REVISITING GRAPH CONVOLUTIONAL NETWORKS WITH MINI-BATCH SAMPLING FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Danfeng Hong[1], *Lianru Gao*[2,*], *Xin Wu*[3], *Jing Yao*[2], *Bing Zhang*[2,4]

[1] Remote Sensing Technology Institute, Germany Aerospace Center, Wessling, Germany
[2] Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China
[3] School of Information and Electronics, Beijing Institute of Technology, Beijing, China
[4] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China

ABSTRACT

Graph convolutional networks (GCNs) have been successfully and widely applied in computer vision and machine learning fields. As a powerful tool, GCNs have recently received increasing attention in the remote sensing community, e.g., hyperspectral image (HSI) classification. However, the application ability of GCNs in identifying the materials via spectral signatures remains limited, since traditional GCNs fail to extract node features for large-scale graphs efficiently. Also, simultaneous consideration of all samples in GCNs tends to obtain poor representations, possibly due to the vanishing gradient problem. To this end, we in this paper develop a novel mini-batch GCN (miniGCN) for HS image classification. More importantly, miniGCN not only can effectively train the network via mini-batch sampling in a supervised way, but also directly infer new samples (out-of-sample) without re-training GCNs. Experiments conducted on two commonly-used HSI datasets demonstrate the superiority of miniGCN over other state-of-the-art network architectures. The codes of this work are available at https://github.com/danfenghong/IEEE_TGRS_GCN for the sake of reproducibility.

Index Terms— Classification, deep learning, graph convolutional network, hyperspectral image, mini-batch.

1. INTRODUCTION

Owing to rich spectral information, the hyperspectral image (HSI) is capable of detecting or identifying the materials or objects of interest more easily and accurately [1]. Over the past decades, both hand-crafted feature extraction, e.g., filtering [2], morphological profiles [3], image descriptors [4], and subspace learning-based feature representation, e.g., sparse representation [5], manifold learning [6], approaches have been largely developed and successfully applied in the HSI classification task. Nevertheless, these methods inevitably meet the performance bottleneck due to the lack of more powerful spectral representability.

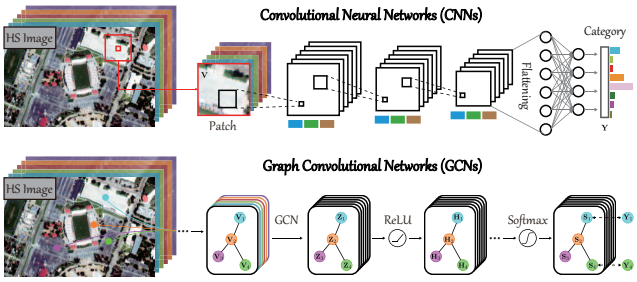

Fig. 1: An illustration to compare the basic network architecture of CNNs and GCNs in HSI classification.

Inspired by the recent success of deep learning (DL) techniques [7, 8], many advanced deep networks have been proposed and obtain better HSI classification results, such as fully connected network, (FCN), autoencoder (AE), convolutional neural network (CNN), recurrent neural network (RNN), generative adversarial network (GAN), etc. Yet they only regard the HSI as grid data and extract feature representations in the Euclidean space. This, to a great extent, fails to consider and measure relations between samples in feature extraction and classification. In recent years, graph convolutional networks (GCNs) [9] have been garnering growing interest in HSI classification, due to its strong ability in handling graph structure data by effectively modeling long- and short-range relations between samples or spectral signatures. Fig. 1 illustrates the differences between GCNs and CNNs in HSI classification. However, some potential problems that lie in GCNs limit the classification performance of HSIs to be further improved and its practical applications, e.g., high computational cost (large graphs), out-of-sample issue (fail to predict new samples without re-training), large memory costs and slow gradient descent (full-batch training).

To overcome these difficulties, we propose to use mini-batch sampling for GCNs, mini-batch GCNs (miniGCNs) for short, which can be trained in a mini-batch fashion for HSI

* Corresponding author, e-mail: gaolr@aircas.ac.cn.

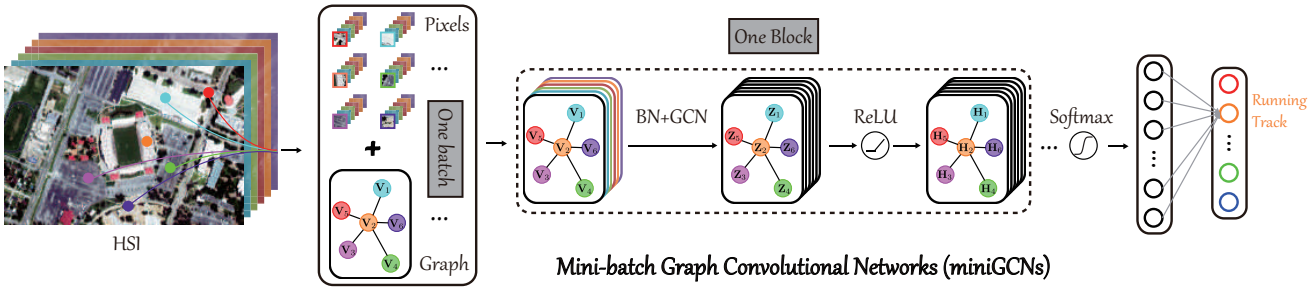

Fig. 2: Network architecture illustration of the proposed miniGCNs for HSI classification.

classification, being the same as CNNs. Note that journal version has been published in IEEE TGRS, where the readers can find more details [10]. More specifically, the main contributions of this paper can be summarized as follows

- The proposed miniGCNs not only can address the issue of high computational cost caused by full-batch training, tending to find a better local optimum, but also can infer new samples out of the training set without re-training GCNs.
- To our best knowledge, this is the first time to propose to use GCNs with mini-batch training for the HSI classification task. This might provide a new sight to guide how to use GCNs to classify the HSI more effectively.

The rest of this paper is organized as follows. Section 2 details the network architecture and clarifies how to sample the whole graph. Experiential results are conducted on the two hyperspectral datasets in comparison with representative baseline methods in Section 3. Finally, Section 4 draws a conclusion with a possible future outlook.

2. MINI-BATCH GRAPH CONVOLUTIONAL NETWORKS

2.1. A Brief Review of GCNs

The graph $\mathcal{G}$ is the core structure in GCNs, which aims to depict the relationship (via edges $\mathcal{E}$) between nodes $\mathcal{V}$ in a non-Euclidean space. Accordingly, the pixels in HSIs and their similarities can be directly seen as nodes and edges in a graph, respectively. The adjacency matrix $\mathbf{A}$ [11, 12] is usually used to define the graph structure to represent the similarities between pixels, which can be computed by

$$\mathbf{A}_{i,j} = \exp(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}), \tag{1}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ denote two spectral signatures corresponding to two different HS pixels, and σ is the width of the radial basis function. Using Eq. (1), the Laplacian matrix $\mathbf{L}$ can be obtained by $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{D}$ denotes the diagonal matrix given by $\mathbf{D}_{i,i} = \sum_j \mathbf{A}_{i,j}$.

According to [9], the propagation rule from the l-th layer to $l+1$-th layer in GCNs can be deducted as follows:

$$\mathbf{H}^{(\ell+1)} = h(\mathbf{A}\mathbf{H}^{(\ell)}\mathbf{W}^{(\ell)} + \mathbf{b}^{(\ell)}), \tag{2}$$

where $\mathbf{H}^{(l)}$ denotes the hidden representations in the l-th layer with respect to the variables $\mathbf{W}$ and $\mathbf{b}$. Moreover, to avoid the scaling factors in the propagation process between layers, the normalized adjacency matrix ($\widetilde{\mathbf{A}}$) is used to replace $\mathbf{A}$ in Eq. (2), we then have the following updated rule:

$$\mathbf{H}^{(\ell+1)} = h(\widetilde{\mathbf{D}}^{-\frac{1}{2}}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(\ell)}\mathbf{W}^{(\ell)} + \mathbf{b}^{(\ell)}). \tag{3}$$

Using Eq. (4) to replace the traditional layer-wise propagation rule in FCNs or CNNs, the GCNs can be also used for the classification task.

2.2. Proposed MiniGCNs

The traditional GCN is a semi-supervised learning approach with full-batch network training since it has to consider all samples in a large graph form in the training process. Training GCNs with such a large graph not only needs high computational cost and memory storage, which can not usually be performed in the PC, but also tends to meet the performance bottleneck. More importantly, the trained network parameters of GCNs fail to be directly used to predict the new samples without re-training.

Through the above analysis around the drawbacks and limitations of GCNs, we expect to make GCNs trainable in a mini-batch fashion. For this purpose, we propose miniGCNs, which is trained in a supervised manner, i.e., without any information of unlabeled samples. This also makes it possible for testing samples to be directly inferred by using the trained networks. The network architecture of the proposed miniGCNs for HSI classification is given in Fig. 2.

Intuitively, our miniGCNs sample the whole graph to perform mini-batch training of GCNs. Supported by the proposition given in [13], the update rule of graph convolutions in the s-th batch can be rewritten as

$$\widetilde{\mathbf{H}}_s^{(\ell+1)} = h(\widetilde{\mathbf{D}}_s^{-\frac{1}{2}}\widetilde{\mathbf{A}}_s\widetilde{\mathbf{D}}_s^{-\frac{1}{2}}\widetilde{\mathbf{H}}_s^{(\ell)}\mathbf{W}^{(\ell)} + \mathbf{b}_s^{(\ell)}). \tag{4}$$

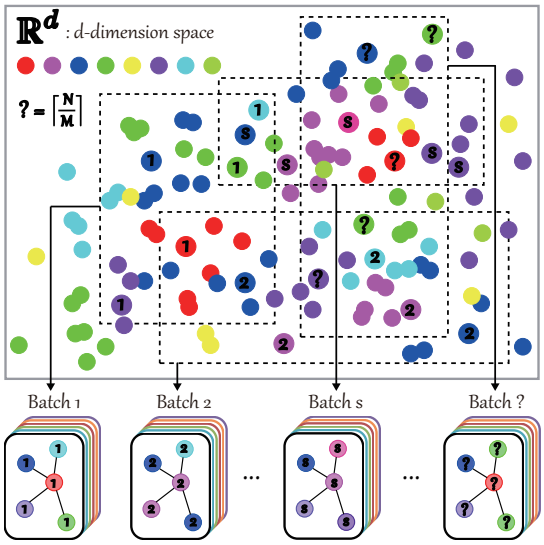

Fig. 3: An example to show how can miniGCNs sample the whole graph to achieve the mini-batch training in GCNs.

Considering a special case, i.e., randomly sampling without replacement in the whole graph [14], the hidden representations of the $(\ell+1)^{th}$ layer in one epoch can be then obtained by collecting results of all batches:

$$\mathbf{H}^{(\ell+1)} = [\widetilde{\mathbf{H}}_1^{(\ell+1)}, \cdots, \widetilde{\mathbf{H}}_s^{(\ell+1)}, \cdots, \widetilde{\mathbf{H}}_n^{(\ell+1)}], \quad (5)$$

where n is the number of sampled batches. Fig. 3 illustrates the mini-batch sampling process in the proposed miniGCNs.

3. EXPERIMENTS

3.1. Data Preparation

3.1.1. Indian Pines Data

The first dataset has been widely used in the HSI classification, which was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor, covering the Indian Pines forest, Indiana, USA. The image consists of 145×145 pixels and 220 spectral bands ranging from 400 nm to 2500 nm. However, only 200 channels are retained due to the effects of noises and water absorption in 20 bands of all ones, e.g., 104-108, 150-163, and 220. In the studied scene, the details about the number of training and testing samples for each category can be found in [15], and a false-color image and the spatial distribution of training and testing samples are shown in Fig. 4.

3.1.2. Houston2013 Data

The second HSI was provided by the 2013 IEEE GRSS data fusion contest, which was acquired by the ITRES CASI-1500 sensor. The image scene consists of 349×1905 pixels and 144 spectral channels in the range of 364 nm to 1046 nm.

Table 1: Performance comparison of HSI classification methods on the Indian Pines data. The best one is shown in bold.

Class No.	KNN	RF	SVM	1-D CNN	GCNs	miniGCNs
1	45.45	57.80	67.34	47.83	65.97	**72.54**
2	46.94	56.51	67.86	42.35	**72.70**	55.99
3	77.72	80.98	**93.48**	60.87	87.50	92.93
4	84.56	85.68	**94.63**	89.49	93.74	92.62
5	80.06	79.34	88.52	92.40	91.39	**94.98**
6	97.49	95.44	94.76	97.04	97.49	**98.63**
7	64.81	**77.56**	73.86	59.69	75.38	64.71
8	48.68	58.85	52.07	65.38	51.70	**68.78**
9	44.33	62.23	72.70	**93.44**	62.7	69.33
10	96.30	95.06	98.77	**99.38**	96.91	98.77
11	74.28	**88.75**	86.17	84.00	86.25	87.78
12	15.45	54.24	71.82	**86.06**	66.97	50.00
13	91.11	97.78	95.56	91.11	95.56	**100.00**
14	33.33	56.41	82.05	**84.62**	71.79	48.72
15	81.82	81.82	90.91	**100.00**	81.82	72.73
16	40.00	**100.00**	**100.00**	80.00	**100.00**	80.00
OA (%)	59.17	69.80	72.36	70.43	71.97	**75.11**
AA (%)	63.90	76.78	83.16	79.60	**81.12**	78.03
κ	0.5395	0.6591	0.6888	0.6642	0.6852	**0.7164**

Note that this dataset is processed by removing the effects of clouds using illumination-related threshold methods based on the similarities between spectral signatures. This cloud-free hyperspectral product was generated by Prof. N. Yokoya from the University of Tokyo and RIKEN AIP. Fig. 5 shows a false-color image of the studied hyperspectral scene and the spatial distribution of the used training and testing sets.

3.2. Experimental Setup

3.2.1. Network Configuration

The networks, including the proposed miniGCNs and compared models are implemented on the Tensorflow platform. The Adam [16] is selected as the optimizer to update the network parameters. Also, an "exponential" policy is adopted to dynamically update the learning rate at intervals of 50 epochs by multiplying a basic learning rate (0.001 in our case), i.e., $(1 - \frac{iter}{maxIter})^{0.5}$. Moreover, we set the maximum number of epochs as 200 and the ℓ_2-norm regularization parameter as 0.001. To speed up the network learning and reduce the effects of over-fitting, batch normalization (BN) [17] is used with 0.9 momentum and 32 batch size.

3.2.2. State-of-the-art Compared Methods

To verify the classification performance of the proposed miniGCNs, we compare with several well-known and representative state-of-the-art baselines, such as K-nearest neighbor (KNN) classifier, random forest (RF), support vector machine (SVM), 1-D CNN, GCNs, in terms of three quantitative indices: *Overall Accuracy (OA)*, *Average Accuracy (AA)*, *Kappa Coefficient (κ)*.

Table 2: Performance comparison of HSI classification methods on the Houston2013 data. The best one is shown in bold.

Class No.	KNN	RF	SVM	1-D CNN	GCNs	miniGCNs
1	83.19	83.38	83.00	87.27	90.14	**98.39**
2	95.68	98.40	98.40	98.21	**99.08**	92.11
3	99.41	98.02	99.60	**100.00**	79.94	99.60
4	97.92	97.54	**98.48**	92.99	96.69	96.78
5	96.12	96.40	**97.82**	97.35	86.18	97.73
6	92.31	**97.20**	90.91	95.10	33.33	95.10
7	80.88	82.09	90.39	77.33	**97.09**	57.28
8	48.62	40.65	40.46	51.38	**71.63**	68.09
9	**72.05**	69.78	41.93	27.95	70.93	53.92
10	53.19	57.63	62.64	**90.83**	72.17	77.41
11	**86.24**	76.09	75.43	79.32	85.22	84.91
12	44.48	49.38	60.04	76.56	63.41	**77.23**
13	28.42	61.40	49.47	**69.47**	62.34	50.88
14	97.57	**99.60**	98.79	99.19	89.73	98.38
15	98.10	97.67	97.46	98.10	**99.36**	98.52
OA (%)	77.30	77.48	76.91	80.04	81.19	**81.71**
AA (%)	78.28	80.35	78.99	82.74	79.82	**83.09**
κ	0.7538	0.7564	74.94	0.7835	0.7962	**0.8018**

3.3. Results and Analysis

Tables 1 and 2 quantify the classification accuracies of different HSI classification algorithms on both datasets, while Figs. 4 and 5 correspondingly visualize the classification maps for qualitative comparison.

Overall, the performance using the traditional classifiers (e.g., KNN, RF, SVM) is inferior to that using deep networks. It should be noted, however, that SVM sightly performs better than other classifiers as well as 1-D CNN and GCNs on the Indian Pines data. The GCNs tend to obtain higher classification accuracies compared to 1-D CNN, due to the use of unlabeled samples and the embedding of relations between samples, nearly 1%-2% OAs on both datasets. As expected, our proposed miniGCNs outperform other compared methods observably, particularly on the Indian Pines data, despite only a supervised network learned on the training set. Although the miniGCNs only achieve a slight performance improvement (about 0.5% OA) over GCNs on the Houston2013 dataset, yet this can also show the effectiveness of the mini-batch sampling in method. In other words, the performance is comparable at least without using additional unlabeled samples and the complete graph structure. In addition, there are similar visual trend in Figs. 4 and 5, where miniGCNs, to some extent, are capable of identifying and classifying the materials, demonstrating more realistic classification results.

4. CONCLUSION

In this paper, we propose a novel GCN version, called miniGCN, which can be trained in a mini-batch fashion, just like CNN, for HSI classification. MiniGCNs have the advantages over traditional GCNs, which not only can train the GCNs more robust and faster, due to the utilization of mini-batch sampling, but also can infer the out of new samples directly without the need of re-training a new GCN. In the future, we would like to further investigate different sampling strategies for graphs and also develop more advanced GCNs on both performance and efficiency.

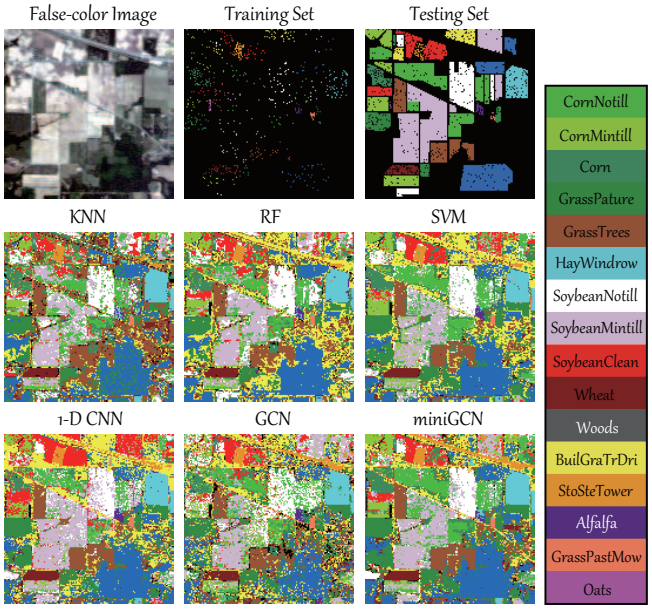

Fig. 4: Classification maps of different HSI classification methods on the Indian Pines data.

5. ACKNOWLEDEGEMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 42030111 and Grant 41722108, and in part by the AXA Research Fund.

6. REFERENCES

[1] D. Hong, N. Yokoya, J. Chanussot, and X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. on Image Process.*, vol. 28, no. 4, pp. 1923–1938, 2019.

[2] X. Cao, B. Ji, Y. Ji, L. Wang, and L. Jiao, "Hyperspectral image classification based on filtering: a comparative study," *J. Appl. Remote Sens.*, vol. 11, no. 3, p. 035007, 2017.

[3] A. Samat, E. Li, W. Wang, S. Liu, C. Lin, and J. Abuduwaili, "Meta-xgboost for hyperspectral image classification using extended mser-guided morphological profiles," *Remote Sens.*, vol. 12, no. 12, p. 1973, 2020.

[4] D. Hong, W. Liu, J. Su, Z. Pan, and G. Wang, "A novel hierarchical approach for multispectral palmprint recognition," *Neurocomputing*, vol. 151, pp. 511–521, 2015.

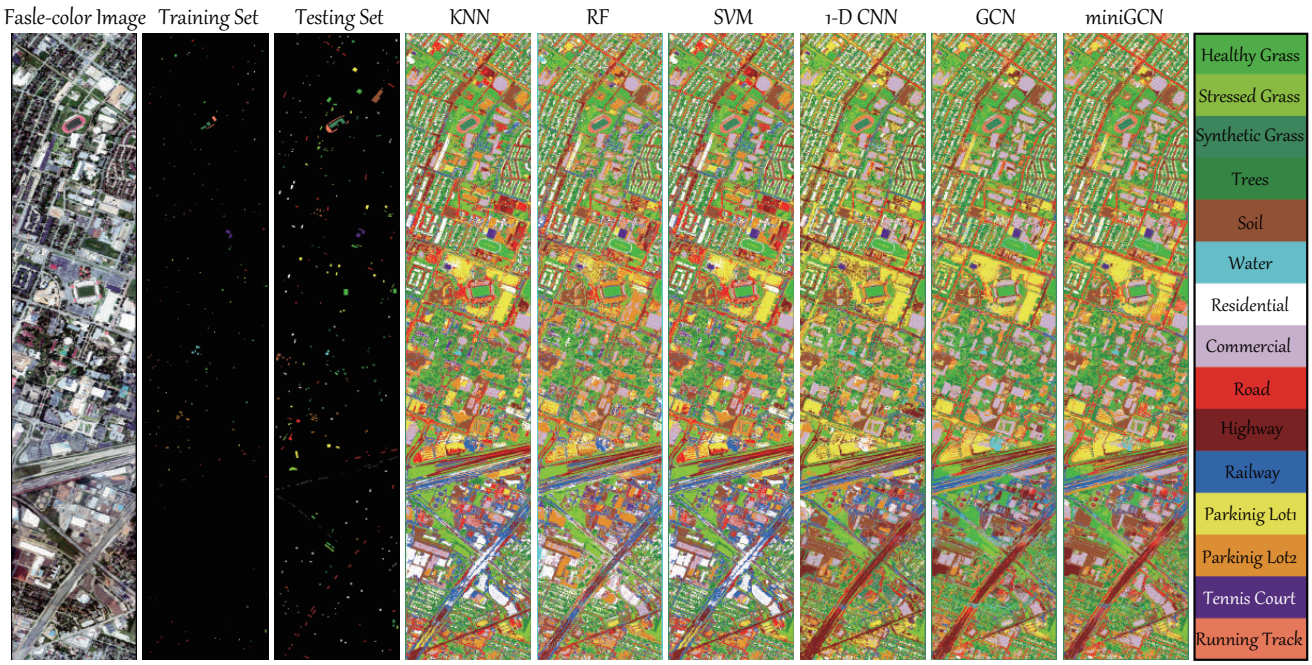

Fig. 5: Classification maps of different HSI classification methods on the Houston2013 data.

[5] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, 2014.

[6] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. Zhu, "Learning to propagate labels on graphs: An iterative multitask regression framework for semi-supervised hyperspectral dimensionality reduction," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 35–49, 2019.

[7] B. Rasti, D. Hong, R. Hang, P. Ghamisi, X. Kang, J. Chanussot, and J. Benediktsson, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, 2020.

[8] D. Hong, L. Gao, N. Yokoya, J. Yao, J. Chanussot, D. Qian, and B. Zhang, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, 2020. DOI: 10.1109/TGRS.2020.3016820.

[9] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[10] D. Hong, L. Gao, J. Yao, B. Zhang, P. Antonio, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, 2020. DOI: 10.1109/TGRS.2020.3015157.

[11] D. Hong, N. Yokoya, J. Chanussot, and X. Zhu, "CoSpace: Common subspace learning from hyperspectral-multispectral correspondences," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4349–4359, 2019.

[12] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. Zhu, "Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 147, pp. 193–205, 2019.

[13] H. Zeng, H. Zhou, A. Srivastava, R. Kannan, and V. Prasanna, "Graphsaint: Graph sampling based inductive learning method," *arXiv preprint arXiv:1907.04931*, 2019.

[14] D. G. Horvitz and D. J. Thompson, "A generalization of sampling without replacement from a finite universe," *J. Am. Stat. Assoc.*, vol. 47, no. 260, pp. 663–685, 1952.

[15] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. Zhu, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3791–3808, 2020.

[16] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[17] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.