

Edge Enhanced Channel Attention-Based Graph Convolution Network for Scene Classification of Complex Landscapes

Haoyi Wang , Xianju Li , *Member, IEEE*, Gaodian Zhou , Weitao Chen , *Member, IEEE*,
and Lizhe Wang , *Fellow, IEEE*

Abstract—Monitoring the land covers in complex landscapes is of great significance for the sustainable development of mine geoenvironments. As most existing remote sensing scene datasets are composed of RGB images, there is a lack of multimodal datasets for complex landscapes with mining land covers (MLCs) at a fine-scale. In this study, a new dataset was created by the China University of Geosciences (CUG), Wuhan (named CUG-MLCs) using ZiYuan-3 imagery-based multispectral and topographic data. Moreover, the characteristics of multisize objects, irregular or blurred edges, and spectral-spatial-topographic heterogeneity and variability limited the classification accuracy. Therefore, an edge enhanced channel attention-based graph convolution network (ECA-GCN) was proposed and tested. The proposed ECA-GCN includes three key modules. 1) Multiscale and shallow feature fusion, used to fuse the multiscale convolutional features and shallow features, which helps present the MLC features with various scales; 2) edge enhanced channel attention, used to further select effective channels after a spatial edge feature enhancement, which helps identify irregular or blurred MLCs; and 3) edge detection-based GCN, used for edge feature-based adjacency matrix and feature maps from (2) to construct GCN, which can obtain edge node relation and global contextual information. This framework improved the representation of complex landscape characteristics. The proposed ECA-GCN achieved an overall accuracy of $66.60\% \pm 1.39\%$, averaged accuracy of $36.25\% \pm 1.50\%$, and Kappa of $55.91\% \pm 2.05\%$, thus, outperforming other models. In general, the proposed dataset and model were positive for the fine classification of complex landscapes.

Index Terms—Attention mechanism, feature fusion, graph convolution network (GCN), remote sensing, scene classification, ZiYuan-3.

Manuscript received 7 November 2022; revised 23 January 2023 and 17 March 2023; accepted 5 April 2023. Date of publication 10 April 2023; date of current version 24 April 2023. This work was supported in part by the Natural Science Foundation of China under Grant 42071430 and Grant U21A2013, in part by the Department of Natural Resources of Hubei Province under Grant ZRZY2021KJ04, in part by the Opening Fund of Key Laboratory of Geological Survey and Evaluation of Ministry of Education under Grant GLAB2020ZR14 and Grant CUG2022ZR02, and in part by the Fundamental Research Funds for the Central Universities. (*Corresponding author: Xianju Li.*)

The authors are with the Faculty of Computer Science and Key Laboratory of Geological Survey and Evaluation of Ministry of Education, China University of Geosciences, Wuhan 430074, China (e-mail: wanghaoyi.d@gmail.com; ddwhlxj@cug.edu.cn; zhougadiao@cug.edu.cn; wtchen@cug.edu.cn; lizhe.wang@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2023.3265677

I. INTRODUCTION

SINCE the Industrial Revolution, greenhouse gas emissions were associated with numerous issues [1], [2], [3]. Achieving carbon peaks [4] and carbon neutralization [5], [6] can effectively prevent the increase in greenhouse effects and mitigate the risk of climate change [7]. Technologies related to carbon neutralization have become a research hotspot in China and other major economic countries.

The mining areas are special, artificial, or semi-artificial terrestrial ecosystems, whose core is the mining operation area [8]. Mining areas and surrounding farmland, woodland, and other mining land cover constitute a complex geological environment. In these regions, plenty of geological and environmental issues may occur due to mining activities [9], [10], [11], [12], [13]. The consequences of the inflicted soil and vegetation damage are usually irreversible; therefore, it is important to restore and protect the environment in mining areas.

Land cover classification is a hot topic in the sphere of remote sensing due to various issues and challenges [14], [15], [16], [17], [18], [19]. Furthermore, monitoring land cover types in mining areas is important for ecological, environmental, and social development [20].

Pixel-oriented [21] and object-oriented [22], [23] methods have achieved impressive performances in MLC classification tasks; however, they have little semantic meaning. To obtain a semantic-level understanding of the meaning and the content of remote sensing images [24], we performed scene classification in complex landscapes. Although there are multiple available datasets such as the UC Merced Land-Use dataset [25], WHU-RS19 dataset [26], and SIRI-WHU dataset [27], the existing remote sensing scene datasets are basically composed of red, green, and blue bands (i.e., RGB images). Only a few datasets, such as SAT-4 airborne [31] and SAT-6 airborne datasets [31] are composed of RGB and near infrared (NIR) images. Table I shows a comparison of remote sensing scene classification datasets.

With respect to the classification methods of MLCs at the fine-scale, there are some advances. Complex landscape features in mining areas, especially the remarkable stereo topographic features and spectral-spatial variability, severely restrict the improvement of MLC classification [33], [34], [35], [36].

Previous research on pixel-oriented classification shows that topographic features are one of the most important features of

TABLE I
COMPARISON OF REMOTE SENSING SCENE CLASSIFICATION DATASETS

Dataset	Number of samples	Number of categories	Total	Spatial resolution	Image size	Year	Image type
UC Merced Land-Use Data Set [25]	100	21	2100	0.3	256×256	2010	RGB
WHU-RS19 [26]	~50	19	1005	up to 0.5	600×600	2012	RGB
SIRI-WHU [27]	200	12	2400	2	200×200	2016	RGB
RSSCN7 [28]	400	7	2800	--	400×400	2015	RGB
RSC11 [29]	~100	11	1232	0.2	512×512	2016	RGB
AID [30]	220-420	30	10000	--	600×600	2017	RGB
NWPU-RESISC45 [24]	700	45	31500	~30 to 0.2	256×256	2016	RGB
SAT-4 airborne datasets [31]	125000	4	500000	--	28×28	2015	RGB+NIR
SAT-6 airborne datasets [31]	67500	6	405000	--	28×28	2015	RGB+NIR
Brazilian Coffee Scene [32]	1438	2	2876	--	64×64	2015	RGB

complex landscape areas. For MLC tasks, if we can ensure high spatial resolution as well as a rich number of images while introducing NIR and digital elevation model (DEM) data bands, the classification precision for complex landscapes will be further improved. Therefore, a multimodal scene dataset for the classification of MLCs at the fine-scale is necessary.

Nevertheless, the combination of artificial feature calculation methods and machine learning algorithms (MLAs) [37], [38], [39] is insufficient to achieve satisfactory classification accuracy. Compared with traditional MLAs, deep learning methods are much more accurate and efficient, and have been extensively used in the remote sensing domain.

For example, Li et al. [33] developed deep belief network-based models (DBN) for fine classification of MLCs using ZiYuan-3 (ZY-3) imagery based multispectral and topographic data. They further developed a multilevel output-based DBN model [39]. Compared with DBN, a convolution neural network (CNN) can extract spatial features more effectively, and it performs well in remote sensing classification tasks. For instance, Zhao et al. [40] proposed a dense connection and dilated convolution-based model to capture more comprehensive spatial information.

Alternatively, some studies have investigated multiscale feature fusion strategies. Liu et al. [41] proposed a context-aware spectral-spatial feature extraction module to capture the multiscale features of scale invariance. Xia et al. [42] proposed a multiscale feature fusion network with a series of redesigned

skip pathways. Zhang et al. [43] constructed a three-branch feature fusion network that uses Dual-Anchor triplet loss and nonlocal operation. Mei et al. [44] proposed a multilevel features fusion framework based on sparse representation. Gu et al. [45] proposed a generative adversarial networks (GANs) structure with a pyramidal multiscale structure. They achieve good classification results by multiscale feature fusion, multibranch fusion, or loss function design. There is a consensus that with the increase of network depth, features become more and more abstract. However, shallow features are also important for complex landscapes. Therefore, the integration of multiscale features and shallow features is worth exploring.

Some researchers attach importance to the attention mechanism. Tong et al. [46] proposed a densely connected convolutional network (DenseNet) with channel attention and label-smoothing-based cross-entropy loss function. Ouyang et al. [47] propose multichannel-feature-fusion landform recognition networks based on channel attention. Chen et al. [48] used global context spatial attention and DenseNet to obtain multiscale global scene features. Liu et al. [49] proposed a multidimensional CNN model with improved channel-spatial attention. Guo et al. [50] proposed a self-attention GAN with similarity loss.

MLCs are characterized by large edge differences. For example, a mining catchment has highly irregular edges and a concentrator has blurred edges. Therefore, the effective extraction of edge information will effectively improve the performance of MLC classification. Ma et al. [51] proposed a foreground

activation framework with a dual-branch decoder and collaborative probability loss. It distinguishes foreground objects from the background, achieving similar effects to using edge information. Yang et al. [52] proposed an architecture with a block shuffle structure, super-pixel branch and self-boosting method to obtain precise edge contour. Liang et al. [53] proposed a dual-stream system structure that combines global visual features with positional functions to improve feature representation by using edge node relationships. Zhang et al. [54] proposed an architecture that uses an edge guidance module to learn edge attention representation and aggregate it with feature information. Wang et al. [55] proposed an architecture to extract multigranularity edge features, and jointly learn the segmentation object mask and edge detection. Zhang et al. [56] proposed an architecture that uses soft boundary detection to transform raw data features and obtain global context information. As mentioned above, combining edge information to improve attention is a direction that can be explored.

In recent years, increasingly more attention has been focused on graph convolution networks (GCN) which utilize the correlation between land cover categories by encoding remote sensing images to form maps. Compared with CNN, GCN improves the inapplicability of translation invariance on non-matrix structured data, and its essence is to extract the spatial features of topological graphs. Liu et al. [57] proposed a heterogeneous deep network combining CNN with GCN for pixel and super-pixel feature fusion. Zhou et al. [58] proposed a depth-wise separable GCN model in which the feature graph adjacency matrix was constructed using a Sobel operator.

In addition to the above models, a deep learning model called transformer has gradually been applied in the field of remote sensing. The model divides the image into blocks. The context is captured using the relationship between image blocks. Bazi et al. [59] applied an attention mechanism to focus on different areas of the image and integrate global information. Tang et al. [60] proposed a transformer that used multilevel features to mine the potential context information of remote sensing scenes. However, we believe that in complex landscapes, the transformer model has limitations for feature modeling and high computational complexity.

In this study, a new multimodal dataset was constructed by the China University of Geosciences (CUG), Wuhan (named CUG-MLCs), using ZY-3 imagery. Moreover, in order to extract complex landscape features of multisize objects, irregular or blurred edges, and spectral-spatial-topographic heterogeneity and variability, we proposed and tested an edge enhanced channel attention-based graph convolution network (ECA-GCN). ECA-GCN has the following three key structures:

- 1) multiscale and shallow feature fusion, which fuses the multiscale convolutional features and the shallow features, thus, aiding the presentation of MLC features with various scales;
- 2) an edge enhanced channel attention that further selects effective channels after a spatial edge feature enhancement, which helps identify irregular or blurred MLCs;
- 3) an edge detection-based GCN that uses the edge feature-based adjacency matrix and feature maps from edge

enhanced channel attention to constructing GCN, which can obtain edge node relation and global contextual information. This framework allowed the improvement of the representation of complex landscape characteristics.

Other parts of the study are as follows: Section II introduces the CUG-MLCs dataset; Section III introduces the details of ECA-GCN; Section IV illustrates the experimental settings and classification results; Section V includes the discussion regarding the algorithm and dataset; and Section VI includes the conclusion of the article and the outlook for future work.

II. CUG-MLCs DATASET

This section introduces the content, production process, and classification system of the CUG-MLCs dataset. Furthermore, the characteristics of the CUG-MLCs dataset are described.

A. Description of the Proposed CUG-MLCs Dataset

The CUG-MLCs dataset developed for this study contains 6125 images. It is a multimodal dataset meant for MLCs at fine-scale. It contains five channels, namely DEM, RGB, and NIR bands. All images have a size of 64 pixels \times 64 pixels. The dataset consists of 20 land covers. The number of samples in each class of the CUG-MLCs dataset varies from 19 to 2333. Table II shows the basic information on the CUG-MLCs dataset. Fig. 1 illustrates example images from the dataset.

The CUG-MLCs dataset is divided into training, validation, and test sets in a 6:2:2 ratio. The CUG-MLCs dataset can be downloaded from the following website: https://drive.google.com/drive/folders/1cLpdTNLIQxpR5_ytYD-TqRtiwTP8JDeS?usp=sharing.

B. Remote Sensing Data Acquisition and Preprocessing

The study area is located in Wuhan City, China. The range of longitude and latitude is: 114°16' E-114°20' E and 30°16' N-30°18' N. The study area is 109.4 km² and presents typical surface mining and agricultural landscape characteristics [61]. Fig. 2 shows the RGB imagery of the study area and division of scene data.

ZY-3 satellite imagery was selected as the data source and the images used are from scenes obtained with different cameras on June 20, 2012. Multispectral images were obtained using some preprocessing methods.

A DEM with a resolution of 10 m was generated using the stereo image pair data. Subsequently, the DEM was resampled to 2.1 m to match the following multispectral image.

Combining the generated DEM data and the rational polynomial coefficients, orthorectification was conducted on the panchromatic and multispectral data. Next, based on corrected panchromatic images, geometric registration of multispectral data was performed using a quadratic polynomial function. The resampling error was controlled within 0.5 pixels using the cubic convolution method. Finally, the Gram-Schmidt method was used to fuse the registered panchromatic and multispectral data. A fused multispectral image with a resolution of 2.1 m was obtained.

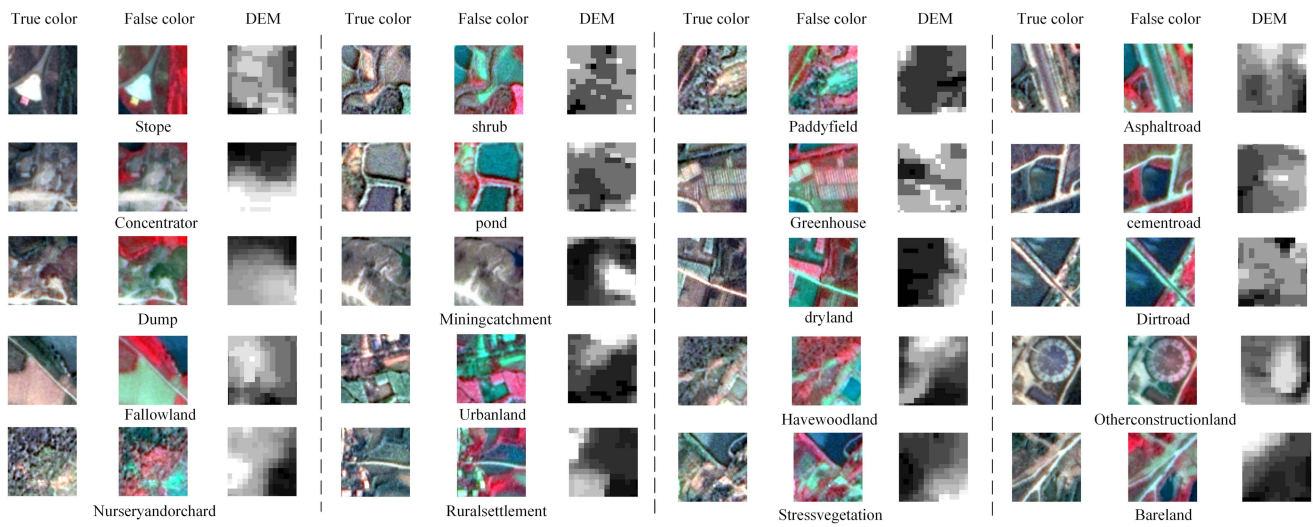


Fig. 1. Example images of various land covers in the CUG-MLCs dataset.

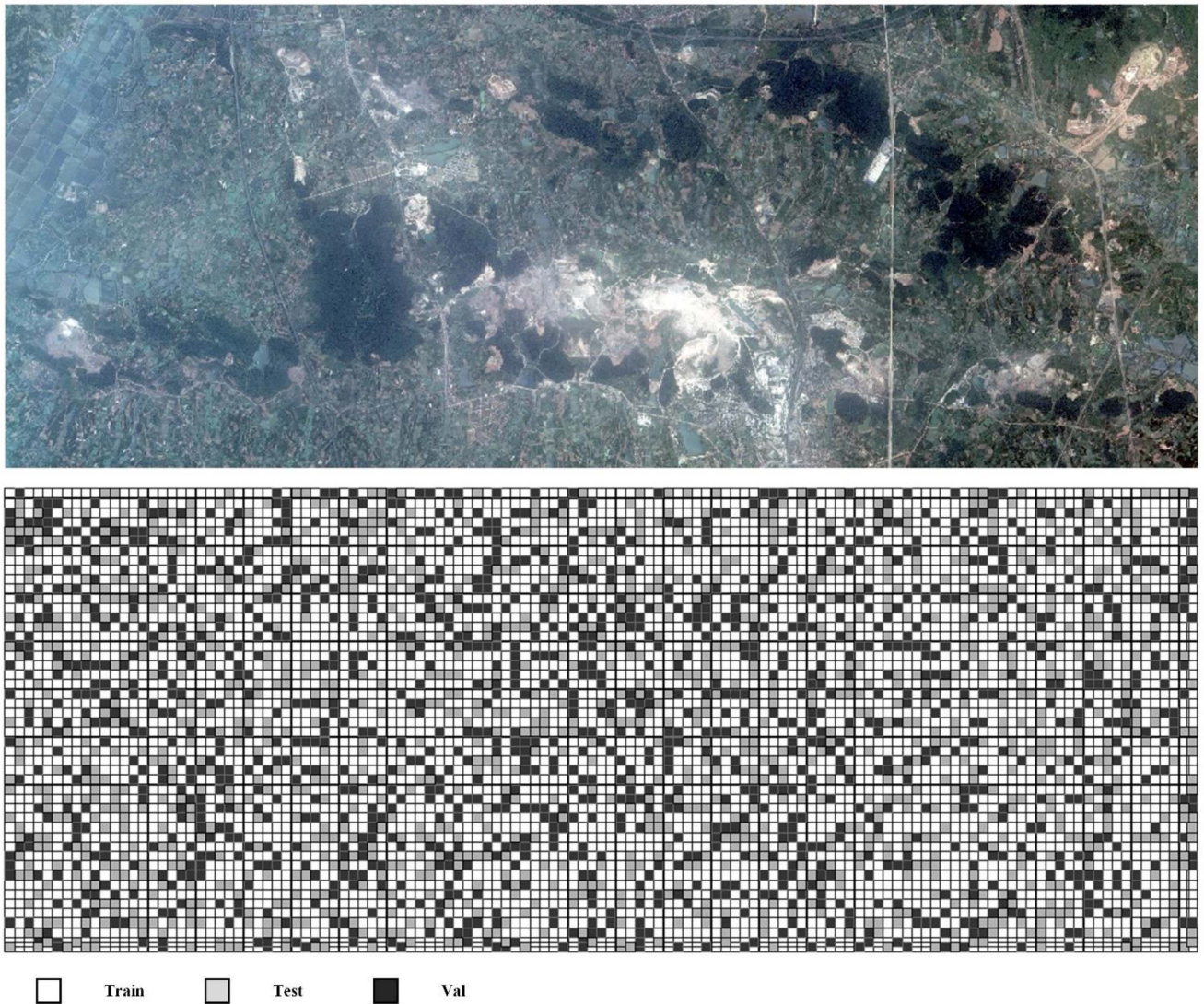


Fig. 2. RGB imagery of the study area and division of scene data.

TABLE II
BASIC INFORMATION OF CUG-MLCS DATASET

Data combination	Sample No.	Class No.	Total No.	Resolution	Image size	Image source
RGB (True color image)	19~2333	20	6125 (with training, validation, and test sample numbers of 3666, 1244, and 1215)	2.1 m	64 pixels × 64 pixels	ZiYuan-3 satellite imagery
NIRRG (False color image)						
Five channels						
Two data (multispectral image and DEM)						
Three data (RGB, NIRRG, and DEM)						

C. Visual Interpretation, Cropping, and Classification System

Manual visual interpretation and image clipping were performed on the preprocessed data to form the final CUG-MLCs dataset. In manual visual interpretation, the representative and significant land cover types investigated in this study were selected. In the clipping process, the image was sequentially clipped from the upper left corner of the image. The image size was 64 pixels × 64 pixels. Finally, a total of 6125 images were divided. Table III shows the description of the MLC classification system and the number of images.

D. CUG-MLCs Dataset Size

Take stope and asphaltroad as examples (as shown in Fig. 3, the pictures on the right are the ground truth of the corresponding area after manual visual interpretation). In a complex landscape, images with different sizes in the same area contain multiple categories. For example, pictures labeled as a stope. Although images with the size of 48 × 48, the number of categories is small, but the proportion of stope is small. Larger size images such as 128 × 128, 224 × 224, 256 × 256, although stope account for a large proportion, there are too many categories. This may make feature extraction difficult. While pictures are labeled as asphaltroad, the category and proportion of 48 × 48 and 64 × 64 size pictures are similar. Larger size images face the same problems as stope pictures. Therefore, based on qualitative and quantitative considerations, we choose 64 × 64 as the image size.

E. CUG-MLCs Dataset Characteristics

- 1) *Complex landscapes*: There are numerous types of land covers including mining areas with heterogeneous terrain and agricultural development landscapes at different phenological stages. There is a large difference in the number of different classes, which has the nature of the class-imbalance.
- 2) *Multimodal*: The CUG-MLCs dataset contains DEM, RGB, and NIR bands. DEM data can effectively improve the classification performance of MLCs.
- 3) *Size of objects*: The size of MLCs varies widely. Some images have large background information, which, if not handled, can affect the classification accuracy.
- 4) *Edges of objects*: The edge of MLCs is irregular or blurred. The effective use of edge node information will help to improve the classification accuracy at the fine scale.

III. METHODS

The network structure of the ECA-GCN is shown in Fig. 4. It contains the following four parts.

- 1) *Multiscale and shallow feature fusion*: The size of land cover types in different mining areas varies greatly, and there is much background information; therefore, we extracted multiscale information using a multiscale convolution kernel to focus on multiscale feature representation. With the increase in network layers, the deep features will inevitably lead to the loss of some features. This may affect the accuracy of small-sized objects. Therefore, we added the branch of shallow features to fuse the deep features with the shallow features.
- 2) *Edge enhanced channel attention*: The importance of multiscale features differs from that of shallow features; thus, after obtaining the characteristics of (1), we added the channel attention module. In addition, to further highlight the importance of local edge information, we enhanced edge information in the attention module. Edge-dependent design highlighted the importance of local image details.
- 3) *Edge detection-based graph convolution network*: We used the Canny operator to detect edges. After processing, it was used as the adjacency matrix of the GCN. After propagating through each layer of the GCN, all features were fused. GCN received the features from the above modules, and used the edge node relationship to further capture the global context information.
- 4) *Classification*: After GCN, we stacked three convolutional pooling structures to reduce channel dimensions and feature map size. Layer-by-layer dimension reduction ensures the preservation of representative features. Ultimately, the classification result is output through 1 × 1 convolution.

TABLE III
DESCRIPTION OF MLC CLASSIFICATION SYSTEM AND THE NUMBER OF IMAGES

Label	Name	Description	Number
0	Stope	It refers to the area where mining operations are carried out, including the mining and abandoned pits.	168
1	Concentrator	The ore dressing activity area, including ore dressing product stacking area and ore stacking area.	68
2	Dump	The place where the topsoil and overburden are collectively stacked in the process of mine exploration and development.	54
3	Paddyfield	Cultivated land with irrigation facilities and water sources. It can grow rice, lotus root and other aquatic crops.	66
4	Greenhouse	It is covered with a layer of white plastic film with high surface reflectivity and a regular shape.	34
5	Dryland	Cultivated land relying on natural precipitation, without irrigation facilities and water sources.	2333
6	Fallowland	There is no crop growth at this stage, and rapeseed and wheat have just been harvested in the study area.	614
7	Nurseryandorchard	The chlorophyll content is high, including protective forest, timber forest, economic forest, etc.	398
8	Havewoodland	Shrubs that are low in height and have many stems, usually not more than 2 meters.	1120
9	Stressvegetation	Affected by mining activities, the forest is under pressure. There is a large amount of mineral dust deposited around the surface of mining land, which grows poorly.	123
10	Shrub	There are no obvious trunks and clumps of relatively small trees less than 2m.	74
11	Pond	A fish pond with a regular shape.	607
12	Miningcatchment	It refers to a pond with an irregular shape. Mainly formed during mining and after the pit closure.	27
13	Asphaltroad	Black roads on true color images.	78
14	Cementroad	White roads on true color images, including the roads in the mining area.	32
15	Dirtroad	Gray roads on true color images.	35
16	Urbanland	Land used by cities of different sizes and built-up areas above county and town.	19
17	Ruralsettlement	Settlements in rural areas. Some of them are centralized, while others are scattered.	191
18	Otherconstructionland	Industrial area land. It is independent of the cities and towns.	12
19	Bareland	Land with soil coverage and vegetation coverage of less than 5%.	72

A. Multiscale and Shallow Feature Fusion

The deep learning method exhibits improved feature extraction as well as classification performance when compared to traditional machine learning methods, and is characterized by increased robustness and easy migration. CNN is an important part of a deep learning method, and its strong representational learning ability has attracted wide attention.

The land cover classification network of the mining area accepts RGB + NIR + DEM images as input. With increasing convolution layers, convolution layers can retrieve features from fine to rough. Nevertheless, classical CNN usually extracts features by stacking the same convolution layers. Furthermore, convolution kernels of different sizes acquire multiscale features from different scales, effectively expanding the information flow, which is helpful for the recognition of small-sized objects. To avoid losing significant image detail information with the

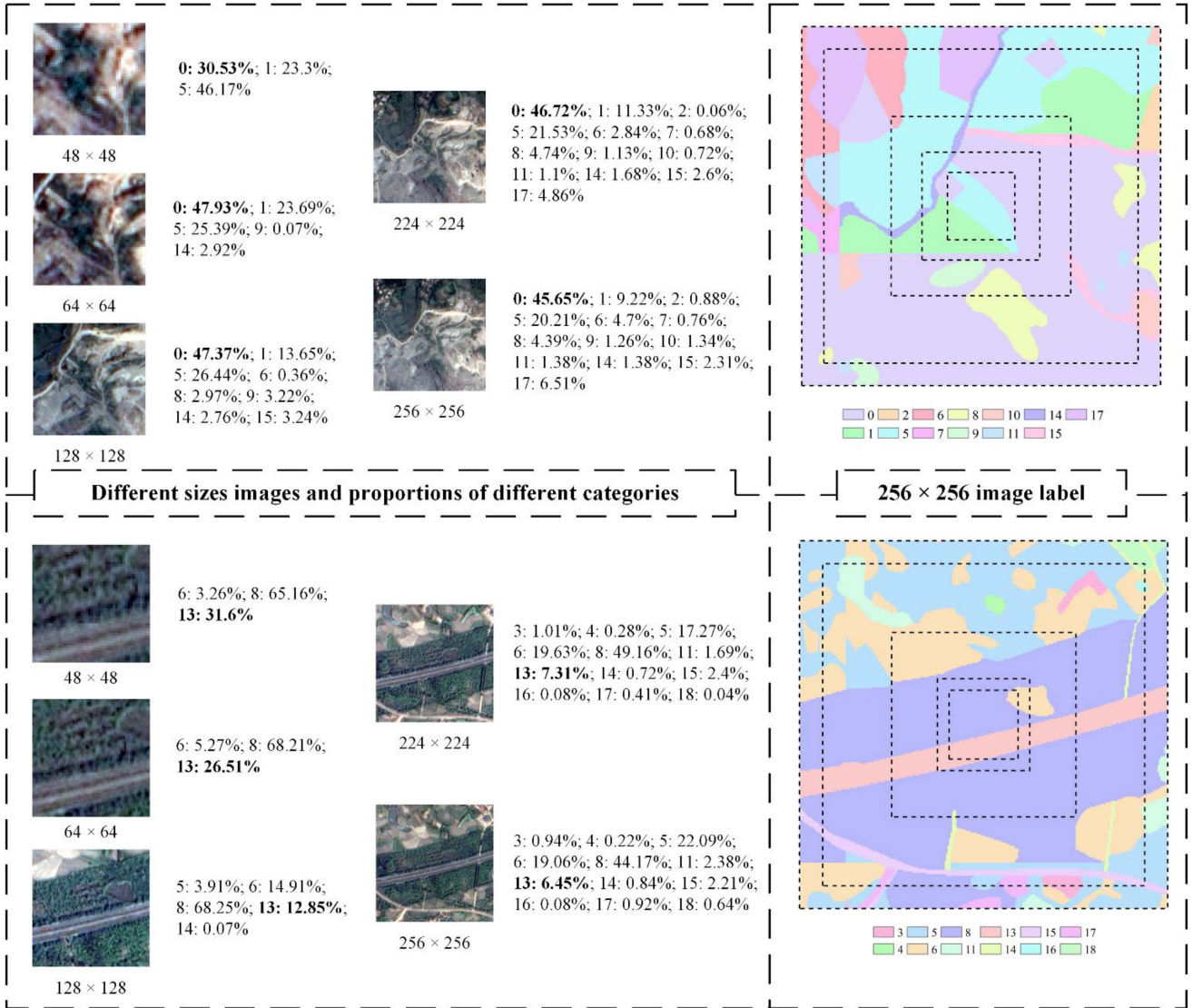


Fig. 3. Display of images and labels of different sizes in the same area (take stope and asphalt road as examples).

increase in depth, we added a shallow feature extraction branch to obtain multiscale information on deep and shallow feature fusion.

The multiscale feature extraction part of the multiscale and shallow feature fusion module comprised four branches. The first branch used a 1×1 convolution kernel. The second branch used a 3×3 convolution kernel, padding was set to 1 to ensure that the size of the feature map was unchanged. The third branch used a 5×5 convolution kernel, with padding set to 2. The fourth branch used two 3×3 convolution kernels. After the convolution layer, a BN layer and the ReLU activation function were added. Finally, a 2×2 MaxPool is used to reduce dimension. The shallow feature extraction part of the multiscale and shallow feature fusion module consists of a 1×1 convolution and MaxPool. The 1×1 convolution is used to normalize the number of feature channels, and the MaxPool was used for feature dimensionality reduction. Ultimately, we fuse the deep and shallow features through the channel-stacking operation to

obtain the module output. The structure of the multiscale and shallow feature fusion module is shown in Fig. 5.

Here, we stack three multiscale and shallow feature fusion modules as our feature extraction structure. This is mainly because of the size of the input image of 64 pixels \times 64 pixels. After pooling-induced dimension reduction, the size of the feature map decreases. To maintain the size of the feature map moderate and consider the network depth, we choose to stack three modules.

B. Edge Enhanced Channel Attention

SENet [62] is a brand-new image recognition structure published in 2017. Through learning the correlation between characteristic channels, SENet strengthens the channels with strong presentation ability and weakens the secondary channels. It alleviates the loss caused by different channel representation capabilities.

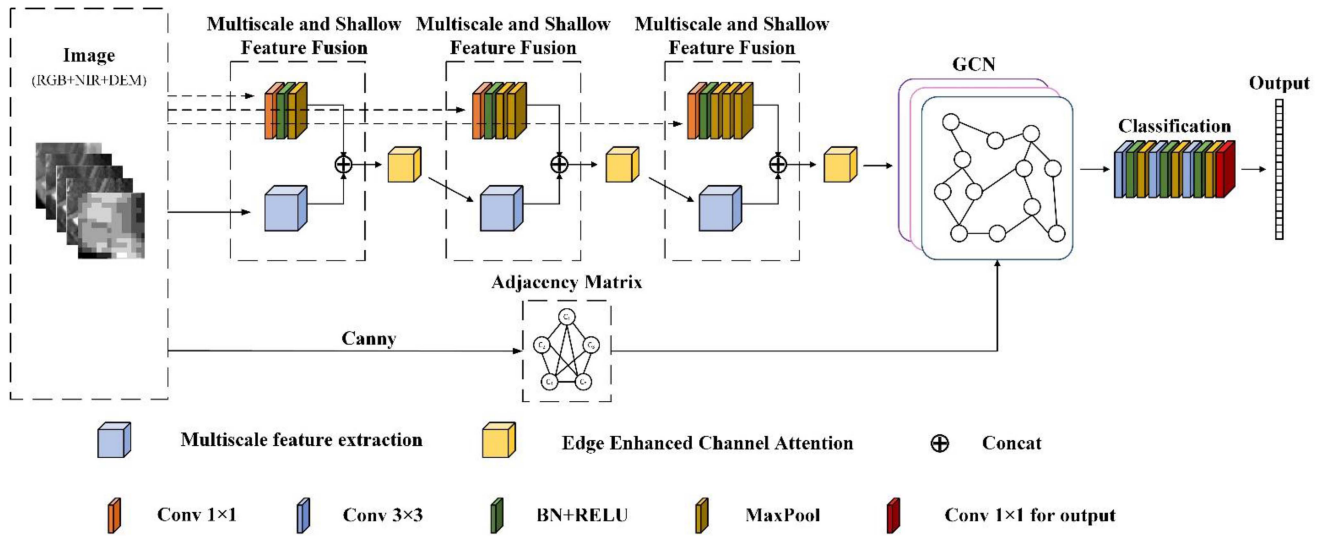


Fig. 4. ECA-GCN network structure.

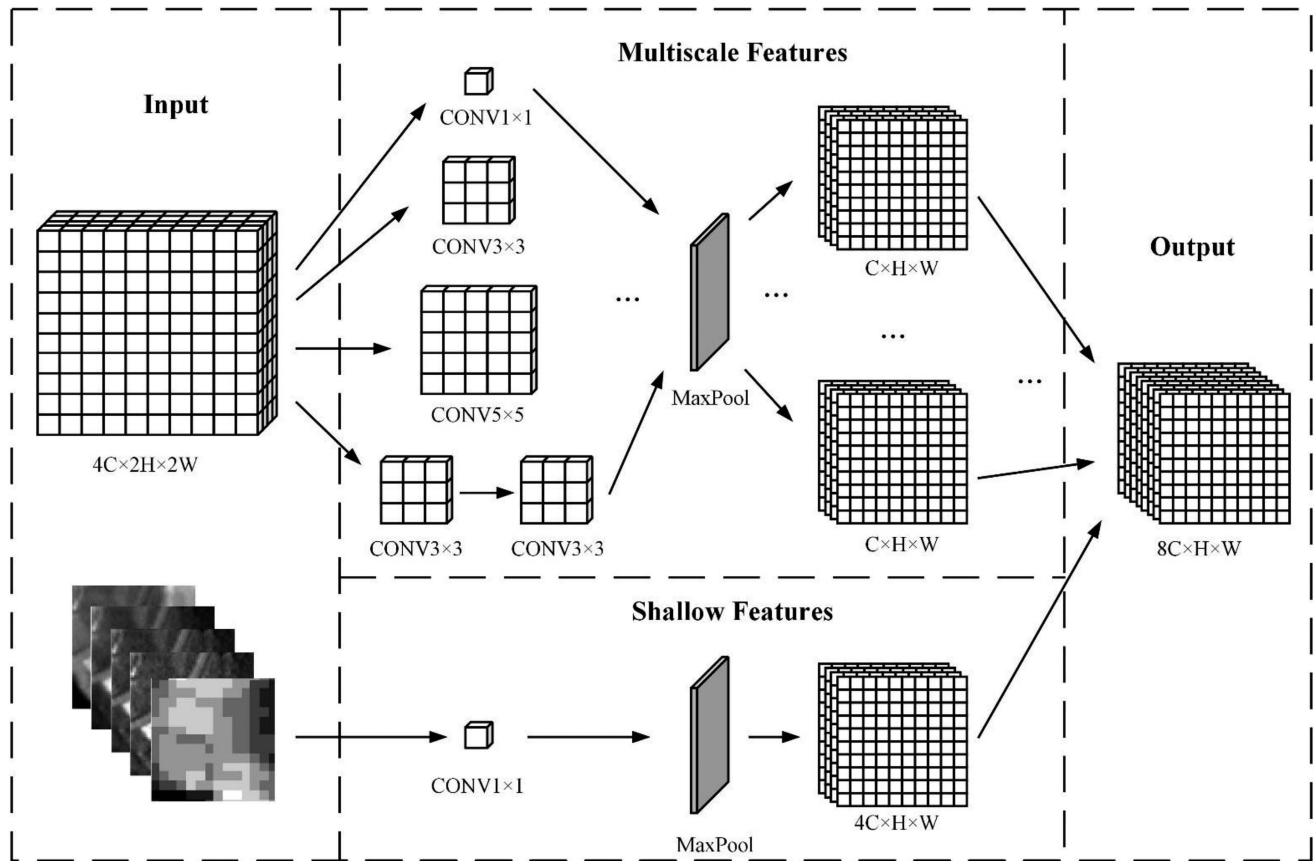


Fig. 5. Multiscale and shallow feature fusion module.

A convolutional block attention module (CBAM) [63] is proposed based on SE attention. The spatial attention module is added after the channel attention module. The attention weight is derived from the channel and spatial dimensions. In this way, it can learn the importance of features and spatial positions, respectively.

The fusion module combines multiscale and shallow features. We screen channels with strong presentation ability by adjusting the weights of channels. To adjust in time, we add an edge enhancement channel attention mechanism after each multiscale and shallow feature fusion module. The improved edge enhanced channel attention is shown in Fig. 6.

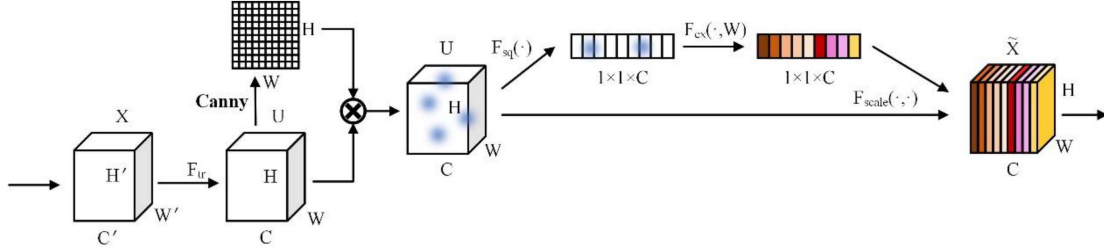


Fig. 6. Edge enhanced channel attention module.

SE attention includes two stages which are squeezing and excitation. In the squeeze phase, the input feature map size is $C \times H \times W$, where C represents the number of channels, and H and W represent the height and width of the feature map, respectively. We used the AdaptiveAvgPool to compress each feature map into a single value; thus, the feature map becomes a $C \times 1 \times 1$ vector, where the output of the squeeze phase can be calculated as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H u_c(i, j). \quad (1)$$

In the excitation stage, to effectively use the local descriptors obtained following the squeeze, we ensure that the channel-wise dependencies are fully captured. Two nonlinear fully connected layers are used. The ReLU activation function is added after the first fully connected layer. After the second fully connected layer, Sigmoid is added. The output of the excitation stage can be calculated as follows:

$$s_c = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)). \quad (2)$$

Next, we assigned corresponding weights to each channel and obtained the final output through channel wise multiplexing

$$F_{scale}(u_c, s_c) = s_c \cdot u_c. \quad (3)$$

The edge enhancement mechanism used the Canny operator to extract the edge information from each channel feature map. The edge enhancement matrix is formed through superposition and normalization. The edge information of the original feature map is enhanced through matrix multiplication. Then, the squeeze and excitation phases commenced. Refer to Section III-C for the calculation process of the Canny operator.

C. Edge Detection-Based GCN

GCN utilizes the graph structure and aggregates node information from the neighborhoods [64]. Therefore, GCN performs well in modeling long-range spatial relations [65].

The formula of graph convolution used in this article is defined as follows:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W\right) \quad (4)$$

where $\sigma(\cdot)$ denotes an activation function and \tilde{A} represents the adjacency matrix, $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$. W is a layer-specific trainable weight matrix and $H^{(l)} \in R^{N \times D}$ is the matrix of activations in

the l th layer. The edge detection-based GCN uses the Canny operator to extract the edge information and construct the graph structure, allowing it to achieve a better feature representation.

The key of GCNs is to generate the adjacency matrix to input tensor $H \in R^{N \times D}$. In the complex landscape, the road presents a linear structure in the image; although the proportion is small, the edge is relatively regular. The mining catchment occupies a small area in the image and has no stable shape. We believe that for the type of multi size MLCs with irregular edges, edge nodes contain more efficient information. The edge nodes can highlight the characteristics of multiscale feature types from the background information, thus, indicating the efficiency of building edges between these nodes. The edge detection-based GCN is shown in Fig. 7.

1) *Graph Construction*: The Canny operator is a commonly used edge detection filter. The calculation process can be divided into four stages:

a) *Image filtering*: First, the Canny operator uses a blur filter to eliminate noise from the input image. Here we use a Gaussian filter, and the Gaussian kernel size is set to 3×3

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (5)$$

where σ represents the parameter of the Gaussian filter, which controls the blurring degree of the image (set to 1).

b) *Image gradient calculation*: The amplitude and direction of the image gradient should be calculated. We use the Sobel filter with a kernel size of 3×3 decomposed into two filters. The first kernel is used to extract the horizontal gradient and the second kernel is used to extract the vertical gradient. The gradient size G can be calculated as follows:

$$G = \sqrt{G_x^2 + G_y^2}. \quad (6)$$

The template of the Sobel operator gradient calculation operator is as follows:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (7)$$

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}. \quad (8)$$

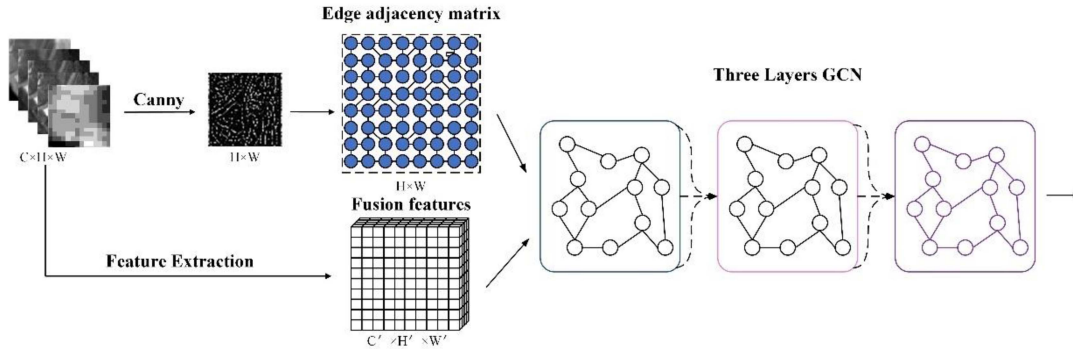


Fig. 7. Edge detection-based GCN.

The gradient direction calculation formula is as follows:

$$\theta = \arctan\left(\frac{G_y}{G_x}\right). \quad (9)$$

c) *Nonmaximum suppression (NMS)*: Edge refinement is performed using the NMS method, ensuring that each edge is a single pixel in width. This step needs to lead to the detection of 8 neighborhoods. If a pixel has the maximum intensity compared to its neighbor, it is the local maximum, and the pixel is retained. First, we create the kernel NMS_{0° in 0° direction with a size of 3×3 and constructed the direction matrix $R(\theta)$

$$NMS_{0^\circ} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \quad (10)$$

$$R(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}. \quad (11)$$

d) *Checking and connecting edges*: Using the double threshold method to select the edge points after NMS. Pixel whose gradient amplitude is lower than the low threshold is selected as nonedge points. Pixels whose gradient amplitude is higher than the high threshold are selected as edge points. Pixel whose gradient amplitude is lower than the high threshold and higher than the low threshold are selected as candidate edge points. When the candidate edge point is directly connected with the edge points, it is considered a part of the edge; otherwise this point will be discarded.

Therefore, the adjacency matrix of graph convolution can be calculated by the following:

$$\tilde{A} = \text{Softmax}(F(i)) \quad (12)$$

where F represents the Canny operator, and i represents the original image.

2) *Graph Convolution*: After using the Canny operator to generate the adjacency matrix, the fully connected graph was generated and then the extracted multiscale features after edge enhanced channel attention were inputted into the GCN. Finally, we inputted the designed classifier and obtained the classification results. The GCN should not be too deep; generally, two to three layers were used. The GCN proposed in this

article uses a three-layer graph convolution. Compared with the standard convolution, GCN can expand the receptive field, obtain long-distance dependence relationships, and efficiently exchange information in a larger range. Therefore, GCN has better feature expression ability.

D. Classification

To effectively utilize the representative features, we carefully designed the classification structure. The classification module accepts feature maps from GCN. We use a 3×3 convolution to reduce the number of channels while maintaining the feature map constant. Further, we reduce the size of the feature map by using maximum pooling, with each convolution layer being followed by the BN layer and ReLU activation function. To reduce the size of the signature map to 1, we stacked three convolution pooling structures. Ultimately, a 1×1 convolution layer was used to classify of land cover in mining areas.

IV. RESULTS

A. Experimental Settings

1) *Machine Configuration*: The experiment was under the Centos7 system. The scene classification algorithms were implemented by using the PyTorch framework. The hardware configuration had 128 GB, the GPU was RTX2080ti with 11 GB of memory, and the CPU was Inter Xeon(R) E5-2620 v4.

2) *Parameter Optimization*: At the beginning of the experiment, the images were normalized. Using the accuracy of the verification set to adjust parameters. We set the learning rate to 0.0001. The number of training samples in each batch was 64 or 32 for different algorithms. Each experimental training iteration was 200 times. We used Cross Entropy as the loss function. Each experiment was performed 5 times to obtain the average value. In addition, the model with the best validation accuracy was the optimal model and was saved. Fig. A1 shows the optimization of the parameters of the selected algorithm.

3) *Model Comparison*: We conducted a comparative experiment between the proposed ECA-GCN and other scene classification models. Four classical CNN models VGGNet-16,

ResNet-18, ResNet-101 and DenseNet-121 were used to conduct experiments on the CUG-MLCs dataset to provide benchmark results for subsequent studies. Based on classical networks, two attention mechanisms SE and CBAM, were used to add the above four convolutional neural network models to form an attention network for experiments. For the VGGNet-16 network, SE and CBAM modules were respectively added before the first full connection layer. For ResNet-18 and ResNet-101 networks, the CBAM module was added to the last residual block, and the SE module was added to the residual block located after the second BN layer. For the DenseNet-121 network, SE and CBAM modules will be added before the last full connection layer.

Data fusion is the current research hotspot in the field of remote sensing. Therefore, based on the classical CNN, dual-stream and three-stream fusion network experiments were carried out. For the dual-stream network, the first branch inputted a multispectral image, while the second branch inputted DEM. For the three-stream network, the first branch inputted RGB, the second branch inputted NIRRG, and the third branch inputted DEM. Among them, VGGNet-16, ResNet-18, ResNet-101, and DenseNet-121 were respectively used as backbone networks for each branch, and the fusion mode of the feature cascade was used for model building. Among them, the VGGNet-16 network performs data fusion before the last full connection layer, ResNet-18 and ResNet-101 networks perform data fusion after the first convolution layer, and the DenseNet-121 network performs data fusion before the last full connection layer. Furthermore, we selected some recent models such as ShuffleNet V2 [66], EfficientNet [67], CAD [46], MF²CNet [68], PDCNet [40], GCSANet [48], and EMTCAL [57] to compare the effectiveness of our proposed model.

We calculated the overall accuracy (OA), average accuracy (AA), and Kappa values of the test set.

B. Results of Accuracy Assessment

Table IV shows the evaluation indicators of the CUG-MLCs dataset. Among all models, the proposed model achieved an OA of 66.60%, AA of 36.25%, and Kappa of 55.91%, the best performance among the analyzed datasets on all evaluation indicators. The results show the effectiveness of our proposed multiscale and shallow feature fusion module, edge enhancement channel attention module, and GCN.

1) *RGB*: Among the four classical convolutional neural networks, VGGNet-16 showed the best classification effect. It achieves an OA of 57.65%, AA of 21.15%, and Kappa of 43.46%. VGGNet performed better on the CUG-MLCs dataset than ResNet and DenseNet. This might be due to the obvious imbalances between the categories of this dataset. In the mining area scenario, the shallow features have a greater impact on the classification of land cover categories in the mining area than deep features. Therefore, when designing the network, considering the deep and shallow features will effectively improve the classification accuracy. Fig. A2 shows the confusion matrix.

2) *NIRRG*: Among the four classical CNNs, VGGNet-16 showed the best classification effect. It achieved an OA of

60.74%, AA of 26.52%, and Kappa of 48.53%. Compared with the RGB band, the OA and Kappa increased by 3.09% and 5.07%, respectively. We hypothesize this is because the NIR band has advantages in reflecting the difference of infrared characteristics reflected or radiated by vegetation and other ground objects.

3) *Five Channels*: The classical CNN with the same network structure only changes the input data from 3-band to 5-band. According to our observations, the OA of VGGNet-16, ResNet-18, ResNet-101, and DenseNet-121 increased by 4.00%, 10.59%, 6.58%, and 6.83%, respectively. It can be concluded that the network with the same structure can effectively improve the model accuracy by adding data sources, that is, adding NIR and DEM bands to their RGB basis.

The dynamic selection of feature information was achieved through the attention mechanism. The SE module focuses on the relationship between different channels and aids the model in automatically learning the importance of different channel characteristics. After the channel attention module, the CBAM module was connected to the spatial attention module to achieve the dual mechanism of channel attention and spatial attention. At this point, the CBAM module was no longer using a single maximum pooling or average pooling, but rather the addition or stacking of maximum pooling and average pooling. Among them, the channel attention modules were added and the spatial attention modules were stacked. Two attention mechanisms were added to the four networks and compared with the source network. Compared with VGGNet-16, VGGNet-16-SE, and VGGNet-16-CBAM led to no significant accuracy improvement. Alternatively, the OA of ResNet-18-SE and ResNet-18-CBAM increased by 1.69% and 2.61%, while the Kappa increased by 1.59% and 3.08%, respectively, when compared to ResNet-18. Compared with ResNet-101, the OA of ResNet-101-SE and ResNet-101-CBAM increased by 1.2% and 1.32%, while Kappa increased by 1.25% and 1.09%, respectively. Compared with DenseNet-121, the OA of DenseNet-121-SE and DenseNet-121-CBAM increased by 0.99% and 0.7%, respectively, while their Kappa increased by 1.18% and 0.43%, respectively. This indicates that except for the VGGNet-16, the classification accuracy of other networks was improved after adding the SE and CBAM modules, with the most significant improvements in accuracy being associated with the ResNet-18 network. Alternatively, the accuracy of VGGNet-16 has not been improved. We hypothesize that the model was already in the over fitting state before adding attention; therefore, adding these parameters exacerbated the over fitting problem, resulting in no improvement in performance.

The dual-stream network uses different branches to extract the characteristics of different bands and obtain difference information. Feature fusion was used to realize feature complementation, with each branch of the four dual-stream networks using a network with the same structure for feature extraction. Compared with the network in which only RGB band images were inputted, the accuracy of the dual-stream VGGNet-16 has not been significantly improved. Alternatively, the OA of dual-stream ResNet-18 increased by 10.39% and Kappa increased by 14.51%. The OA of dual-stream ResNet-101 increased by

TABLE IV
QUANTITATIVE COMPARISON OF DIFFERENT NETWORKS USING CUG-MLCS DATASETS

Methods	Image type	OA (%)	AA (%)	Kappa (%)
VGGNet-16	RGB (True color image)	57.65 ± 0.88	21.15 ± 1.62	43.46 ± 1.90
ResNet-18		44.79 ± 2.65	17.05 ± 2.16	26.86 ± 2.96
ResNet-101		49.32 ± 1.02	16.06 ± 1.90	30.14 ± 1.02
DenseNet-121		53.79 ± 0.97	21.11 ± 0.87	37.40 ± 1.05
VGGNet-16	NIRRG (False color image)	60.74 ± 1.17	26.52 ± 2.44	48.53 ± 1.65
ResNet-18		52.97 ± 0.69	20.49 ± 2.66	36.77 ± 1.50
ResNet-101		53.74 ± 1.51	20.85 ± 2.50	38.55 ± 1.52
DenseNet-121		58.57 ± 1.16	27.72 ± 1.50	45.96 ± 1.39
Dual-Stream VGGNet-16	Two data (multispectral image and DEM)	56.69 ± 1.62	21.95 ± 1.95	43.24 ± 1.85
Dual-Stream ResNet-18		55.18 ± 1.16	23.55 ± 3.03	41.37 ± 1.39
Dual-Stream ResNet-101		56.03 ± 1.57	22.45 ± 1.34	41.77 ± 1.26
Dual-Stream DenseNet-121		60.43 ± 0.89	25.34 ± 1.55	47.03 ± 1.58
Three-Stream VGGNet-16	Three data (RGB, NIRRG, and DEM)	40.46 ± 5.45	8.89 ± 2.44	15.03 ± 7.36
Three-Stream ResNet-18		51.97 ± 3.44	20.13 ± 1.27	34.63 ± 2.02
Three-Stream ResNet-101		54.21 ± 1.99	22.88 ± 1.83	38.49 ± 1.38
Three-Stream DenseNet-121		57.12 ± 1.16	22.20 ± 0.60	42.04 ± 1.72
VGGNet-16	Five channels	61.65 ± 0.48	27.35 ± 2.29	48.91 ± 1.59
ResNet-18		55.38 ± 1.54	21.51 ± 2.92	39.82 ± 2.83
ResNet-101		55.90 ± 2.39	24.48 ± 4.11	41.84 ± 2.87
DenseNet-121		60.62 ± 0.83	27.39 ± 1.64	47.48 ± 1.85
VGGNet-16-SE		61.48 ± 0.20	24.74 ± 1.52	48.49 ± 0.79
ResNet-18-SE		57.07 ± 0.60	21.53 ± 1.82	41.41 ± 1.40
ResNet-101-SE		57.10 ± 1.45	23.01 ± 2.05	43.09 ± 1.79
DenseNet-121-SE		61.61 ± 0.40	27.23 ± 0.76	48.66 ± 0.21
VGGNet-16-CBAM		61.36 ± 0.73	25.27 ± 2.02	49.33 ± 1.17
ResNet-18-CBAM		57.99 ± 0.73	22.42 ± 1.27	42.90 ± 1.26
ResNet-101-CBAM		57.22 ± 0.79	21.95 ± 2.42	42.93 ± 1.35
DenseNet-121-CBAM		61.32 ± 0.35	25.91 ± 1.07	47.91 ± 0.61
ShuffleNet V2		51.49 ± 2.06	16.20 ± 1.89	32.78 ± 3.47
EfficientNet		52.07 ± 1.70	16.45 ± 1.02	33.43 ± 3.64
CAD		59.05 ± 1.92	26.53 ± 2.06	45.31 ± 2.19
MF ² CNet		57.29 ± 1.18	24.57 ± 2.29	42.77 ± 2.21
PDCNet	65.73 ± 0.68	35.96 ± 1.52	54.62 ± 0.90	
GCSANet	59.31 ± 0.87	26.36 ± 2.37	45.09 ± 1.47	
EMTCAL	52.86 ± 2.67	20.31 ± 2.85	36.84 ± 2.41	
ECA-GCN (without GCN)	58.98 ± 0.82	27.75 ± 0.85	45.70 ± 1.31	
ECA-GCN	66.60 ± 1.39	36.25 ± 1.50	55.91 ± 2.05	

6.71%, and Kappa increased by 11.63%. The OA of dual-stream DenseNet-121 increased by 6.64% and Kappa increased by 9.63%. Compared with the single branch network of input five bands, the OA of VGGNet-16 was 4.96% higher than that of the dual-stream VGGNet-16. The OA of ResNet-18 was 0.2% higher than that of dual-stream ResNet-18. The OA of dual-stream ResNet-101 was 0.13% higher than that of ResNet-101. The OA of DenseNet-121 was 0.19% higher than that of the dual-stream DenseNet-121. The fusion strategy of the dual-stream network we used was simple channel superposition. It can be seen that the effect of directly inputting 5-band images is better than that of double branches. This may be because different data bands have different importance for land cover classification in the mining area. The fusion method of channel superposition is equivalent to giving the same weight to the features extracted

by the two branches, which makes the classification effect not ideal.

Identical to the dual-stream network, the three-stream network uses different branches to extract differential information. Compared with the network in which only RGB band images were inputted, the accuracy of the three-stream VGGNet-16 was not significantly improved. The OA of three-stream ResNet-18 increased by 7.18% and Kappa increased by 7.77%. The OA of three-stream ResNet-101 increased by 4.89%, and Kappa increased by 8.35%. The OA of three-stream DenseNet-121 increased by 3.33% and Kappa increased by 4.64%. Compared with the dual-stream network, the accuracy of the three-stream network has not been improved. We hypothesize that this is because the three-stream network has a large number of low-efficiency redundancy characteristics. Thus, to extract more

representative features, inefficient features can be removed, or efficient features can be enhanced.

Compared to the analyzed networks, the ECA-GCN proposed in this article performed best. Compared with ShuffleNet V2, the OA of ECA-GCN increased by 15.11% and its Kappa increased by 23.13%. Compared with EfficientNet, the OA of ECA-GCN increased by 14.53% and its Kappa increased by 22.48%. Compared with CAD, the OA of ECA-GCN increased by 7.55%, and its Kappa increased by 10.6%. Compared with MF²CNet, the OA of ECA-GCN increased by 9.31%, and its Kappa increased by 13.14%. Compared with PDCNet, the OA of ECA-GCN increased by 0.87% and its Kappa increased by 1.29%. Compared with GCSANet, the OA of ECA-GCN increased by 7.29% and its Kappa increased by 10.82%. Compared with EMTCAL, the OA of ECA-GCN increased by 13.74% and its Kappa increased by 19.07%. We verified the effectiveness of the GCN network through ablation experiments and concluded that: Compared with ECA-GCN(without GCN), the OA of ECA-GCN increased by 7.62%, and its Kappa increased by 10.21%. This indicates the effectiveness of the ECA-GCN proposed in this article. We used multiscale and shallow feature fusion to obtain multiscale information on deep as well as shallow features and filtered important features using the edge enhanced channel attention. The GCN was constructed on the relationship of image edge nodes. Finally, the designed classifier was inputted for classification. ECA-GCN can effectively improve the classification accuracy of difficult land cover samples of the mining area, such as concentrator, dump, paddy field, mining catchment, asphalt road, and dirt road. As shown in Fig. A3, the accuracy of ECA-GCN in dryland was 91%, the accuracy for the pond was 87%, and the accuracy of nursery and orchard was 63%. Competitive classification accuracy was maintained in the land cover categories of a large number of mining areas. The accuracy of ECA-GCN in concentrator was 54%, while the accuracies for dump, paddy field, and rural settlement were 10%, 38%, and 55%, respectively. Our structure effectively extracts the edge information and improves the classification effect of a few categories. However, there are some categories, for which the performance of ECA-GCN remains unsatisfactory. For example, the greenhouse was wrongly divided into dry land and rural settlement. Cement road was wrongly classified as dry land. Urban land was wrongly divided into dry land and rural settlement. This is due to the small number of samples and the extremely unbalanced categories. Conversely, these parts of the mining area represent small land cover areas with vast background information. Other construction land was wrongly classified as rural settlements. This is due to the high similarity of the characteristics of these two types. Although the color characteristics are relatively different, other characteristics are relatively consistent, thus, allowing for misclassifications.

In order to further analyze the classification of MLCs under complex landscapes, we select some test set samples (such as asphaltroad, miningcatchment, and dump) and output the classification probability of each category. Refer to Table V for details, where the red box in the sample image is the general range of MLCs. In addition to the asphaltroad, the first image

also contains dryland, havewoodland, cementroad, and other categories; however, the area of the asphaltroad is relatively small. However, from the classification probability, we can see that the ECA-GCN can effectively enhance the features of the asphaltroad and improve classification accuracy. In addition to the miningcatchment, the second image also contains stope, dump and other categories. From the perspective of classification probability and classification results, it is not ideal. This may be because the miningcatchment is not only small in area, but also has similar characteristics with other categories (such as concentrator, fallowland, etc.). We believe that we can try to add geoscientific prior knowledge to the network in the future to improve the classification accuracy of difficult-to-distinguish samples. The third image includes stressvegetation, cementroad, and other categories in addition to dump. The dump is very similar to the background, and the edge is very irregular. It can also be proved from the classification probability that feature extraction is difficult. However, the classification of the ECA-GCN is correct, which proves the effectiveness and pertinence of the algorithm.

V. DISCUSSIONS

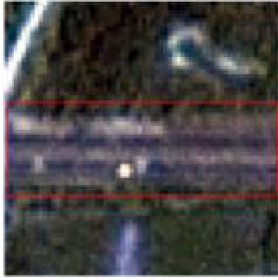


This section discusses the ablation experiments of different modules of the proposed ECA-GCN to verify the effectiveness of the proposed modules. The limitations of the CUG-MLCs dataset established in this article are also discussed.

A. Effectiveness of the Multiscale and Shallow Feature Fusion

The earliest multiscale module should be the inception module [69]. There are two main methods to improve network performance, increasing the width and/or depth of the network. However, the deeper and wider the network, the larger the number of parameters. When the dataset size is small, the network is easy to over fit. Alternatively, when the network is deep, it is easy to cause gradient disappearance and gradient explosion. This restricts the development of CNNs. The inception module solves these problems well as there are two main contributors to the inception: One is to use 1×1 convolution to increase and reduce dimensions; the other is to simultaneously perform convolution re-polymerization on multiple dimensions. Drawing on this idea, we refer to and design the effectiveness of the multiscale and shallow feature fusion module described in this article.

To verify the effectiveness of the multiscale and shallow feature fusion module, we used single branch, double branch, and four branch structures to carry out experiments. Wherein the single branch was composed of two convolution layers with a size of 3×3 and one maximum pooling layer, the first branch of the double branch consisted of two convolution layers with a size of 3×3 and one maximum pooling layer, while the second branch consisted of a convolution layer with a size of 1×1 convolution layer and one maximum pooling layer. The BN layer and ReLU activation function were added after each convolution layer.

TABLE V
CLASSIFICATION PROBABILITY OF EACH CATEGORY AND RESULT OF SAMPLE IMAGE

MLCs	Sample image	Classification probability				Result
		Label	probability	Label	probability	
Asphaltroad		0	0	10	0.01%	True
		1	0	11	0.35%	
		2	0	12	0	
		3	0	13	80.73%	
		4	0	14	0.05%	
		5	2.92%	15	0.77%	
		6	0.02%	16	0	
		7	1.33%	17	0	
		8	13.8%	18	0	
		9	0	19	0	
Miningcatchment		0	11.31%	10	0.40%	False
		1	22.66%	11	15.01%	
		2	5.02%	12	10.31%	
		3	3.31%	13	0.14%	
		4	0.83%	14	3.56%	
		5	0.1%	15	3.61%	
		6	20.92%	16	0.43%	
		7	0.17%	17	1.94%	
		8	0.02%	18	0.02%	
		9	0.10%	19	0.13%	
Dump		0	22.08%	10	17.23%	True
		1	6.94%	11	0.03%	
		2	22.62%	12	10.91%	
		3	0.06%	13	0.06%	
		4	0.92%	14	0.19%	
		5	2.44%	15	0.06%	
		6	6.02%	16	0.03%	
		7	0.04%	17	0.67%	
		8	0.85%	18	0.01%	
		9	5.58%	19	3.25%	

Bold values represent the labels and probabilities corresponding to MLCS.

TABLE VI
EXPERIMENTS ON DIFFERENT BRANCHES OF MULTISCALE AND SHALLOW FEATURE FUSION MODULE

Methods	Number of branches	OA (%)	AA (%)	Kappa (%)
ECA-GCN	1	65.86 ± 0.29	34.13 ± 1.66	54.53 ± 0.55
ECA-GCN	2	65.74 ± 0.85	34.00 ± 2.69	54.55 ± 1.30
ECA-GCN	4	66.60 ± 1.39	36.25 ± 1.50	55.91 ± 2.05

As shown in Table VI the four-branch structure of the multiscale and shallow feature fusion proposed in this article exhibited the best classification effect, which was characterized by 0.74% higher OA than that of the single branch and 0.86% higher OA than that of the double branch. This strongly proves that convolution kernels of different sizes acquire multiscale features from different scales, which is helpful for multiscale object recognition.

B. Effectiveness of the Edge Enhanced Channel Attention

To verify the effectiveness of the edge enhanced channel attention, we compared the designed module with the original SE attention and channel shuffling methods.

Channel shuffle [70] is used to solve the problem of accuracy loss caused by full constraints between channels induced by point-by-point convolution in small networks. Channel shuffle effectively strengthens the information flow between channel groups and the information representation ability. This in turn enhances the channel representation capability from another perspective. Therefore, we used it as a comparison strategy for experiments.

As shown in Table VII the edge enhanced channel attention proposed in this article has the best classification effect, showing a 1.1%, 0.63%, and 1.1% increase in OA when compared to no enhancement strategy, the use of SE attention, and the use of channel shuffle. This shows that the feature map is enhanced by the edge enhancement strategy. Fig. 8 shows the attention

TABLE VII
EXPERIMENTS OF DIFFERENT ENHANCEMENT STRATEGIES

Methods	Strategy	OA (%)	AA (%)	Kappa (%)
ECA-GCN	None	65.50 ± 1.01	35.89 ± 2.02	54.01 ± 1.40
ECA-GCN	SE Attention	65.97 ± 1.29	34.04 ± 3.03	54.28 ± 1.69
ECA-GCN	Channel Shuffle	65.50 ± 1.34	33.63 ± 2.38	53.96 ± 2.17
ECA-GCN	Edge enhanced channel attention	66.60 ± 1.39	36.25 ± 1.50	55.91 ± 2.05

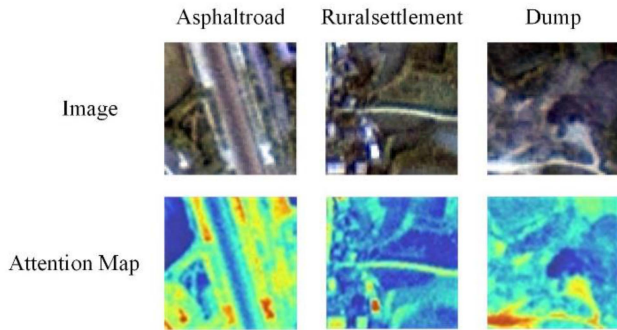


Fig. 8. Original images and the edge enhanced channel attention maps derived from Grad-CAM++.

maps of edge enhancement channel attention. We choose three scenarios: asphaltroad, ruralsettlement, and dump. It can be seen that the edge enhancement channel attention effectively captures MLCs while enhancing edge features. Thus, we showed that the method of edge enhanced channel attention is effective in strengthening the characteristics of land cover categories in mining areas with multi size and irregular edges.

C. Limitations of CUG-MLCs Dataset

Our CUG-MLCs dataset has some limitations. As exposed by the aforementioned experiments, although the proposed ECA-GCN performs best in OA, AA, and Kappa, the accuracy of all models is poor. From a dataset perspective, we identified three reasons that limit the accuracy of the model.

1) *Visual Interpretation Deviation*: Before the dataset was established in this article, labels should be given according to a representative and significant land cover types. Due to the use of large images, we have divided several areas, set up a team to interpret by block, and provided manual interpretations, which are associated with potential errors. Therefore, the accuracy of land cover types in some mining areas is affected.

2) *Dataset Size*: Because dataset production is laborious and time-consuming, we only used one remote sensing image and the corresponding DEM for dataset construction. This inevitably leads to a small dataset. However, due to the characteristics of the mining scene itself, the number of categories is extremely unbalanced, which also limits the capacity to improve the model accuracy.

3) *Strong Homogeneity*: In the mining area scenario, some categories are highly homogeneous, such as dump and fallow-land. In the case of consistent visual features, targeted modules

must be designed to effectively improve classification accuracy, which is part of our future work.

To address the above limitations, we will make the following improvements:

a) *Dataset improvement*: We will use ZY-3 images in different time phases and GF-7 images in the study area to expand the dataset. In addition, there are polymetallic mining areas in the southeast of Hubei Province and phosphate mining areas in the west of Hubei Province, which will be integrated into the scope of our research.

b) *Model improvement*: We will use a multibranch fusion strategy for experiments, thus, exploring the possibility of using multisource data to improve the accuracy of MLC classification. An important strategy to improve class imbalance is data augmentation, in addition to translation, rotation, and other strategies. As GAN is a hot topic at the moment, we will consider using GAN to expand the small number of categories and improve classification accuracy. In order to avoid excessive computation, the GCN is behind the feature extraction structure. We will study the influence of GCN location on model accuracy.

VI. CONCLUSION

Due to the urgent demand for semantic-level understanding of MLCs at the fine-scale, the study constructed a multimodal dataset named CUG-MLCs. In view of the multisize and irregular edge of the MLCs, we proposed an ECA-GCN with the following key points.

- 1) Multiscale and shallow feature fusion module, which was used to extract multiscale information and fuse the multiscale convolutional features with the shallow features.
- 2) Furthermore, the edge enhanced channel attention was used to select effective channels after a spatial edge feature enhancement.
- 3) Lastly, the edge detection-based GCN was used to construct an adjacency matrix that uses edge node relationships and learns the global contextual information.

Our results indicate that the ECA-GCN constructed in this study achieved an OA of 66.60%, AA of 36.25%, and Kappa of 55.91% on the CUG-MLCs dataset, and outperformed the classical CNN and recent networks. Thus, the proposed model is adequate for the fine classification of complex landscapes. In the future, on one hand, we will aim to improve the dataset. On the other hand, we will focus on multimodal and multi-branch feature learning and fusion, class imbalance learning, and multilabel scene classification.

APPENDIX

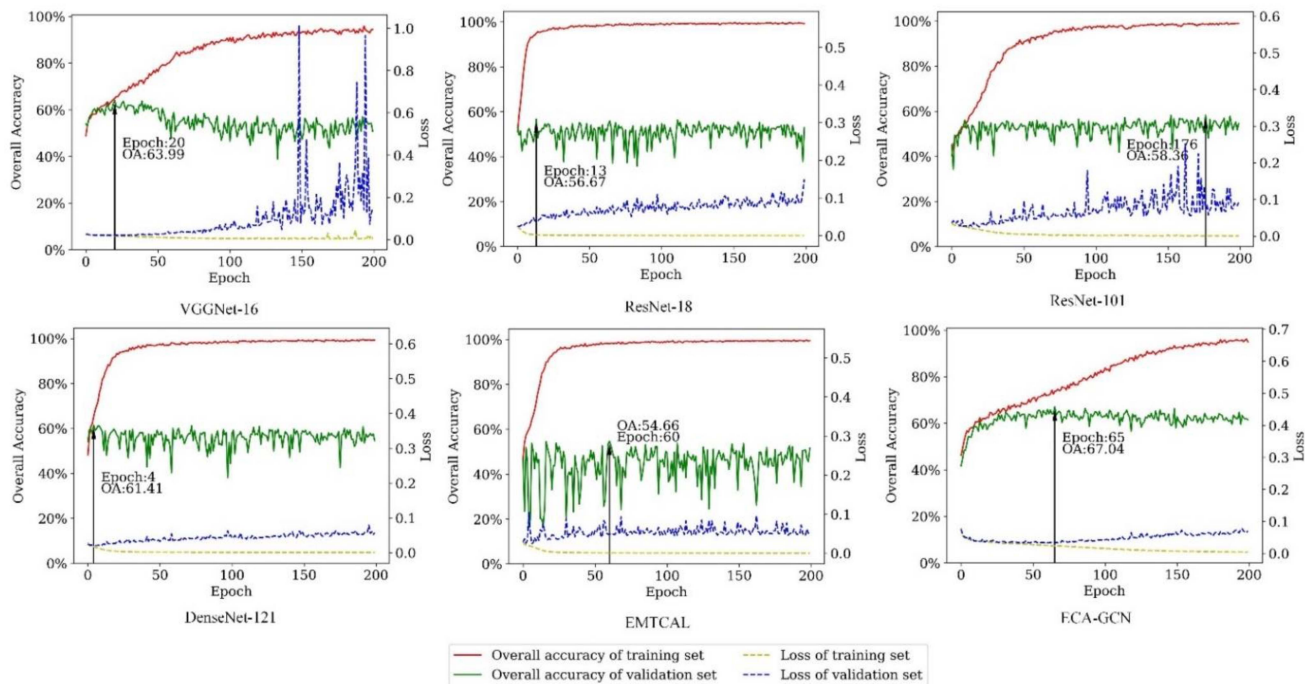


Fig. A1. Optimization of the parameters of the selected algorithm. VGGNet-16, ResNet-18, ResNet-101, DenseNet-121, EMTCAL, and ECA-GCN were assessed.

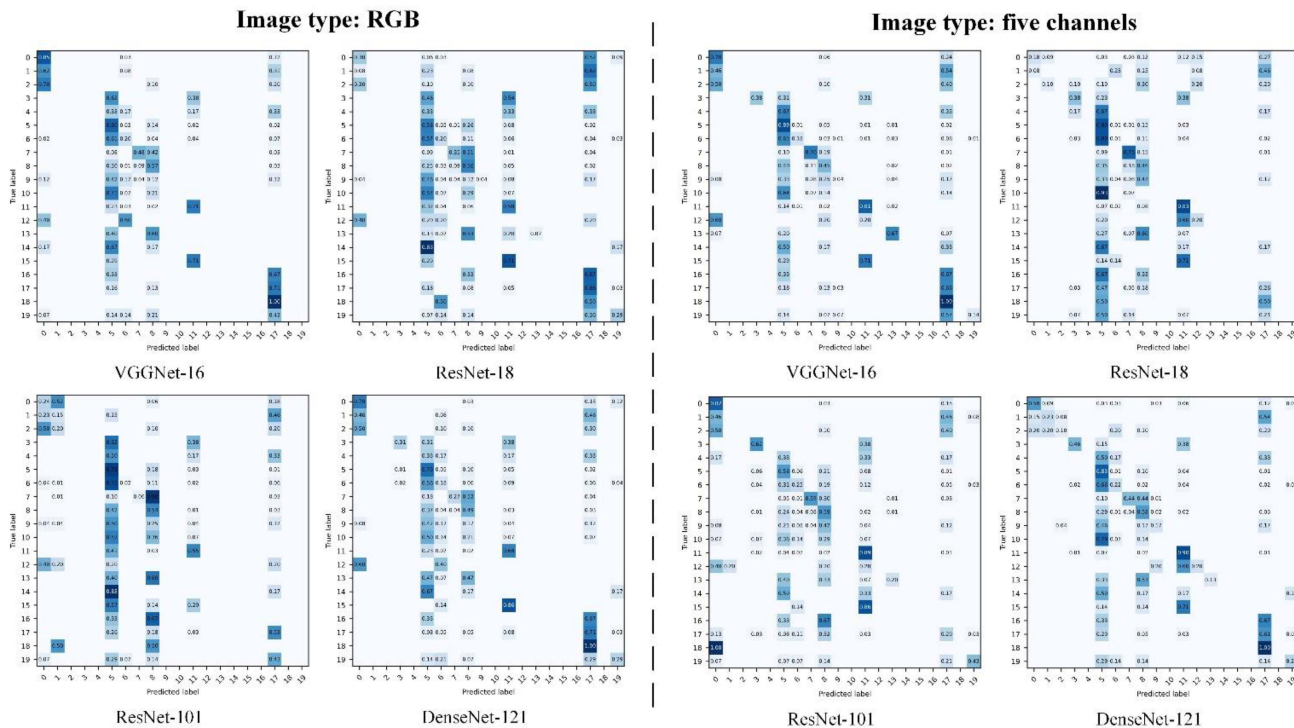


Fig. A2. VGGNet-16, ResNet-18, ResNet-101, DenseNet-121 confusion matrix (image type: RGB and five channels).

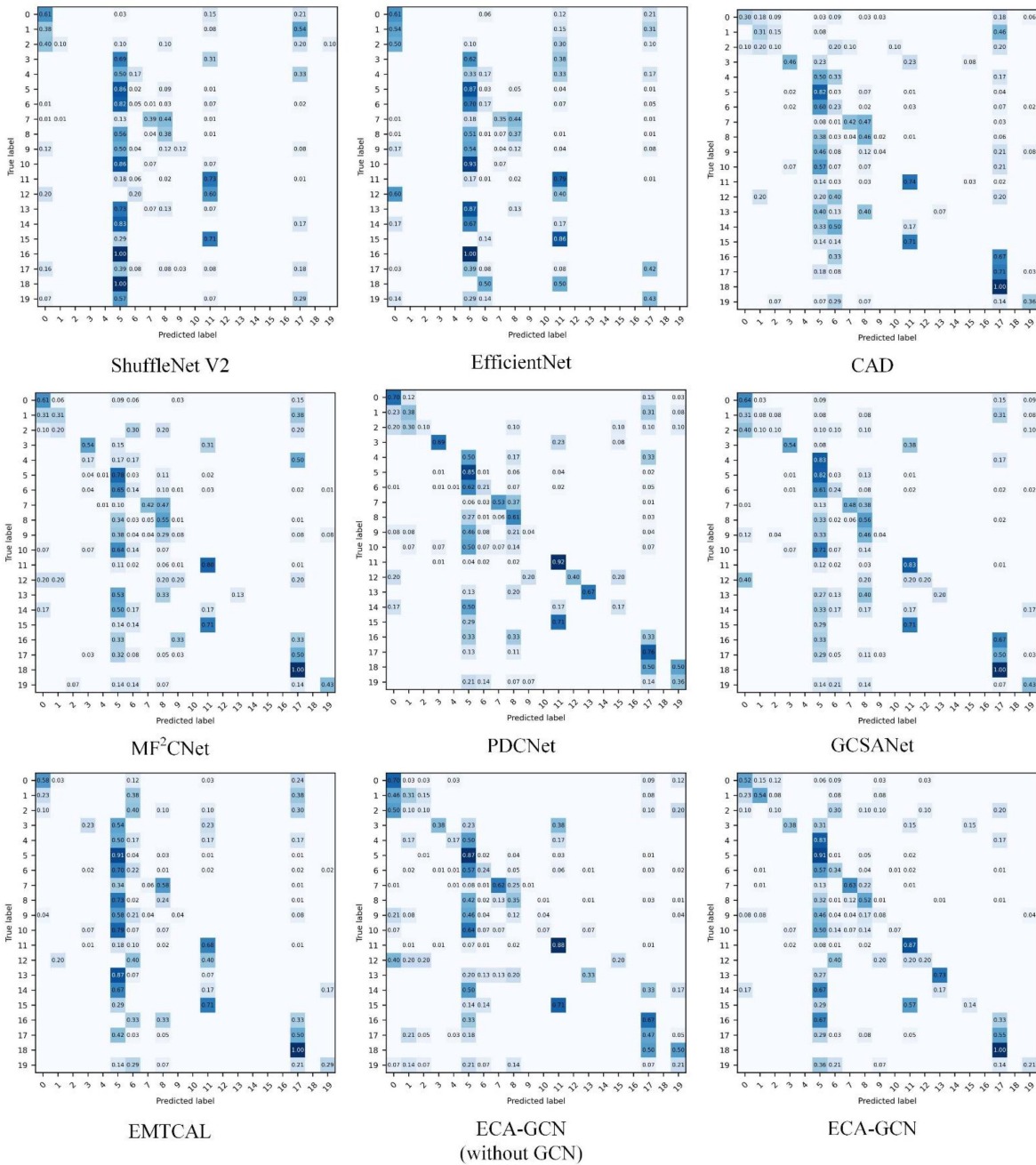


Fig. A3. ShuffleNet V2, EfficientNet, CAD, MF²CNet, PDCNet, GCSANet, EMTCAL, ECA-GCN(without GCN) and ECA-GCN Confusion Matrix (image type: five channels).

ACKNOWLEDGMENT

The authors would like to thank Z. Dai, who studied at the University of Electronic Science and Technology of China, and gave advice on drawing.

REFERENCES

[1] P. H. Abelson, "Constraints on greenhouse gas emission," *Science*, vol. 278, no. 5339, 1997, Art. no. 783.

[2] D. Touma, S. Stevenson, F. Lehner, and S. Coats, "Human-driven greenhouse gas and aerosol emissions cause distinct regional impacts on extreme fire weather," *Nature Commun.*, vol. 12, no. 1, pp. 1–8, 2021.

[3] K. R. Gurney, J. Liang, G. Roest, Y. Song, K. Mueller, and T. Lauvaux, "Under-reporting of greenhouse gas emissions in US cities," *Nature Commun.*, vol. 12, no. 1, pp. 1–7, 2021.

[4] D. Li, G. Huang, S. Zhu, L. Chen, and J. Wang, "How to peak carbon emissions of provincial construction industry? Scenario analysis of Jiangsu Province," *Renewable Sustain. Energy Rev.*, vol. 144, 2021, Art. no. 110953.

[5] M. Finkbeiner and V. Bach, "Life cycle assessment of decarbonization options—Towards scientifically robust carbon neutrality," *Int. J. Life Cycle Assessment*, vol. 26, no. 4, pp. 635–639, 2021.

[6] D. Niu, G. Wu, Z. Ji, D. Wang, Y. Li, and T. Gao, "Evaluation of provincial carbon neutrality capacity of China based on combined weight and improved TOPSIS model," *Sustainability*, vol. 13, no. 5, 2021, Art. no. 2777.

- [7] M. Salvia et al., "Will climate mitigation ambitions lead to carbon neutrality? An analysis of the local-level plans of 327 cities in the EU," *Renewable Sustain. Energy Rev.*, vol. 135, 2021, Art. no. 110253.
- [8] Z. Bai and J. Zhao, "On several problems of land reclamation of open-pit mine and ecology rebuilding," *Metal Mine Des. Construction*, vol. 1, pp. 33–37, 2000.
- [9] L. Sonter et al., "Mining drives extensive deforestation in the Brazilian Amazon," *Nature Commun.*, vol. 8, no. 1, pp. 1–7, 2017.
- [10] K. Johansen, P. Erskine, and M. McCabe, "Using unmanned aerial vehicles to assess the rehabilitation performance of open cut coal mines," *J. Cleaner Prod.*, vol. 209, pp. 819–833, 2019.
- [11] X. Hou et al., "Variability of environmental factors and the effects on vegetation diversity with different restoration years in a large open-pit phosphorite mine," *Ecological Eng.*, vol. 127, pp. 245–253, 2019.
- [12] S. Li et al., "Incorporating ecological risk index in the multi-process MCRE model to optimize the ecological security pattern in a semi-arid area with intensive coal mining: A case study in Northern China," *J. Cleaner Prod.*, vol. 247, 2020, Art. no. 119143.
- [13] C. Liu, W. Xiong, Y. Zhang, Y. Sun, M. Xiong, and C. Guo, "Research on invulnerability technology of node attack in space-based information network based on complex network," *Electronics*, vol. 8, no. 5, 2019, Art. no. 507.
- [14] R. DeFries and A. Belward, "Global and regional land cover characterization from satellite data: An introduction to the special issue," *Int. J. Remote Sens.*, vol. 21, nos. 6/7, pp. 1083–1092, 2000.
- [15] J. Cihlar, "Land cover mapping of large areas from satellites: Status and research priorities," *Int. J. Remote Sens.*, vol. 21, nos. 6/7, pp. 1093–1114, 2000.
- [16] J. Chen et al., "Global land cover mapping at 30 m resolution: A POK-based operational approach," *Int. Soc. Photogrammetry Remote Sens., J. Photogrammetry Remote Sens.*, vol. 103, pp. 7–27, 2015.
- [17] P. Gong et al., "Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data," *Int. J. Remote Sens.*, vol. 34, nos. 7/8, pp. 2607–2654, 2013.
- [18] P. Sellers et al., "Remote sensing of the land surface for studies of global change: Models-algorithms-experiments," *Remote Sens. Environ.*, vol. 51, no. 1, pp. 3–26, 1995.
- [19] Q. Lv, Y. Dou, X. Niu, J. Xu, and B. Li, "Classification of land cover based on deep belief networks using polarimetric RADARSAT-2 data," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2014, pp. 4679–4682.
- [20] A. Madasa, I. Orimoloye, and O. Olofade, "Application of geospatial indices for mapping land cover/use change detection in a mining area," *J. Afr. Earth Sci.*, vol. 175, 2021, Art. no. 104108.
- [21] W. Chen, X. Li, and L. Wang, "Fine land cover classification in an open pit mining area using optimized support vector machine and worldview-3 imagery," *Remote Sens.*, vol. 12, no. 1, 2019, Art. no. 82.
- [22] T. Chen et al., "Object-oriented open-pit mine mapping using Gaofen-2 satellite image and convolutional neural network, for the Yuzhou City, China," *Remote Sens.*, vol. 12, no. 23, 2020, Art. no. 3895.
- [23] T. Chen, J. Trinder, and R. Niu, "Object-oriented landslide mapping using ZY-3 satellite imagery, random forest and mathematical morphology, for the Three-Gorges Reservoir, China," *Remote Sens.*, vol. 9, no. 4, 2017, Art. no. 333.
- [24] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
- [25] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2010, pp. 270–279.
- [26] F. Hu, G. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14680–14707, 2015.
- [27] B. Zhao, Y. Zhong, G.-S. Xia, and L. Zhang, "Dirichlet-derived multiple topic scene classification model fusing heterogeneous features for high resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2108–2123, Apr. 2016.
- [28] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2321–2325, Nov. 2015.
- [29] L. Zhao, P. Tang, and L. Huo, "Feature significance-based multibag-of-visual-words model for remote sensing image scene classification," *J. Appl. Remote Sens.*, vol. 10, no. 3, pp. 035004–035004, 2016.
- [30] G.-S. Xia et al., "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.
- [31] S. Basu, S. Ganguly, S. Mukhopadhyay, R. Dibiano, M. Karki, and R. Nemani, "DeepSAT—A learning framework for satellite imagery," in *Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2015, pp. 1–10.
- [32] O. Penatti, K. Nogueira, and J. Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2015, pp. 44–51.
- [33] X. Li et al., "Multimodal and multi-model deep fusion for fine classification of regional complex landscape areas using ZiYuan-3 imagery," *Remote Sens.*, vol. 11, no. 22, 2019, Art. no. 2716.
- [34] W. Chen et al., "A review of fine-scale land use and land cover classification in open-pit mining areas by remote sensing techniques," *Remote Sens.*, vol. 10, no. 1, 2017, Art. no. 15.
- [35] W. Chen, X. Li, H. He, and L. Wang, "Assessing different feature sets' effects on land cover classification in complex surface-mined landscapes by ZiYuan-3 satellite imagery," *Remote Sens.*, vol. 10, no. 1, 2017, Art. no. 23.
- [36] Z. Zhang, G. He, M. Wang, Z. Wang, T. Long, and Y. Peng, "Detecting decadal land cover changes in mining regions based on satellite remotely sensed imagery: A case study of the stone mining area in Luoyuan County, SE China," *Photogrammetric Eng. Remote Sens.*, vol. 81, no. 9, pp. 745–751, 2015.
- [37] Z. Kassouk, J. Thouret, A. Gupta, A. Solikhin, and S. Liew, "Object-oriented classification of a high-spatial resolution SPOT5 image for mapping geology and landforms of active volcanoes: Semeru case study, Indonesia," *Geomorphology*, vol. 221, pp. 18–33, 2014.
- [38] X. Zeng, Z. Liu, C. He, Q. Ma, and J. Wu, "Detecting surface coal mining areas from remote sensing imagery: An approach based on object-oriented decision trees," *J. Appl. Remote Sens.*, vol. 11, 2017, Art. no. 015025.
- [39] M. Li, Z. Tang, W. Tong, X. Li, W. Chen, and L. Wang, "A multi-level output-based DBN model for fine classification of complex geoenvironments area using ZiYuan-3 TMS imagery," *Sensors*, vol. 21, no. 6, 2021, Art. no. 2089.
- [40] F. Zhao, J. Zhang, Z. Meng, and H. Liu, "Densely connected pyramidal dilated convolutional network for hyperspectral image classification," *Remote Sens.*, vol. 13, no. 17, 2021, Art. no. 3396.
- [41] D. Liu et al., "A discriminative spectral-spatial-semantic feature network based on shuffle and frequency attention mechanisms for hyperspectral image classification," *Remote Sens.*, vol. 14, 2022, Art. no. 2678.
- [42] M. Xia, Y. Cui, Y. Zhang, Y. Xu, J. Liu, and Y. Xu, "DAU-Net: A novel water areas segmentation structure for remote sensing image," *Int. J. Remote Sens.*, vol. 42, no. 7, pp. 2594–2621, 2021.
- [43] M. Zhang, Q. Cheng, F. Luo, and L. Ye, "A triplet nonlocal neural network with dual-anchor triplet loss for high-resolution remote sensing image retrieval," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2711–2723, 2021.
- [44] S. Mei, K. Yan, M. Ma, X. Chen, S. Zhang, and Q. Du, "Remote sensing scene classification using sparse representation-based framework with deep feature fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5867–5878, 2021.
- [45] S. Gu, R. Zhang, H. Luo, M. Li, H. Feng, and X. Tang, "Improved SinGAN integrated with an attentional mechanism for remote sensing image classification," *Remote Sens.*, vol. 13, 2021, Art. no. 1713.
- [46] W. Tong, W. Chen, W. Han, X. Li, and L. Wang, "Channel-attention-based DenseNet network for remote sensing image scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4121–4132, 2020.
- [47] S. Ouyang, J. Xu, W. Chen, Y. Dong, X. Li, and J. Li, "A fine-grained genetic landform classification network based on multimodal feature extraction and regional geological context," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4511914.
- [48] W. Chen, S. Ouyang, W. Tong, X. Li, X. Zheng, and L. Wang, "GCSANet: A global context spatial attention deep learning network for remote sensing scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1150–1162, 2022.
- [49] J. Liu et al., "An investigation of a multidimensional CNN combined with an attention mechanism model to resolve small-sample problems in hyperspectral image classification," *Remote Sens.*, vol. 14, 2022, Art. no. 785.
- [50] D. Guo, Y. Xia, and X. Luo, "Self-supervised GANs with similarity loss for remote sensing image scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2508–2521, 2021.
- [51] A. Ma, J. Wang, Y. Zhong, and Z. Zheng, "Factseg: Foreground activation-driven small object semantic segmentation in large-scale remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5606216.

- [52] X. Yang, Z. Chen, B. Zhang, B. Li, Y. Bai, and P. Chen, "A block shuffle network with superpixel optimization for Landsat image semantic segmentation," *Remote Sens.*, vol. 14, no. 6, 2022, Art. no. 1432.
- [53] J. Liang, Y. Deng, and D. Zeng, "A deep neural network combined CNN and GCN for remote sensing scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4325–4338, 2020.
- [54] Z. Zhang et al., "Et-net: A generic edge-attention guidance network for medical image segmentation," in *Proc. Med. Image Comput. Comput. Assist. Interv. 22nd Int. Conf.*, 2019, pp. 13–17.
- [55] R. Wang et al., "Boundary-aware context neural network for medical image segmentation," *Med. Image Anal.*, vol. 78, 2022, Art. no. 102395.
- [56] Z. Zhang et al., "Collaborative boundary-aware context encoding networks for error map prediction," *Pattern Recognit.*, vol. 125, 2022, Art. no. 108515, doi: [10.1016/j.patcog.2021.108515](https://doi.org/10.1016/j.patcog.2021.108515).
- [57] Q. Liu, L. Xiao, J. Yang, and Z. Wei, "CNN-enhanced graph convolutional network with pixel-and superpixel-level feature fusion for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8657–8671, Oct. 2021.
- [58] G. Zhou, W. Chen, Q. Gui, X. Li, and L. Wang, "Split depth-wise separable graph-convolution network for road extraction in complex environments from high-resolution remote-sensing Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5614115.
- [59] Y. Bazi, L. Bashmal, M. Rahhal, R. Dayil, and N. Ajlan, "Vision transformers for remote sensing image classification," *Remote Sens.*, vol. 13, no. 3, 2021, Art. no. 516.
- [60] X. Tang, M. Li, J. Ma, X. Zhang, F. Liu, and L. Jiao, "EMTCAL: Efficient multiscale transformer and cross-level attention learning for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5626915.
- [61] X. Li, W. Chen, X. Cheng, and L. Wang, "A comparison of machine learning algorithms for mapping of complex surface-mined and agricultural landscapes using ZiYuan-3 stereo satellite imagery," *Remote Sens.*, vol. 8, 2016, Art. no. 514.
- [62] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
- [63] S. Woo, J. Park, J. Lee, and I. Kweon, "Cbam: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.
- [64] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: A comprehensive review," *Comput. Social Netw.*, vol. 6, no. 1, pp. 1–23, 2019.
- [65] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.
- [66] N. Ma, X. Zhang, H. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 116–131.
- [67] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn. Mach. Learn. Res.*, 2019, pp. 6105–6114.
- [68] L. Bai, Q. Liu, C. Li, Z. Ye, M. Hui, and X. Jia, "Remote sensing image scene classification using multiscale feature fusion covariance network with octave convolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5620214.
- [69] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.
- [70] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6848–6856.



Haoyi Wang received the B.Eng. degree in computer science and technology from the China University of Geosciences, Wuhan, China, in 2016. He is currently working toward the master's degree in the Faculty of Computer Science, China University of Geoscience, Wuhan, China.

His research interests include remote sensing image process, computer vision, and deep learning.



Xianju Li (Member, IEEE) received the B.S. degree in geomatics engineering, the M.S. degree in geodesy and survey engineering, and the Ph.D. degree in surveying and mapping from China University of Geoscience, Wuhan, China, in 2009, 2012, and 2016, respectively.

Since 2016, he has been an Associate Professor with the School of Computer Science, China University of Geosciences. He has authored and coauthored more than 20 papers. His main research interests include remote sensing image processing and analysis, computer vision, and machine learning.



Gaodian Zhou received the B.S. in resource prospecting engineering and the M.S. degree in mineral resource prospecting and exploration from the China University of Geosciences, Wuhan, China, in 2014 and 2017, respectively. He is currently working toward the Ph.D. degree with the School of Computer Science, China University of Geoscience, Wuhan, China.

His research interests include semantic segmentation, remote sensing image process, and big data.



Weitao Chen (Member, IEEE) was born in Wugang, China. He received the B.E. degree in land resource management from Jiaozuo Institute of Technology, Jiaozuo, China, in 2003, the M.E. degree in quaternary geology from China University of Geoscience, Wuhan, China, in 2006, and the Ph.D. degree in environmental science and engineering from China University of Geosciences, Wuhan, China, in 2012.

He is a Professor with the School of Computer Science, China University of Geosciences. He has authored and coauthored more than 30 papers. His main research interests include machine learning and remote sensing of environment.



Lizhe Wang (Fellow, IEEE) received the B.E. and M.E. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1998 and 2001, respectively, and the Doctor of Engineering degree from the University Karlsruhe (Magna Cum Laude), Karlsruhe, Germany, in 2008.

He is a ChuTian Chair Professor with the School of Computer Science, China University of Geosciences, Wuhan, China. His research interests include HPC, e-Science, and remote sensing image processing.

Dr. Wang is a Fellow of IET and British Computer

Society.