A Dual-Channel Fully Convolutional Network for Land Cover Classification Using Multifeature Information

Ziwei Liu, Mingchang Wang, Fengyan Wang, Xue Ji, and Zhiguo Meng, *Member, IEEE*

***Abstract*—High-resolution remote sensing images have the advantage of timeliness, and they can display feature information in more detail. Deep learning embodies its unique characteristics in land cover classification, target recognition, and other fields, which can automatically learn the in-depth feature information of images and make accurate classification decisions. However, when deep learning models extract high-dimensional abstract feature information, they often ignore and lose part of the underlying features essential for classification accuracy. This article proposes a dual-channel fully convolutional network (D-FCN), whose two channels, respectively, take image data and low-level features such as color, texture, and shape as the different input data to combine the underlying features with high-dimensional abstract features. To reduce the complexity of the model, we add a large number of skip connections between the model and make full use of the advantages of weight sharing and local connections to connect spatial context information. We used multifeature information as the model input and compared and analyzed the impact of different features on the land cover classification accuracy, and finally obtained the most suitable combination of multifeature information. In addition, we provide a small-scale land cover classification dataset with labels to verify the applicability and transferability of the D-FCN, and use the optimal combination of multifeature information to conduct comparative experiments on the small-scale dataset. The experimental results show that D-FCN has outstanding applicability and transferability. Compared with other state-of-the-art models, D-FCN has a more challenging performance and greatly reduces model complexity.**

***Index Terms*—Deep learning, fully convolutional network, land cover classification, multifeature information, remote sensing.**

I. Introduction

REMOTE sensing images are formed by sensors receiving electromagnetic waves reflected from objects. As an essential means to analyze and predict the changes in the earth's surface, remote sensing images have been widely used in land use surveys, urban planning, precision agriculture, and atmospheric research [1]–[3]. Target recognition and classification is an important research content of remote sensing technology. However, since high-resolution remote sensing images have a large amount of complex feature information, they will also be affected by various external or internal factors during the classification process [4], [5]. In image classification tasks, traditional visual interpretation methods rely too much on the professional's classification experience and professional knowledge, which is time-consuming, inefficient, and more susceptible to subjective awareness [6].

Deep learning technology provides a new research direction in the field of remote sensing image classification. The remote sensing image classification method based on artificial intelligence algorithms promotes the development of image classification technology toward intelligence and automation [7], [8]. Compared with some models with shallow structures, deep learning models can use a large number of complex parameters to fit the feature information of remote sensing images and mine the in-depth features of the original data, thereby improving the accuracy of the classification [9]. Deep learning technology has achieved excellent research results in the field of image processing. Many researchers apply deep learning technology to land cover classification and have constructed many deep learning models with outstanding performance [10]–[12]. Carbonneau *et al.* [13] proposed a CNN-supervised classification (CSC) model for land cover classification of river images. The experimental results show that the CSC method uses deep learning techniques to obtain more accurate classification results compared with other traditional supervised classification methods. Dong *et al.* [14] proposed a novel feature ensemble network using a multiscale feature integration method, using the Resnet-101 backbone to extract feature maps at different scales, and finally obtained higher classification accuracy. Kampffmeyer *et al.* [15] constructed a convolutional neural network model for urban land use classification, replacing missing feature information during the training process and achieving feature fusion even when data modalities are lost.

The increase in the resolution of remote sensing images makes the surface information appear highly detailed [16]. Due to the imaging conditions and the complexity of the composition of surface features, the interclass similarity is increasing, and the intraclass variability and distinguishability of different features

Manuscript received December 10, 2021; revised January 19, 2022 and February 9, 2022; accepted February 18, 2022. Date of publication February 24, 2022; date of current version March 9, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 42171407 and Grant 42077242; in part by the Natural Science Foundation of Jilin Province under Grant 20210101098JC; in part by the National Key R&D Program of China under Grant 2021YFD1500100; in part by the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, MNR under Grant KF-2020-05-024; and in part by the Scientific Research Project of the 13th Five-Year Plan of Jilin Province Education Department under Grant JJKH20200999KJ. *(Corresponding author: Mingchang Wang.)*

The authors are with the College of Geoexploration Science and Technology, Jilin University, Changchun 130026, China (e-mail: lzw21@mails.jlu.edu.cn; wangmc@jlu.edu.cn; wangfy@jlu.edu.cn; jixuesdqd@jlu.edu.cn; mengzg@jlu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3153287

are decreasing [17]–[19]. The classification of ground features based only on spectral radiation characteristics or gray information cannot effectively distinguish different ground features with very similar spectral characteristics [20], [21]. To solve the above problems, more and more researchers are focusing on characteristic information such as spatio-temporal, texture, color, and edge information. Combined with the spectral characteristics of remote sensing images, multifeature fusion technology can obtain an ideal classification result by using an appropriate classifier [22]–[24]. Xiu *et al.* [25] combined principal component analysis and boosting naive Bayesian tree to construct a rotating forest model and used multitemporal remote sensing data for land use classification. Li *et al.* [26] proposed a three-channel CNN-based model, which extracts spatial features and elevation features from the hyperspectral and DSM data generated by LiDAR, respectively. The extracted features and image data are input into the three-channel model, and classification research is carried out through the SVM and the extreme learning machine classifier. Tarabalka *et al.* [27] combined the spatial and spectral characteristics of remote sensing images and used SVM to classify hyperspectral data. Compared with the traditional classification method that only uses spectral information, this method is more suitable for situations with large spots and similar spectral characteristics after adding spatial features.

The deep learning model can automatically extract and abstract deep feature information. However, as the depth of the model increases, it is often easier to ignore the underlying feature information, resulting in too low classification accuracy, and the classification accuracy cannot reach the ideal state. To remedy the above issues, this article proposes a dual-channel fully convolutional network (D-FCN) to combine high-dimensional features and low-level feature information. D-FCN uses a dual-encoder structure to improve the feature extraction capability of the network, uses multiple skip connections to enhance spatial context information and network performance, and reduces the complexity of the model. In addition, to verify the applicability and transferability of the D-FCN proposed in this article, we constructed a small-scale dataset with labels for land cover classification. Compared with other state-of-the-art models, the D-FCN proposed in this article has outstanding performance. The main contributions of this article are as follows.

1) The D-FCN uses a dual-encoder structure and multiple skip connections to reduce the complexity of the model, which makes full use of the advantages of weight sharing and local connection to improve the generalization ability and robustness of the model.
2) The low-level feature information such as color, texture, and shape is combined with deep semantic features to assist the classification of high-resolution remote sensing images in this article. The classification accuracy is effectively improved by combining the low-level features with the in-depth feature information extracted by the deep learning model.
3) We provide a small-scale land cover classification dataset with labels to verify the applicability and transferability of the D-FCN. Compared with other state-of-the-art network models, the D-FCN proposed in this article has a more challenging performance.

The rest of this article is organized as follows. Section II explains some details of the method used in this article. Section III introduces the experimental dataset, experimental settings, and experimental results. In Section IV, we discuss the network performance and the impact of multifeature information on classification. Finally, Section V concludes this article.

II. Methodology

A. Structure of the D-FCN

In recent years, semantic segmentation models represented by U-Net [28], SegNet [29], and DeepLab [30] have been frequently used in the tasks of land cover classification. Most of these models take the framework of encoder–decoder. The role of the encoder is to extract the deep-level feature information of the input data through the convolution and pooling layer. The function of the decoder is to restore the extracted feature information layer by layer to the same size as the input data and then realize the pixel-level classification of remote sensing images. The experimental model in this article uses encoder–decoder architecture to construct a D-FCN for land cover classification.

The VGG was proposed by the Visual Geometry Group from Oxford University. The network includes six configurations: A, A-LRN, B, C, D, and E. Among them, VGG-16 (configuration D) is the most commonly used framework of the encoder. The characteristic of the VGG-16 is that a convolutional block composed of multiple small-scale convolutional kernels (3×3) is used to replace the larger scale convolutional kernels (such as 11×11, 7×7, 5×5). Multilayer nonlinear processing technology can increase the depth of the network while ensuring the learning ability of the model within a given range of receptive fields.

Due to the unique advantages of VGG-16, many excellent semantic segmentation models take the VGG-16 or the optimized VGG network as the encoder to extract the deep feature information of the input data [31], [32]. In this experiment, the first 13 layers of VGG-16 were selected as the encoder structure. However, because VGG-16 has a deeper network structure, it includes more parameters and occupies more computing resources. In building the model, we refer to the idea of forming the AlexNet network, dividing the encoder into two parts equally, and input sample data from two channels simultaneously to speed up the model's training speed.

Comprehensively considering the advantages of the VGG-16 and the AlexNet, the model encoder proposed in this article is constructed using a dual-channel input method. The two channels of the model take the image and low-level feature as input data, respectively. The encoder consists of five symmetrical convolution blocks. Each convolution block contains a different number of convolutional kernels whose spatial size and number are 3×3 and 2^{i+4} (i is the sequence of the convolution block), respectively. We added the batch normalization layer after the convolutional layer to reduce overfitting. Besides, the max-pooling method is used for dimensionality reduction between each convolutional block. The structure of the model is shown in Fig. 1. The function of the decoder is to enlarge

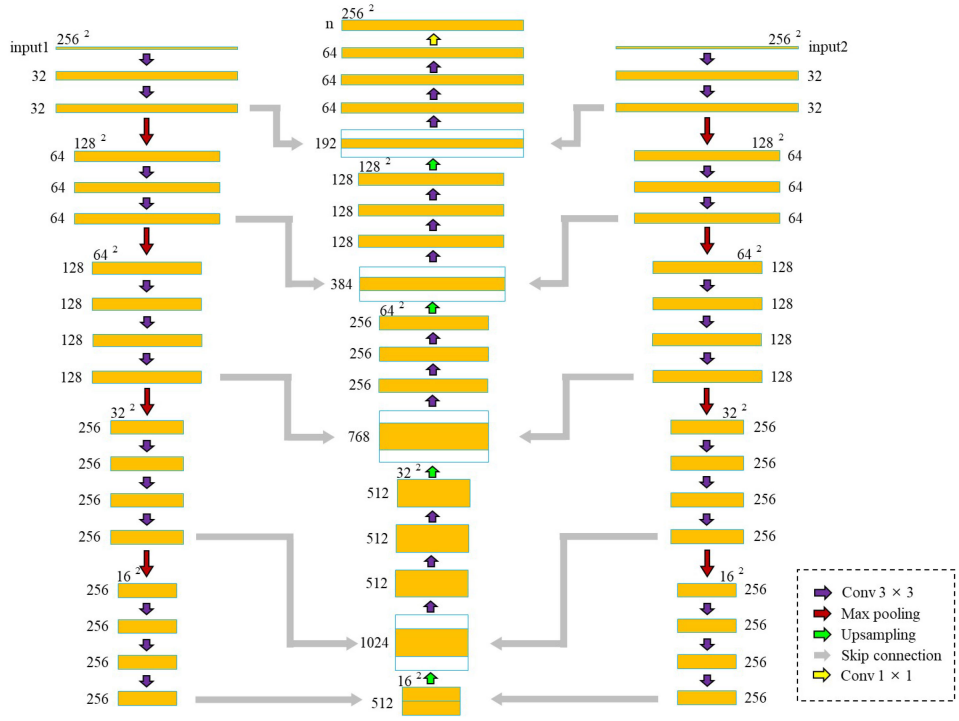

Fig. 1. Network architecture of the D-FCN.

the feature map data extracted by the encoder layer by layer until it is restored to the same size as the input data. A suitable classifier is selected to determine the data category. The decoder structure in this experiment corresponds to the encoder, using four expanded convolution blocks to restore the feature map data layer by layer. Each convolution block is composed of an upsampling layer and several convolution layers with a spatial size of 3 × 3. The activation function of the last layer uses the softmax as the classifier, whose spatial size of the convolution kernel is 1 × 1, the number of convolution kernels is *n* (*n* is the number of sample categories).

B. Extraction of Multifeature Information

With the improvement of spatial resolution, the feature information provided by remote sensing images is becoming increasingly abundant, and complex features cannot be distinguished effectively based on spectral information alone [33], [34]. The classification of high-resolution remote sensing images can be assisted by introducing other feature information [35]. Single feature information can only describe part of the attributes of the image, and it does not represent the image information comprehensively and cannot provide enough distinguishing information [36]. The feature information obtained by different algorithms is fused, and the fused feature data solves the problem of the one-sided details of the single feature to a certain extent [37]. The information described by various features is aggregated together to form a more comprehensive and rich multifeature data to assist in high-resolution image classification, thereby improving image classification accuracy [38].

1) Texture Feature: The texture feature is a quantitative description of the texture properties of an image. It is a visual feature that does not depend on color or brightness information to reflect the homogeneity of the image, containing important arrangement information of the image's surface structure and its relationship with the surrounding environment [39]. The texture information of the image cannot be expressed in words or language. Therefore, it is necessary to extract the information data that can describe the texture characteristics from the two aspects of the texture structure and statistics of the image [40]. The gray-level co-occurrence matrix (GLCM) is a matrix function that can perform statistical investigations on all image pixels according to the distance and angle between each pixel in the image. GLCM extracts texture features in a conditional probability manner to describe the spatial correlation of image pixels on gray levels. This article uses the three characteristic values of energy, contrast, and homogeneity extracted based on GLCM to describe the texture features of the image

$$Ene = \sum_i \sum_j g(i,j)^2 \tag{1}$$

$$Con = \sum_i \sum_j (i-j)^2 g(i,j) \tag{2}$$

$$Hom = \sum_i \sum_j \frac{g(i,j)}{1+(i-j)^2}. \tag{3}$$

2) Shape Feature: The shape feature is a high-level visual feature in the image. Under certain extreme conditions where other ground features do not exist, the object's category can still be identified through the shape and contour of the object [41].

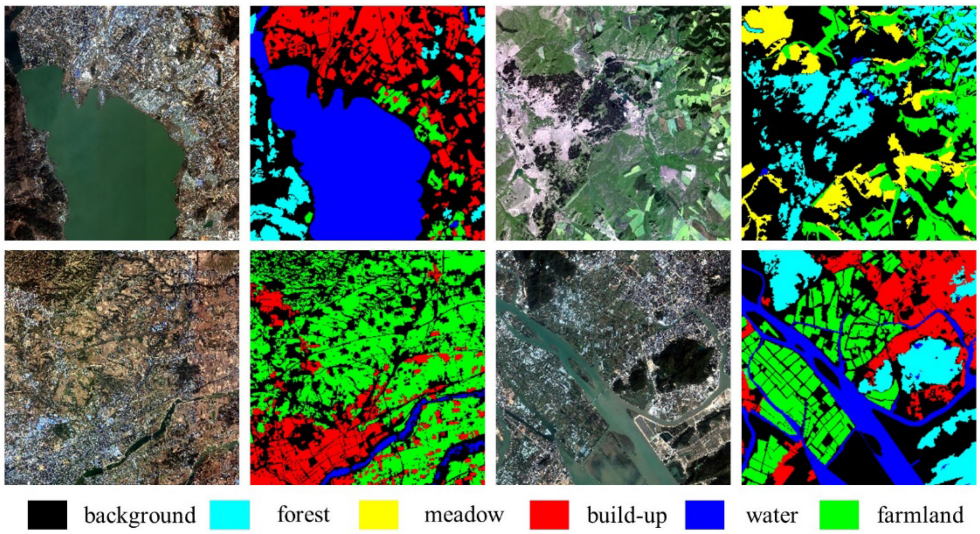

Fig. 2. Image and label data of the GID.

For image data with complex features and a large amount of data, the edge detection method is used to describe the shape features of the object, which can retain detailed local and global information in space [42]. For some local noise-sensitive areas, the unevenness of the shape contour can be used to describe the shape characteristics of the object. In this article, the Canny operator with outstanding extraction performance is used to extract the edge information of the image as the shape feature. This method is a type of contour feature extraction based on shape.

3) Color Feature: The color feature is the most commonly used feature information in the field of image retrieval. Color can intuitively reflect the connection and difference between different objects and is less affected by image size and shooting angle, possessing a simple extraction method and strong robustness [43]. Color features are based on feature values or feature vectors extracted from different color spaces, which are not affected by the rotation and translation of the image. However, they often lack the spatial distribution information of the image itself. In this article, the HSV color space distribution is selected as the color characteristics of the image. While preserving the image space information, the HSV color space can better reflect the human eye's ability to perceive and discriminate color.

III. Experiment

To verify the performance of the D-FCN proposed in this article, in this section, we performed classification experiments on the Gaofen Image Dataset (GID) and the Gaofen Changchun Dataset (GCD) that we have marked manually. And the result obtained by D-FCN is compared with other state-of-the-art models. In Section III-A, we introduced the details of the two datasets. The hyperparameter settings of the experiment are described in Section III-B. In Section III-C, we presented and analyzed the two experiments in detail.

A. Dataset

1) Gaofen Image Dataset: The high-resolution dataset, the GID, produced by Wuhan University, is selected for the experiment to verify the effectiveness of the model proposed in this article. The dataset is derived from the Gaofen-2 remote sensing satellite, including 150 remote sensing images, and the total coverage area exceeds 50 000 km^2. Each image has corresponding label data. After preprocessing, each image has a spatial resolution of 1 m and a spatial size of 7200 $\times$ 6800, including six categories of build-up, farmland, forest, meadow, water, and background. The image and label data are shown in Fig. 2.

2) Gaofen Changchun Dataset: The GF-2 satellite is equipped with two cameras used to shoot multispectral images and panchromatic images. We preprocessed the original data such as radiometric calibration, atmospheric correction, geometric correction, image fusion, and cropping, and finally obtained a small-scale land cover classification dataset with labels, the GCD, covering an area of 49 km^2. Each image includes five categories of water, build-up, road, vegetation, and background. The image and label data of the GCD are shown in Fig. 3.

B. Experimental Setup

For the GID, we selected ten images with better quality from 150 images as experimental data. Eight images were used as training images, and two were used as testing images. After sample selection and data enhancement, we obtained a new dataset including 30 000 images with a spatial size of 256 $\times$ 256. The dataset is divided into a training set including 22 500 images and a test set of 7500 images using a random division method.

For the GCD, after reprocessing the original GF-2 image, we obtained four image data with a size of 3500 $\times$ 3500. Among the four images, three are used for training, and the other is used

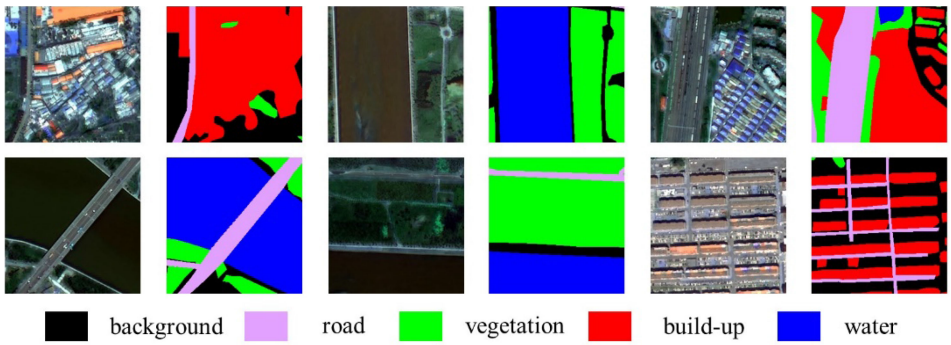

Fig. 3. Image and label data of the GCD.

for testing. The original dataset was enhanced and cropped to obtain a new dataset with a spatial size of 256 × 256, including 9600 training images and 2400 testing images. We added multilayer batch normalization to the D-FCN model and used a more significant initial learning rate of 0.01 to improve the network convergence speed. According to the experimental conditions, stochastic gradient descent is taken as the optimizer with a batch size of 4. The method of adaptively adjusting the learning rate is used in the experiment to enhance the efficiency of model training. According to the situation of model training, the learning rate is automatically reduced to improve the convergence speed of the model. Specifically, when the test loss value of the model does not decrease for four epochs, the learning rate is multiplied by a factor of 0.5. To prevent the occurrence of overtraining, that is, the phenomenon of fluctuations in classification accuracy due to overtraining, the strategy of terminating training early is used. Specifically, when the test loss value of the model does not decrease for 20 epochs, the training is stopped. TensorFlow is taken as the backend, and Keras is used to implement all network models trained on NVIDIA GTX1660 GPU.

C. Results

In this section, we carried out two parts of the experiment. In the first part of the experiment, we verified the performance of D-FCN on the GID. Specifically, we compared and analyzed the impact of the different features on classification accuracy and selected the most suitable combination of feature information for the D-FCN. In the second part of the experiment, we verified the applicability and transferability of the D-FCN. We provide a small-scale land cover classification dataset named GCD and compare the results using the D-FCN under the most optimal combination with other state-of-the-art models such as U-Net and SegNet.

1) Classification Results Using GID: High-resolution remote sensing image data contains a wealth of feature information. The texture, shape, color, and other feature information extracted from it can represent the feature information of the image at different levels. Using different types of feature information to assist the classification of high-resolution remote sensing images will have different results. To analyze the impact of various features on the classification results, we use one and multiple feature fusion to perform classification experiments, respectively. The most suitable feature combination method is obtained through the qualitative and quantitative comparison of the classification results. All experiments were performed under the same classifier and the same sample dataset and experimental parameters were used to perform classification experiments on different feature combinations. The classification accuracy is displayed in Table I.

TABLE I
CLASSIFICATION ACCURACY FOR DIFFERENT FEATURE COMBINATIONS

Feature Information	OA	Kappa
Texture feature	84.21%	0.7785
Color feature	77.88%	0.6783
Shape feature	80.81%	0.7244
Texture and color feature	79.19%	0.7018
Texture and shape feature	85.96%	0.8030
Color and shape feature	80.18%	0.7128
Texture, color, and shape feature	81.15%	0.7345

As displayed in Table I, the classification accuracy of the three feature combinations is lower than that of the texture and shape feature combinations. The OA of the classification results obtained using only the shape features can reach 80.81%, but the OA decreases after adding the color features. The classification accuracy of the texture and shape features combination is the best among all feature combinations. The OA can reach 85.96%, and the Kappa coefficient can reach 0.8030. And the classification result obtained by using the texture feature is second only to the classification result of the texture and shape feature combination, whose OA and Kappa coefficient are 84.21% and 0.7785, respectively. The texture features can effectively express the texture information of image spatial scale and spatial structure, which can enhance the heterogeneity between different categories, especially for features with obvious texture information, such as farmland and roads. Texture features help to enhance the classification performance of some hybrid image elements. For example, certain grassland areas and background can be accurately distinguished by extracting texture features. The contour-based shape features can accurately the boundary information of the features and represent the shape contour of the features completely. In this article, we use Canny to extract the edge

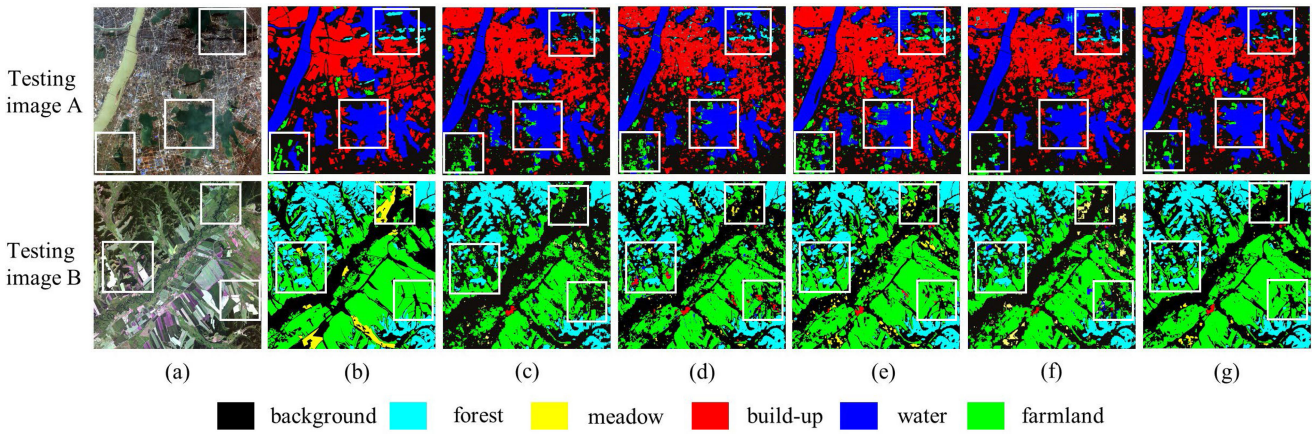

Fig. 4. Visual comparison of different models on the GID. (a) Satellite image. (b) Ground truth. (c) FCN. (d) SegNet. (e) U-Net. (f) D-LinkNet. (g) D-FCN. The white boxes in the figure represent area (1), area (2), and area (3) from top to bottom.

information of images, which effectively describes the boundary information of different features. In addition, the combination of texture and shape features can improve the classification performance of features with texture and boundary information. Although HSV color space is more suitable for image processing and analysis, data redundancy also occurs when used with RGB data, so our incorporation of color features based on shape and texture features can lead to degraded classification performance. The experimental results show that the texture features extracted based on the GLCM can effectively improve the classification accuracy. Utilizing a single feature, two feature combinations, and three feature combinations have different effects on the classification results. The number of feature combinations is not a decisive factor affecting classification accuracy.

We quantitatively compared and analyzed the effects of texture, color, and shape on the classification results in Table I. Among them, the method of texture and shape feature combination has the best classification effect. To verify the effectiveness of this method, FCN [44], SegNet [29], U-Net [28], and D-LinkNet [45] were used for comparative experiments.

The visual comparison of the D-FCN proposed and the other models is shown in Fig. 4.

For the testing image A in Fig. 4(the first row), when the U-Net is used for classification, in area (1), the two categories of forest and farmland show obvious misclassification in the classification results. For the FCN, there is also some misclassification of background categories and forest categories. In area (2), there is also a misclassification of the water category for all models. The part of the background is divided into forest categories in area (3). For SegNet, the misclassification phenomenon has been improved to a certain extent, but there are a large number of small fragments in the result. Many scattered meadow categories appear in the build-up category. The visual results obtained using the D-LinkNet model outperformed the results of U-Net, FCN, and SegNet. However, when the D-FCN is used for classification, the classification result is almost the same as the ground truth, and there is no obvious misclassification. At the same time, the number of broken image spots is reduced, and the classification effect is better than other models. For the testing image B in Fig. 4 (the second row), in area (1), all models have poor extraction effects for the meadow category. However, they cannot effectively distinguish the boundary of the meadow and background categories. In areas (2) and (3), using FCN, SegNet, U-Net, and D-LinkNet for classification, some water and build-up categories appear in the result, which does not exist in the ground truth. The misclassification phenomenon is more obvious compared with the D-FCN. However, in the classification results obtained by the D-FCN, the misclassification phenomenon has been effectively improved. Considering the classification results of the two testing images, the overall effect of the results using the D-FCN is significantly better than the other models, reducing the number of fragments and improving the misclassification to a certain extent.

It can be seen from Table II that the classification effect using the D-FCN is the best, with the OA of 85.96% and the Kappa coefficient of 0.8030. Compared with U-Net and SegNet, the OA has increased by 1.91% and 3.73%, and the Kappa coefficient has increased by 0.0212 and 0.0512. Compared with FCN and D-LinkNet, the OA has increased by 5.93% and 4.1%, and the Kappa coefficient has increased by 0.0856 and 0.0572. In addition, when using the D-FCN method for classification, the precisions of the five categories of water, build-up, farmland, forest, and meadow are all higher than the other models. Although all models have not reached the ideal situation for the extraction accuracy of the meadow category, the D-FCN uses the dual-channel structure to enhance the network feature extraction ability and effectively improve the precision of the meadow category.

Comprehensive analysis of the qualitative and quantitative comparisons of the different models, the D-FCN proposed in this article outperforms the other models. It improves the phenomenon of fragmented patterns and gaps, eliminates more obvious segmentation marks, and effectively enhances the extraction accuracy of various categories.

2) Classification Results Using the GCD: In this section, the experiment uses FCN, U-Net, SegNet, and D-LinkNet models as the comparative experiment to verify the effectiveness and transferability of the D-FCN. The classification vision results

TABLE II
QUANTITATIVE COMPARISON OF DIFFERENT MODELS USING GID

Method	Background	Water	Build-up	Farmland	Forest	Meadow	OA	Kappa
FCN	73.15%	96.27%	81.49%	83.84%	94.69%	27.78%	80.03%	0.7174
SegNet	77.47%	93.61%	79.92%	87.72%	88.48%	36.44%	82.23%	0.7518
U-Net	83.23%	95.27%	78.54%	82.38%	92.32%	36.71%	84.05%	0.7818
D-LinkNet	76.20%	93.19%	83.98%	85.97%	92.64%	41.31%	81.86%	0.7458
D-FCN	80.21%	97.18%	87.01%	90.01%	94.05%	68.27%	85.96%	0.8030

The precision of each category, OA, and Kappa coefficient are given in the table.

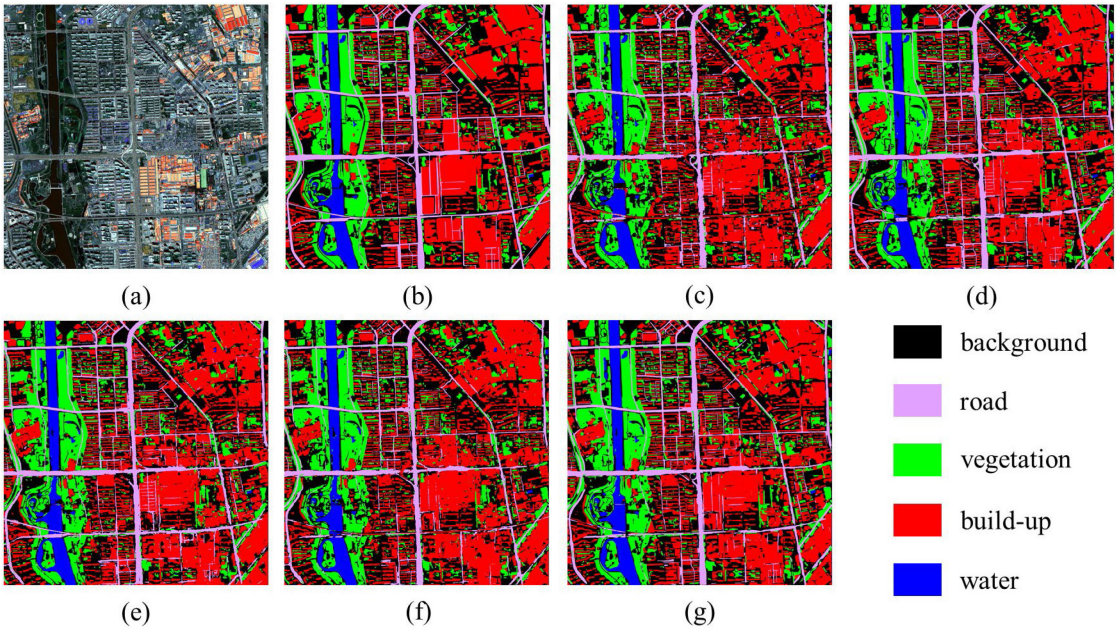

Fig. 5. Visual comparison of different models on the GCD. (a) Satellite image. (b) Ground truth. (c) FCN. (d) SegNet. (e) U-Net. (f) D-LinkNet. (g) D-FCN.

TABLE III
QUANTITATIVE COMPARISON OF DIFFERENT MODELS USING THE GCD

Method	Background	Build-up	Water	Road	Vegetable	OA	Kappa
FCN	77.80%	81.88%	96.74%	79.76%	85.25%	81.29%	0.7392
SegNet	82.11%	84.81%	95.28%	80.45%	86.77%	84.14%	0.7802
U-Net	82.91%	83.71%	94.86%	78.29%	86.57%	83.76%	0.7760
D-LinkNet	80.17%	83.03%	94.36%	83.46%	83.65%	82.62%	0.7583
D-FCN	87.42%	89.06%	94.08%	85.55%	86.90%	87.96%	0.8336

The precision of each category, OA, and Kappa coefficient are given in the table.

are shown in Fig. 5. For the category of build-up, the result of the D-FCN shows an outstanding visual outcome, obtaining a continuous and complete region. However, the results of the build-up extracted by FCN, U-Net SegNet, and D-LinkNet have more hollow areas. For the road category, the D-FCN obtains a more consistent result with the label data, and the accuracy is better than the other models. For other categories, all the models have no apparent difference in visual comparison.

Table III lists the precision of each category, OA, and Kappa coefficient obtained using different classification methods. The accuracy of the classification results is evaluated from the perspective of quantitative analysis. It can be seen from Table III that, on the whole, the OA and Kappa coefficient of the D-FCN are higher than other models. Compared with other four models, the OA of D-FCN increases by 6.67%, 3.82%, 4.2%, and 5.34%, and the Kappa coefficient increases by 0.0944, 0.0534, 0.0576, and 0.0753. From a local point of view, the extraction effect of the D-FCN on the three categories of background, build-up, and road is significantly better than the other models, and the precision is 87.42%, 89.06%, and 85.55%, respectively. For the

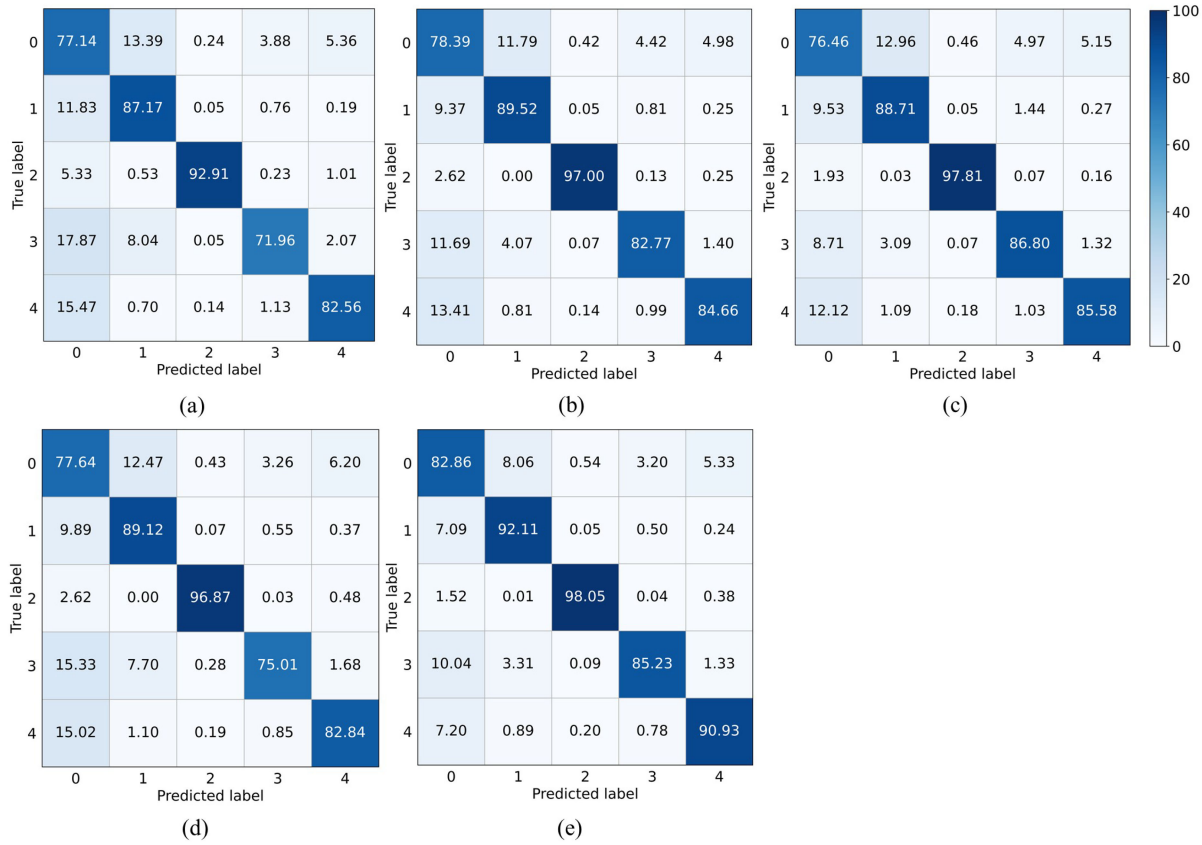

Fig. 6. Confusion matrices of three models. (a) FCN. (b) SegNet. (c) U-Net. (d) D-LinkNet. (e) D-FCN.

two categories of water and vegetation, the precision of the all models is not much different.

Combining the evaluation metrics of precision and recall can more objectively evaluate the effectiveness of the result. After a specific calculation of the confusion matrix, the diagonal elements can be expressed as the recall of each category. The confusion matrices (%) of the all models are shown in Fig. 6. Comprehensive analysis of Table III and Fig. 6, the D-FCN proposed in this article considers texture and shape features to improve the build-up extraction effect. The precision and recall reached 89.06% and 92.11%, respectively. The all models have the best effect on water extraction, with the precision of over 94% and recall of over 92%. Above all, when the D-FCN proposed in this article was applied to land cover classification in Changchun City by combining the underlying feature information to assist image classification, it achieved better classification results and effectively improved classification accuracy. The experimental results show that the D-FCN using texture and shape features to assist classification has certain applicability and transferability.

IV. Discussion

In Section III, we performed classification experiments on the GID. For all categories, the accuracy of the meadow and background categories is slightly lower, especially the meadow category. The background category contains complex land information, including redundant data such as clutter, noise, clouds, and other unrecognizable or too few samples except build-up, farmland, forest, meadow, and water. As shown in Fig. 7, the two categories of background and meadow show strong similarities in spectral characteristic. The meadow in the image is difficult to distinguish from part of the background area, and the separability is poor, resulting in low classification accuracy of the meadow category. The D-FCN proposed in this article uses texture and shape features to assist image classification, which effectively improves the classification accuracy, especially for the grass category. The improvement effect is the most obvious. Experimental results show that the D-FCN model proposed in this article uses a dual-channel structure, which improves the feature extraction ability of the encoder and enhances the generalization ability and robustness of the model. In addition, the use of features such as texture and shape of the image can effectively improve classification accuracy, which plays an essential role in the classification task of complex features.

The boundary of the build-up extracted by D-FCN is more consistent with the label data, and there is no obvious misclassification. However, in the three local visual results in Fig. 8, the background category is misclassified from other categories in some areas. We divide the other areas, except for buildings, water bodies, roads, and vegetation, into the background category when defining the feature category. In these areas, some types of ground objects cannot meet the requirements of model training due to too few pixels. The spectral characteristics of these ground objects are similar to the other four categories in some areas, making the model unable to distinguish the background from some nonbackground areas effectively.

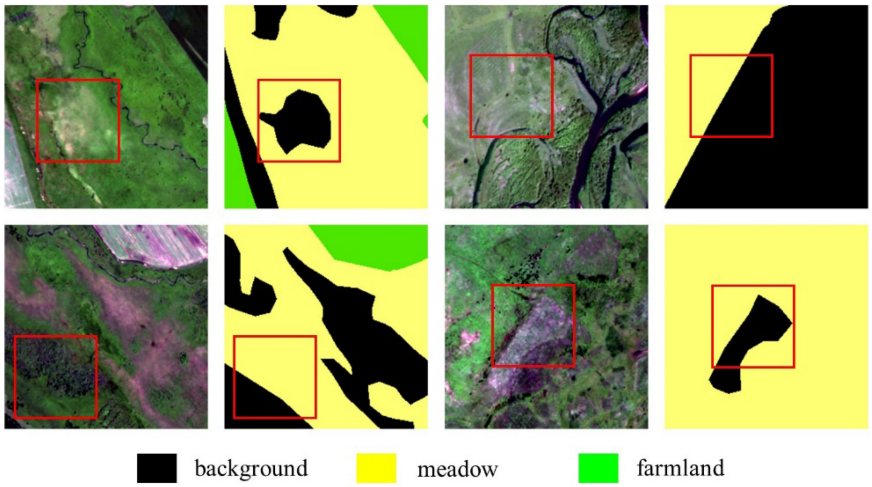

Fig. 7. Confusing areas of meadow and background categories.

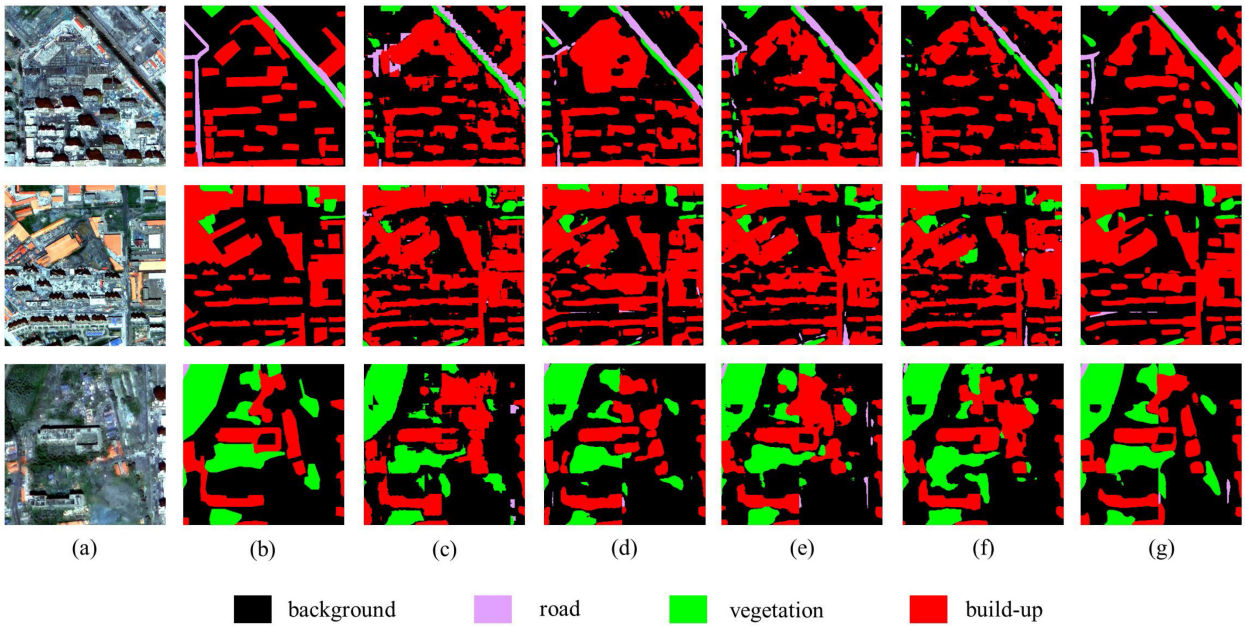

Fig. 8. Local visual comparison of classification results. (a) Satellite image. (b) Ground truth. (c) FCN. (d) SegNet. (e) U-Net. (f) D-LinkNet. (g) D-FCN.

The D-FCN proposed in this article uses a dual-encoder structure to enhance the feature extraction capability of the model. The dual-encoder structure consists of two symmetric encoders that have the same organization but with different weights. Compared with other models, D-FCN uses two symmetric encoders to extract the feature maps of the image and the underlying features separately to enhance the feature characterization capability of the network. Since both symmetry encoders have the same depth, the feature map data extracted by the symmetry encoder have the same dimensional information and similar depth semantics. The skip connection can use these feature maps to connect the spatial context information. Symmetric encoders can effectively improve the feature extraction capability of the network and avoid missing information due to the difference in depth. In addition, the high spatial resolution images contain rich feature information, and the dual-encoder structure can utilize them more effectively and extract deep semantic information. To avoid wasting computational power, each encoder parameter of D-FCN is only half of that of the U-Net model, which speeds up the convergence of the model and improves the robustness and generalization ability of the model by using more jump connections to link spatial context information.

We conducted classification experiments on two datasets to verify the performance and transferability of the D-FCN. On the GID, we use different feature information as input, use the dual-channel structure to improve the feature extraction ability of the model and verify the impact of the underlying feature information such as color, texture, and shape on image classification. The experimental results show that we can significantly improve the classification accuracy by combining the underlying features with the in-depth features. However, the transferability of the model is also an essential factor in determining the

performance of the deep learning model, so we use the D-FCN on the GCD for classification experiments. Compared with other state-of-the-art models, the D-FCN achieves higher classification results, proving that the D-FCN has good transferability.

V. Conclusion

Remote sensing image classification is a crucial technology in remote sensing data extraction and analysis and an important research direction in remote sensing image processing. The accuracy of image classification can be effectively improved by taking advantage of the high spatial resolution of remote sensing images and comprehensively using multiple feature information such as spectrum, texture, shape, and color to assist image classification. This article proposes a dual-channel fully convolutional neural network (D-FCN) model to realize the multisource input of the model. The two input channels of image and multifeature information extract feature maps simultaneously, which can make full use of the characteristics of weight sharing and local connection of the convolutional layer to accelerate the convergence speed of the model. The public dataset GID and the small-scale land cover classification dataset GCD are used to verify the performance of the D-FCN and the influence of multifeature information on the classification results. Experimental results show that different types of single feature information have different effects on image classification results. Among single feature information, the texture feature has the best effect on improving classification accuracy, and the combination of various features has unique effects on image classification accuracy. The number of fused features is not a decisive factor in improving classification accuracy. The classification result obtained by the combination of texture and shape features is the best. The D-FCN proposed in this article uses a dual-channel structure to take advantage of high-resolution remote sensing images and makes full use of the feature information provided by the image to improve the classification accuracy effectively.

Three-dimensional (3-D) data such as point cloud, DEM, and other data can provide helpful information, but the feature information selected in this article only considers two-dimensional data. In the next stage of work, we will try to introduce 3-D data to improve classification accuracy. In addition, the large number of parameters in the model is still the main problem that limits the interpretability of deep learning. In future work, we will try to use more intelligent and more explanatory models to solve the characteristic "black box" of deep learning.

References

[1] X. Niu and Y. F. Ban, "Multi-temporal RADARSAT-2 polarimetric SAR data for urban land-cover classification using an object-based support vector machine and a rule-based approach," *Int. J. Remote Sens.*, vol. 34, no. 1, pp. 1–26, May 2013.

[2] Q. H. Weng, "Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends," *Remote Sens. Environ.*, vol. 117, pp. 34–49, Feb. 2012.

[3] M. C. Wang, Y. L. Song, F. Y. Wang, and Z. G. Meng, "Boundary extraction of urban built-up area based on luminance value correction of NTL image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7466–7477, Jul. 2021.

[4] L. Ma, M. C. Li, X. X. Ma, L. Cheng, P. J. Du, and Y. X. Liu, "A review of supervised object-based land-cover image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 277–293, Aug. 2017.

[5] X. Lv, D. Ming, Y. Chen, and M. Wang, "Very high resolution remote sensing image classification with seeds-CNN and scale effect analysis for superpixel CNN classification," *Int. J. Remote Sens.*, vol. 40, no. 2, pp. 506–531, Sep. 2018.

[6] B. Huang, B. Zhao, and Y. M. Song, "Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery," *Remote Sens. Environ.*, vol. 214, pp. 73–86, Sep. 2018.

[7] M. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanesh, and Y. Zhang, "Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery," *Remote Sens.*, vol. 10, no. 7, pp. 1119–1139, Jul. 2018.

[8] P. Helber, B. Bischke, A. Dengel, and D. Borth, "EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2217–2226, Jul. 2019.

[9] C. Lin *et al.*, "Deep learning network intensification for preventing noisy-labeled samples for remote sensing classification," *Remote Sens.*, vol. 13, no. 9, pp. 1689–1706, May 2021.

[10] S. Wang, W. Chen, S. M. Xie, G. Azzari, and D. B. Lobell, "Weakly supervised deep learning for segmentation of remote sensing imagery," *Remote Sens.*, vol. 12, no. 2, pp. 207–231, Jan. 2020.

[11] S. J. Liu, H. W. Luo, and Q. Shi, "Active ensemble deep learning for polarimetric synthetic aperture radar image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 9, pp. 1580–1584, Sep. 2021.

[12] K. B. Larson and A. R. Tuor, "Deep learning classification of cheatgrass invasion in the western United States using biophysical and remote sensing data," *Remote Sens.*, vol. 13, no. 7, pp. 1246–1266, Apr. 2021.

[13] P. E. Carbonneau *et al.*, "Adopting deep learning methods for airborne RGB fluvial scene classification," *Remote Sens. Environ.*, vol. 251, Dec. 2020, Art. no. 112107.

[14] S. Dong, Y. Zhuang, Z. X. Yang, L. Pang, H. Chen, and T. Long, "Land cover classification from VHR optical remote sensing images by feature ensemble deep learning network," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1396–1400, Aug. 2020.

[15] M. Kampffmeyer, A. B. Salberg, and R. Jenssen, "Urban land cover classification with missing data modalities using deep convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 6, pp. 1758–1768, Jun. 2018.

[16] X. J. Guo, C. C. Zhang, W. R. Luo, J. Yang, and M. Yang, "Urban impervious surface extraction based on multi-features and random forest," *IEEE Access*, vol. 8, pp. 226609–226623, 2020.

[17] R. Y. Fan, R. Y. Feng, L. Z. Wang, J. N. Yan, and X. H. Zhang, "Semi-MCNN: A semisupervised multi-CNN ensemble learning method for urban land cover classification using submeter HRRS images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4973–4987, Aug. 2020.

[18] A. Sharma, X. Liu, and X. Yang, "Land cover classification from multi-temporal, multi-spectral remotely sensed imagery using patch-based recurrent neural networks," *Neural Netw.*, vol. 105, pp. 346–355, Sep. 2018.

[19] Y. Y. Chen, D. P. Ming, and X. W. Lv, "Superpixel based land cover classification of VHR satellite image combining multi-scale CNN and scale parameter estimation," *Earth Sci. Informat.*, vol. 12, no. 3, pp. 341–363, Sep. 2019.

[20] M. Carranza-Garcia, J. Garcia-Gutierrez, and J. C. Riquelme, "A framework for evaluating land use and land cover classification using convolutional neural networks," *Remote Sens.*, vol. 11, no. 3, pp. 274–296, Feb. 2019.

[21] T. Yamakita, F. Sodeyama, N. Whanpetch, K. Watanabe, and M. Nakaoka, "Application of deep learning techniques for determining the spatial extent and classification of seagrass beds, Trang, Thailand," *Botanica Marina*, vol. 62, no. 4, pp. 291–307, Aug. 2019.

[22] C. Zhang *et al.*, "Joint deep learning for land cover and land use classification," *Remote Sens. Environ.*, vol. 221, pp. 173–187, Feb. 2019.

[23] Y. Zhu, C. Geiss, E. M. So, and Y. Jin, "Multitemporal relearning with convolutional LSTM models for land use classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3251–3265, Feb. 2021.

[24] Y. M. Luo, Y. Ouyang, R. C. Zhang, and H. M. Feng, "Multi-feature joint sparse model for the classification of mangrove remote sensing images," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 6, pp. 177–185, Jun. 2017.

[25] Y. C. Xiu, W. B. Liu, and W. J. Yang, "An improved rotation forest for multi-feature remote-sensing imagery classification," *Remote Sens.*, vol. 9, no. 11, pp. 1205–1221, Nov. 2017.
[26] H. Li, P. Ghamisi, U. Soergel, and X. X. Zhu, "Hyperspectral and LiDAR fusion using deep three-stream convolutional neural networks," *Remote Sens.*, vol. 10, no. 10, pp. 1649–1668, Oct. 2018.
[27] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Jul. 2009.
[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, Munich, Germany, 2015, pp. 234–241.
[29] 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.
[30] 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," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.
[31] Y. H. Liu, J. Yao, X. H. Lu, M. H. Xia, X. B. Wang, and Y. Liu, "RoadNet: Learning to comprehensively analyze road networks in complex urban scenes from high-resolution remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2043–2056, Apr. 2019.
[32] P. Zhang, Y. Ke, Z. Zhang, M. Wang, P. Li, and S. Zhang, "Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery," *Sensors*, vol. 18, no. 11, pp. 3717–3737, Nov. 2018.
[33] L. F. Zhang, L. P. Zhang, D. C. Tao, and X. Huang, "A modified stochastic neighbor embedding for multi-feature dimension reduction of remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 83, pp. 30–39, Sep. 2013.
[34] F. Karsli, M. Dihkan, H. Acar, and A. Ozturk, "Automatic building extraction from very high-resolution image and LiDAR data with SVM algorithm," *Arabian J. Geosci.*, vol. 9, no. 14, pp. 635–646, Sep. 2016.
[35] T. Bai, D. R. Li, K. M. Sun, Y. P. Chen, and W. Z. Li, "Cloud detection for high-resolution satellite imagery using machine learning and multi-feature fusion," *Remote Sens.*, vol. 8, no. 9, pp. 715–735, Sep. 2016.
[36] Z. W. Liu, M. C. Wang, F. Y. Wang, and X. Ji, "A residual attention and local context-aware network for road extraction from high-resolution remote sensing imagery," *Remote Sens.*, vol. 13, no. 24, Dec. 2021, Art. no. 4958.
[37] M. C. Wang, H. M. Zhang, W. W. Sun, S. Li, F. Y. Wang, and G. D. Yang, "A coarse-to-fine deep learning based land use change detection method for high-resolution remote sensing images," *Remote Sens.*, vol. 12, no. 12, pp. 1933–1953, Jun. 2020.
[38] M. C. Wang, X. Y. Zhang, X. F. Niu, F. Y. Wang, and X. Q. Zhang, "Scene classification of high-resolution remotely sensed image based on ResNet," *J. Geovisualization Spatial Anal.*, vol. 3, no. 2, pp. 16–24, Dec. 2019.
[39] V. Risojevic and Z. Babic, "Fusion of global and local descriptors for remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 836–840, Jul. 2013.
[40] Q. Q. Zhu, Y. F. Zhong, B. Zhao, G. S. Xia, and L. P. Zhang, "Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 747–751, Jun. 2016.
[41] H. F. Liu, M. H. Yang, J. Chen, J. L. Hou, and M. Deng, "Line-constrained shape feature for building change detection in VHR remote sensing imagery," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 10, pp. 410–428, Oct. 2018.
[42] H. M. Zhang *et al.*, "A novel squeeze-and-excitation W-Net for 2D and 3D building change detection with multi-source and multi-feature remote sensing data," *Remote Sens.*, vol. 13, no. 3, pp. 440–465, Feb. 2021.
[43] H. Yu, W. Yang, G. S. Xia, and G. Liu, "A color-texture-structure descriptor for high-resolution satellite image classification," *Remote Sens.*, vol. 8, no. 3, pp. 259–282, Mar. 2016.
[44] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 1337–1342.
[45] L. Zhou, C. Zhang, and M. Wu, "D-LinkNet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, 2018, pp. 192–196.

Ziwei Liu received the master's degree in cartography and geographical information engineering, in 2021, from the College of Geo-Exploration Science and Technology, Jilin University, Changchun, China, where he is currently working toward the Ph.D. degree in resources and environment.

His research interests include arable land quality assessment and land use/cover change.

Mingchang Wang received the Ph.D. degree in remote sensing and geographic information system from the College of Geo-Exploration Science and Technology, Jilin University, Changchun, China, in 2008.

From 2012 to 2013, he was a Visiting Scholar with the Chiba Institute of Technology, Narashino, Japan. He is currently a Professor with the College of Geo-Exploration Science and Technology, Jilin University, teaching cartography and map data processing and other related courses. His current research focuses on remote sensing analysis and mapping.

Fengyan Wang received the Ph.D. degree in geological engineering from Jilin University, Changchun, China, in 2006.

She is currently a Professor with the College of Geo-Exploration Science and Technology, Jilin University. She is mainly engaged in the teaching and research of engineering surveying, deformation monitoring data processing, camera calibration, surveying and fine description of complex rock structures.

Xue Ji received the M.S. degree in environmental sciences from The First Institute of Oceanography, Ministry of Natural Resources, Qingdao, China, in 2017, and the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan, China, in 2021.

She is currently a Lecturer with Jilin University, Changchun, China. Her research interests include hydrographic surveying and charting, subsea topography refinement analysis, and airborne LiDAR bathymetry processing.

Zhiguo Meng (Member, IEEE) received the Ph.D. degree in geophysics from Jilin University, Changchun, China, in 2008.

He is currently a Professor with the College of Geo-Exploration Science and Technology, Jilin University. From 2011 to 2012, he was a Visiting Scholar with the University of California, Santa Barbara. His research interests include the application and development of microwave remote sensing technology in the planetary science, primarily the microwave measurement of the lunar regolith. He has authored/coauthored more than 70 scientific papers.