

HYPERSPECTRAL IMAGE CLASSIFICATION WITH LOCALIZED SPECTRAL FILTERING-BASED GRAPH ATTENTION NETWORK

Shengliang Pu^{1,2,3,*}, Yining Song¹, Yingyao Chen¹, Yating Li¹, Jinyan Zhang¹, Qinglong Lin¹, Xunda Zhu¹, Yongquan Chen¹, Hao Zeng¹, Kexin Liao¹, Hongmei Yu¹, Jincheng Yuan¹, Songling Yu¹, Wenting Zuo^{1,4}

¹ Faculty of Geomatics, East China University of Technology, Nanchang 330013, China - shengliangpu@163.com

² Key Laboratory of Mine Environmental Monitoring and Improving around Poyang Lake, Ministry of Natural Resources, Nanchang 330013, China - shengliangpu@ecut.edu.cn

³ Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China - pusl@aircas.ac.cn

⁴ State Key Laboratory of Nuclear Resources and Environment, Nanchang 330013, China - wentingzuo@ecut.edu.cn

Commission 3, WG III/4

KEY WORDS: Hyperspectral image classification, Graph attention network, Self-attention, Localized spectral filtering, Graph-based deep learning.

ABSTRACT:

Graph-based deep learning has been proved a promising approach that has an apparent superiority for learning graph data and modeling spatial topological relations between features. In particular, graph attention networks (GATs) are good at efficiently processing the graph-structured hyperspectral data by leveraging masked self-attention layers to address the known shortcomings of previous frameworks based on graph convolutions or their approximations. In this study, we proposed a novel approach that combines localized spectral filtering and GAT for the hyperspectral image classification task. First, we conducted unsupervised *t*-SNE (*t*-distributed stochastic neighbor embedding) manifold learning-based feature dimensionality reduction to create localized hyperspectral data cubes. Then, these feature cubes combined with localized adjacent matrices were fed into a shallow graph attention network in a supervised learning manner. Finally, we obtained credible classification results and promising classification performance in distinguishing diversified land covers through reducing the possible redundancy of spectral information and enhancing the expression of local spatial-spectral information. Experiments on two real hyperspectral data sets (that is, Indian Pines-A (IA) and Huanghekou (HH) data sets) demonstrated that the presented approach offers promising classification performance, that is, the GAT using *t*-SNE acquires superior performance than that of using PCA (principal component analysis), and also proves the great importance of combining spatial- and spectral information for hyperspectral image classification.

* Corresponding author

1. INTRODUCTION

Hyperspectral imaging is particularly useful for per-pixel thematic classification by mining unique spectral signatures of land surface materials to obtain attribute map regarding diverse land covers. Hyperspectral image (HSI) classification categorizes each three-dimensional (3-D) pixel cube into a set of groundtruth classes, which has been a very active research area in recent years (Deng et al., 2018; Pu et al., 2021a). Recently, the graph-based deep learning represented by GCNs has received more and more attention in describing class boundaries and modeling topological relations among samples in the irregular graph domains converted from hyperspectral data (Pu et al., 2021b).

Graphs are a kind of universal representation of non-Euclidean structured data, which could encode complex geometric structures (Chung and Graham, 1997). Graph-based representations can be used to model a variety of problems and domains (Avelar et al., 2020). In this context, the recent novelties regarding graph-based deep learning have attracted growing attention from the scientific community. The graph-based deep learning methods as a kind of deep learning technique have undoubtedly brought enormous prosperity in hyperspectral remote sensing intelligent information extraction (Yang et al., 2018; Pu et al., 2021b). In particular, HSIs from regular grids (or call image domains) into irregular graph domains could adapt to the superiority of graph-based deep learning for preferably illustrating class boundaries and modeling feature relationships (Pu et al., 2021b).

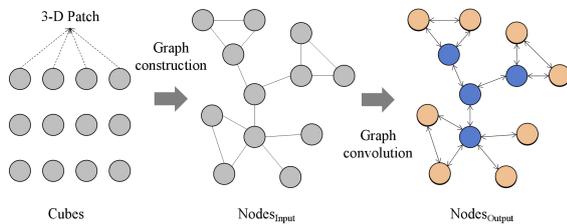


Figure 1. Message passing in a typical GCN.

Graph Neural Networks (GNNs) are a class of graph-based deep learning methods designed to perform inference on data described by graphs (Wu et al., 2019). Graph convolution networks (GCNs) have received more and more attention in quantifying nonlinear or topological features in irregular graphs converted from hyperspectral data (Pu et al., 2021b). Typically, a simple GCN plays the functional roles depending on the message-passing, pooling, and fully-connected layers. The crucial role is the message-passing layer, which is to compute and update a representation of each node in the graph while leveraging local information from its neighbours (referring to Figure 1).

The graph-based convolutional neural networks (that is, a kind of GCNs) showcase distinctive characteristics (Wan et al., 2020), that is, competently process the irregular regions in the non-Euclidean (or non-grid) graph data structure, and even incorporate multiple graph inputs which can be dynamically updated and refined with the multi-scale neighborhood. However, the GCN-based methods might be difficult to aggregate the newly joined nodes, and then these methods might fail to understand the global and contextual information of the graph scenario (Ding et al., 2021). In this regard, attention mechanisms focus on the most relevant parts of the graph-based input to make effective decisions allowing for handling

variable-sized inputs.

Graph attention networks (GATs) are a kind of graph-based neural network architectures that operate on graph-structured data, leveraging masked self-attention layers to address the known shortcomings of prior methods based on graph convolutions or their approximations (Velickovi et al., 2017). Different from many related works paid attention to the graph-based semi-supervised learning methods for HSI classification, which make the graph input built on the full graph (that is, one image corresponding to a full graph), which combines the labeled and unlabeled nodes by employing a graph Laplacian regularizer (Pu et al., 2021b). The unlabeled nodes are completely observed during training or test a deep learning model, whereas the standard formulation of semi-supervised learning paradigm requires the independent and identically distributed (i.i.d) assumption between the labeled and unlabeled nodes (Hamilton et al., 2020).

In this study, we also follow the abovementioned supervised setting to train a GAT. Continuing the previous works on spectral graph-based convolutional neural networks (CNNs) for HSI classification (Defferrard et al., 2016; Pu et al., 2021b; Pu et al., 2022), the main contributions of the presented study are summarized below.

- We introduced a graph attention-based deep learning architecture to perform node classification of the graph-structured hyperspectral data.
- We employed unsupervised t -SNE (t -distributed stochastic neighbor embedding) manifold learning-based feature dimensionality reduction to collect the patch-based feature cubes and to create localized graph adjacent matrices, subsequently used the usual supervised setting to fit the presented graph-based deep learning model.
- We used two graph attention layers to learn the spatially local graph representation and to represent the localized topological patterns of the graph node and its neighboring nodes.

The rest of this paper is organized as follows. The technical details of the proposed approach are presented in Section 2. Next, we analyze the experimental results and discuss the derived findings in Section 3. Finally, some concluding remarks are given in Section 4.

2. PROPOSED APPROACH

2.1 t -SNE

High-dimensional data such as HSIs might not always have good separability. Since feature dimensionality reduction methods based on t -SNE (t -distributed stochastic neighbor embedding) manifold learning have been known to have excellent properties than those based on principal component analysis (PCA), hence we tried t -SNE for reducing dimensions of hyperspectral data in this study (see Figure 2). In this regard, t -SNE can be regarded as one of the effective dimensionality reduction and visualization methods for high-dimensional data preprocessing in the scientific community. Besides, t -SNE is an embedded model in essence, which can transfer data from high-dimensional space to low dimensional-space, while retaining local characteristics of data (Van der Maaten and Hinton, 2008).

Given a high-dimensional HSI matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{C \times N}$ with N spectral vectors of

size C (that is, the columns of \mathbf{X} , and $N = h \times w$), herein h is the height and w means the width. Considering t -SNE transforms the affinities between data points into conditional probabilities, and the similarity of data points in the original space is represented by Gaussian joint distribution.

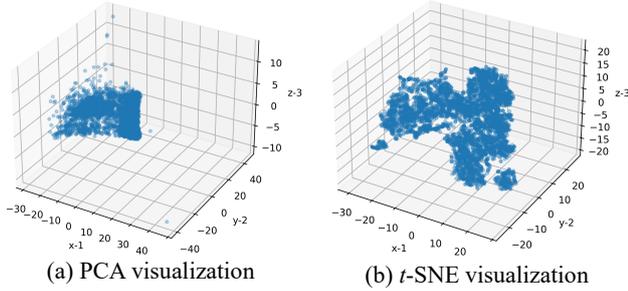


Figure 2. Comparison of PCA and t -SNE for feature dimensionality reduction of the Indian Pines-A (IA) data set.

Therefore, the t -SNE algorithm converts \mathbf{X} into a low-dimensional matrix $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N] \in \mathbb{R}^{D \times N}$. C represents the number of bands, and D denotes the number of components, both subject to $D < C$. The t -SNE algorithm can obtain the optimal dimensionality reduction result by calculating the minimum value of Kullback-Leibler (KL) divergence between the joint conditional probability of the original space and the embedded space. t -SNE focuses on local data structure by identifying the patterns based on the similarity of data points with multiple features. Intuitively, after t -SNE transformation, the obtained data components will become separable in the low-dimensional space. In the meantime, the KL divergence function can be regarded as a loss function minimized by the stochastic gradient descent (SGD) algorithm.

2.2 Localized spectral filtering

The formulation of CNNs in the context of spectral graph theory provides the necessary mathematical background and efficient numerical schemes to design fast localized spectral filters on graphs (Pu et al., 2021b). Importantly, such a technique offers the same linear computational complexity and constant learning complexity as classical CNNs, while being universal to any graph structure (Defferrard et al., 2016). Then, we give \mathbf{G} by using a binary adjacency matrix $\mathbf{A} \in \{0, 1\}^{N \times N}$, node features $\mathbf{X} \in \mathbb{R}^{N \times F}$, and edge features $\mathbf{E} \in \mathbb{R}^{N \times N \times S}$. For each sample, we build a K -nearest neighbor (abbreviated as K -NN) adjacency matrix with

$$\mathbf{W}_{ij}^{K-NN} = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma_x^2}\right), \quad (1)$$

where x_i, x_j are the 2-D coordinates of pixels i, j , and σ_x is the scale parameter defined as the averaged distance x_K of the K nearest neighbors for each node.

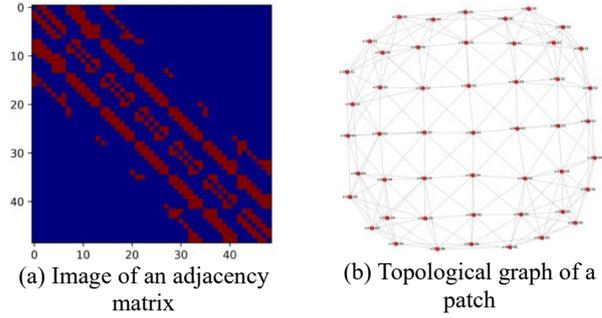


Figure 3. Visualization of localized adjacency matrix (without adding additional noises) in two different forms.

The algorithm for constructing a localized adjacency matrix (See Figure 3) could be summarized as follows.

Firstly, given data matrix 3-D $\mathbf{X} \in \mathbb{R}^{h \times w \times b}$ and its label vector $\mathbf{y} \in \mathbb{R}^{h \times w}$, and assign two parameters: the number of neighbors K and the length of patch l , subsequently calculate $s = 2l + 1$ and ready to get the adjacency matrix \mathbf{A} for the K -NN graph;

Secondly, we use the size of each patch S as the size of two same coordinate vectors $\mathbf{s} \in \{0, 1\}$ to construct grid-like sampling coordinate matrices $\mathbf{M} \in \mathbb{R}^{s^2 \times 2}$, then return 2D coordinates for a square grid of equally spaced nodes;

Thirdly, we compute the weighted graph \mathbf{G} of K -neighbors with the number of neighbors K for points in \mathbf{M} using the above Eq. (1) and return the resultant adjacent matrix \mathbf{A} ;

Finally, it will be ready for creating HSI cubes, and completes data preparation for GNNs.

2.3 Graph attention network

As known to us, graph attention networks (GATs) introduce attention-based deep learning architectures to perform node classification for graph-structured data (Veličković et al., 2017). By learning the importance weight of each node to the classified node, the graph attention mechanism makes the important nodes have greater weight, and hence global and contextual information can be learned from the graph via attention mechanism (Ding et al., 2021). The graph convolution output \mathbf{h}_i^l of each node can be expressed as follows:

$$\mathbf{h}_i^l = \sigma\left(\sum_{j \in N_i} a_{ij} \cdot \mathbf{W}^T \mathbf{h}_j^{l-1}\right), \quad (2)$$

where σ denotes the activation function, N_i is the size of the neighboring set of node i , and a_{ij} is the learned attention weight. The attention coefficient a_{ij} is generated dynamically just depending on the local neighborhood and rearranges the neighbors by their importance, which makes the model more flexible to the specific input sample. \mathbf{W} is a filtering matrix for dimensionality reduction and feature extraction of HSI data.

Consequently, some important neighbors could be emphasized in the summation (Sha et al., 2020).

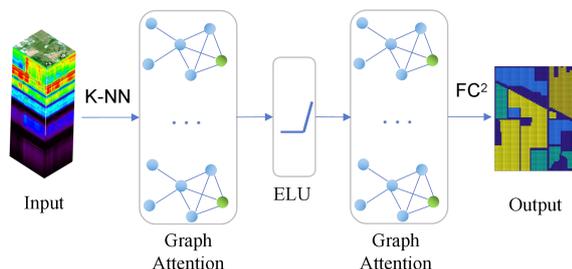


Figure 4. Flowchart of the designed graph attention network.

The proposed HSI classification model based on GAT is schematically shown in Figure 4. The main idea for graph attention-based HSI classification is to refine the graph signal on the spatial-spectral domain via a localized spectral filter, implemented by a GAT composed of two graph attention layers and two fully-connected (FC²) layers. After the multilayer perceptron, the output $\mathbf{Z} \in \mathbb{R}^{N \times M}$ for the whole graph was obtained, and only the labeled nodes were used as supervised regression in the cross-entropy loss:

$$L = - \sum_{s \in \mathbf{Y}_{\text{labeled}}} \sum_{m=1}^M \mathbf{y}_{sm} \ln \mathbf{Z}_{sm}, \quad (3)$$

where \mathbf{y}_{sm} is the true label of training data and M is the number of object classes. There is a particularity in the graph-based learning methods that the eventual predictions for labeled and unlabeled samples are iterated simultaneously until a stable status is reached, and therefore, no extra test process is required. Note that, “ELU” rather than “ReLU” is used for GNN; the 1st FC layer is with $\text{nb_classes} \times 10$ (that is, 10 times the number of classes) units, while the 2nd FC layer with nb_classes (that is, the number of classes) units in this study.

3. EXPERIMENTS

3.1 Environmental setting and data sets

The cloud-based experimental platform used in this study is the “Planetary Computer” platform built on Microsoft Azure cloud, which is a development environment that provides access to its data resources and application programming interfaces (APIs) through open-source tools and allows users to easily extend the experimental analysis process using the power of Azure computing. For an independent virtual machine with a 4-core CPU, 28 GB of RAM, and a T4 GPU (graphics processing unit) (NVIDIA Tesla T4 GPUs, 16 GB of graphic memory). These virtual machines are ideal for deploying artificial intelligence (AI) services, responding to user-generated requests in real-time, or using NVIDIA’s grid driver and virtual GPU technology for interactive graphics and visualization workloads.

Indian Pines (IP) scene was gathered by the 224-band AVIRIS (airborne visible/infrared imaging spectrometer) sensor in the wavelength range 400 to 2500 nm at a 20 m spatial resolution (that is, 20 meters/pixel, or abbreviated as 20 m/p), in north-western Indiana. Indian Pines-A (IA) data set (see Figure 5 was a subset of the IP dataset, which consisted of 86×69 pixels and contained 200 spectral reflectance bands by removing bands

covering the region of water absorption.

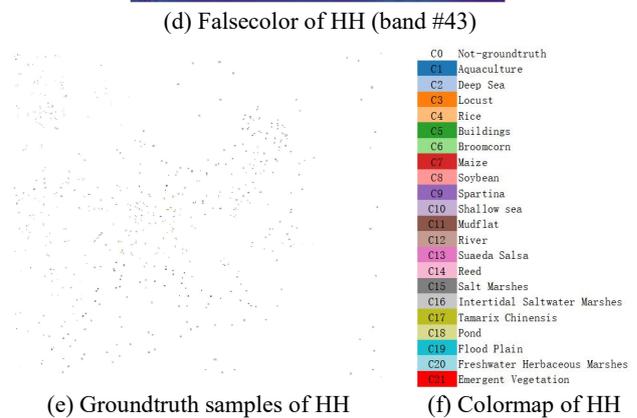
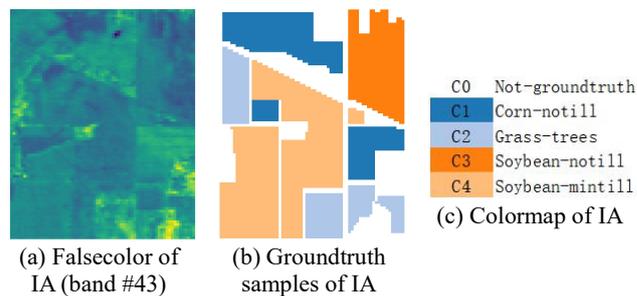


Figure 5. Indian Pines-A (IA) and Huanghekou (HH) data sets.

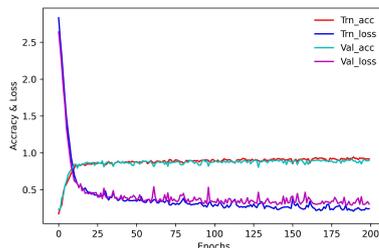
Huanghekou (HH) data set is released by Jiao et al. (2019), which includes 21 classes, and its material types of overwhelming landscapes incorporate water bodies, grassland, forest land, buildings, bare land, etc., which are different from the traditional land cover types. Specifically, a GF-5 AHSI (advanced hyperspectral imager) sensor was adopted with a spatial resolution of 30 m and a spectral range covering VNIR (visible and near-infrared: 390-1029 nm) and SWIR (short-wave infrared: 1005-2513 nm). There are 150 bands in the VNIR spectral range (visible and near-infrared, without excluding band 1) and 180 bands in the SWIR spectral range (short-wave infrared, without excluding bands 42-53, 96-115, 119-121, 172-173, 175-180, etc.), so there are 285 bands left after removing a number of bad bands from a total of 330 bands. The spatial size of the HH data set was 1185 rows and 1342 columns, and its capturing time was November 1st, 2018. Note that, the spectral resolution is 5 nm of VNIR (visible and near-infrared) and 10 nm of SWIR (short-wave infrared). Note that, the spatial distribution of groundtruth samples of HH appears very sparse than that of the IA data set. That's why there's a lot of white space in the scene.

3.2 Parameters and training

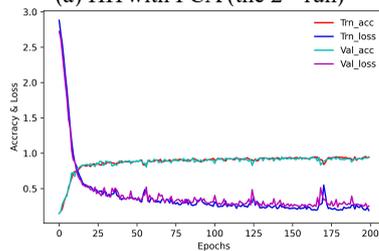
In order to facilitate parameter analysis of the experiments, all parameters have been divided into three groups (see Table 1). The first group contains parameters for training a network, the second group relates to sampling a data set and preparing data for training, the third one involves extra control parameters for each instance of the experiments. For the classification experiments, the size of the training sample set of different hyperspectral data sets was set to 35 samples for each category, and there are 5 independent random runs were performed in total. The training procedure continues 200 epochs. Note that, the values listed are finally determined, which are not necessarily the optimal value for acquiring the best classification performance.

Groups	Parameters	Values
#1 Training (network)	Number of graph attention layers	2
	Number of neighbors	8
	Number of attention heads	4
	Regularization rate for l2	5e-4
	Learning rate for stochastic gradient descent (SGD)	1e-3
	Number of training epochs	200
	Patience for early stopping	100
2# Sampling (data set)	Batch size	32
	Number of training samples	35
	Patch length	3
	Patch size	7
3# Controlling (instance)	Noise level (affinity matrix)	0.0
	Number of components	3
	Random noise (note that adding before feature reduction)	0%
	Number of repetitions (iterations)	5

Table 1. Parameters and setting.



(a) HH with PCA (the 2nd run)



(b) HH with *t*-SNE (the 1st run)

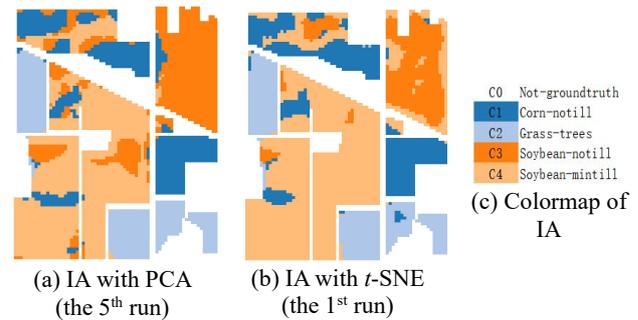
Figure 6. Accuracy and loss curves of fitting and validating the designed GAT model using the HH data set.

In this study, a fixed number of samples were used for each category (that is, each category has an equal number of training samples). Then, the validation set was designed in the same way to make it the same size as the training set. Finally, the training set and validation set were removed from all samples, and the

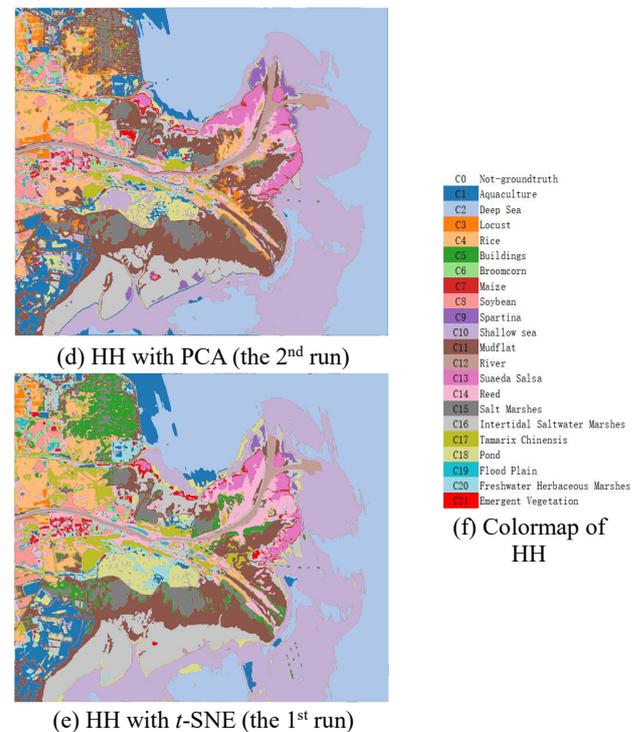
remaining samples were treated as the test set. If the aforementioned condition is not met, the existing samples in hand will be simply reproduced by creating multiple copies to meet the requirements of the due amount of samples. In the following, the robustness of graph-based deep learning models mainly depends on fine-tuning hyperparameters in an iterative optimization.

The most representative method is to draw the accuracy and loss curves when training and validating in the designed graph-based deep learning models. It can reflect several aspects of the model's robustness, that is, whether a final convergence, whether a smooth convergence, and a general number of epochs required for convergence, etc. Therefore, the accuracy and loss curves of fitting and validating the designed GAT model with 200 epochs using HH data set were illustrated in this study. It can be seen from Figure 6 that the GAT model with different feature dimensionality reduction algorithms (that is, PCA and *t*-SNE) have relatively good characteristics of convergence. Note that, for the HH data set, the 2nd run using PCA dimensionality reduction and the 1st run using *t*-SNE dimensionality reduction indicate the best classification results in the total 5 random runs.

3.3 Classification results



(a) IA with PCA (the 5th run) (b) IA with *t*-SNE (the 1st run)



(d) HH with PCA (the 2nd run) (e) HH with *t*-SNE (the 1st run) (f) Colormap of HH

Figure 7. Classification maps of the presented experiments using the IA and HH data sets.

Experiments using two real hyperspectral data sets (note that, the first one IA is a small data set, and the latter one HH is a large data set) show that the proposed approach has promising classification performance using three classification accuracy metrics of overall accuracy (OA), average accuracy (AA), and Kappa (K), indicating that the combination of t -SNE and local spectral graph convolution filtering can indeed improve the discriminant ability of the graph-based deep learning model for identifying every pixel' category, and also prove the importance of combining spatial- and spectral information for hyperspectral remote sensing image classification. Particularly, probability maps are essentially a kind of heat map with scale normalization. It can be seen from the probability maps that the graph-based deep learning models also have weak prediction phenomena for the pixels residing on the boundary between categories.

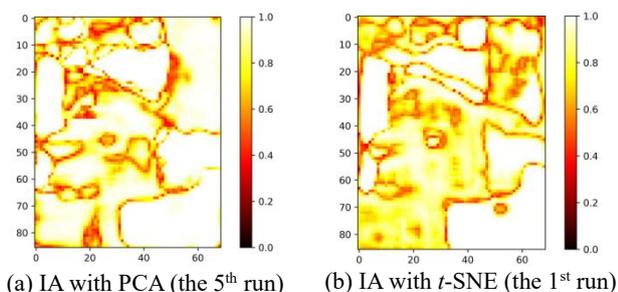


Figure 8. Predictive maximum probability maps of the presented experiments using the IA data set.

By observing classification maps in Figure 7 (a)-(c), probability maps in Figure 8, and confusion matrices in Figure 9, for experiments using IA data set, most misclassifications occur between class 1 (corn-notill) and class 3 (soybean-notill), class 1 (corn-notill) and class 4 (soybean-mintill), and between class 3 (soybean-notill) and class 4 (soybean-mintill). At the same time, class 2 (grass-trees) maintains good separability from other categories, which means small intra-class differences. Note that, as confusion matrices have been disclosed here, therefore the per-class accuracies will not be included in the statistics of classification accuracies. In a sense, the effect of both is equivalent. Note that, see Figure 9, the names of every class are as follows, class 1 (corn-notill) class 2 (grass-trees), class 3 (soybean-notill), and class 4 (soybean-mintill).

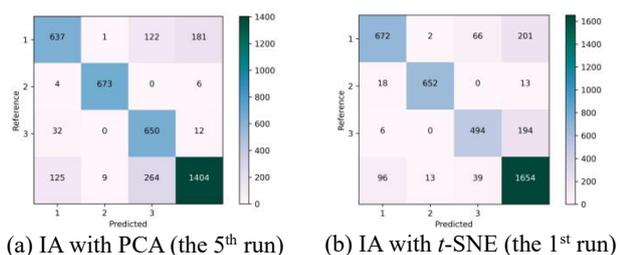


Figure 9. Confusion matrices of the presented experiments using the IA data set.

Also, by observing classification maps in Figure 7 (d)-(f), for the presented experiments using the HH data set, some categories like, classes 2-4 (deep sea, locust, and rice), class 6 (broomcorn), classes 8-9 (soybean and spartina), class 12 (river), classes 15-17 (salt marshes, intertidal saltwater marshes, and

Tamarix Chinensis), classes 20-21 (freshwater herbaceous marshes and emergent vegetation) maintain good separability from other categories.

Due to the non-strict production of data sets in many cases, that would make the labeled samples much smaller in size or sparsely distributed in the geospatial distribution, or even the labeled area of groundtruth samples has certain generalization processing. That is, there might be mislabeling issues at the boundary of or inside a certain category. Therefore, the reliability of the graph-based deep learning model will inevitably be adversely affected. In other words, the difference or variance within an individual class might be relatively small. The reason is that, on the one hand, there might be some dirty (or wrong) labels in the labeling procedure; on the other hand, the definition of categories might not consider the difference of materials, or the boundary between categories might go through generalized processing. The aforementioned two reasons may lead to the decrease of inter-class separability. In addition to the above reasons, strictly speaking, there might be mixed categories in the definition of some categories, or the intra-class differences might be over influential leading to misclassification to some extent.

Experiments	K	OA	AA	Training time (s)	Test time (s)
IA with PCA	0.7242 ± 0.0137	0.8102 ± 0.0081	0.7951 ± 0.0253	41.2	0.9
IA with t -SNE	0.7346 ± 0.0241	0.8171 ± 0.0165	0.8073 ± 0.0148	40.7	0.9
HH with PCA	0.8781 ± 0.0140	0.8875 ± 0.0130	0.8241 ± 0.0261	129.7	1.1
HH with t -SNE	0.9287 ± 0.0074	0.9343 ± 0.0068	0.8751 ± 0.0156	130.9	1.1

Table 2. Classification accuracies and running time statistics.

It also can be seen from Table 2 that, introducing t -SNE dimensionality reduction has brought up a better performance in feature learning and predicting labels of GNNs. The reason is that t -SNE can effectively improve the separability among classes in the feature space. Such excellent performance is not invariable in deep learning but is possibly affected by parameter settings and changes in the network structure. At the same time, the different number of categories, class separability, and scene complexity, all would lead to possible uncertainty when measuring the final classification performance. In terms of running time, since having the same number of components after the feature dimensionality reduction of raw HSI data, therefore their computational costs are fairly approximate for the same data set. Note that, the derived evaluation metrics (that is, OA, AA, K) are based on test samples rather than the whole image. At the same time, the records of training and test times don't include the processing times of t -SNE or PCA transformation. Therefore, it is the reason why fitting a GAT using t -SNE could be faster than that of using PCA.

4. CONCLUSIONS

Graph-based deep learning model has been widely used in HSI classification, and has attracted more and more attention due to its strong expression ability. In particular, the emerging graph representation learning and graph neural networks show good characteristics in processing and analyzing graph-structured data. In this study, we proposed a novel approach that combines localized spectral filtering and *t*-SNE manifold learning based on graph attention mechanism. Newly HH data set has poor separability between categories, yet better classification performance could be obtained. It indicates that class separability might not be a critical issue, and the spatial texture or contextual information might somehow reduce the sensitivity of spectral differences to the properties of land surface materials. For *t*-SNE based feature learning that contains a large number of iterative calculations would lead to huge computational cost, so we have to adopt GPU acceleration to derive feature data and saves them to the storage medium, then loads and resamples them again, further to effectively avoid huge memory consumption. In conclusion, with the emergence of more and more intelligent algorithms for HSI information extraction, graph-based deep learning models such as GATs would drive the advanced development of future HSI classification research.

ACKNOWLEDGEMENTS

This work was funded by the Research Funding of the East China University of Technology (No. DHBK2019192). In addition, the authors thank four anonymous reviewers for their insightful comments and helpful suggestions, which undoubtedly improve the quality of this manuscript.

REFERENCES

- Avelar, P.H., Tavares, A.R., da Silveira, T.L., Jung, C.R., Lamb, L.C., 2020. Superpixel image classification with graph attention networks. Presented at the 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), IEEE, pp. 203 – 209.
- Chung, F.R., Graham, F.C., 1997. Spectral graph theory. American Mathematical Soc.
- Defferrard, M., Bresson, X., Vandergheynst, P., 2016. Convolutional neural networks on graphs with fast localized spectral filtering. *Adv. Neural Inf. Process. Syst.* 29.
- Deng, F., Pu, Shengliang, Chen, X., Shi, Y., Yuan, T., Pu, Shengyan, 2018. Hyperspectral image classification with capsule network using limited training samples. *Sensors* 18, 3153.
- Ding, Y., Zhao, X., Zhang, Z., Cai, W., Yang, N., 2021. Graph sample and aggregate-attention network for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* 19, 1 – 5.
- Hamilton, W.L., 2020. Graph representation learning. *Synth. Lect. Artificial Intell. Mach. Learn.* 14, 1 – 159.
- Jiao, L., Sun, W., Yang, G., Ren, G., Liu, Y., 2019. A hierarchical classification framework of satellite multispectral/hyperspectral images for mapping coastal wetlands. *Remote Sens.* 11, 2238.
- Pu, S., Gao, L., Song, Y., Chen, Y., Li, Y., Luo, L., Xu, G., Xie, X., Nie, Y., 2021a. Hyperspectral image classification with residual learning networks. Presented at the AOPC 2021: Optical Sensing and Imaging Technology, SPIE, pp. 322 – 327.
- Pu, S., Wu, Y., Sun, Xu, Sun, Xiaotong, 2021b. Hyperspectral image classification with localized graph convolutional filtering. *Remote Sens.* 13, 526.
- Pu, S., 2022. Spectral graph neural networks with manifold-learning-based feature extraction for hyperspectral image classification. *Qeios* 1 – 4. <https://doi.org/10.32388/B899UZ>
- Sha, A., Wang, B., Wu, X., Zhang, L., 2020. Semisupervised classification for hyperspectral images using graph attention networks. *IEEE Geosci. Remote Sens. Lett.* 18, 157 – 161.
- Van der Maaten, L., Hinton, G., 2008. Visualizing data using *t*-SNE. *J. Mach. Learn. Res.* 9.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., 2017. Graph attention networks. *ArXiv Prepr. ArXiv171010903*.
- Wan, S., Gong, C., Zhong, P., Pan, S., Li, G., Yang, J., 2020. Hyperspectral image classification with context-aware dynamic graph convolutional network. *IEEE Trans. Geosci. Remote Sens.* 59, 597 – 612.
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Philip, S.Y., 2020. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 4 – 24.
- Yang, X., Ye, Y., Li, X., Lau, R.Y., Zhang, X., Huang, X., 2018. Hyperspectral image classification with deep learning models. *IEEE Trans. Geosci. Remote Sens.* 56, 5408 – 5423.