This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LGRS.2022.3177778, IEEE Geoscience and Remote Sensing Letters

Contextual-aware Land Cover Classification with U-shaped Object Graph Neural Network(U-OGNN)

Wenzhi Zhao, Shu Peng, and Jiage Chen, Rui Peng,

***Abstract*—The timely and accurate land cover mapping with remote sensing images played a huge role in ecosystem monitoring. However, due to the spectral variability and spatial complexity of high-resolution remote sensing images, it is often difficult to find an efficient method to achieve accurate land cover classification. To explore useful contextual and hierarchical features in remote sensing images, this paper proposes a U-shaped object graph neural network (U-OGNN), which is mainly composed of self-adaptive graph construction (SAGC), hierarchical graph encoder, and decoder. For self-adaptive graph construction, the similarity measurement is applied to generate contextual-aware graph structure, by feeding deep features extracted from convolution and multi-layer attention operations. Graph encoder and decoder fuse multi-level information over different scales by capturing hierarchical features of adjacent objects. In this way, the proposed method is able to predict land-cover types by considering multi-level contextual information accurately. Experiments on GID land-cover classification datasets, the overall accuracies of the U-OGNN reach 87.81 %.**

***Index Terms*—land-cover classification, graph neural network (GNN), OBIA, Self-adaptive graph construction (SAGC), Graph encoder and decoder.**

I. Introduction

LAND cover information refers to the coverage formed by the current natural and human influence on the earth surface, reflecting the natural state of the earth surface, such as forests, meadow, and bare land, etc[1]. Land cover maps deepen our understanding of the morphological changes of the Earth's surface and play a significant role in the Earth's ecosystems and environmental monitoring[2-4]. However, with considerable progress in processing Earth observation data, the large-scale land cover classification remains a formidable task for high-resolution satellite imagery[5, 6]. In higher spatial resolution images, inter-class spectral similarity and intra-class difference[7, 8] are more prominent, which aggravates the difficulty of efficient land-cover mapping.

Over the past decades, remote sensing classification methods have been developed rapidly in terms of remote sensing information extraction. These methods can be roughly divided into two categories (pixels and objects) according to the processing units. Pixel-level methods[9] mainly rely on spectral characteristics to classify interesting targets. Still, they are usually difficult to achieve good accuracy and have serious salt and pepper noise due to limited spatial pattern awareness. Therefore, the sliding window analysis technique was applied to extract spatial features (i. e., texture information)[10, 11]and fuse them with spectral features for target identification. However, it is worth noting that such mentioned above methods are often difficult to capture the shape feature of interesting targets, let alone to formulate the spatial relationship between ground targets. The object-based image analysis (OBIA) method is proposed[12–14] to tackle this problem. This method generates relatively uniform objects (i.e., candidates of interested targets) according to prior spectral, geometric, and spatial characteristics. Then, the attributes of these objects can be fed into the classifier to distinguish various types of land covers. However, the classification accuracy for both pixel-based and object-based methods depends on the robustness of hand-crafted image features.

To handle the problem of robust feature generation and selection, the deep convolutional neural network (DCNN) as an automatic feature extraction method has received significant attention. In this scope, FCN [15] and U-Net[16] are the most representative model structures in robust feature extraction and land cover classification. For example, G. Fu et al.[17] designed a multi-scale fully connected network by introducing 2D convolution and jump connection, significantly improving classification accuracy in high-resolution remote sensing imagery interpretation. Meanwhile, ResUnet[18] is also able to capture high-level semantic features across different scales with the residual unit and 2D convolution. Although these DCNNs can extract high-level representative features, the spatial relationship (i.e., contextual information) between ground targets is still unexploited. For example, buildings are usually surrounded by city roads and far away from farmland. By formulating such contextual information, it is possible to perform efficient land cover mapping and reasoning.[19,20]

W. Zhao, and R. Peng are with the State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Aerospace Information Research Institute of Chinese Academy of Sciences, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, and Beijing Engineering Research Center for Global Land Remote Sensing Products, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, and National Geomatics Center of China, Beijing 100830, China. (corresponding author: Shu Peng, Jiage Chen)

S. Peng and J. Chen are with the National Geomatics Center of China, Beijing 100830, China.

1545-598X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Beijing Normal University. Downloaded on May 27,2022 at 00:44:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LGRS.2022.3177778, IEEE Geoscience and Remote Sensing Letters

In order to consider the critical relationship between adjacent targets, graph neural networks (GNN)[21] have been intensively studied in recent years. It is mainly divided into two categories: the spectral- and the spatial-domain GNN. The spectral domain graph neural network converts the original data into the frequency domain for convolution processing and then restored to the spatial domain. For instance, Hong et al. proposed a Graph Convolutional Networks for Hyperspectral Image Classification [22], which replaces the convolution kernel of the spectral domain. However, the spectral-domain network has the problems of time-consuming in terms of Laplacian matrix decomposition calculation. Also, it is unable to ensure local connection in the spectral domain. Therefore, ChebNet[23] uses Chebyshev polynomial to replace the convolution kernel in the spectral domain, which reduced the computational complexity of the model parameters greatly. Moreover, the GCN[24] further simplified convolution operation based on ChebNet, as only considering first-order Chebyshev polynomial for convolution simplification. Similarly, in order to extract contextual information from the spatial domain, GNN[25] uses a random walk strategy to select a fixed number of neighbor nodes and sort them according to the expected size. To relax the constraint of fixed adjacent nodes, the SAGE[26] considers neighborhood nodes by using a uniform sampling strategy and then gather the information of neighborhood nodes to predict the label in general. However, Veličković et al. [27] believe that all the neighborhood nodes share convolution kernel parameters that will limit the model's performance due to the different correlations between each node and the central node. Therefore, graph attention networks(GAT) utilize the attention mechanism to calculate the correlation across nodes, and then adjacent nodes are aggregated to complete the graph convolution process.

In these graph neural networks, topological spatial relationships can be well constructed between different nodes, which provides a possibility for intelligent interpretation of remote sensing images. Compared with CNN, graph neural networks can effectively use the global representation and remote context dependence in remote sensing to a certain extent. Therefore, Liu et al.[28] proposed a self-constructed graph neural network (SCG-Net) for pixel-level semantic segmentation. However, GCNs usually suffer from a huge computational cost in large-scale remote sensing image interpretation. The minibatch GCN [29] is proposed to lift this problem, as it allows to train large-scale GCNs in a minibatch fashion. Neverthless, there are still two challenges that hinder further improvement on large-scale remote sensing classification. For one thing, long-range dependency is difficult to be formulated at the pixel level. For another, hierarchical feature representation has not been evaluated in the field of graph-based remote sensing classification.

To solve the above problem, this paper proposes a U-shaped object-based graph neural network(U-OGNN) to capture hierarchical contextual information for efficient land-cover classification. Specifically, the image object is firstly generated by jointly considering spectral and geometric features. Then, the image object is regarded as the individual node to be fed into the U-OGNN for contextual relationship formulation. The innovation of this study can be summarized as follows: 1) A self-adaptive graph construction (SAGC) strategy is proposed, it formulates the contextual information without increasing the number of trainable parameters. 2) The U-shaped structure is developed to capture hierarchical contextual features for multi-level image interpretation. 3) The U-shaped object-oriented graph neural network (U-OGNN) is generated to jointly explore both the object-based features and adjacent contextual information for efficient land-cover classification.

The rest of this paper is organized as follows: Section 2 describes the implementation process of U-OGNN in detail. In section 3, the research data source and experiments are introduced. The conclusion is drawn in the last section.

II. The Proposed U-OGNN

In this section, we propose a U-OGNN method for land-cover classification, as shown in Fig.1. The workflow of this method includes the following steps: 1) Node feature extraction and self-adaptive construction graph (SAGC). 2) A U-shaped object-oriented graph neural network is designed to extract hierarchical contextual features (U-OGNN).

A. Node feature extraction and self-adaptive graph construction (SAGC)

The proposed method starts with the initial image segmentation to realize object-based image representation. SLIC algorithm[30], as a super-pixel segmentation method, is used to divide the image into homologous objects with input spectral features. Extract geometric features from different objects, such as s area, perimeter, and solidity. The initial edges in the graph are constructed by capturing the object's first-order neighbors, and the object's spectral and geometric features are taken as the initial node attributes.

To be specific, the convolution layer (1×1 Conv) and the multi-headed attention block (MAB) are used to extract the node features of the high level. To fuse the spectral and geometric features within the nodes, the 1×1 Conv is used to integrate the attribute information before attention-based refinement linearly.

To generate attention-refined deep features, the conventional self-attention[31] has three different matrices that derived from generated features, namely Query matrix ($\mathrm{Q} \in \mathbb{R}^{n\times d_q}$), Key matrix ($\mathrm{K} \in \mathbb{R}^{n\times d_k}$) and Value matrix ($V \in \mathbb{R}^{n\times d_v}$). They are obtained by multiplying the features of the nodes by three different weight matrices $W^q \in \mathbb{R}^{d\times d_q}$, $W^k \in \mathbb{R}^{d\times d_k}$ and $W^v \in \mathbb{R}^{d\times d_v}$. We calculate the single attention function as:

$$\text{Att }(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (1)$$

In order to learn robust information from different representation subspaces, multi-head attention[30, 31] can be repeated by several times. Each additional layer of attention can extract different representation information from these objects. The formula is defined as follows :

1545-598X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Authorized licensed use limited to: Beijing Normal University. Downloaded on May 27,2022 at 00:44:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LGRS.2022.3177778, IEEE Geoscience and Remote Sensing Letters

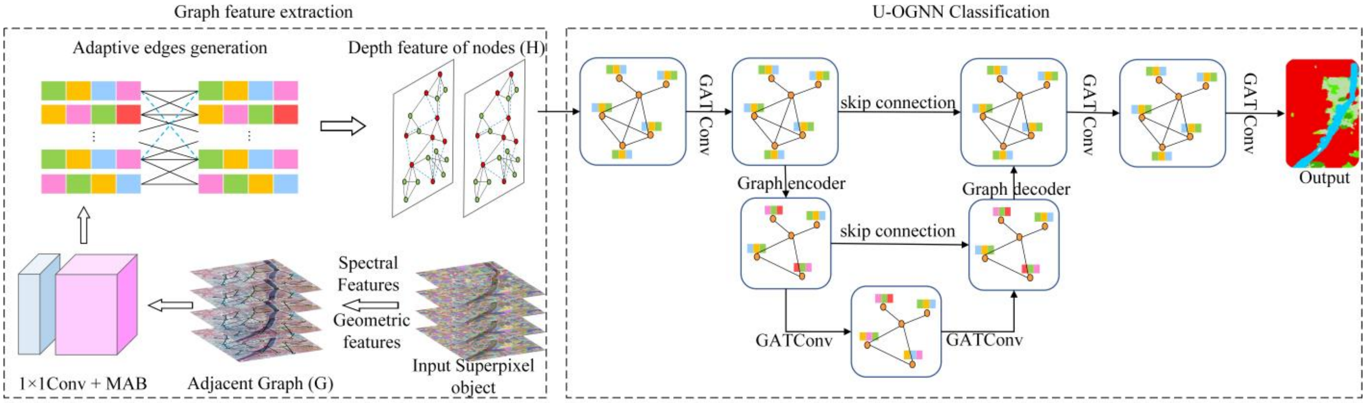

Fig. 1. An overview of our Object-oriented U-shaped graph neural network (U-OGNN).

$$\text{MultiHead}(Q,K,V)=\text{Cat}\left(Head_1,\cdots,Head_i\right)W^{O} \qquad (2)$$

$$\text{where } Head_j=\text{Att}\left(QW_j^{Q},KW_j^{K},VW_j^{V}\right) \qquad (3)$$

It is worth noting that $W_j^Q \in \mathbb{R}^{d_q\times d_q^M}$, $W_j^K \in \mathbb{R}^{d_k\times d_k^M}$, $W_j^V \in \mathbb{R}^{d_v\times d_v^M}$ and $W^o \in \mathbb{R}^{id_q^M\times d}$ in multi-head attention represent merging all attention into a larger feature matrix.

Since the deep feature is generated by combining convolution layer and multi-head attention mechanism, the graph construction by adaptively assigning edges across adjacent nodes is proposed. Specifically, in order to achieve the purpose of adaptive graph edge determination, the Kernel depth information angle mapping (KDAM) automatically generates graph edges by measuring the similarity between nodes' features. Different from the spatially adjacent edges, it maintains the relationship of similar nodes regardless of the spatial ranges. As shown in Fig.2, the depth feature $H \in \mathbb{R}^{n\times d}$ of the four nodes is represented by different colours. It can be seen that the characteristics of h_1 and h_3 are very similar. The relationship between remote objects is captured by constructing the edge of feature similarity to integrate global context features.

Algorithm 1 Self-adaptive graph construction (SAGC) based on the node depth feature

```
Input: G, X   # Graph, Attribute of the initial node
Output: G,H  # Adaptive generation graph, Node feature
Strat
# Using 1 × 1 convolution layer ( $1\times1Conv$ ).
$H = 1\times1Conv(X)$
# Using Multi-Head Attention blocks(MAB).
$H = MAB(G, H)$
$n,d = H.shape$
# Compute depth feature similarity between nodes.
$Y_1,Y_2 = H.reshape(n,1,d), H.reshape(1,n,d)$
# Using the broadcast mechanism where $\sigma^2 = 0.5$ .
$T = th.\exp\left(-th.abs\left(Y_1,Y_2\right).reshape\left(n,n,d\right)^2/\left(2\sigma^2\right)\right)$
$S = th.\arccos\left(T\right).sum\left(axis=2\right).reshape\left(n,n\right)$
# Find edges with high similarity.
$E = th.nonzero\left(S + th.eye(n) < 0.1\right).t()$
# Add edges to the graph structure.
$G.add\_edges\left(E[0],E[1]\right)$
return G,H
End
```

Therefore, we use the KDAM strategy to generate blue values in the adjacency matrix. Its formula can be formulated as:

$$S(Y_1,Y_2)=\arccos\left(\frac{k(Y_1,Y_2)}{\| k(Y_1,Y_2)\|\cdot\| k(Y_1,Y_2)\|}\right) \qquad (4)$$

$$k(Y_1,Y_2)=\exp\left(-\| Y_1-Y_2\|^2/2\sigma^2\right) \qquad (5)$$

$Y_1 \in \mathbb{R}^{n\times 1\times d}$ and $Y_2 \in \mathbb{R}^{1\times n\times d}$ are the same node features but have different dimensions. $S \in \mathbb{R}^{n\times n}$ is a similarity matrix, where the smaller the element $S_{ij} \in [0, \pi/2]$ represents the higher similarity.

The SAGC module is mainly composed of the above two steps, including node feature extraction based on multi-layer attention blocks and adaptive edge generation based on KDAM. Details are illustrated in Algorithm 1.

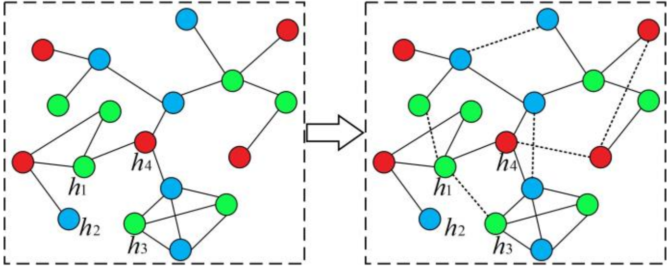

Fig. 2.Adaptive generation graph structure based on deep features of node.

B. U-OGNN

A U-shaped object-oriented graph neural network is proposed for land cover classification with hierarchical contextual information. This framework mainly includes the graph attention network convolutional layer (GATConv), graph encoder, decoder modules, and skip connection. The purpose of the graph encoder is to learn node feature representation and graph structure information at different levels. The graph decoder aligns the super-node features with the original node features by a reflection to ensure the fusion of different levels of features. In order to maintain contextual information at different levels, skip connection is applied to transfer node connectivity and keep detailed information.

Graph Encoder. To achieve the purposed of efficient classification, we developed a graph encoder to construct hierarchical feature representation. In order to build abstract feature representations at different levels, we introduced the super-node technique to aggregate low-level nodes. However, the traditional adjacency matrix is difficult to be aggregated

1545-598X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Beijing Normal University. Downloaded on May 27,2022 at 00:44:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LGRS.2022.3177778, IEEE Geoscience and Remote Sensing Letters

with 0 and 1 values. To handle this problem, we construct the weighted adjacency matrix by the measuring similarity of the adjacent objects with GATConv.

$$\mathrm{H}^l, \mathrm{Z} = GATConv(G, H^{l-1}) \tag{6}$$

$$A' = dense\left(sparse\left(E, Z\right)\right) \tag{7}$$

Where Z is the sparse edge weight learned from GATconv layer of the U-OGNN graph network, and A' is the weighted adjacency matrix constructed by Z .

To generate a higher-level feature map, the proposed U-OGNN uses the distribution matrix $P \in \mathbb{R}^{m\times n}$ to perform a sub-sampling operation. Therefore, the graph encoder introduces a learnable super-node clustering module as:

$$\hat{A} = PA'P^T \tag{8}$$

$$H' = PH^l \tag{9}$$

Where $\hat{A} \in \mathbb{R}^{m\times m}$ is the adjacency matrix of the super-nodes and $H' \in \mathbb{R}^{m\times d(l)}$ is the feature of super-nodes.

Graph Decoder. In this section, the purpose of the graph decoder is to align and fuse the features of the super-node with those of the original node and to map the features of the super node back to the dimension before compression through the allocation matrix $P' \in \mathbb{R}^{n\times m}$. This formula can be described as:

$$H^{l+1} = P'H' \tag{10}$$

III. EXPERIMENTS

A. Data set and Experimental setup

In this study, we tested the U-OGNN model with the public Gaofen image dataset (GID)[32]. The GID dataset contains 150 GF Nir-RGB images, and each image has a dimension of 6800×7200 pixels with a spatial resolution of 1m. To reduce the computational cost of graph-based U-OGNN and facilitate the comparison with Unet and ResUNet, we crop them into 256×256 patches and choose 60 % of the samples for training and the rest for testing.

The U-OGNN model is implemented by Pytorch version 1.8.0 and the Deep Graph Library framework. The experimental computer is Intel (R) Xeon (R) Silver 4210R CPU-2.40GHz, NVIDIA RTX 3090 graphics processor. This model uses the high-level semantic features extracted by CNN and MAB. At the same time, we use the SAGC and U-OGNN modules to enhance the migration ability of the network. In the training stage, the Adam algorithm and cross-entropy are used as optimizer and loss function respectively. The initial learning rate is set to 0.001, and the learning rate is adaptively adjusted by ReduceLROnPlateau strategy with at least 300 epochs.

TABLE I
COMPARISON OF DIFFERENT LAND-COVER CLASSIFICATION METHODS ON GID (IN PERCENTAGE). OCNN (I), OGCN (II), OSAGE (III), OGAT (IV), Unet (V), ResUNet(VI), U-OGCN (VII), U-OGNN (VIII)

Class	I	II	III	IV	V	VI	VII	VIII
1	80.12	76.85	80.15	81.73	83.39	**88.24**	82.18	87.32
2	78.51	80.21	80.31	81.90	85.95	84.23	81.22	**84.27**
3	74.67	77.13	76.78	77.53	83.48	85.67	81.43	**86.54**
4	65.59	63.48	67.26	68.19	60.49	69.67	73.64	**80.63**
5	93.75	92.67	94.64	95.04	92.23	94.43	95.75	**96.73**
6	82.65	81.22	85.63	84.63	87.6	83.84	87.87	**89.92**
OA	79.98	78.44	81.00	81.95	83.32	86.35	83.52	**87.81**
AA	79.21	78.59	80.80	81.51	82.19	84.35	83.68	**87.57**
Kappa	71.37	68.61	72.61	74.15	75.49	80.29	76.22	**82.55**

TABLE II
TEST PERFORMANCE OF DIFFERENT SETTINGS ON GID.
(OA: OVERALL ACCURACY, AA: AVERAGE ACCURACY)

	SAGC	U-shaped structure	GID		
			OA	OA	OA
OGCN	—	—	78.44	78.59	68.61
	√	—	78.81	79.35	69.10
	—	√	82.57	82.57	75.05
	√	√	**83.52**	**83.68**	**76.22**
OGAT	—	—	81.95	81.51	74.51
	√	—	83.13	81.90	75.89
	—	√	85.74	85.71	79.52
	√	√	**87.81**	**87.57**	**82.55**

In ablation experiments, we explored SAGC, graph encoder and decoder components in the framework of OGCN and OGAT and measured how they affect the final performance of the model. In the study, we selected two OGCN and OGAT models with multi-layer attention blocks and evaluated them on GID dataset. We evaluate the effectiveness of the proposed SAGC and U-shaped structure. The comparison of test performance is shown in Table II. In OGCN, we find that the connection of highly similar semantic features through SAGC is conducive to capturing the global information in the image, thereby improving the model's performance. At the same time, according to the graph encoder and decoder module, the U-shaped structure network is designed to capture hierarchical features, which improves the accuracy of 3 – 4 % on OGCN and OGAT, respectively.

B. Quantitative evaluation and visual comparison

The classification performance of the U-OGNN method is compared with U-OGCN and benchmark OCNN, OGCN[24], OSAGE[26], OGAT[27], Unet and ResUNet.

The U-OGNN is applied to classify six land-cover types with image segments. Meanwhile, we compare our classification results with other competitive methods. The accuracies are reported in Tables I. In the table, 1, 2, 3, 4, 5and 6 indicate background, building, farmland, meadow, water and forest respectively. The overall accuracy of U-OGCN is 87.81 %, which is superior to the benchmark methods. For the object-based OCNN and OGCN, the classification accuracies are lower than 80% in terms of OA. Specifically, the forest and farmland without regular shapes are difficult to perform

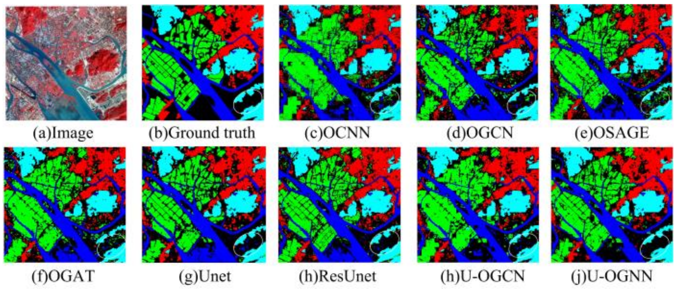
(a)Image (b)Ground truth (c)OCNN (d)OGCN (e)OSAGE
(f)OGAT (g)Unet (h)ResUnet (h)U-OGCN (j)U-OGNN

Fig. 3. The classification results were obtained through different algorithms on the Gaofen image dataset

1545-598X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Beijing Normal University. Downloaded on May 27,2022 at 00:44:05 UTC from IEEE Xplore. Restrictions apply.

accurate classification. However, pixel-based classification methods such as Unet and ResUNet are slightly better than object-based methods, regardless of inaccurate image segmentation. With the help of a U-shaped structure, the classification accuracies on forest, meadow, and farmland are improved by around 5%. To visualize the results, we mapped the GID dataset and the classification maps with U-OGNN and other competitive method. Fig. 3 (a)-(b) shows the high-resolution GF-images obtained in 2015 and their reference land-cover labels. Fig. 3 (c)-(j) are the results generated by OCNN, OGCN, OSAGE, OGAT, Unet, ResUnet, U-OGCN, and U-OGNN achieved the best classification results in terms of detailed information preserving and class identification. A representative area circled with a yellow line is included for comparison. For the methods, OCNN, OGCN, OGAT, and OSAGE constantly misclassified background with farmland, as similar spectral-spatial features can be observed. Differently, Unet and ResUNet are difficult in identifying the class of river and background without considering contextual information. With the help of a U-shaped structure, the misclassification is minimized by introducing multi-level contextual information.

IV. CONCLUSION

In order to capture multi-level contextual information at the global scale, we developed a new U-shaped graph structure named U-OGNN for land cover classification. Firstly, the self-adaptive graph construction (SAGC) algorithm is proposed based on image objects using KDAM measurement in the feature space. Then, hierarchical characteristics at different levels can be extracted by the U-shaped graph structure which proved to be effective in terms of contextual information extraction. To validate the effectiveness of the proposed U-OGNN, one challenging dataset have been included for land cover classification. Compared with the baseline model, our U-OGNN achieves competitive performance on publicly available GID datasets.

REFERENCES

[1] P. Fisher, A. J. Comber, and R. Wadsworth, “Land use and land cover: contradiction or complement,” Re-Present. GIS, pp. 85–98, 2005.

[2] S. Jin, L. Yang, P. Danielson, C. Homer, J. Fry, and G. Xian, “A comprehensive change detection method for updating the National Land Cover Database to circa 2011,” Remote Sens. Environ., vol. 132, pp. 159–175, 2013.

[3] J. Chen et al., “Global land cover mapping at 30 m resolution: A POK-based operational approach,” ISPRS J. Photogramm. Remote Sens., vol. 103, pp. 7–27, 2015.

[4] R. Jalal et al., “Toward efficient land cover mapping: an overview of the national land representation system and land cover map 2015 of Bangladesh,” IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 12, no. 10, pp. 3852–3861, 2019.

[5] B. Yifang, P. Gong, and C. Gini, “Global land cover mapping using Earth observation satellite data: Recent progresses and challenges,” ISPRS J. Photogramm. Remote Sens. Print, vol. 103, no. 1, pp. 1–6, 2015.

[6] V. S. Martins, A. L. Kaleita, B. K. Gelder, H. L. da Silveira, and C. A. Abe, “Exploring multiscale object-based convolutional neural network (multi-OCNN) for remote sensing image classification at high spatial resolution,” ISPRS J. Photogramm. Remote Sens., vol. 168, pp. 56–73, 2020.

[7] Y. Xu, J. Du, and J. Shi, “Endmember classes determination using spectral similarity analysis for MODIS reflectance channels,” 2014, pp. 2989–2992.

[8] Y. Wang, D. Yu, S. Ji, Q. Cheng, and M. Luo, “The Joint Spatial and Radiometric Transformer for Remote Sensing Image Retrieval,” Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., vol. 43, pp. 227–231, 2020.

[9] H. Yan, “Remote sensing image classification based on svm classifier,” 2011, vol. 1, pp. 30–33.

[10] X. Huang and L. Zhang, “An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery,” IEEE Trans. Geosci. Remote Sens., vol. 51, no. 1, pp. 257–272, 2012.

[11] B. Zhao, Y. Zhong, and L. Zhang, “A spectral–structural bag-of-features scene classifier for very high spatial resolution remote sensing imagery,” ISPRS J. Photogramm. Remote Sens., vol. 116, pp. 73–85, 2016.

[12] T. Blaschke, “Object based image analysis for remote sensing,” ISPRS J. Photogramm. Remote Sens., vol. 65, no. 1, pp. 2–16, 2010.

[13] H. Costa, G. M. Foody, and D. S. Boyd, “Supervised methods of image segmentation accuracy assessment in land cover mapping,” Remote Sens. Environ., vol. 205, pp. 338–351, 2018.

[14] M. D. Hossain and D. Chen, “Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective,” ISPRS J. Photogramm. Remote Sens., vol. 150, pp. 115–134, 2019.

[15] J. Sherrah, “Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery,” ArXiv Prepr. ArXiv160602585, 2016.

[16] Z. Zhang, Q. Liu, and Y. Wang, “Road extraction by deep residual u-net,” IEEE Geosci. Remote Sens. Lett., vol. 15, no. 5, pp. 749–753, 2018.

[17] G. Fu, C. Liu, R. Zhou, T. Sun, and Q. Zhang, “Classification for high resolution remote sensing imagery using a fully convolutional network,” Remote Sens., vol. 9, no. 5, p. 498, 2017.

[18] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, “ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data,” ISPRS J. Photogramm. Remote Sens., vol. 162, pp. 94–114, 2020

[19] F. Zhou, R. Hang, H. Shuai, and Q. Liu, "Hierarchical Context Network for Airborne Image Segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

[20] F. Zhou, R. Hang, and Q. Liu, "Class-guided feature decoupling network for airborne image segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2245-2255, 2020.

[21] J. Zhou et al., “Graph neural networks: A review of methods and applications,” AI Open, vol. 1, pp. 57–81, 2020.

[22] Danfeng Hong, Lianru Gao, Jing Yao, Bing Zhang, Antonio Plaza, Jocelyn Chanussot. "Graph Convolutional Networks for Hyperspectral Image Classification," IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(7): 5966-5978. DOI: 10.1109/TGRS.2020.3015157.

[23] M. Defferrard, X. Bresson, and P. Vandergheynst, “Convolutional neural networks on graphs with fast localized spectral filtering,” Adv. Neural Inf. Process. Syst., vol. 29, pp. 3844–3852, 2016.

[24] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” ArXiv Prepr. ArXiv160902907, 2016.

[25] Y. Hechtlinger, P. Chakravarti, and J. Qin, “A Generalization of Convolutional Neural Networks to Graph-Structured Data,” stat, vol. 1050, p. 26, 2017.

[26] W. Hamilton, Z. Ying, and J. Leskovec, “Inductive representation learning on large graphs,” 2017, pp. 1024–1034.

[27] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, “Graph attention networks,” ArXiv Prepr. ArXiv171010903, 2017.

[28] Q. Liu, M. Kampffmeyer, R. Jenssen, and A.-B. Salberg, “SCG-Net: Self-Constructing Graph Neural Networks for Semantic Segmentation,” ArXiv Prepr. ArXiv200901599, 2020.

[29] 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, 2021.

[30] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pp. 2274–2282, 2012.

[31] A. Vaswani et al., “Attention is all you need,” 2017, pp. 5998–6008.

[32] X.-Y. Tong et al., “Land-cover classification with high-resolution remote sensing images using transferable deep models,” Remote Sens. Environ., vol. 237, p. 111322, 2020.

1545-598X (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Beijing Normal University. Downloaded on May 27,2022 at 00:44:05 UTC from IEEE Xplore. Restrictions apply.