Multifrequency Graph Convolutional Network With Cross-Modality Mutual Enhancement for Multisource Remote Sensing Data Classification

Jin-Yu Yang, Heng-Chao Li, *Senior Member, IEEE*, Jing-Hua Yang, *Member, IEEE*, Lei Pan, Qian Du, *Fellow, IEEE*, and Antonio Plaza, *Fellow, IEEE*

***Abstract*— The mining of meaningful features and effective fusion of multisource remote sensing (RS) data have always been the challenging research problems in the joint classification of hyperspectral image (HSI) and light detection and ranging (LiDAR) data. In this article, we propose a multifrequency graph convolutional network with cross-modality mutual enhancement (MFGCN-CME) for multisource RS data classification. Specifically, we design an adaptive multifrequency graph feature learning module to capture the low- and high-frequency multiscale features of HSI and LiDAR in parallel and further adaptively aggregate them. Then, we propose a bipartite graph (BG) enhancement learning module to obtain the spatial-enhanced HSI features and spectral-enhanced LiDAR features by propagating intermodality information. To the best of our knowledge, the BG is first used to multisource RS data classification task. Furthermore, compared with traditional fusion methods, a gated fusion module is used to fully explore the complementarity of two data sources. Finally, a joint loss function combing a classification loss and a semisupervised contrastive loss is developed to improve the model robustness. Comprehensive experiments on different HSI and LiDAR datasets demonstrate that our proposed method can yield better performance compared with several state-of-the-art multisource RS data classification methods.**

***Index Terms*— Bipartite graph (BG), contrastive learning, gated fusion, graph convolutional neural networks (CNNs), multifrequency, multisource remote sensing (RS) data classification.**

I. Introduction

With the rapid development of satellite sensor technology, remote sensing (RS) images, i.e., hyperspectral image (HSI) [1], [2], light detection and ranging (LiDAR), multispectral image (MSI) [3], and synthetic aperture radar (SAR), are acquired by various sensors, which could measure different aspects of the same object on the Earth's surface. Multisource RS data classification task plays a significant role in Earth observation mission applications, such as plant monitoring [4] and urban development [5]. Among these multisource RS data, HSI provides plenty of spectral information of land covers, but its passive imaging mode makes it susceptible to cloudy weather. LiDAR is an active RS technology, providing elevation and spatial information robust to weather and illumination variations. Currently, the comprehensive utilization of HSI and LiDAR data has received extensive attention for accurate land cover classification.

The HSI classification task, which assigns each pixel to a unique category by using the spatial and spectral information [6], [7], is useful for a variety of applications, i.e., military target detection [8], vegetation monitoring [9], and disaster prevention and control [10]. Due to the complex spatial distribution and the spectral heterogeneity of objects, the HSI classification is still a challenging task. Some traditional methods for HSI classification, such as support vector machine (SVM) [11], random forest [12], sparse-collaborative representation [13], multiple logistic regression [14], and spectral clustering [15], extract the spatial–spectral features to improve the final classification accuracy. However, these techniques are quite empirical and depend heavily on professional expertise. Inspired by the success of deep learning methods to automatically extract discriminative and abstract features, deep neural networks, such as autoencoder networks [16], [17], convolutional neural networks (CNNs) [18], [19], [20], deep belief networks [21], capsule networks [22], recurrent neural networks (RNNs) [23], and long short-term memory networks [24], have been rapidly introduced into the HSI classification task and have produced satisfactory results. However, these deep learning methods often ignore the intrinsic correlation between adjacent land covers in HSIs.

To address the above problem, graph convolutional networks (GCNs) aggregate and transform information from the neighbors of each pixel in HSIs to better model the

Manuscript received 8 October 2023; revised 2 January 2024; accepted 14 January 2024. Date of publication 19 January 2024; date of current version 1 February 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62271418 and Grant 62001437, and in part by the Natural Science Foundation of Sichuan Province under Grant 2023NSFSC0030 and Grant 24NSFSC7075. *(Corresponding authors: Heng-Chao Li; Jing-Hua Yang.)*

Jin-Yu Yang, Heng-Chao Li, and Jing-Hua Yang are with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China (e-mail: hcli@home.swjtu.edu.cn; yangjinghua110@126.com).

Lei Pan is with the Southwest Institute of Electronic Technology, Chengdu 610036, China.

Qian Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA.

Antonio Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politecnica, University of Extremadura, 10003 Cáceres, Spain.

Digital Object Identifier 10.1109/TGRS.2024.3356510

spatial contextual structure for HSI classification [25], [26]. According to the usage of labeled and unlabeled data, GCN-based HSI classification methods can generally be divided into three categories, i.e., the supervised GCN, the unsupervised GCN, and the semisupervised GCN. In the supervised GCN models, the labeled data are used to construct a downsampled graph (or topological structure) for the training of the network for HSI classification [27], [28]. Nevertheless, collection of labeled data is a highly time-consuming, expensive, and a complicated process. Unlabelled samples are largely available at no cost. The unsupervised GCN learning framework is trained by a given set of unlabeled data for HSI classification task [29]. However, the two different types of GCN-based HSI classification models mentioned above separately involve labeled and unlabeled data in train stage, which fail to fully utilize the effective sample information and result in the poor generalization ability. This has motivated researchers to develop the semisupervised GCN model for HSI classification task, which encodes the set of labeled and unlabeled samples into a graph based on their spectral similarity or spatial distance to train the HSI classification model and improve the classification performance [30], [31], [32], [33].

Although the above HSI classification methods have achieved good performance, the presence of numerous mixed pixels in the low-resolution HSI poses a significant challenge in distinguishing objects with comparable spectral reflectance due to the limitations of hardware imaging equipment. Unlike HSI data, LiDAR data bring high-precision digital elevation models and information on the target altitude [34], which can provide complementary information for the HSI data to satisfy different application requirements. Therefore, more and more researchers take advantage of HSI and LiDAR to interpret ground objects at more detailed and precise level. Recently, GCNs have also been used to extract the structural information of HSI and LiDAR for multisource RS data classification task. For example, in [35], a dual GCN extracted the spectral features with spectrum internal connection and a dense network acquired the spatial feature, which are fused by a fully connected network for HSI and LiDAR classification. In addition, a dual-coupled CNN-GCN that contains a coupled CNN and a coupled GCN was proposed to achieve the complementarity of spatial–spectral features and structural features from hyperspectral and LiDAR data for classification task [36]. Xiu et al. [37] combined the self-supervised feature extraction module and the semisupervised graph attention network module to mitigate the dependence on annotated samples in the joint land cover classification with the HSI and LiDAR data. These GCN-based methods are beneficial for feature extraction, but the effective fusion of HSI and LiDAR features is also a key issue for multisource RS classification task. The feature fusion combines the features from different data sources while preserving the sufficient semantic information, and finally sends the fusion feature to a classifier to complete classification. For instance, a graph fusion strategy fused the graph features from different sources by a transition graph [38]. By fusing the weight matrices of HSI and LiDAR data, Du et al. [39] built a multimodal graph to effectively fuse the feature information between multisource RS data for classification task. Zhang et al. [40] proposed a novel SOTNet, which transmits structural information and aligns physical properties, for effectively improving the collaborative performance of multisource RS. In the unified multimodal deep learning framework [41], a novel fusion strategy, namely, cross fusion, was proposed to learn more compact representations across modalities by interactively updating the parameters of different subnetworks for better classification performance.

Despite the fact that the above GCN-based multisource RS data classification methods have obtained promising results in the joint classification of HSI and LiDAR, four key issues still need to be solved.

1) Existing methods which rely on low-pass filters produce the poor classification accuracy on tiny objects and boundary regions.
2) Existing multisource RS data classification methods often use the weight sharing strategy to explore the shared feature representations, which are hard to achieve the interaction of information from multisource data and may obtain nondiscriminative feature representations.
3) The above multisource RS data classification methods often use the additive, elementwise multiplicative, and concatenation strategies to fuse the HSI and LiDAR data, which are difficult to fully explore the complementarity of these two different data.
4) Most methods only consider semantic matching on corresponding positions and ignore topology and label information of HSI and LiDAR data, leading to the suboptimal classification results.

Therefore, in this article, we propose a multifrequency graph convolutional network with cross-modality mutual enhancement (MFGCN-CME) for multisource RS data classification task. Experiments demonstrate the effectiveness of the proposed method. The main contributions of this article are presented as follows.

1) We develop an adaptive multifrequency graph feature learning module (AMGFLM) to learn the low- and high-frequency multiscale features of HSI and LiDAR data by graph convolutional and graph wavelet convolutional operations, and further use attention mechanism to fuse the multiscale features, respectively, which addresses the above issue (1).
2) We design a bipartite graph enhancement learning module (BGELM) to achieve cross-modality mutual enhancement of HSI and LiDAR features, which solves the limitation (2). The bipartite graph (BG) is first introduced into multisource RS classification. Furthermore, a gated fusion module is introduced to fully integrate different types of information from HSI and LiDAR data sources and attack the limitation (3).
3) We design a semisupervised contrastive loss to promote representation learning by constructing positive and negative sample pairs based on the geometrical and label information between HSI and LiDAR data, thus boosting MFGCN-CME effectiveness and robustness. The module alleviates the above limitation (4).

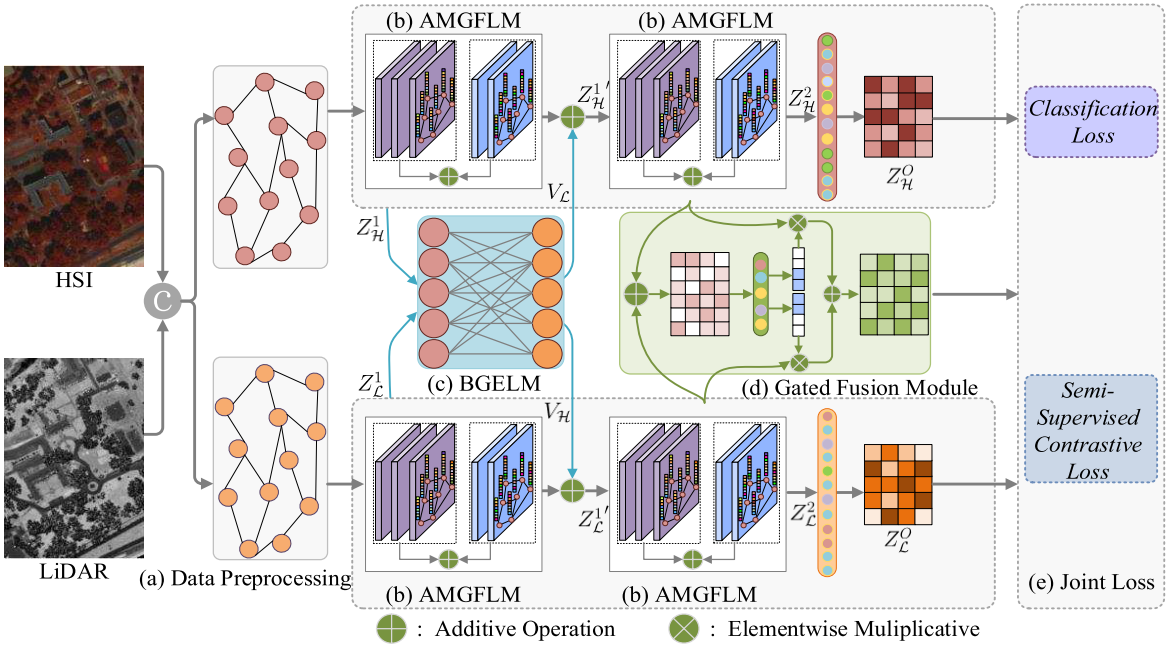


Fig. 1. Illustration of our proposed model. (a) Data preprocessing constructs HSI and LiDAR graphs. (b) AMGFLM is designed to capture the low-frequency multiscale features by graph convolutional operation in the purple modules and high-frequency multiscale features by graph wavelet convolutional operation in the blue modules, and further adaptively aggregate them. (c) BGELM is used to obtain the spatial-enhanced feature representations of HSI data by LiDAR features and spectral-enhanced feature representations of LiDAR data by HSI features. (d) Gated fusion module adaptively integrates the features from HSI and LiDAR. (e) Joint loss function that combines the classification loss and the semisupervised contrastive loss improves model robustness.

The remainder of this article is organized as follows. In Section II, the proposed method is described in detail. Comprehensive quantitative analysis and evaluation of this method are implemented in Section III. This article is finally concluded in Section IV.

II. Proposed Method

A. Method Overview

HSI data contain many bands carrying a lot of spectral information, and LiDAR data are rich in height (spatial) information. These motivate us to design an MFGCN-CME for better joint HSI and LiDAR data classification performance. The overall architecture of the proposed method is represented in Fig. 1. Specifically, the input HSI and LiDAR data are cascaded and sent to the simple linear iterative clustering (SLIC). HSI and LiDAR graphs are constructed in Section II-B. After that, the AMGFLM in Section II-C learns the fusion of low- and high-frequency multiscale features. Then, the BGELM can enhance spatial representations of HSI data using LiDAR features and enhance spectral representations of LiDAR data using HSI features, as shown in Section II-D. Next, the gated fusion module explores the complementarity of different features from HSI and LiDAR in Section II-E. Finally, Section II-F introduces the joint loss function to optimize all of the parameters in the proposed method. The details of these modules are described as follows.

B. Data Preprocessing

Assuming that $X_{\mathcal{H}} \in \mathbb{R}^{H_{\mathcal{HL}} \times W_{\mathcal{HL}} \times B_{\mathcal{H}}}$ and $X_{\mathcal{L}} \in \mathbb{R}^{H_{\mathcal{HL}} \times W_{\mathcal{HL}} \times B_{\mathcal{L}}}$ represent the HSI and LiDAR data, respectively, where $H_{\mathcal{HL}}$ and $W_{\mathcal{HL}}$ denote the height and width of both data, $B_{\mathcal{H}}$ is the number of spectral bands in the HSI data, and $B_{\mathcal{L}}$ is the number of channels in the LiDAR data. These data usually contain a large number of pixels $N = H_{\mathcal{HL}} \times W_{\mathcal{HL}}$ in the spatial dimension. In the construction of HSI and LiDAR graphs, if each pixel of HSI and LiDAR data corresponds to a graph node, the large-scale adjacency matrix will require large storage space and high computational complexity. Therefore, the input HSI and LiDAR data are cascaded along the channel dimension. Next, the SLIC segments the cascaded image into a small number of compact homogeneous superpixels, which is defined as

$$\hat{X} = f_{\text{SLIC}}(X_{\mathcal{H}} || X_{\mathcal{L}}) \tag{1}$$

where $\hat{X} \in \mathbb{R}^{N_C \times (B_{\mathcal{H}} + B_{\mathcal{L}})}$ ($N_C \ll N$) represents the segmentation results, $f_{\text{SLIC}}(\cdot)$ denotes the SLIC, and $||$ is the concatenation operation.

As shown in Fig. 1(a), to determine the dual branch MFGCN-CME for HSI or LiDAR data, $\hat{X}$ is split into $\hat{X_{\mathcal{H}}} \in \mathbb{R}^{N_C \times B_{\mathcal{H}}}$ and $\hat{X_{\mathcal{L}}} \in \mathbb{R}^{N_C \times B_{\mathcal{L}}}$ along the channel dimension. Graph structured data contain two matrices: a feature matrix and an adjacency matrix. For HSI data, each above-obtained HSI superpixel is regarded as a node in HSI graph. The feature $\hat{x}_{\mathcal{H}_i}$ of the ith superpixel is obtained by the average of its all pixels, and $\hat{X}_{\mathcal{H}} = \{\hat{x}_{\mathcal{H}_1}, \hat{x}_{\mathcal{H}_2}, \ldots, \hat{x}_{\mathcal{H}_{N_C}}\}$ is referred as the feature matrix of HSI. The adjacency matrix $A_{\mathcal{H}}$ of HSI indicates the relationship between superpixels, which is calculated as

$$A_{\mathcal{H}_{ij}} = \begin{cases} e^{-0.2\|\hat{x}_{\mathcal{H}_i} - \hat{x}_{\mathcal{H}_j}\|^2}, & \text{if } \hat{x}_{\mathcal{H}_j} \in \mathcal{N}(\hat{x}_{\mathcal{H}_i}) \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where $\hat{x}_{\mathcal{H}_i}$ represents the ith superpixel in HSI, and $\mathcal{N}(\hat{x}_{\mathcal{H}_i})$ is the set of neighbors of $\hat{x}_{\mathcal{H}_i}$. In addition, the HSI graph is

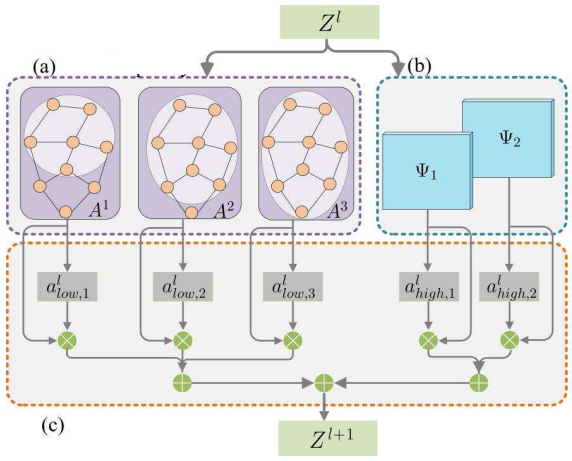


Fig. 2. Proposed AMGFLM utilizes graph convolutional and graph wavelet convolutional operations to learn the low-frequency and high-frequency multiscale features, and further introduces attention mechanism to adaptively aggregate them. (a) Low frequency. (b) High frequency. (c) Attention mechanism.

dynamically updated to explore the optimal graph and produce the discriminative embedded features [31]. Similar to HSI data, $\hat{X}_{\mathcal{L}}$ is the feature matrix of LiDAR, and the adjacency matrix $A_{\mathcal{L}}$ of LiDAR data is obtained by (2).

C. Adaptive Multifrequency Graph Feature Learning Module

In general, the frequency information in the image reflects the speed of changes in pixel values in the spatial domain. For HSI and LiDAR data, the smooth areas in the spatial domain represent their low-frequency components, corresponding to the main body of each class in the image. The high-frequency components imply a higher change of pixel in the image, which can be considered as the class boundary. We design an AMGFLM to capture the low- and high-frequency multiscale features in HSI and LiDAR by graph convolutional operation and graph wavelet convolutional operation, and use attention mechanism to aggregate these multiscale features, respectively, which is depicted in Fig. 1(b). Therefore, the proposed module can correctly classify the main part and the boundary of each class simultaneously, which can address the issue (1) in Introduction. The detailed flowchart of our proposed AMGFLM is described in Fig. 2.

The proposed AMGFLM utilizes graph convolutional operation and graph wavelet convolutional operation to learn the low-frequency and high-frequency multiscale features. Specifically, GCN utilizes Laplacian smoothing filter to make adjacent nodes on the graph similar, so the node features in the graph domain should be smooth [42]. Thus, we employ GCN to extract the low-frequency features. Moreover, considering that the high-order neighbor information contains valuable hidden relationship between nodes, a multiorder GCN is proposed to effectively capture the low-frequency multiscale features from multiorder neighbors. Further, only low-frequency features are not enough, we introduce graph wavelet convolutional operation, which defines the convolutional operation via the wavelet transform to learn high-frequency features [43]. Thus, by introducing multiscale wavelets, graph wavelet convolutional operation is used to obtain high-frequency multiscale features.

In the lth layer, the low-frequency features $H^{l'}_{\text{low}_s}$ and the high-frequency features $H^{l'}_{\text{high}_s}$ of the sth scale can be computed by

$$H^{l'}_s = \tilde{P}^l_s Z^l W^l_s \tag{3}$$

where $H^{l'}_s$ is the convolutional result on Z^l when Z^0 is $\hat{X}_{\mathcal{H}}$ or $\hat{X}_{\mathcal{L}}$. W^l_s is the sth scale's learnable parameter for feature transformation. $\tilde{P}^l_s$ in (3) is described as

$$\tilde{P}^l_s = \begin{cases} \tilde{D}_s{}^{-\frac{1}{2}} \tilde{A}^s \tilde{D}_s{}^{-\frac{1}{2}} & \text{(4a)} \\ \Psi_s F^l_s \Psi_s^\top & \text{(4b)} \end{cases}$$

where $\tilde{D}_s{}^{-(1/2)} \tilde{A}^s \tilde{D}_s{}^{-(1/2)}$ denotes the normalized adjacency matrix with self-connections in which $\tilde{D}_s$ is the diagonal degree matrix of $\tilde{A}^s = A^s + I$. In addition, Ψ_s is a set of wavelets as bases [43]. F^l_s indicates a learned diagonal filter matrix. Eqs. (4a) and (4b) are used to learn the low- and high-frequency features, respectively.

The attention mechanism assigns different weights to multiscale features so that their useful features are selectively emphasized and useless features are suppressed [44]. The low-frequency and high-frequency multiscale features are obtained separately by (4a) and (4b). Therefore, as one of the low-frequency multiscale features, we introduce attention mechanism to obtain attention weights for the low-frequency multiscale features, which are calculated as

$$a^l_{\text{low}_s} = \frac{\exp\left(H^{l'}_{\text{low}_s}\right)}{\sum_{j=1}^{C_{\text{low}}} \exp\left(H^{l'}_{\text{low}_j}\right)} \tag{5}$$

where C_{low} is the number of scales.

Next, the low-frequency features of different order neighbors are aggregated by the obtained weights to get the new representation, which can be learned by

$$H^{l'}_{\text{low}} = \sum_{s=1}^{C_{\text{low}}} a^l_{\text{low}_s} H^{l'}_{\text{low}_s}. \tag{6}$$

Similar to the fusion of low-frequency multiscale information mentioned above, attention mechanism is also introduced to calculate the weights of high-frequency multiscale information and fuse them. To ensure the integrity of the information, a communication mechanism is established between two frequencies to compliment the diverse information corresponding to low- and high-frequency features mutually

$$Z^{l+1} = \sigma\left(\lambda H^{l'}_{\text{low}} + (1-\lambda) H^{l'}_{\text{high}}\right) \tag{7}$$

where λ ($0 \le \lambda \le 1$) is the ratio of these two types of features. $\sigma(\cdot)$ is defined as ReLU$(\cdot)$ in this equation.

D. Bipartite Graph Enhancement Learning Module

In this section, to achieve spatial enhancement from LiDAR to HSI data and spectral enhancement from HSI to LiDAR data, we design a BGELM based on the BG and graph attention mechanism, as shown in Fig. 1(c). To the best of our knowledge, it is the first attempt to introduce the BG to

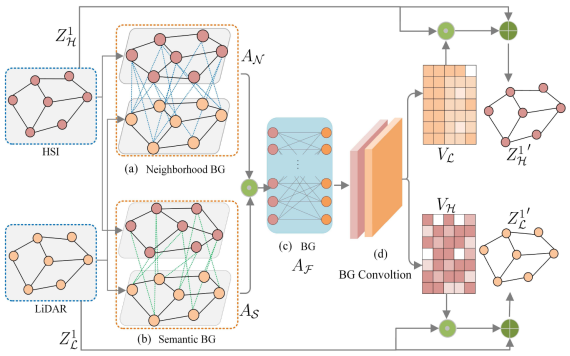


Fig. 3. Detailed architecture of the proposed BGELM. The proposed BGELM introduces the BG to model the relationship between HSI and LiDAR data, and uses BG convolution to propagate the spatial information of LiDAR into the HSI domain for enhancing the spatial features of HSI, and then employs residual learning to fuse the enhanced features and original features for obtaining $Z_{\mathcal{H}}^{1\prime}$. The same operation is used to obtain the spectral-enhanced information of LiDAR data $Z_{\mathcal{L}}^{1\prime}$. (a) Neighborhood BG. (b) Semantic BG. (c) BG. (d) BG convolution.

model the relationship between HSI and LiDAR data. Based on the constructed BG, the proposed module can propagate and aggregate intermodality information to enhance the spatial features of HSI data by LiDAR features and enhance the spectral features of LiDAR data by HSI features, thereby achieving the interaction of information from multisource data more effectively and obtaining discriminative feature representations. The module is used to solve the problem (2) in Introduction. The architecture of our proposed BGELM is shown in Fig. 3.

This module combines HSI and LiDAR data into one BG to emphasize the information with strong correlation and suppresses the weak one. The BG edge aims to reveal such underlying relationship between samples of HSI and LiDAR data, and the edge weights are increased to highlight their importance. Thus, we construct the BG in the following neighborhood and semantic relationships.

The neighborhood BG is designed to model the local neighborhood correlation between samples of HSI and LiDAR data. Specifically, consider that the samples at the same spatial position from different sources are the same object. In addition, the adjacent samples in each source image are also the same object. Based on these image characteristics, each edge in the neighborhood BG connects the two samples in the corresponding neighborhood from different sources. As illustrated in Fig. 3(a), the construction of the neighborhood BG is expressed as

$$A_{\mathcal{N}_{ij}} = \begin{cases} 1, & \text{if } z_{\mathcal{L}_j}^1 \in \mathcal{N}\left(z_{\mathcal{H}_i}^1\right) \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

where $z_{\mathcal{H}_i}^1$ and $z_{\mathcal{L}_j}^1$ are the output of the 1st layer from the HSI and LiDAR branches, respectively. The weight between the ith sample in HSI and its nearby jth samples in LiDAR is set as 1.

In addition, as shown in Fig. 3(b), the semantic BG is designed to learn the feature similarity of HSI and LiDAR data, which is described as

$$A_{\mathcal{S}_{ij}} = \text{sim}\left(z_{\mathcal{H}_i}^1, z_{\mathcal{L}_j}^1\right) \tag{9}$$

where $\text{sim}(\cdot)$ defines the cosine similarity between $z_{\mathcal{H}_i}^1$ and $z_{\mathcal{L}_j}^1$.

The elementwise multiplication operation fuses the neighborhood BG $A_{\mathcal{N}}$ and the semantic BG $A_{\mathcal{S}}$. And the normalized fusion graph is obtained by

$$\begin{aligned} A_{\mathcal{F}} &= A_{\mathcal{N}} \odot A_{\mathcal{S}} \\ \tilde{A}_{\mathcal{F}_{ij}} &= \frac{\exp\left(A_{\mathcal{F}_{ij}}\right)}{\sum_{k=1}^{N_c} \exp\left(A_{\mathcal{F}_{ik}}\right)} \end{aligned} \tag{10}$$

where $\odot$ indicates the elementwise dot.

Finally, the proposed BGELM based on the predefined BG fully takes advantage of the cross-modality reasoning procedure for feature enhancement. Specifically, the BG convolutional operation propagates spectral information from the HSI to LiDAR domain and enhances the features of LiDAR while propagating spatial information from the LiDAR to HSI domain and enhance the features of HSI. When HSI and LiDAR data are regarded as the main modality and as the auxiliary modality, the main idea of this module is that the spatial information of LiDAR data are mapped to the HSI domain in order to enhance the spatial features of HSI data. In particular, inspired by graph attention mechanism, the BG convolutional operation fuses all the neighboring nodes in the auxiliary modality with the learned weights to get the guidance matrix $V_{\mathcal{L}}$

$$v_{\mathcal{L}_i} = \sigma\left(\sum_{j=1}^{N_c} \tilde{A}_{\mathcal{F}_{ij}} z_{\mathcal{L}_j}^1\right) \tag{11}$$

where $v_{\mathcal{L}_i}$ is the features of the ith sample in $V_{\mathcal{L}}$. Furthermore, to avoid the degradation problem, residual learning is also utilized to fuse the enhanced features and original backbone features, which is expressed as

$$Z_{\mathcal{H}}^{1\prime} = \left(Z_{\mathcal{H}}^1 \odot V_{\mathcal{L}} + Z_{\mathcal{H}}^1\right) W_{\mathcal{H}} \tag{12}$$

where $W_{\mathcal{H}}$ is the trained weight matrix. $Z_{\mathcal{H}}^{1\prime}$ is the spatial-enhanced features of HSI data.

Similarly, we also use the proposed BGELM to get the spectral-enhanced LiDAR features. When LiDAR data are considered as the main modality, our proposed module propagates the spatial–spectral information of HSI to the LiDAR domain to obtain the guide matrix $V_{\mathcal{H}}$, which is used to enhance feature representation for LiDAR data. The enhanced features are added to original features calculated as follows:

$$\begin{aligned} v_{\mathcal{H}_j} &= \sigma\left(\sum_{i=1}^{N_c} \tilde{A}_{\mathcal{F}_{ji}}^{\top} z_{\mathcal{H}_i}^1\right) \\ Z_{\mathcal{L}}^{1\prime} &= \left(Z_{\mathcal{L}}^1 \odot V_{\mathcal{H}} + Z_{\mathcal{L}}^1\right) W_{\mathcal{L}} \end{aligned} \tag{13}$$

where $v_{\mathcal{H}_j}$ is the features of the jth sample in $V_{\mathcal{H}}$, $W_{\mathcal{L}}$ is the trained weight matrix, and $Z_{\mathcal{L}}^{1\prime}$ is the spectral-enhanced information of LiDAR data.

E. Gated Fusion Module

After getting the output features from the HSI and LiDAR branches, how to combine them becomes another critical issue. However, most of the existing deep learning models stack them together by using additive fusion, elementwise multiplicative fusion, and concatenation fusion, which make it challenging to explore the complementarity of HSI and

Algorithm 1 MFGCN-CME for Multisource RS Data Classification.

Input: The HSI data $X_{\mathcal{H}}$; The LiDAR data $X_{\mathcal{L}}$; Ground truth Y;
1: Segment the whole image into superpixels via SLIC algorithm;
2: Construct $\hat{X}_{\mathcal{H}}$, $A_{\mathcal{H}}$, $\hat{X}_{\mathcal{L}}$, and $A_{\mathcal{L}}$ according to Eqs. (1) and (2);
3: Parameter setting and weights initialization;
4: **for** epoch$\leq$10000 **do**
5: Calculate $Z^1_{\mathcal{H}}$ and $Z^1_{\mathcal{L}}$ according to Eq. (7);
6: Obtain the spatial-enhanced representation $Z^{1'}_{\mathcal{H}}$ of HSI data and the spectral-enhanced representation $Z^{1'}_{\mathcal{L}}$ of LiDAR data by Eqs. (11), (12), and (13);
7: Calculate $Z^2_{\mathcal{H}}$ and $Z^2_{\mathcal{L}}$ by Eq. (7);
8: Obtian $Z^O_{\mathcal{H}}$ and $Z^O_{\mathcal{L}}$ by fully connected layer and fuse $Z^2_{\mathcal{H}}$ and $Z^2_{\mathcal{L}}$ to gain $Z^O_{\mathcal{H}+\mathcal{L}}$ by Eq. (15);
9: Compute the HSI output $O_{\mathcal{H}}$, the LiDAR output $O_{\mathcal{L}}$, and the fusion output $O_{\mathcal{H}+\mathcal{L}}$;
10: Miniminze the joint loss function Eq. (16);
11: Back propagation and update the parameters by Adam optimizer;
12:**End for**
Output: Classification map $O_{\mathcal{H}+\mathcal{L}}$.

LiDAR features. In order to solve the problem that also is the issue (3) in Introduction, we propose a gated fusion module to dynamically fuse the spatial–spectral features of HSI data and spatial features of LiDAR data in our model by designating contribution of each branch with regard to their current characteristics. As a result, our model can pay more attention to discriminative features and suppress noise. As shown in Fig. 1(d), we first add the output features (i.e., $Z^2_{\mathcal{H}}$ and $Z^2_{\mathcal{L}}$) of the 2nd layer from the HSI and LiDAR branches to obtain the fusion features, and then input it to a fully connected layer to generate the gated vector g, which can be formulated as

$$g = \partial\left(W_g\left(Z^2_{\mathcal{H}} + Z^2_{\mathcal{L}}\right) + b_g\right) \tag{14}$$

where W_g is the trainable parameter. b_g is the bias. ∂ means the sigmoid function. The gate vector g is used to learn the fusion features $Z^O_{\mathcal{H}+\mathcal{L}}$ from HSI and LiDAR data sources as follows:

$$Z^O_{\mathcal{H}+\mathcal{L}} = g \odot Z^2_{\mathcal{H}} + (1-g) \odot Z^2_{\mathcal{L}}. \tag{15}$$

F. Loss Function

The proposed method requires an effective loss function to guide model training. By combining contrastive learning [45], [46], we propose a joint loss function that combines a classification loss and a semisupervised contrastive loss to improve model robustness during training step, which can be described as

$$\text{Loss} = \gamma \text{Loss}_{\text{clas}} + (1-\gamma)\text{Loss}_{\text{con}} \tag{16}$$

where γ is a balance factor to control the impact of $\text{Loss}_{\text{clas}}$ and Loss_{con} on the overall objective.

1) Classification Loss: All weights and biases in the proposed model need to be learned. The output classification results can be obtained by Softmax$(\cdot)$. Then, the cross-entropy loss is used to calculate the classification error among the network output and the labels of the original labeled samples, which is

$$\text{Loss}_O = -\sum_{g\in y_G}\sum_{f=1}^{C} Y_{\text{gf}} ln\, O \tag{17}$$

where y_G is the set of indices corresponding to the labeled samples, C denotes the number of classes, Y represents the label matrix, and Loss_O is the classification loss. Similar to [31], the network parameters are learned by using full-batch gradient descent.

In the proposed MFGCN-CME, the network output is composed of the HSI output $O_{\mathcal{H}}$, the LiDAR output $O_{\mathcal{L}}$, and the fusion output $O_{\mathcal{H}+\mathcal{L}}$, which are entered into (17) to obtain the three classification losses $\text{Loss}_{O_{\mathcal{H}}}$, $\text{Loss}_{O_{\mathcal{L}}}$, and $\text{Loss}_{O_{\mathcal{H}+\mathcal{L}}}$. The total classification loss $\text{Loss}_{\text{clas}}$ can be expressed as

$$\text{Loss}_{\text{clas}} = \text{Loss}_{O_{\mathcal{H}}} + \text{Loss}_{O_{\mathcal{L}}} + \text{Loss}_{O_{\mathcal{H}+\mathcal{L}}}. \tag{18}$$

2) Semisupervised Contrastive Loss: Contrastive learning extracts feature representations by minimizing the distances between features of similar samples and maximizing the distances between those of dissimilar samples, which can address the issue of small numbers of training samples in the field of HSI classification [47], [48], [49]. However, existing multisource RS data classification methods often only consider semantic matching on corresponding positions. In this article, a semisupervised contrastive loss is designed to explore the rich contrastive relations of all of the samples (i.e., large amounts of labeled and unlabeled samples) between HSI and LiDAR based on the topological structure information and label information between HSI and LiDAR data to learn discriminative features

$$\text{Loss}_{\text{con}} = \text{Loss}_{\text{gc}} + \text{Loss}_{\text{lc}} \tag{19}$$

where Loss_{gc} and Loss_{lc} are the graph contrastive loss and label contrastive loss, respectively. Thus, the proposed semisupervised contrastive loss is used to solve the issue (4) in Introduction.

Considering that the characteristics of the same object among different modalities are similar, we propose graph contrastive loss Loss_{gc} to construct extra positive pair based on the topological information among different modalities and minimize the cosine similarity between them. Specifically, the samples in HSI as the anchor and its own neighboring samples in LiDAR are considered as a positive sample pair, while the nodes embedding and the other nodes are regarded as the negative sample pair. sim$(\cdot)$ is used to pull closer the representations of the positive sample pair and push the representations of the negative sample pair, which can be defined by

$$\mathcal{L}_{\text{gc}}\left(z^O_{\mathcal{H}_i}\right) = -\log \frac{\sum_{k=1} A_{\mathcal{N}_{ik}} \exp\left(\text{sim}\left(z^O_{\mathcal{H}_i}, z^O_{\mathcal{L}_k}\right)/\tau\right)}{\sum_{j=1}^{N_C} \exp\left(\text{sim}\left(z^O_{\mathcal{H}_i}, z^O_{\mathcal{L}_j}\right)/\tau\right)} \tag{20}$$

$$\mathcal{L}_{\text{gc}}\left(z^O_{\mathcal{L}_i}\right) = -\log \frac{\sum_{k=1} A_{\mathcal{N}_{ik}} \exp\left(\text{sim}\left(z^O_{\mathcal{L}_i}, z^O_{\mathcal{H}_k}\right)/\tau\right)}{\sum_{j=1}^{N_C} \exp\left(\text{sim}\left(z^O_{\mathcal{L}_i}, z^O_{\mathcal{H}_j}\right)/\tau\right)} \tag{21}$$

$$\text{Loss}_{\text{gc}} = \frac{1}{2N_C}\sum_{i=1}^{N_C}\left(\mathcal{L}_{\text{gc}}\left(z^O_{\mathcal{H}_i}\right) + \mathcal{L}_{\text{gc}}\left(z^O_{\mathcal{L}_i}\right)\right) \tag{22}$$

where $\mathcal{L}_{\mathrm{gc}}(z^O_{\mathcal{H}_i})$ and $\mathcal{L}_{\mathrm{gc}}(z^O_{\mathcal{L}_i})$ are the two symmetric losses, which represent graph contrastive on different data sources. $A_{\mathcal{N}}$ connects each positive sample pair with edge in Section II-D.

Further, in the proposed semisupervised model, a training dataset includes both labeled and unlabeled samples. The pseudo-labels of unlabeled samples can be obtained with predicted labels in the training phase. Thus, a large number of unlabeled samples are added to the training set for promoting the performance of MFGCN-CME. A new contrastive loss uses the label information between HSI and LiDAR to construct positive and negative sample pairs. In particular, HSI and LiDAR samples with the same label are treated as positive sample pairs, while HSI and LiDAR samples with different labels are treated as negative sample pairs. It is defined as

$$\mathcal{L}_{\mathrm{lc}}\left(z^O_{\mathcal{H}_i}\right) = -\log \frac{\sum_{m=1} \exp\left(\mathrm{sim}\left(z^O_{\mathcal{H}_i}, z^O_{\mathcal{L}_m}\right)/\tau\right)}{\sum_{j=1}^{N_{\mathcal{C}}} \exp\left(\mathrm{sim}\left(z^O_{\mathcal{H}_i}, z^O_{\mathcal{L}_j}\right)/\tau\right)} \tag{23}$$

$$\mathcal{L}_{\mathrm{lc}}\left(z^O_{\mathcal{L}_i}\right) = -\log \frac{\sum_{m=1} \exp\left(\mathrm{sim}\left(z^O_{\mathcal{L}_i}, z^O_{\mathcal{H}_m}\right)/\tau\right)}{\sum_{j=1}^{N_{\mathcal{C}}} \exp\left(\mathrm{sim}\left(z^O_{\mathcal{L}_i}, z^O_{\mathcal{H}_j}\right)/\tau\right)} \tag{24}$$

$$\mathrm{Loss}_{\mathrm{lc}} = \frac{1}{2N_{\mathcal{C}}} \sum_{i=1}^{N_{\mathcal{C}}} \left(\mathcal{L}_{\mathrm{lc}}\left(z^O_{\mathcal{H}_i}\right) + \mathcal{L}_{\mathrm{lc}}\left(z^O_{\mathcal{L}_i}\right)\right) \tag{25}$$

where $z^O_{\mathcal{L}_m}$ is a positive sample for $z^O_{\mathcal{H}_i}$ and all other examples are negatives. Similarly, $z^O_{\mathcal{H}_m}$ is a positive example for $z_{\mathcal{L}_i}$ and all other examples are negatives. The samples in a positive pair have the same labels. $\mathcal{L}_{\mathrm{lc}}(z^O_{\mathcal{H}_i})$ and $\mathcal{L}_{\mathrm{lc}}(z^O_{\mathcal{L}_i})$ are the label contrastive loss on different modalities. The implementation details of our MFGCN-CME are shown in Algorithm 1.

III. Experimental Results

In this section, the performance of the proposed method is evaluated on three multisource datasets consisting of HSI and LiDAR data. Specifically, we first describe the characteristics of three multisource datasets, i.e., Trento,[1] Houston,[2] and MUUFL[3] datasets, eight baseline methods, and four quantitative evaluation indicators. We analyze the parameter settings of the proposed MFGCN-CME. Next, the quantitative and qualitative comparisons between our method and these baseline methods are used to verify the effectiveness of the proposed MFGCN-CME. Finally, a series of ablation studies is conducted to analyze the influence of each module in our method.

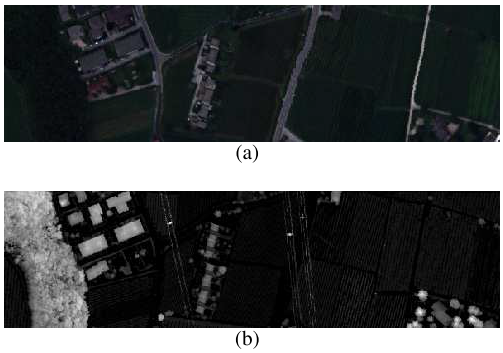
A. Experimental Datasets

For the comprehensive comparison, three multisource datasets are used to evaluate the performance of our model, whose detailed information is described as follows.

1) Trento: These data were captured by the AISA Eagle sensor for the HSI data and the Optech ALTM 3100EA sensor for the LiDAR data over a rural area in the south of Trento, Italy. The dataset consists of 166 × 600 pixels with the spatial resolution of 1 m per pixel, including six categories of land cover. The number of spectral bands for the hyperspectral data is 63, covering the wavelength from 0.42 to 0.99 μm. Fig. 4 shows the false-color map of the Trento HSI data and the gray image of the Trento LiDAR data. The numbers of labeled pixels and total pixels for different classes are listed in Table I.

2) Houston: These data were gathered with the Compact Airborne Spectrographic Imager (CASI) sensor over the University of Houston, Houston, TX, USA, and the surrounding

[1]https://download.csdn.net/download
[2]https://hyperspectral.ee.uh.edu
[3]https://github.com/GatorSense/MUUFLGulfport

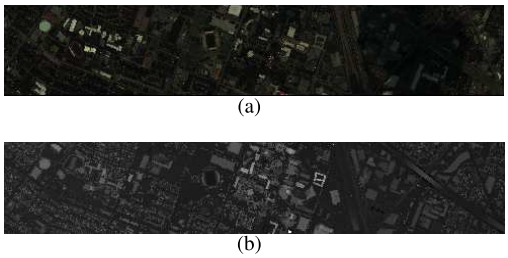


(a) (b)

Fig. 4. (a) False-color map (using bands 25, 15, and 2) of the Trento HSI data. (b) Grayscale representation of the Trento LiDAR data.

TABLE I
NUMBERS OF TRAINING AND TOTAL PIXELS OF ALL CLASSES IN THE TRENTO DATASET

NO.	Color	Class	Training	Total
1		Apple trees	129	4034
2		Buildings	125	2903
3		Ground	105	479
4		Wood	154	9123
5		Vineyard	184	10501
6		Roads	122	3374
Total			819	30414

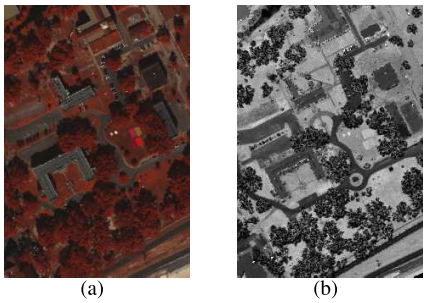


(a) (b)

Fig. 5. (a) False-color map (using bands 57, 27, and 17) of the Houston HSI data. (b) Grayscale representation of the Houston LiDAR data.

TABLE II
NUMBERS OF TRAINING AND TOTAL PIXELS OF ALL CLASSES IN THE HOUSTON DATASET

NO.	Color	Class	Training	Total
1		Healthy grass	198	1251
2		Stressed grass	190	1254
3		Synthetic grass	192	697
4		Tree	188	1244
5		Soil	186	1242
6		Water	182	325
7		Residential	196	1268
8		Commercial	191	1244
9		Road	193	1252
10		Highway	191	1227
11		Railway	181	1235
12		Parking lot 1	192	1233
13		Parking lot 2	184	469
14		Tennis court	181	428
15		Running track	187	660
Total			2832	15029



(a) (b)

Fig. 6. (a) False-color map (using bands 31, 16, and 6) of the MUUFL HSI data. (b) Grayscale representation of the MUUFL LiDAR data.

TABLE III
NUMBERS OF TRAINING AND TOTAL PIXELS OF ALL CLASSES IN THE MUUFL DATASET

NO.	Color	Class	Training	Total
1		Trees	150	23096
2		Mostly grass	150	4420
3		Mixed ground surface	150	6882
4		Dirt and sand	150	1826
5		Road	150	6687
6		Water	150	466
7		Building shadow	150	2233
8		Building	150	6240
9		Sidewalk	150	1385
10		Yellow curb	150	183
11		Cloth panels	150	269
Total			1650	53687

area on June 2012, which were introduced in the 2013 IEEE Geoscience and Remote Sensing Society (GRSS) Data Fusion Contest. Both HSI and LiDAR data have a spatial size of 349 × 1905 pixels with a spatial resolution of 2.5 m. For the hyperspectral data, there exist 144 spectral bands with the wavelength ranging from 0.38 to 1.05 μm. The Houston dataset contains 15 land cover classes, such as "Tree," "Water," and "Commercial," as shown in Fig. 5. In addition, Table II reports the numbers of training pixels and total pixels for each class.

3) MUUFL: The MUUFL dataset was obtained over the campus area of Gulf Park University of Southern Mississippi, Long Beach, MS, USA, by the Reflective Optics System Imaging Spectrometer sensor in November 2010. In the MUUFL dataset, the HSI data comprise 72 spectral bands ranging from 0.38 to 1.05 μm, while the LiDAR data contain two rasters with a 1.06 μm wavelength. The first and last eight spectral bands were removed due to excessive noise. This dataset consists of 325 × 220 pixels, totaling 11 different land-cover classes. Fig. 6 shows the false-color map of the MUUFL HSI data and the gray image of the MUUFL LiDAR data. Table III lists the distribution of training and total samples for the MUUFL dataset.

To verify the classification performance of the proposed method, some classical and state-of-the-art classification methods are selected as the comparison methods.

a) SVM [11]: SVM is the most classical classification method, which is performed on the well-known libsvm toolbox3 in our case.

b) CapsNet [22]: The 3-D capsule network first introduces the maximum correntropy criterion for addressing the noise and outlier problem in HSIs, in which the spatial–spectral information of HSI and the elevation information of LiDAR can be efficiently fused to extract the discriminative features for the classification.

c) SSCL3DNN [50]: The SSCL3DNN extends ConvLSTM to the 3-D version for better preserving the intrinsic structure of HSI data. In the model, the local 3-D cube is decomposed into a spectral sequence and then input to each memory cell band by band to the effectively fused spatial and spectral features.

d) FGCN [49]: The FGCN integrates multiscale spatial features, directional texture features, frequency variation features, and spectral features for accurate multisource joint classification.

e) A^3CLNN [51]: The model develops an effective three-level fusion strategy and a novel stepwise training strategy to fully integrate the spatial and spectral information contained in the HSI and LiDAR data, exploiting their complementarity.

f) SepDGC [52]: The SepDGC automatically learns a multistream network structure within a single-stream architecture for multisource RS data classification.

g) AM^3Net [53]: The AM^3Net uses the spectral–spatial feature learning module, the spectral–spatial mutual-guided module, and the adaptive multistage feature fusion module for realizing the fusion and classification of multisource RS data.

h) NNCNet [54]: The NNCNet integrates a nearest neighbor-based data augmentation scheme into the contrastive learning framework to capture intermodal semantic alignments more accurately and uses a bilinear attention module to exploit the second-order feature interactions between the HSI and LiDAR data.

Moreover, four evaluation indexes, i.e., per-class accuracy, overall accuracy (OA), average accuracy (AA), and kappa coefficient (κ), are adopted to evaluate and analyze the classification results of the proposed method and other related works, whose higher values indicate better classification performance.

B. Experimental Settings

All experiments are ran on a computer with an Intel[4] Xeon[4] CPU E5-1650 V4 with 3.6 GHz, 64 GB, and an NVIDIA GeForce RTX 2080 Ti graphical processing unit (GPU). For software system configuration, we adopt Windows 10 X64 as our operating system for all experiments. CUDA 9.0 and cuDNN 7.0.5, Tensorflow-GPU with 1.12, and Python 3.5.6 are the main programming environment. Especially, all methods involved in our experiments are completed in Anaconda 4.5.11.

We run each method ten times and report the average results. In the proposed model, the Adam optimization is selected to train the architecture for optimal performance. The training epoch number is 10 000. The learning rate is reduced by 0.1 every 500 epochs, starting with an initial learning rate of 0.001. The classification performance is also affected by two specific parameters in exception to the basic parameters mentioned above. One of the factor in MFGCN-CME is the ratio λ, which represents the balance

[4]Registered trademark.

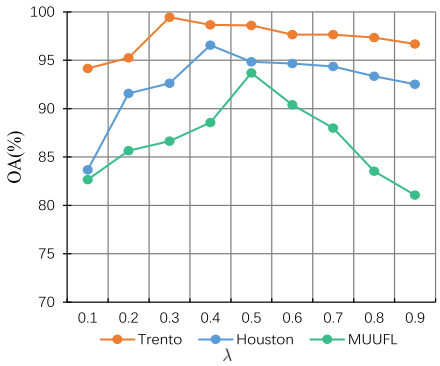


Fig. 7. Classification accuracy of the proposed method using different λ on three datasets.

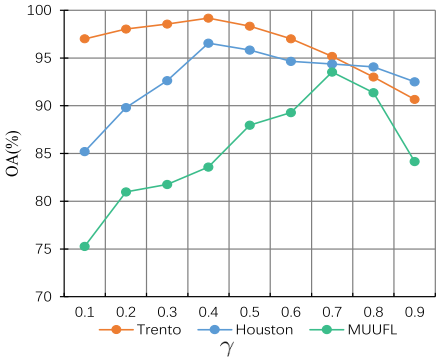


Fig. 8. Classification accuracy of the proposed method using different γ on three datasets.

between the low-frequency and high-frequency multiscale features. The parameter γ is an additional component that establishes the classification loss's relative importance in relation to the semisupervised contrastive loss. The details of parameter analysis are presented as follows.

First, we study the effects of the parameter λ for the proposed method. Fig. 7 exhibits the OA results on three datasets when λ in the range $\{0.1, 0.2, \ldots, 0.8, 0.9\}$, where the abscissa represents different values of λ, and the ordinate represents the classification accuracy. We observe that lower values result in the loss of low-frequency multiscale features, which seriously affects the classification performance. In addition, a high λ leads to the loss of high-frequency multiscale information, making it difficult to classify object boundaries. As shown in Fig. 7, the optimal results can be obtained in the Trento, Houston, and MUUFL datasets when λ is set to 0.3, 0.4, and 0.5, respectively.

In addition, the proposed method uses the hyperparameter γ to control the magnitude of contrastive learning. To investigate its influence, we conduct experiments on these three datasets with γ values from 0.1 to 0.9 at 0.1 intervals. As shown in Fig. 8, our model performs best on the Trento and Houston datasets when the γ value is 0.4, while for the MUUFL dataset, the best performance can be obtained with γ=0.7. Moreover, we observe that the OA results gradually improve with γ values ranging from 0.1 to 0.3 in the Trento and Houston datasets and from 0.1 to 0.6 in the MUUFL dataset, because a suitable value of γ allows the classification and semisupervised contrastive learning constraint in (16) to learn the more discriminative intermediate representations.

C. Experimental Results

To evaluate the performance of the proposed model, Tables IV–VI present detailed classification results on three benchmark datasets. The best result for each column is bolded to draw attention to it. We can see that the proposed method outperforms eight baseline methods on the Trento, Houston, and MUUFL datasets. This appearance can be summarized by the following factors. First, different from the compared methods, the designed feature extraction module simultaneously learns the low-frequency features and high-frequency features. Second, the proposed BGELM enhances the spatial representations of HSI data by LiDAR features and the spectral information of LiDAR data by HSI features. In addition, not simply fusing HSI and LiDAR data, a novel combination strategy with the gate attention mechanism can adaptively explore the complementary information of HSI and LiDAR data. To further realize the potential of our model, we introduce semisupervised contrastive loss across different modalities to enrich the self-supervision signals during the optimization.

More specifically, from Tables IV–VI, we can see that, for the Trento, Houston, and MUUFL datasets, the proposed model yields highly competitive classification accuracies of 99.40%, 95.60%, and 93.17%, respectively, with a gain over 6% with respect to that achieved by the SVM. And it can also be found that the classification performance of deep learning methods is superior to SVM in terms of OA, AA, and κ in most cases. The above experimental results show that for traditional models, such as SVM, the way of converting the HSI and LiDAR data into vectors leads to the loss of the spatial and geometric structure information. In addition, the OA results of the proposed model on three datasets have significantly improved compared with CapsNet, which is due to the fact that CapsNet may not fully learn the spectral and scale information contained in different classes. From Tables IV–VI, the proposed model can improve the model accuracy by 4.55%, 20.56%, and 6.16% on three benchmark datasets compared with FGCN, respectively. FGCN is theoretically robust to changes in semantic information, such as shift, rotation, sensor noises, or distortions. However, it fails to focus on the latent representation extraction by using the relationship of the paired multimodal data (i.e., HSI and LiDAR). In contrast to these above comparison methods, SSCL3DNN and A^{3}CLNN are the ConvLSTM-based models for joint learning of spatial–spectral features by modeling long-term dependencies in the spectral field. In particular, compared with SSCL3DNN, the dual-channel A^{3}CLNN model can improve the classification accuracy in almost all cases, respectively, benefiting from the developed three-level fusion and stepwise training strategies. Our method represents 5.09%, 12.41%, and 4.96% relative improvements over SepDGC on the Trento, Houston, and MUUFL datasets, respectively. Moreover, on three datasets, our approach outperforms AM^{3}Net by about 1.05%, 2.58%, and 14.91% on OA, respectively. This improvement in the performance of MFGCN-CME is due to the multifrequency feature extraction method. In addition, the classification accuracies of our method and NNCNet are higher than other

TABLE IV
CLASSIFICATION RESULTS FOR THE TRENTO DATASET

Class	SVM	CapsNet	SSCL3DNN	FGCN	A^{3}CLNN	SepDGC	AM^{3}Net	NNCNet	MFGCN-CME
1	97.69	97.15	98.32	93.04	98.92	94.57	99.21	99.94	**100.00**
2	87.36	99.07	96.88	98.94	99.14	96.00	99.13	96.58	**99.28**
3	87.06	97.29	82.19	81.28	98.12	76.82	97.86	**100.00**	**100.00**
4	99.80	**100.00**	99.81	92.55	**100.00**	95.72	99.98	99.98	**100.00**
5	93.36	94.62	96.74	99.41	**99.95**	96.79	99.25	99.79	99.17
6	69.34	91.71	85.34	86.41	90.57	82.82	88.82	89.54	**97.70**
OA	92.69	96.75	96.46	94.85	98.73	94.31	98.35	98.50	**99.40**
AA	89.10	96.64	93.21	91.94	97.78	90.45	97.37	97.64	**99.36**
κ	90.22	95.69	95.30	93.14	98.31	92.46	97.80	98.00	**99.20**

TABLE V
CLASSIFICATION RESULTS FOR THE HOUSTON DATASET

Class	SVM	CapsNet	SSCL3DNN	FGCN	A^{3}CLNN	SepDGC	AM^{3}Net	NNCNet	MFGCN-CME
1	82.53	81.39	82.05	80.81	81.73	80.73	92.21	80.91	**94.85**
2	84.77	83.08	80.98	68.98	84.43	86.36	90.69	**99.90**	94.68
3	86.93	97.43	89.44	62.30	91.49	96.97	**98.61**	97.02	93.86
4	95.83	88.64	90.85	87.97	96.72	82.95	88.25	98.48	**100.00**
5	97.54	**100.00**	99.78	97.91	99.97	**100.00**	94.59	**100.00**	90.81
6	88.81	95.10	87.18	85.91	97.90	95.37	95.74	98.60	**100.00**
7	81.16	91.23	91.51	88.05	87.06	75.55	78.99	84.04	**97.38**
8	44.92	92.40	93.32	79.96	96.93	76.28	88.68	92.21	**99.74**
9	86.40	80.64	78.88	73.08	87.88	74.52	98.29	93.48	**99.86**
10	59.75	65.54	55.60	48.06	70.82	71.80	97.58	**98.35**	89.44
11	71.82	88.99	90.83	54.36	**98.13**	80.97	97.72	93.07	92.33
12	92.41	87.42	91.80	80.01	94.65	83.06	94.14	**98.65**	94.97
13	85.96	62.46	85.96	71.57	96.02	81.02	**100.00**	85.61	97.47
14	83.00	95.95	78.41	83.80	97.30	92.05	**100.00**	97.16	**100.00**
15	74.21	96.41	94.86	61.68	96.05	95.74	99.36	**100.00**	96.82
OA	80.15	86.61	86.01	75.04	90.55	83.19	93.02	94.18	**95.60**
AA	81.07	87.11	86.10	74.98	91.81	84.89	94.32	94.50	**96.15**
κ	78.58	85.50	84.84	72.88	89.75	81.87	92.42	93.68	**95.23**

TABLE VI
CLASSIFICATION RESULTS FOR THE MUUFL DATASET

Class	SVM	CapsNet	SSCL3DNN	FGCN	A^{3}CLNN	SepDGC	AM^{3}Net	NNCNet	MFGCN-CME
1	85.78	87.06	89.76	90.05	89.93	93.37	91.16	91.34	**96.56**
2	76.26	47.52	83.22	84.41	83.25	76.39	80.33	81.79	**89.57**
3	74.40	80.70	75.69	78.72	83.89	79.30	63.27	81.77	**91.31**
4	90.87	95.34	96.06	94.26	93.19	88.82	79.16	**96.65**	94.89
5	81.56	90.14	88.15	80.41	87.22	92.24	62.92	**93.80**	88.45
6	99.68	99.05	**100.00**	99.18	**100.00**	80.20	99.68	**100.00**	99.05
7	92.31	63.94	91.83	**93.57**	89.82	80.41	77.42	83.82	90.85
8	82.69	96.10	92.46	89.46	**96.73**	96.13	76.67	93.79	94.64
9	76.68	69.31	73.19	**81.01**	77.65	71.28	35.89	57.81	69.83
10	**96.96**	69.69	81.81	91.56	84.84	39.64	56.25	93.93	75.75
11	91.59	86.55	91.59	98.22	91.59	88.70	**100.00**	98.31	94.95
OA	82.97	83.53	87.49	87.01	88.95	88.21	79.26	89.08	**93.17**
AA	86.25	80.49	87.62	89.17	88.92	80.59	74.80	88.46	**89.62**
κ	77.92	78.44	83.66	83.13	85.56	84.79	72.15	85.63	**90.91**

seven baseline methods because these two methods introduce contrastive learning to solve the problem of limited training samples. Finally, as shown in Tables IV–VI, the OA results of MFGCN-CME on three datasets are 0.9%, 1.42%, and 4.09% higher than the suboptimal results, respectively. The reason is that our method can extract frequency-sdomain features for HSI and LiDAR, and achieve mutual enhancement and effective fusion of the two types of features.

For visual comparison, the classification maps on the Trento dataset obtained by all methods are displayed in Fig. 9, respectively. We observe that the classification maps of the proposed method show more consistent agreement with ground truth than other methods. In particular, from Fig. 9, it is clear that SVM fails to maintain spatial continuity in the classification maps because of the lack of spatial information. In addition, as expected, other deep learning methods generally obtain smooth classification maps. As shown in Fig. 9, the proposed method can suppress the misclassification and noisy scattered points in classification by performing convolution operations to learn the spatial neighborhood information. We can find that our method obtains better boundaries between different objects from the enlarged partial view, especially for classes 3 and 6. The main reason is that the proposed method can adopt graph wavelet convolutional operation to achieve satisfactory results on tiny objects and boundary regions.

D. Ablation Study

Each module in the proposed model plays a pivotal role. In this section, we conduct a series of ablation studies on

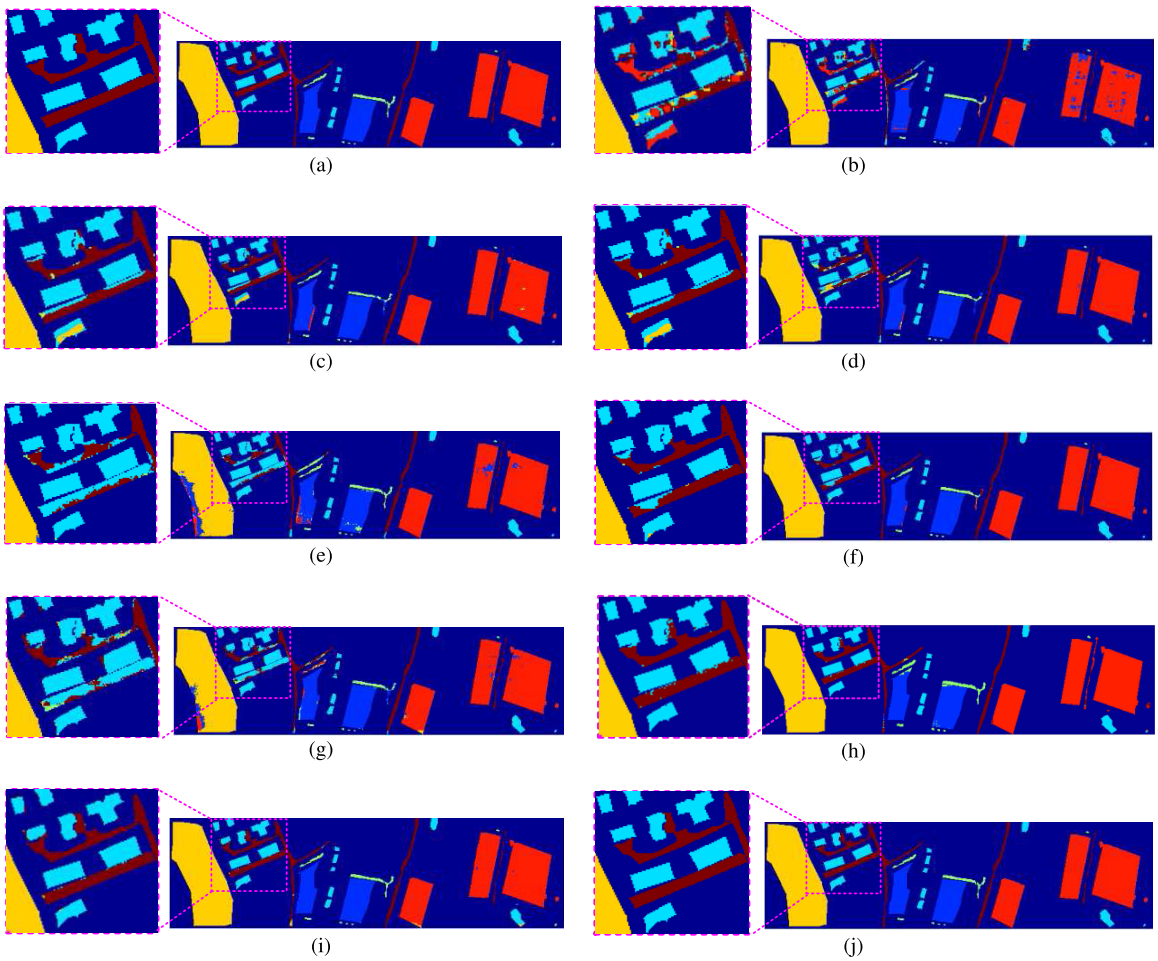


Fig. 9. Classification maps obtained by different methods for the Trento dataset. (a) Ground-truth map (OA; AA; κ). (b) SVM (92.69; 89.10; 90.22). (c) CapsNet (96.75; 96.64; 95.69). (d) SSCL3DNN (96.46; 93.21; 95.30). (e) FGCN (91.27; 89.23; 88.39). (f) A^{3}CLNN (98.73; 97.78; 98.31). (g) SepDGC (94.31; 90.45; 92.46). (h) AM^{3}Net (98.35; 97.37; 97.80). (i) NNCNet (98.50; 97.64; 98.00). (j) **MFGCN-CME (99.40; 99.36; 99.20)**.

TABLE VII
ABLATION RESULTS WITH/WITHOUT THE AMGFLM

Method	Trento			Houston			MUUFL		
	OA	AA	κ	OA	AA	κ	OA	AA	κ
GCN	90.20	86.37	87.05	50.48	54.67	46.85	79.57	80.94	73.85
MGCN	96.89	92.19	95.83	51.53	56.29	47.90	87.47	83.46	83.40
GWCN	85.40	86.59	81.18	72.39	74.34	70.27	71.29	75.75	64.19
MGWCN	88.82	93.92	85.61	82.89	86.16	81.42	79.28	81.54	73.47
MFGCN-CME	**99.40**	**99.36**	**99.20**	**95.60**	**96.15**	**95.23**	**93.17**	**89.62**	**90.91**

TABLE VIII
ABLATION RESULTS WITH/WITHOUT THE BGELM

Method	Trento			Houston			MUUFL		
	OA	AA	κ	OA	AA	κ	OA	AA	κ
w/o BGELM	99.28	93.05	99.04	91.98	93.56	91.30	84.96	83.72	80.59
MFGCN-CME	**99.40**	**99.36**	**99.20**	**95.60**	**96.15**	**95.23**	**93.17**	**89.62**	**90.91**

three datasets to discuss the rationality of these proposed components, i.e., the AMGFLM, the BGELM, the gated fusion module, and two contrastive losses, in improving the performance.

1) Influence of the AMGFLM: One difference between our method with standard GCN-based HSI classification methods is that the proposed method can learn the effective fusion of low-frequency and high-frequency features. To quantitatively investigate the effectiveness of the multifrequency feature learning strategy, the AMGFLM is replaced with GCN, multiorder graph convolutional network (MGCN), graph wavelet convolutional network (GWCN), and multiscale graph wavelet convolutional network (MGWCN) to build four methods (i.e., GCN, MGCN, GWCN, and MGWCN). The experimental results of these four variants above and the MFGCN-CME are reported in Table VII. We can find from Table VII that compared with GCN, MGCN achieves the average gains of 6.69%, 1.05%, and 7.9%, respectively, because the extracted low-frequency multiscale features contain the high-order neighbor information. MGWCN outperforms GWCN by an increase of at least 3%, which suggests the usefulness of incorporating the multiscale information into the graph feature extraction. In addition, compared with the four ablation models, the proposed MFGCN-CME method achieves decent performance improvements for multisource RS classification task. The results in Table VII demonstrate that simultaneously considering frequency features (i.e., low-frequency features and high-frequency features) boosts the model performance.

2) Influence of the BGELM: To examine the effect of the BGELM that aims for the mutual enhancement of HSI and LiDAR data, we compare the classification accuracies of the proposed method and a variant of MFGCN-CME without the BGELM module (i.e., w/o BGELM) on the Trento,

TABLE IX
ABLATION RESULTS WITH/WITHOUT THE GATED FUSION MODULE

Method	Trento			Houston			MUUFL		
	OA	AA	κ	OA	AA	κ	OA	AA	κ
MFGCN-CMEAF	88.06	88.18	84.58	94.84	95.60	94.43	83.53	80.34	78.11
MFGCN-CMEEF	90.77	91.91	87.96	88.25	89.10	87.25	83.79	80.90	77.65
MFGCN-CMECF	98.97	91.26	98.63	94.30	95.04	93.81	80.54	79.67	74.76
MFGCN-CME	**99.40**	**99.36**	**99.20**	**95.60**	**96.15**	**95.23**	**93.17**	**89.62**	**90.91**

TABLE X
ABLATION RESULTS WITH/WITHOUT THE GRAPH CONTRASTIVE LOSS AND LABEL CONTRASTIVE LOSS

Method	Trento			Houston			MUUFL		
	OA	AA	κ	OA	AA	κ	OA	AA	κ
w/o GCL	86.78	75.59	81.57	87.93	90.38	86.89	85.07	85.25	80.72
w/o LCL	94.17	90.77	92.19	92.83	91.13	92.22	89.28	85.78	85.82
w/o GCL+LCL	83.58	79.85	78.47	85.15	83.50	84.02	83.59	82.02	78.56
MFGCN-CME	**99.40**	**99.36**	**99.20**	**95.60**	**96.15**	**95.23**	**93.17**	**89.62**	**90.91**

Houston, and MUUFL datasets. As listed in Table VIII, the proposed BGELM brings performance gains of 0.12%, 3.62%, and 8.21% for the three datasets in terms of OA results, respectively. One possible explanation is that the BGELM can implement the mutual enhancement of LiDAR and HSI features.

3) Influence of the Gated Fusion Module: The proposed MFGCN-CME method adopts the gated fusion module to adaptively fuse the spatial–spectral features of HSI and the spatial features of LiDAR. We carry out the experiments for analyzing its influence on the classification performance, whose results are reported in Table IX. Here, MFGCN-CMEAF, MFGCN-CMEEF, and MFGCN-CMECF are three ablation models by using the additive fusion, elementwise multiplicative fusion, and concatenation fusion instead of the gated fusion module. The OA results show that the proposed method applying the gated fusion module achieves the highest accuracy, which yields at least 0.43%, 0.76%, and 9.38% gains for the three datasets compared with these ablation methods, respectively. It demonstrates the effectiveness of the gated fusion module in exploring the complementary of HSI and LiDAR data.

4) Influence of Graph Contrastive Loss and Label Contrastive Loss: In our method, graph contrastive loss and label contrastive loss can optimize the parameters of the model by maximizing the consistency between HSI and LiDAR data. To study the effectiveness of these two contrastive losses, we design three variants: w/o GCL refers to the proposed MFGCN-CME method removing graph contrastive loss; w/o LCL denotes the proposed MFGCN-CME method excluding label contrastive loss; and w/o GCL+LCL is the proposed MFGCN-CME method simultaneously removing graph contrastive loss and label contrastive loss. As observed from Table X, compared to w/o GCL, the proposed MFGCN-CME increases the OA results by 12.62%, 7.67%, and 7.67% for the Trento, Houston, and MUUFL datasets, respectively. In addition, the proposed MFGCN-CME improves the OA values by 5.23%, 2.77%, and 3.89% for the three datasets compared with w/o LCL, respectively. In addition, the proposed MFGCN-CME yields 15.82%, 10.45%, and 9.58% gains in OA for the Trento, Houston, and MUUFL datasets compared with w/o GCL+LCL, respectively. The OA, AA, and κ results of MFGCN-CME significantly increase compared with three variants, which indicate the positive impact of the graph contrastive loss and label contrastive loss in improving classification performance.

IV. CONCLUSION

This article proposes an MFGCN-CME for the classification of multisource RS data. In the proposed method, the AMGFLM is designed as the feature extractor to fully capture the low- and high-frequency multiscale features of HSI and LiDAR data in parallel and further adaptively aggregate them. Furthermore, the BGELM is proposed to enhance the spatial features of HSI data by LiDAR features and the spectral features of LiDAR data by HSI features, aiming at obtaining more comprehensive and discriminative feature representation. In addition, the gated fusion module is introduced to explore the complementarity of two different data sources in an adaptive way. Finally, a joint loss function combines a classification loss and a semisupervised contrastive loss, which leverages the geometric relationships and label information between HSI and LiDAR to improve the model robustness. Extensive experiments show that our method is more efficient than some baseline methods and produces state-of-the-art classification results. However, the proposed method, as the semisupervised method, requires a certain labeled samples for the joint classification of HSI and LiDAR data. In the future, extending the proposed method for unsupervised learning requires more research. In addition, we will extend the proposed method to other multisource RS data classification tasks, such as HSI and MSI, and HSI and SAR.

ACKNOWLEDGMENT

The authors would like to thank the Editor-in-Chief, Associate Editor, and Anonymous Reviewers, for their valuable comments and suggestions.

REFERENCES

[1] C. Lu et al., "A novel hyperspectral remote sensing technique with hour-hectometer level horizontal distribution of trace gases: To accurately identify emission sources," *J. Remote Sens.*, vol. 3, p. 0098, Nov. 2023.

[2] M. Ding, X. Fu, T.-Z. Huang, J. Wang, and X.-L. Zhao, "Hyperspectral super-resolution via interpretable block-term tensor modeling," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 641–656, Apr. 2021.

[3] Y. Gao et al., "Hyperspectral and multispectral classification for coastal wetland using depthwise feature interaction network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5512615.

[4] P. Kumar Yadav et al., "Detecting volunteer cotton plants in a corn field with deep learning on UAV remote-sensing imagery," *Comput. Electron. Agricult.*, vol. 204, Jan. 2023, Art. no. 107551.

[5] J. Chen et al., "Urban built environment assessment based on scene understanding of high-resolution remote sensing imagery," *Remote Sens.*, vol. 15, no. 5, p. 1436, Mar. 2023.

[6] K. Y. Ma and C.-I. Chang, "Kernel-based constrained energy minimization for hyperspectral mixed pixel classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5510723.

[7] W.-S. Hu, H.-C. Li, Y.-J. Deng, X. Sun, Q. Du, and A. Plaza, "Lightweight tensor attention-driven ConvLSTM neural network for hyperspectral image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 734–745, Apr. 2021.

[8] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[9] A. Ghiyamat and H. Z. M. Shafri, "A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment," *Int. J. Remote Sens.*, vol. 31, no. 7, pp. 1837–1856, Apr. 2010.

[10] Y. Tan, L. Lu, L. Bruzzone, R. Guan, Z. Chang, and C. Yang, "Hyperspectral band selection for lithologic discrimination and geological mapping," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 471–486, 2020.

[11] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.

[12] Z. Liu, B. Tang, X. He, Q. Qiu, and F. Liu, "Class-specific random forest with cross-correlation constraints for spectral–spatial hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 2, pp. 257–261, Feb. 2017.

[13] J. Peng and Q. Du, "Robust joint sparse representation based on maximum correntropy criterion for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7152–7164, Dec. 2017.

[14] X. Wang, "Kronecker factorization-based multinomial logistic regression for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[15] Y. Su, L. Gao, M. Jiang, A. Plaza, X. Sun, and B. Zhang, "NSCKL: Normalized spectral clustering with kernel-based learning for semisupervised hyperspectral image classification," *IEEE Trans. Cybern.*, vol. 53, no. 10, pp. 6649–6662, 2023.

[16] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.

[17] M. Li, J. Li, Y. Liu, and F. Liu, "Detail injection-based convolutional auto-encoder for pansharpening," *J. Remote Sens.*, vol. 2022, p. 0004, Dec. 2022.

[18] H. Tulapurkar, B. Banerjee, and K. M. Buddhiraju, "Multi-head attention with CNN and wavelet for classification of hyperspectral image," *Neural Comput. Appl.*, vol. 35, no. 10, pp. 7595–7609, Apr. 2023.

[19] S. Mei, R. Jiang, M. Ma, and C. Song, "Rotation-invariant feature learning via convolutional neural network with cyclic polar coordinates convolutional layer," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5600713.

[20] Q. Liu, J. Peng, G. Zhang, W. Sun, and Q. Du, "Deep contrastive learning network for small-sample hyperspectral image classification," *J. Remote Sens.*, vol. 3, p. 0025, Jan. 2023.

[21] C. Chen, Y. Ma, and G. Ren, "Hyperspectral classification using deep belief networks based on conjugate gradient update and pixel-centric spectral block features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4060–4069, 2020.

[22] H.-C. Li, W.-Y. Wang, L. Pan, W. Li, Q. Du, and R. Tao, "Robust capsule network based on maximum correntropy criterion for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 738–751, 2020.

[23] C. Shi and C.-M. Pun, "Multiscale superpixel-based hyperspectral image classification using recurrent neural networks with stacked autoencoders," *IEEE Trans. Multimedia*, vol. 22, no. 2, pp. 487–501, Feb. 2020.

[24] S. Mei, X. Li, X. Liu, H. Cai, and Q. Du, "Hyperspectral image classification using attention-based bidirectional long short-term memory network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5509612.

[25] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Apr. 2017, pp. 1–14.

[26] L. Mou, X. Lu, X. Li, and X. X. Zhu, "Nonlocal graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8246–8257, Dec. 2020.

[27] 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.

[28] X. Zhang, S. Chen, P. Zhu, X. Tang, J. Feng, and L. Jiao, "Spatial pooling graph convolutional network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5521315.

[29] Y. Cheng, Y. Chen, Y. Kong, and X. Wang, "Soft instance-level domain adaptation with virtual classifier for unsupervised hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5509013.

[30] A. Qin, Z. Shang, J. Tian, Y. Wang, T. Zhang, and Y. Y. Tang, "Spectral–Spatial graph convolutional networks for semisupervised hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 241–245, Feb. 2019.

[31] S. Wan, C. Gong, P. Zhong, B. Du, L. Zhang, and J. Yang, "Multiscale dynamic graph convolutional network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3162–3177, May 2020.

[32] 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.

[33] Y. Su et al., "ACGT-Net: Adaptive cuckoo refinement-based graph transfer network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5521314.

[34] B. Rasti, P. Ghamisi, and R. Gloaguen, "Hyperspectral and LiDAR fusion using extinction profiles and total variation component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3997–4007, Jul. 2017.

[35] F. Guo, Z. Li, Q. Meng, L. Wang, and J. Zhang, "Dual graph convolution joint dense networks for hyperspectral and LiDAR data classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jun. 2022, pp. 1141–1144.

[36] L. Wang and X. Wang, "Dual-coupled CNN-GCN-based classification for hyperspectral and LiDAR data," *Sensors*, vol. 22, no. 15, p. 5735, Jul. 2022.

[37] D. Xiu, Z. Pan, Y. Wu, and Y. Hu, "MAGE: Multisource attention network with discriminative graph and informative entities for classification of hyperspectral and LiDAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5539714.

[38] W. Li, J. Wang, Y. Gao, M. Zhang, R. Tao, and B. Zhang, "Graph-feature-enhanced selective assignment network for hyperspectral and multispectral data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5526914.

[39] X. Du, X. Zheng, X. Lu, and A. A. Doudkin, "Multisource remote sensing data classification with graph fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10062–10072, Dec. 2021.

[40] M. Zhang, W. Li, Y. Zhang, R. Tao, and Q. Du, "Hyperspectral and LiDAR data classification based on structural optimization transmission," *IEEE Trans. Cybern.*, vol. 53, no. 5, pp. 3153–3164, May 2023.

[41] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.

[42] Y. Min, F. Wenkel, and G. Wolf, "Scattering GCN: Overcoming oversmoothness in graph convolutional networks," in *Proc. Annu. Conf. Neural Inf. Process. Syst. (NIPS)*, Dec. 2020, pp. 1–11.

[43] B. Xu, H. Shen, Q. Cao, Y. Qiu, and X. Cheng, "Graph wavelet neural network," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, May 2019, pp. 1–13.

[44] J. Li, M. D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 996–1010, Apr. 2013.

[45] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.

[46] K. Hassani and A. H. Khasahmadi, "Contrastive multi-view representation learning on graphs," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2020, pp. 4116–4126.

[47] R. Hang, X. Qian, and Q. Liu, "Cross-modality contrastive learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5532812.

[48] S. Jia, X. Zhou, S. Jiang, and R. He, "Collaborative contrastive learning for hyperspectral and LiDAR classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5507714.

[49] X. Zhao, R. Tao, W. Li, W. Philips, and W. Liao, "Fractional Gabor convolutional network for multisource remote sensing data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–18, 2022, Art. no. 5503818.

[50] W.-S. Hu, H.-C. Li, L. Pan, W. Li, R. Tao, and Q. Du, "Spatial–Spectral feature extraction via deep ConvLSTM neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4237–4250, Jun. 2020.

[51] H.-C. Li, W.-S. Hu, W. Li, J. Li, Q. Du, and A. Plaza, "A^{3}CLNN: Spatial, spectral and multiscale attention ConvLSTM neural network for multisource remote sensing data classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 2, pp. 747–761, Feb. 2022.

[52] Y. Yang, D. Zhu, T. Qu, Q. Wang, F. Ren, and C. Cheng, "Single-stream CNN with learnable architecture for multisource remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5409218.

[53] J. Wang, J. Li, Y. Shi, J. Lai, and X. Tan, "AM^{3}Net: Adaptive mutual-learning-based multimodal data fusion network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 8, pp. 5411–5426, Aug. 2022.

[54] M. Wang, F. Gao, J. Dong, H.-C. Li, and Q. Du, "Nearest neighbor-based contrastive learning for hyperspectral and LiDAR data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5501816.

Jin-Yu Yang received the B.Sc. and M.Sc. degrees in software engineering from Xihua University, Chengdu, China, in 2016 and 2019, respectively. She is currently pursuing the Ph.D. degree in information and communication engineering with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu.

Her research interests include deep learning, graph neural network, and remote sensing image classification.

Heng-Chao Li (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from Southwest Jiaotong University, Chengdu, China, in 2001 and 2004, respectively, and the Ph.D. degree from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2008, all in information and communication engineering.

From November 2013 to October 2014, he was a Visiting Scholar working with Prof. William J. Emery with the University of Colorado Boulder, Boulder, CO, USA. He is currently a Professor with the School of Information Science and Technology, Southwest Jiaotong University. His research interests include the statistical analysis of synthetic aperture radar (SAR) images, remote-sensing image processing, and pattern recognition.

Dr. Li was a recipient of the Sichuan Natural Science Award, the CSIG Natural Science Award, the CSGPC Science and Technology Progress Award, and the 2018 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society for his service to IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He is an Associate Editor of IEEE JSTARS and *Journal of Southwest Jiaotong University*. He is an Editorial Board Member of *Journal of Remote Sensing* and *Journal of Radars*.

Jing-Hua Yang (Member, IEEE) received the Ph.D. degree from the Macau University of Science and Technology, Macau, China, in 2023.

She is currently an Assistant Professor with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China. Her research interests include data mining, image processing, tensor analysis, and federated learning.

Lei Pan received the B.Sc. degree in communication engineering from the Shandong University of Science and Technology, Qingdao, China, in 2010, and the M.Sc. degree in communication and information system and the Ph.D. degree in signal and information processing from the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China, in 2013 and 2019, respectively.

He is currently working with the Southwest Institute of Electronic Technology, Chengdu. His research interests include remote sensing image processing, natural language processing, and artificial intelligence.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland, Baltimore, MD, USA, in 2000.

She is currently a Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Antonio Plaza (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in computer engineering from the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura, Cáceres, Spain, in 1999 and 2002, respectively.

He is currently the Head of the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura. His research interests include hyperspectral data processing and parallel computing of remote sensing data.

Dr. Plaza has served as the Editor-in-Chief of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2013 to 2017 and guest-edited ten special issues on hyperspectral remote sensing for different journals.