) The classification result of the trained FCN-16s model. (d) The classification result of the trained FCN-32s model. (e) The classification result of the trained U-Net model. (f) The classification result of the SEFCN, which was the trained FCN-16s model with the initial weight of the trained FCN-32s. (g) The ground-truth image.
TABLE 4. Comparison of the accuracies of different classifiers.

| Model     | Index | Bare land | Building | Road   | Ship   | Vegetation | Water   |
|-----------|-------|-----------|----------|--------|--------|------------|---------|
| FCN-8s    | PA    | 0.4000    | 0.8308   | 0.6176 | 0.6800 | 0.8872     | 0.9574  |
|           | UA    | 0.3393    | 0.9159   | 0.7554 | 0.3696 | 0.9425     | 0.9490  |
| FCN-16s   | PA    | 0.6737    | 0.8377   | 0.7000 | 0.9200 | 0.9106     | 0.9663  |
|           | UA    | 0.8649    | 0.9587   | 0.8322 | 1.0000 | 0.9008     | 0.9698  |
| FCN-32s   | PA    | 0.6000    | 0.8377   | 0.7118 | 1.0000 | 0.8935     | 0.9734  |
|           | UA    | 0.8906    | 0.9179   | 0.8345 | 1.0000 | 0.9159     | 0.9649  |
| U-Net     | PA    | 0.4211    | 0.9057   | 0.1706 | 0.1200 | 0.8593     | 0.8156  |
|           | UA    | 0.1282    | 0.6303   | 0.3152 | 0.0341 | 0.9269     | 0.9935  |
| SEFCN     | PA    | 0.6421    | 0.8488   | 0.8235 | 1.0000 | 0.9099     | 0.9805  |
|           | UA    | 0.8841    | 0.9444   | 0.7910 | 1.0000 | 0.9176     | 0.9634  |

primarily includes water and some boats. The right image is located in the suburbs, and it includes vegetation and some nearby factories. Figure 5e shows the results of U-Net, which indicates that most of the roads and ships are not well recognized, and the classified features are fragmented. However, it can be seen that most of the building outlines are recognized, and most of the vegetation and water are also well identified. The buildings in Figure 5b, which is the FCN-8s classified image, cannot be clearly separated from the ground, while the buildings in Figures 5c and 5d, which are the separately classified FCN-16s and FCN-32s images, cannot be completely identified. However, the buildings in Figure 5f (the SEFCN classified image) can be separated from the ground and can also be completely recognized. Because the relatively thick roads in the FCN-8s classification results are not distinct from the surrounding vegetation and buildings, the relatively thin roads in the classification images of the FCN-16s and FCN-32s are likely to be intermittent and unconnected. By contrast, most of the continuous roads can be clearly distinguished from buildings and vegetation in the SEFCN classification images. In summary, the classification performances of FCN-8s and U-Net are exactly the opposite, the features in FCN-8s are fused together, while those of the U-Net are broken, and the classification results of FCN-16s and FCN-32s are similar and better, while the results of the SEFCN model classification are the closest to the ground-truth images.

The confusion matrix for each category determines the PA and UA of each feature, which are shown in Table 4, and the precision-recall (PR) curves for each category are shown in Figure 6. From the statistics, the buildings, roads, boats, and water performed the best in the SEFCN classification images, with PA values of 0.8488, 0.8235, 1.0, and 0.9805, respectively. When considering vegetation, the PA of the SEFCN was only 0.0097 lower than that of the FCN-16s, while the UA of the SEFCN was 0.0168 higher than that of the FCN-16s. This might have been caused by a deviation in the random verification samples. In addition, the PR curves prove that the SEFCN has the best performance among all classification categories. The statistics indicated that FCN-16s and FCN-32s have similar performances for six kinds of features, and the PR curves proved that the performance of FCN-16s is slightly better than that of FCN-32s, which is also consistent with the previous results. For U-Net, the statistics and the PR curves both indicated that roads and ships are not well recognized, while bare land performs better, and the performances for water and vegetation are close to those of the other models. The PA of the buildings in U-Net is the highest among all classifiers, and the PR curves also prove that the building classification results of U-Net are better than those of FCN-8s and are very close to the other three classifiers. The bare land in all the models performed poorly because the labels were less distributed and uneven. Generally, the class types in the SEFCN model classification images had the highest accuracies, the accuracy levels of those of the FCN-16s and FCN-32s models were similar, and the U-Net classified results were worse for roads and ships but performed well for buildings, vegetation and water, while the results were similar to those of the FCN-8s for bare land.

V. DISCUSSION

A. COMPARISON WITH U-Net

The advantage of the U-Net model is that it allows the classification model to be trained with a small number of samples. However, in the multitype feature classification experiment of high-resolution remote sensing images, the feature types are diverse, the styles vary, and it is difficult for a small number of samples to achieve high-precision classification. In the experiment with U-Net, the initial weight of VGG16 accelerates the convergence of loss during training, and the classification results show a good performance for buildings, vegetation and water.

However, U-Net was unable to identify the ships and roads, and although FCN-8s was better than U-Net, it still obtained a low accuracy in these two categories. The reason behind this result is the long skip connection. FCN-16s has a short skip connection, but its results are very similar to the classification results of FCN-32s. However, FCN-8s and U-Net, with their long skip connections, obviously have a large deviation in the recognition of small targets. This indicates that the feature maps close to the input image level have small receptive fields [50] and mainly recognize small targets. In addition,
FIGURE 6. The precision-recall (PR) curves for each category for the five classifiers on our own dataset.

these feature maps are not fused into the final classified image in an appropriate way, resulting in a semantic gap that makes the classifiers perform worse.

The U-Net performs worse than does the FCN-8s because the FCN-8s uses the summation to add the features of the upsampling pooling layer element by element, while the U-Net uses concatenation to connect the features of downsampling and upsampling. The concatenation of the U-Net structure causes more and longer skip connections between deep features and shallow features, resulting in more serious semantic gaps.

The experiments have proven that the number of upsampling layers is not as much as is possible. The short connection of the FCN-16s proved to be effective in improving the accuracy, and it does not produce a semantic gap. Based on this, SEFCN combines the FCN-32s and FCN-16s to better integrate deep features and shallow features to improve the classification accuracy, and the metrics also prove the effectiveness of the model.

B. COMPARISON WITH OTHER FINE-TUNING FCN-32S MODELS

In the full-training process of FCN-32s models, every group of 4000 iterations generates a fine-tuning FCN-32s model. We employed 23 models with more than 10 thousand iterations as the initial weight to use in the full training of the FCN-16s, provided that the same base learning rate was used with all full trainings. The loss values of these fully trained models are shown in Figure 7, and we chose the seven models with the lowest loss values in our experiment. The fine-tuning FCN-32s model was iterated 68 thousand times and was then input to the FCN-16s model for full training, which produced the SEFCN model. The other six fine-tuning FCN-32s models used for the FCN-16s full training were separately iterated 52 thousand times, 60 thousand times, 84 thousand times, 88 thousand times, 96 thousand times, and 100 thousand times (in the following section, they are denoted by exp-1 to exp-6).

To evaluate the accuracy of these seven fine-tuning FCN-32s models in image classification after the FCN-16s full training, we employed the models to classify the WorldView-2 images of the research area. Table 5 records the index values of the results. Exp-1 represents the classifier of a fully trained FCN-16s model with an initial weight that was iterated 52 thousand times for the FCN-32s models and spent the least amount of time on training. The loss values of the SEFCN model and exp-1 were similar, at 13,315 and 13,321, respectively. However, the results of exp-1 were worse than those of the SEFCN model, as evidenced by the F1 score, precision, OA, and kappa coefficient. The recall of the SEFCN model was 0.0017 lower than that of exp-6, which was iterated twice with full-time training and obtained the lowest loss value, but the F1 score of the SEFCN model was higher than that of exp-6, as were the other three metrics.
FIGURE 7. The loss values of fully trained FCN-16s models with different initial weights of fine-tuning FCN-32s models. The seven models with the lowest loss values are marked by the blue frame, and the red squares represent the loss value of the SEFCN model.

TABLE 5. Comparison of the classification results of the seven fully trained FCN-16s models with different fine-tuning FCN-32s models.

| Model | Precision | Recall | F1  | OA  | Kappa |
|-------|-----------|--------|-----|-----|-------|
| exp-1 | 0.8856    | 0.9374 | 0.9107 | 0.8409 | 0.8848 |
| exp-2 | 0.8832    | 0.9375 | 0.9096 | 0.8393 | 0.8834 |
| exp-3 | 0.8841    | 0.9340 | 0.9084 | 0.8367 | 0.8817 |
| exp-4 | 0.8889    | 0.9311 | 0.9095 | 0.8384 | 0.8826 |
| exp-5 | 0.8742    | 0.9424 | 0.9070 | 0.8356 | 0.8811 |
| exp-6 | 0.8775    | 0.9396 | 0.9075 | 0.8359 | 0.8812 |
| SEFCN | 0.8889    | 0.9379 | 0.9128 | 0.8444 | 0.8872 |

The SEFCN model achieved the highest F1 score among these classifiers, and the findings indicate that the SEFCN is superior to the other six models in the classification of WorldView-2 images.

C. OPTIMIZATION BY A CONDITIONAL RANDOM FIELD

CRF was used to optimize the feature boundaries of the SEFCN classified images. The boundary optimization had no significant effect on irregular borders such as vegetation, bare land and water, but it had a good effect on the optimization of regular roads, buildings and boat boundaries. Figure 8 shows the optimization results of the boundaries of buildings, roads, and ships, which are represented by (a), (b), and (c), respectively. Each section is composed of an ordered sequence of the original image, the SEFCN classified image before the CRF poststep, and the results of the classified image optimized by the CRF. A building’s boundary is usually integrated with those of adjacent buildings or other features such that it cannot be distinguished as a single building. After CRF optimization, the borders of the buildings are clearly separated, and the buildings also become regular and smooth. There are often multiple roads merged in the WorldView-2 images, and the SEFCN usually recognizes these roads as a single road. Separate road boundaries are not visible in the middle picture of Figure 8b, while the optimized road intersections have clear road boundaries in the right picture of Figure 8b. In the optimized picture of Figure 8b, unmerged roads can clearly be seen, and the overall boundaries of the continuous roads become neat and regular. The boundary of a ship on the water of the reservoir is usually not recognized and cannot be classified. After optimization, however, the boundary can be clearly separated from the other ships and the water surface. In the right picture of Figure 8c, the optimized shapes of ships become regular and neat, and very close to those of the ships in the original image.

D. LAND USE ANALYSIS AND LIMITATIONS

When examining the classification images from the research area, there is much bare land, vegetation coverage, water, and space with concentrated buildings and roads in the urban area of Yingde City. Based on this observation, we provide some advice for this research area. First, the issue of land management confusion caused by the problem of land rights in the reservoir [51] is crucial. Until the land ownership rights are determined, the land should be developed rationally and legally. Additionally, since red tea is an economic resource for Yingde City [52] and a large area of farmland exists in the suburbs [53], farmland could be returned to forests and grass to a certain extent. Finally, it is necessary to maximize...
the rational development and utilization of urban and transportation land in vertical three-dimensional construction for large areas of unused land such as bare land, making full use of underground and above-ground resources.

In the SEFCN, the FCN-32s and FCN-16s were combined to make full use of the features of pool 4 and pool 5 and better integrate deep features and shallow features. The fine-tuning training neither prevents the loss value from sufficiently converging nor results in overfitting for the superposition of two full trainings. In comparison with the fully trained FCN-32s and FCN-16s, the SEFCN achieved a higher accuracy and was confirmed to be effective in the WorldView-2 image classification of the research area. However, our experiments explored only the classification of the SEFCN model for WorldView-2 images, and we have further experimental plans to apply the SEFCN model to the classification of other high-resolution remote sensing images. Future work will focus on more data enhancement methods, structure and parameter optimization. Finally, the modified results of the CRF are not perfect. In the optimization of the boundaries of buildings and roads, some roads are not reclassified in places where the boundaries of the buildings are clear, and some buildings are not visible when the road boundaries are clear and complete. Future work will explore better optimization algorithms to obtain complete and clear roads and buildings.

VI. CONCLUSION

In our research, sufficient ground-truth images were made for the training and validation datasets. Then, the FCN-8s, the FCN-16s, the FCN-32s, the U-Net, and the SEFCN were trained based on the datasets. The SEFCN performed the best, yielding a pixel accuracy and a mean accuracy of 0.9171 and 0.9065, respectively. The WorldView-2 images of the study area were classified by the five models. When the classified results of the U-Net, FCN-8s, FCN-16s, and FCN-32s models based on WorldView-2 images in the study area were compared with those of SEFCN, we concluded that the SEFCN obtained the highest classification accuracy, with precision, recall, and F1 scores of 0.8889, 0.9379, and 0.9128, respectively. Additionally, the PA and the UA indicated that the SEFCN performed the best in the classification of most categories, and the PR curve also proved this conclusion. The classification of bare land by the SEFCN was worse, which might have been due to insufficient and uneven labels of bare land, as well as unmarked pieces of this type in the ground-truth images.

In our study, the CRF effectively optimized the feature boundaries, especially for buildings, roads, and ships, in the classification images of the SEFCN model. This helped rectify the feature edges in the classified images and made the features closer to those of the WorldView-2 images. The classified images show that the study area has a large amount of bare land, extensive vegetation coverage, rich waters, concentrated buildings, and roads. Considering the actual situation in the study area, the land ownership rights should be clearly divided. Because red tea is important for the economy of Yingde City, extensive vegetation coverage provides superior natural conditions for its cultivation. Finally, the existence of large areas of unused land, such as bare land in the urban area, indicates that maximizing the rational development and utilization of the urban area is necessary.

Overall, the SEFCN combines the FCN-32s and FCN-16s to better utilize deep features and shallow features in pool 4 and pool 5 to improve the classification accuracy. In addition, the loss value fully converges after enough iterations, and this prevents overfitting caused by the superposition of two full trainings. In comparison with the fully trained FCN-32s and FCN-16s, the SEFCN achieved a higher accuracy and was confirmed to be effective in the WorldView-2 image classification of the research area. However, our experiments still have deficiencies in applying high-resolution remote sensing image classification with different sensors. In future work, we will also focus on exploring better edge optimization algorithms, sample data enhancement methods and structure optimization.

REFERENCES

[1] X. Chen, Y. Lei, and G. Huang, “Evaluation of the status quo of non-point source pollution in the Feilaixia reservoir area of Beijiang,” Water Resour. Protection, vol. 35, no. 2, pp. 44–48, 2019.
[2] G. Huang, X. Chen, and X. Ren, “Characteristic analysis and simulation of non-point source pollution in typical watershed of Feilaixia Reservoir in Beijiang,” Water Resource Protection, vol. 35, no. 4, pp. 9–16, 2019.
[3] X. Li, G. Fang, and S. Huang, “Post-evaluation of national economy of Feilaixia water conservancy project,” People’s Yangtze River, vol. 36, no. 10, pp. 58–68, 2015.
[4] G. Meng and Q. Fu, “Discussion on reservoir resettlement planning under the background of rural revitalization strategy: An example of the reservoir relocation and resettlement of Feilaixia project,” J. Guangdong Tech. College Water Resour. Elect. Power, vol. 17, no. 4, pp. 18–21, 2019.
[5] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, May 2015.
[6] W. Ma, “A novel multi-model decision fusion network for object detection in remote sensing images,” Remote Sens., vol. 11, no. 7, pp. 737–755, 2019.
[7] L. Liu, “Multiscale deep spatial feature extraction using virtual RGB image for hyperspectral imagery classification,” Remote Sens., vol. 12, no. 2, pp. 280–305, Jan. 2020.
[8] S. Ji, “Building instance change detection from large-scale aerial images using convolutional neural networks and simulated samples,” Remote Sens., vol. 11, no. 11, pp. 1343–1363, Jun. 2019.
[9] Z. Wen, “Application research of convolutional neural network in image classification,” M.S. thesis, Dept. Comput., Electr. Sci., Tech. Univ., Sichuan, China, 2015.
[10] H. Wang, “Application of deep convolutional neural network in remote sensing image classification,” Ph.D. dissertation, Dept. Map Carto. Geo. Inf. Eng., Geosci. Univ., Beijing, China, 2018.
[11] P. Ding, “Research on optical remote sensing target detection technology based on deep convolutional neural network,” Ph.D. dissertation, Dept. Changehun Inst. Opt., Fine Mech. Phys., Chin. Acad. Sci. Univ., Jilin, China, 2019.
[12] N. Tziolas, N. Tsakiridis, E. Ben-Dor, and J. Theocharis, “Employing a multi-input deep convolutional neural network to derive soil clay content from a synergy of multi-temporal optical and radar imagery data,” Remote Sens., vol. 12, no. 9, pp. 1389–1415, Apr. 2019.
[13] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. Paragios, “Building detection in very high resolution multispectral data with deep learning features,” in Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), Jul. 2015, pp. 1873–1876.
[14] S. Saito, T. Yamashita, and Y. Aoki, “Multiple object extraction from aerial imagery with convolutional neural networks,” Electron. Imag., vol. 2016, no. 10, pp. 1–9, Feb. 2016.
[15] T. Zhang and X. Zhang, “High-speed ship detection in SAR images based on a grid convolutional neural network,” Remote Sens., vol. 11, no. 10, pp. 1206–1230, May 2019.

[16] Y. Wu, “A coarse-to-fine network for ship detection in optical remote sensing images,” Remote Sens., vol. 12, no. 2, pp. 246–270, Jan. 2020.

[17] J. 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, pp. 1162–1181, May 2019.

[18] G. Wu, X. Shao, Z. Guo, Q. Chen, W. Yuan, X. Shi, and Y. Xu, “Automatic building segmentation of aerial imagery using multi-constraint fully convolutional networks,” Remote Sens., vol. 10, no. 3, pp. 407–425, Mar. 2018.

[19] Z. Xu, C. Hopkins, and L. Chasmer, “Filtering stems and branches from terrestrial laser scanning point clouds using deep 3-D fully convolutional networks,” Remote Sens., vol. 10, no. 8, pp. 1215–1233, Aug. 2018.

[20] A. Song, “Change detection in hyperspectral images using recurrent 3D fully convolutional networks,” Remote Sens., vol. 10, no. 11, pp. 1827–1849, Nov. 2018.

[21] J. Tai, “Application of deep learning in remote sensing image target detection and surface coverage classification,” Ph.D. dissertation, Dept. Phot. Remote Sens., Wuhan Univ., Wuhan, China, 2017.

[22] N. Mboha, “Fully convolutional networks and geographic object-based image analysis for the classification of VHR imagery,” Remote Sens., vol. 11, no. 5, pp. 597–614, Mar. 2019.

[23] X. Li, Z. Yuan, and Q. Wang, “Unsupervised deep noise modeling for hyperspectral image change detection,” Remote Sens., vol. 11, no. 3, pp. 256–275, Jan. 2019.

[24] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A deep convolutional encoder-decoder architecture for image segmentation,” 2015, arXiv:1511.00561. [Online]. Available: http://arxiv.org/abs/1511.00561

[25] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, “The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation,” 2016, arXiv:1611.09126. [Online]. Available: http://arxiv.org/abs/1611.09126

[26] L.-C. 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,” 2016, arXiv:1606.00915. [Online]. Available: http://arxiv.org/abs/1606.00915

[27] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, “ENet: A deep neural network architecture for real-time semantic segmentation,” 2016, arXiv:1606.02147. [Online]. Available: http://arxiv.org/abs/1606.02147

[28] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” 2015, arXiv:1505.04597. [Online]. Available: http://arxiv.org/abs/1505.04597

[29] Z. Zhou, M. Mahfuzur Rahman Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A nested U-Net architecture for medical image segmentation,” 2018, arXiv:1807.10165. [Online]. Available: http://arxiv.org/abs/1807.10165

[30] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, “Convolutional neural networks for medical image analysis: Full training or fine tuning?,” IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1299–1312, May 2016, doi: 10.1109/tmi.2016.2535302.

[31] A. Krizhevsky, I. Sutskever, and G. Hinton, “Imagenet classification with deep convolutional neural networks,” in Proc. NIPS, Lake Tahoe, NV, USA, 2012, pp. 1097–1105.

[32] O. Ghorbanzadeh, T. Blaschke, and K. Gholamnia, “Evaluation of demarcation of water conservancy projects in Guangdong province,” Proc. 24th Interfaces Conf. NIPS, Granada, Spain, 2011, pp. 12–15.

[33] Feilai Xia Water Control Project in Guangdong Province, Guangdong Feilai Xia Water Control Project Management Office, China Water Cult., Hubei, China, 2016, vol. 5.

[34] Guangdong Feilai Xia Water Control Project Management Office, “The management office of Feilai Xia Water Conservancy Project in Guangdong Province. Deepen creation and promote management, concentrate on development,” China Water Cult., vol. 5, p. 41, Oct. 2016.

[35] S. Zhou and Z. Li, “The landscape culture strategy for water system planning in Mountain City: The Case of Yingde City,” Central China Archit., vol. 37, no. 6, pp. 123–126, Jun. 2019.

[36] R. Roni and P. Jia, “An optimal population modeling approach using geographically weighted regression based on high-resolution remote sensing data: A case study in Dhaka City, Bangladesh,” Remote Sens., vol. 12, no. 7, pp. 1184–1199, Apr. 2020.

[37] S. M. M. Karaki, “Invariant feature matching for image registration application based on new dissimilarity of spatial features,” PLoS ONE, vol. 11, no. 3, Mar. 2016, Art. no. e0149710.

[38] L. Lei, “Research on time series classification and change detection method of remote sensing image based on cyclic neural network model,” Ph.D. dissertation, Dept. Inst. Remote Sens. Digit. Earth, Chin. Acad. Sci. Univ., Beijing, China, 2019.

[39] S. Guo, Q. Jun, H. Wang, X. Wang, Y. Wang, and S. Xiang, “Learnable gated convolutional neural network for semantic segmentation in remote sensing images,” Remote Sens., vol. 11, no. 16, pp. 1922–1944, Aug. 2019.

[40] J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement,” 2018, arXiv:1804.02767. [Online]. Available: http://arxiv.org/abs/1804.02767

[41] Y. Yao, S. He, and C. Zhang, “Research on poverty alleviation in agricultural industry from the perspective of precision poverty alleviation: Based on the investigation of tea industry in Yingde City,” Ind. Sci. Forum, vol. 19, no. 1, pp. 95–96, 2020.

[42] X. Chen and Z. Guo, “Analysis of the problems and countermeasures of demarcation of water conservancy projects in Guangdong province,” Guangdong Water Resour. Hydro Power, vol. 8, pp. 100–102, Aug. 2019.

[43] B. Zhang and L. Li, “Analysis and prediction of the driving force of land use spatio-temporal changes in Yingde City,” Tropical Geomech., vol. 39, no. 1, pp. 14–22, Jun. 2018.

PINGHAO WU was born in China, in 1995. She is currently pursuing the Ph.D. degree with the Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou, China. Her research interests include remote sensing interpretation, deep learning, remote sensing applications, and environmental remote sensing.
KAIWEN ZHONG graduated from the Wuhan University of Photogrammetry and Remote Sensing, Wuhan, China. He is currently a Professor with the Guangzhou Institute of Geography, the Director of the Chinese Society of Geographic Information Systems, the Executive Director of the Guangdong Society of Remote Sensing and Geographic Information Systems, the Director of the Guangdong Society of Geography, and the Leader of the Public Laboratory of Guangdong Geospatial Information and the Guangdong Key Laboratory of Remote Sensing and Geographic Information Systems. He has presided over the completion of more than 50 projects, including the National Natural Science Foundation of China, the Guangdong Natural Science Foundation, the Guangdong Province Science and Technology Project, and Guangdong Province and Chinese Academy of Sciences Guangdong Province Science and Technology Project. He has published more than 80 articles in multiple countries, as well as one coauthored monograph, applied for and authorized four software copyrights. His research interests include impermeable water surfaces, urban remote sensing, and agricultural remote sensing. He received five provincial and higher scientific and technological progress awards.

HONGDA HU received the Ph.D. degree from Wuhan University, Wuhan, China. He is currently an Assistant Researcher with the Guangzhou Institute of Geography. His research interests include spatiotemporal statistics and environmental remote sensing. He has received national natural science funds. He has also published science citation index (SCI) papers in prestigious journals, such as Remote Sensing, in the area of impermeable water surfaces. His research interests include road surface models, impermeable water surfaces, and urban remote sensing. He received the Pearl River Talent Award in Guangdong Province.

JIANHUI XU received the Ph.D. degree from Wuhan University, Wuhan, China. He is currently an Associate Professor with the Guangzhou Institute of Geography. He has published numerous SCI papers in prestigious journals, such as Remote Sensing, in the area of impermeable water surfaces. His research interests include road surface models, impermeable water surfaces, and urban remote sensing. He received the Pearl River Talent Award in Guangdong Province.

YUNPENG WANG was born in Shanxi, China, in 1968. He received the B.S. degree in geology from Lanzhou University, Lanzhou, China, in 1990, and the M.S. and Ph.D. degrees in remote sensing and geochemistry from the Guangzhou Institute of Geochemistry, Chinese Academy of Sciences (GIGCAS), Guangzhou, China, in 1992 and 1996, respectively. He has been with GISCAS, since 1996, where he has been a Research Professor, since 2002. He was a Visiting Scholar with Imperial College London, London, U.K., and Brookhaven National Laboratory, Upton, NY, USA, in 2003 and 2004, respectively. He is also currently an Affiliate Professor with the Graduate University, CAS, Beijing, China. He has published more than 100 articles in international journals. His main research interests include remote sensing and applications for the environment, geology, and geochemistry.

Dr. Wang was a recipient of the Second-Rank Award of the National Sciences and Technology Progress of China.

YI ZHAO was born in China, in 1992. She is currently pursuing the Ph.D. degree with the Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou, China. Her research interests include hyperspectral remote sensing and impermeable water surfaces.

***