





Hyperspectral Classification via Global-Local Hierarchical Weighting Fusion Network

Bing Tu , *Member, IEEE*, Wangquan He, *Student Member, IEEE*, Wei He , *Member, IEEE*, Xianfeng Ou , *Member, IEEE*, and Antonio Plaza , *Fellow, IEEE*

Abstract—The fusion of spectral–spatial features based on deep learning has become the focus of research in hyperspectral image (HSI) classification. However, previous deep frameworks based on spectral–spatial fusion usually performed feature aggregation only at the branch ends. Furthermore, only first-order statistical features are considered in the fusion process, which is not conducive to improving the discrimination of spectral–spatial features. This article proposes a global–local hierarchical weighted fusion end-to-end classification architecture. The architecture includes two subnetworks for spectral classification and spatial classification. For the spectral subnetwork, two band-grouping strategies are designed, and bidirectional long short-term memory is used to capture spectral context information from global to local perspectives. For the spatial subnetwork, a pooling strategy based on local attention is combined to construct a global–local pooling fusion module to enhance the discriminability of spatial features learned by a convolutional neural network. For the fusion stage, a hierarchical weighting fusion mechanism is developed to obtain the nonlinear relationship between both spectral and spatial features. The experimental results on four real HSI datasets and a GF-5 satellite dataset demonstrate that the method proposed is more competitive in terms of accuracy and generalization.

Index Terms—Band grouping, deep learning (DL), features fusion, global–local, hyperspectral image (HSI).

I. INTRODUCTION

EACH pixel in hyperspectral images (HSIs) has a continuous and rich spectral curve. Compared with RGB images and multispectral images, HSIs can simultaneously obtain

rich spectral information and spatial information of ground features [1], and have advantages in detecting and identifying ground cover. Therefore, it is widely used in urban remote sensing [2], environmental monitoring [3], precision agriculture [4], and other fields [5]. Around the key issues in these application fields, a large number of research topics have become the focus, such as classification [6]–[9], target detection [10]–[12], super-resolution [13]–[15], etc. Hyperspectral classification predicts the label of each pixel, which is the basic task of most applications and has important research significance [16].

In recent years, the rapid development of advanced pattern recognition methods has extensively promoted the development of HSI classification [17]. Deep learning (DL) captures the advanced features of the original data adaptively through a hierarchical structure. As a powerful feature extraction tool [18], it has been successfully applied in the field of remote sensing [19]–[22]. Chen *et al.* [23] proposed a combination of stacked autoencoder and principal component analysis (PCA) to extract spectral features. Zhong *et al.* [24] developed a deep belief network for hyperspectral classification tasks through regularized pretraining and fine-tuning processes, using diversity to promote priors rather than latent factors. It benefited from the superior performance of the recurrent neural network (RNN) [25] and its variant, long short-term memory (LSTM) [26], in natural language processing. In [27], Mou *et al.* considered the inherent sequence data structure of hyperspectral pixels and introduced an RNN to treat the spectrum of pixels as a 1-D sequence for training. However, there are some problems with these networks, due to they are all fully connected (FC) layers or additional sequence updates are needed to optimize a large number of parameters, the training cost is increased significantly. In addition, these networks did not fully exploit the 2-D spatial information of HSI.

Convolutional neural network (CNN) can more effectively address these problems due to its local connection and parameter sharing [28]. Hu *et al.* [29] used CNNs for feature extraction of spectral vector for the first time and applied it to HSI classification task. Xie *et al.* proposed a densely connected CNN framework [30]. It extracts multiscale patches around the center pixel and then uses dense modules to fuse multiscale information for classification. In [31], a 3-D CNN (3D-CNN) was proposed to simultaneously extract features in the spatial and spectral domains of HSI for classification. In [32], the hybrid 3-D–2-D CNNs were developed to reduce the model complexity caused by using 3-D CNNs alone. Feng *et al.* [33]

Manuscript received September 24, 2021; revised November 2, 2021 and November 26, 2021; accepted December 2, 2021. Date of publication December 9, 2021; date of current version December 24, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61977022, in part by the Science Foundation for Distinguished Young Scholars of Hunan Province under Grant 2020JJ2017, in part by the Key Research and Development Program of Hunan Province under Grant 2019SK2012, in part by the Foundation of Department of Water Resources of Hunan Province under Grant XSKJ2021000-12 and Grant XSKJ2021000-13, in part by the Natural Science Foundation of Hunan Province under Grant 2020JJ4340, and Grant 2021JJ40226, in part by the Foundation of Education Bureau of Hunan Province under Grant 20B257, and Grant 20B266, in part by the Scientific Research Fund of Education Department of Hunan Province under Grant 19A200, in part by the Open Fund of Education Department of Hunan Province under Grant 20K062, and in part by the Postgraduate Scientific Research Innovation Project of Hunan Province under Grant QL20210254. (*Corresponding authors: Wei He.*)

Bing Tu, Wangquan He, Wei He, and Xianfeng Ou are with the School of Information Science and Technology, Hunan Institute of Science and Technology, Yueyang 414000, China (e-mail: tubing@hnist.edu.cn; wangquan_he@vip.hnist.edu.cn; hewei@hnist.edu.cn; ouxf@hnist.edu.cn).

Antonio Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, E-10003 Caceres, Spain (e-mail: aplaza@unex.es).

Digital Object Identifier 10.1109/JSTARS.2021.3133009

constructed a semisupervised CNN network to transform band selection into a reinforcement learning problem. These studies verify the effectiveness of CNN in hyperspectral classification tasks. However, the CNN-based method will inevitably produce the oversmoothing phenomenon, leading to the misallocation of small targets and the loss of image edge details.

Based on the above work, many researchers have further explored the method of hyperspectral feature extraction based on DL. The two-branch network has recently attracted attention in HSI classification. Compared with the single model, the dual-branch network can break the performance bottleneck of the single-branch network and achieve higher performance. Yang *et al.* [34] used 1DCNN and 2DCNN to extract the spectral features and spatial features of HSI, respectively, and cascade fusion classification of the two extracted features. Hong *et al.* [35] developed a combination of mini-graph convolutional networks (miniGCNs) and CNN for hyperspectral classification and explored three fusion strategies (additive fusion, multiplicative fusion, and cascade fusion) to compare the performance gains achieved. A joint BiRNN-based spectral attention network and CNN-based spatial attention network for hyperspectral classification was proposed in [36]. In [37], a gated recurrent units (GRUs)-based HSI classification model of a cascaded RNN is proposed, which designs two cascaded RNN models for fully learning the information between the spectrals, and performs a weighted fusion of the learned features, simultaneously, the convolutional layer is integrated into the proposed model to learn further the spatial characteristics of each band. Although the above methods obtained excellent performance in hyperspectral processing, the internal features in the spectral and spatial domains were not fully explored since they only considered feature aggregation at the end of the branches. In addition, previous works usually used a simple linear fusion strategy for spectral and spatial dimensional features, which reflects the limited discriminability of pixel features and limits the classification accuracy.

To overcome these problems, we propose a global–local hierarchical weighting fusion network (GLH-WFN) for hyperspectral classification. The model includes two data streams used to extract the spectral and spatial features of HSI. In the spectral subnetwork, we design two different band grouping strategies to focus on the local and global information of the bands, and use the BiLSTM as the backbone network to obtain the global–local spectral features. In the spatial subnetwork, a global–local pooling fusion module is designed after each convolutional layer to enhance the complementarity of local and global information in the downsampling process. To better fuse the spectral and spatial features, we use hierarchical weighted fusion to obtain the spectral–spatial second-order statistical features, which can effectively utilize the correlation information between different channels to generate more representative features compared to the first-order statistical features [38],[39]. Moreover, the model uses an end-to-end strategy to train both spectral and spatial subnetworks simultaneously. The main contributions of this article can be summarized as follows.

- 1) A global–local grouping fusion-based BiLSTM (GL-BiLSTM) is proposed to full capture global and local

context information of the spectral vector, which can further improve the separability of spectral features and effectively overcome the training difficulties caused by high-dimensional spectra.

- 2) Attention mechanism is introduced in the pooling layer, and the global–local pooling fusion (GL-PF) module is proposed. The module obtains spatial global and local information by considering the strong complementarity and correlation between different pooling layers, and can adaptively sampling and enhancing important features to effectively overcome interfering information in patches.
- 3) An end-to-end HSI supervised classification network (GLH-WFN) is proposed. Compared with the existing spectral–spatial joint network, GLH-WFN makes full use of the second-order statistical properties of spectral and spatial features. In addition, the model considers the weights of spectral and spatial features in the merged layer.

The rest of this article is organized as follows: Section II briefly reviews related works. Section III introduces the proposed GLH-WFN in detail, including the spectral subnetworks, spatial subnetworks, and hierarchical weighted fusion layers. Section IV reports the experimental results and analysis, and the practical application analysis of the proposed method is presented in Section V. Finally, Section VI concludes the article.

II. RELATED WORKS

In this section, we briefly review the relevant methods used in the proposed framework, including BiLSTM, pooling operation in CNN, and bilinear pooling.

A. Bidirectional Long Short-Term Memory

As a variant of RNN [25], LSTM [26] is proposed to solve the problem of gradient disappearance and gradient explosion caused in the process of long sequences modeling. Compared with RNN, LSTM adds a memory cell and three control gates (forget gate, input gate, and output gate). Specifically, at time t , the three control gates can be expressed as follows:

Forget Gate

$$f_t = \sigma(W_{hf} \cdot h_{t-1} + W_{xf} \cdot x_t + b_f). \quad (1)$$

Input Gate

$$i_t = \sigma(W_{hi} \cdot h_{t-1} + W_{xi} \cdot x_t + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_{hc} \cdot h_{t-1} + W_{xc} \cdot x_t + b_c) \quad (3)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t. \quad (4)$$

Output Gate

$$o_t = \sigma(W_{ho} \cdot h_{t-1} + W_{xo} \cdot x_t + b_o) \quad (5)$$

$$h_t = o_t \odot \tanh(C_t) \quad (6)$$

where x_t , h_t represent input and output of the hidden layer, respectively, and f_t , i_t , and o_t represent the forget gate, input gate, and output gate, respectively. C_t is the cell state, and \tilde{C}_t is the candidate cell value.

$\{(W_{\Delta f}, b_f), (W_{\Delta i}, b_i), (W_{\Delta o}, b_o), (W_{\Delta C}, b_C), \Delta = h, x\}$ are the weight matrices and bias terms of forget gate, input gate, output gate, and candidate cell value, respectively. $\sigma(\cdot)$ is the sigmoid function. “ \odot ” is the elementwise multiplication.

The unique gating mechanism of LSTM can significantly improve the capacity to capture long sequence information. However, the LSTM only considers past information. BiLSTM [40] consists of a forward LSTM layer and a backward LSTM layer. These two LSTM layers trained along opposite directions connect the output layer simultaneously, so that the output is determined by the states from past and future. Specifically, at time t , the output of the hidden layer of the LSTM in two different directions can be expressed as

$$\vec{h}_t = \text{LSTM}(x_t, \vec{h}_{t-1}) \quad (7)$$

$$\overleftarrow{h}_t = \text{LSTM}(x_t, \overleftarrow{h}_{t-1}) \quad (8)$$

where $\overleftarrow{h}_{t-1}, \vec{h}_{t-1}$ represent the forward and backward hidden layer output at $(t-1)$ time, respectively, and LSTM represents LSTM unit. After two LSTM hidden layer outputs in different directions are obtained, they are concatenated to form the final output of the entire BiLSTM hidden layer. The process can be expressed as

$$H_t = [\vec{h}_t, \overleftarrow{h}_t] \quad (9)$$

where $[\cdot, \cdot]$ represents connection operation, and H_t is the final output.

B. Pooling Operation in CNNs

Pooling layers are essential for CNNs in reducing the size of the model and improving operating efficiency. Furthermore, they can also expand the receptive field and improve the robustness of features. Currently, max-pooling, average-pooling, and sum-pooling are used widely in many popular backbone networks. Max-pooling takes the largest value in the feature area as the output, average-pooling, and sum-pooling take the average value and sum in the corresponding area as the output. In recent years, the research on pooling strategy has attracted more attention [38]. In [41], the researchers proposed a random pooling, which selects regional elements randomly according to their probability value. A combination based on max-pooling and average-pooling is proposed in [42]. Saeedan *et al.* [43] proposed a differentiable pooling strategy with higher performance than the max-pooling layer. In contrast to the traditional $N \times N$ local area, a stripe pooling strategy with a narrow convolution kernel ($1 \times N$ or $N \times 1$) is proposed in [44], which is conducive to improving the ability of the backbone network to simulate long-distance relationships. Furthermore, other pooling methods are used to achieve specific tasks. He *et al.* proposed spatial pyramid pooling [45], which uses a pooling layer with asynchronous length and window size to obtain multiscale information, and solves the problem of fixed input size and repeated convolution calculations in the network. These pooling strategies have improved the performance of CNN in different tasks.

C. Bilinear Pooling

Recent works [38], [39] have shown that high-order statistics have achieved exciting performance in computer vision tasks. Bilinear pooling shows great potential for feature fusion by modeling higher order statistics information of features [46]. According to the different sources of feature extraction, it can be divided into multimodal bilinear pooling and homogeneous bilinear pooling (second-order pooling). For two features $f_A(\mathcal{I}, l) \in \mathbb{R}^{D \times M}$ and $f_B(\mathcal{I}, l) \in \mathbb{R}^{D \times N}$ of sample \mathcal{I} at position l , first perform an outer product operation, as follows:

$$Fu(l, \mathcal{I}, f_A, f_B) = f_A^T(l, \mathcal{I})f_B(l, \mathcal{I}) \in \mathbb{R}^{M \times N}. \quad (10)$$

The sum-pooling operation is used for all positions of Fu to obtain the matrix ξ

$$\xi(\mathcal{I}) = \sum_l Fu(l, \mathcal{I}, f_A, f_B) \in \mathbb{R}^{M \times N}. \quad (11)$$

Finally, matrix ξ is spanned into a vector, denoted as bilinear vector α

$$\alpha = \text{vec}(\xi(\mathcal{I})) \in \mathbb{R}^{MN \times 1}. \quad (12)$$

After the matrix normalization operation and L_2 normalization operation are performed on α , the fusion feature γ is obtained

$$\beta = \text{sign}(\alpha)\sqrt{|\alpha|} \in \mathbb{R}^{MN \times 1} \quad (13)$$

$$\gamma = \beta / \|\beta\|_2 \in \mathbb{R}^{MN \times 1} \quad (14)$$

where M and N represent the number of channels.

In general, bilinear pooling can generate more representative features by aggregating second-order statistics between features due to [47]: 1) By describing the relationship between feature vectors, it can make full use of the correlation information between features. 2) The second-order statistic is aggregated by the outer product operation, which can fully capture the global representation between features.

III. PROPOSED METHOD

As shown in Fig. 8, the method proposed in this article includes three parts: 1) Global–local grouping fusion-based BiLSTM, 2) global–local pooling fusion-based CNN, and 3) spectral–spatial hierarchical weighting fusion. The method enhances the spectral–spatial feature representation of the HSI and simultaneously uses the high-order statistical information of the spectral–spatial feature to strengthen their connection, which is more conducive to improving classification accuracy.

A. Spectral: Global–Local-Grouping Fusion-Based BiLSTM

DL has been shown to be significantly effective in dealing with complex spectral features in HSI [37]; vector-based spectral classification methods focus only on the local dependency of the spectrum and do not take into account the complementarity between nonadjacent bands [27]. In contrast, BiLSTM represents the spectrum from a sequence perspective and pays more attention to its contextual information. If the high-dimensional spectral vectors are directly input into the BiLSTM band by band

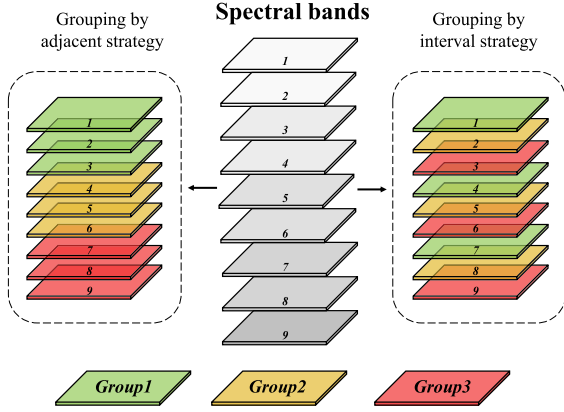


Fig. 1. Illustration of the two spectrum grouping strategies

for training, it will make the network structure too deep, thus making training difficult. Therefore, it is necessary to reasonably group spectral vectors consisting of hundreds of bands.

We design two grouping strategies from the global and local perspectives of the whole spectral band to focus on the contextual information between the spectra. Specifically, for a pixel \mathbf{u} in HSI, the i th reflection value can be described as u_i . Let n be the total number of bands and $g^{(i)}$ be the input sequence group of the i -th time step in BiLSTM. Then, the two grouping strategies can be described as follows.

Adjacent Grouping Strategy

$$\begin{aligned} g_{ad}^{(1)} &= [u_1, u_2, \dots, u_m] \\ g_{ad}^{(2)} &= [u_{m+1}, u_{m+2}, \dots, u_{2m}] \\ &\dots \\ g_{ad}^{(\tau)} &= [u_{(\tau-1)m+1}, u_{(\tau-1)m+2}, \dots, u_n]. \end{aligned} \quad (15)$$

Interval Grouping Strategy

$$\begin{aligned} g_{in}^{(1)} &= [u_1, u_{1+\tau}, \dots, u_{1+\tau(m-1)}] \\ g_{in}^{(2)} &= [u_2, u_{2+\tau}, \dots, u_{2+\tau(m-1)}] \\ &\dots \\ g_{in}^{(\tau)} &= [u_i, u_{i+\tau}, \dots, u_n] \end{aligned} \quad (16)$$

where τ represents the number of time steps in BiLSTM, and m is the length of the input sequence at a time step. Fig. 1 illustrates an example of two different grouping strategies.

As show in Fig. 2, suppose the time step is 3, we plot the correlation matrix between the spectral bands of Indian Pines images (See Section IV-A for details) and the interband correlation matrix between the spectral groups generated by the two grouping strategies. From Fig. 2(b)–(d), we can find that the correlation matrix between spectral groups consisting of adjacent band grouping strategies is diverse, and different groups have different degrees of correlation. The wavelength span of the spectral groups within the same step is short, and this grouping strategy is more concerned with the local characteristics of the

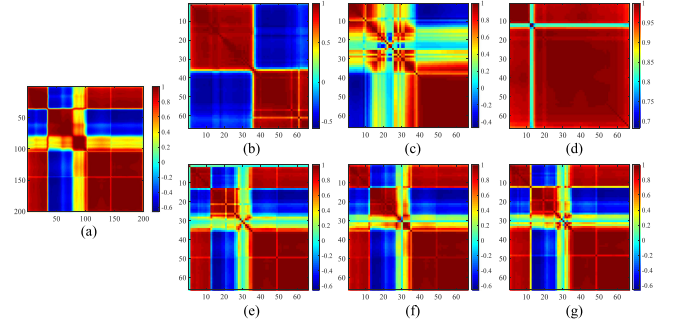


Fig. 2. Spectral correlation matrices of Indian Pines dataset. (a) Original spectral band. (b)–(d) Spectral group generated by adjacent band grouping strategy (i.e., local grouping strategy). (e)–(g) Spectral group generated by interval band grouping strategy (i.e., global grouping strategy).

spectrum. However, Fig. 2(e)–(g) shows that correlation matrices between the spectral groups obtained from interval band grouping strategies are similar and consistent with the original bands [see Fig. 2(a)]. The wavelength span of the spectral groups within the same step is long, and this grouping strategy is more concerned with the global characteristics of the spectrum.

The common point of the two grouping strategies is that they are beneficial to solving the complex of BiLSTM training caused by the excessively deep spectral dimension. In order to fully mine the deep spectral information, we combine the global and local information of the band simultaneously to further strengthen the discrimination of spectral features. As shown in Fig. 3, for the pixel \mathbf{u} , the spectral bands are grouped by two different strategies and fed into two models simultaneously for training, and the features learned by the two strategies are fused using elementwise addition. We consider their contributions to be equivalent and the process is represented as

$$\mathcal{F}_{spe} = \mathcal{F}_{global} \oplus \mathcal{F}_{local} \quad (17)$$

where \mathcal{F}_{global} and \mathcal{F}_{local} represent the outputs of the interval band grouping strategy and the adjacent band grouping strategy through the FC layer of BiLSTM, respectively; \mathcal{F}_{spe} is the fusion feature and \oplus denote elementwise addition. Then, after \mathcal{F}_{spe} passes through an FC layer, the softmax layer determines the label allocation.

B. Spatial: Global–Local–Pooling Fusion-Based CNN

When using CNN for feature extraction of HSI, the neighborhood patch centered on the target pixel is usually selected to replace its spatial information as the input of the network. As shown in Fig. 4, on the one hand, the pooling operation inevitably leads to the loss of mapping information, so it is necessary to consider the most discriminative features of sampling. On the other hand, in complex hyperspectral scenes, there is less spatial information representing the target pixel in patch, and the interfering spatial information may have a greater interference to the expression of pixel labels on the feature map. In the case of limited network structure, a single pooling operation often cannot overcome this problem well. Inspired by [48], we introduce attention learning in pooling layer and propose GL-PF module

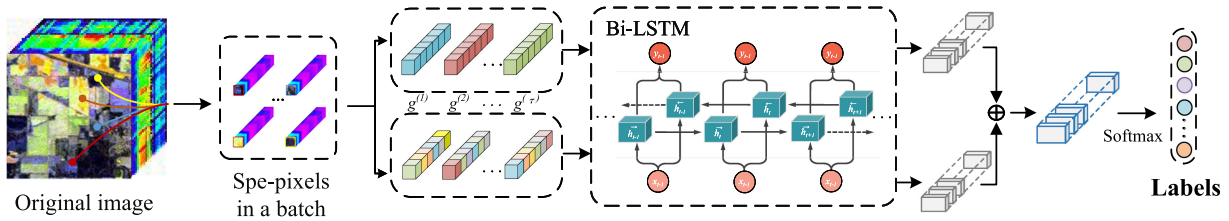


Fig. 3. Flowchart of the proposed GL-BiLSTM.

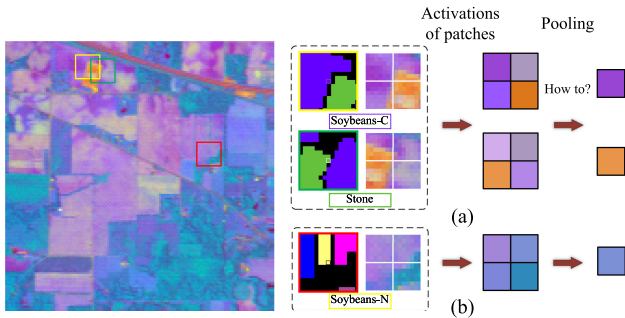


Fig. 4. (From left to right) HSI after dimensionality reduction, the groundtruth of the target pixel and its corresponding neighborhood patch, the corresponding activation in the feature map, and finally the pooling activation we want. (a) Pixels at the boundary of the two classes. The purple tone activation is caused by Soybeans-C, and the orange tone activation is caused by the Stone. We want to get the purple tone activation in the patch above and the orange tone activation in the patch below. (b) The activation process of a target pixel in a more complex scene.

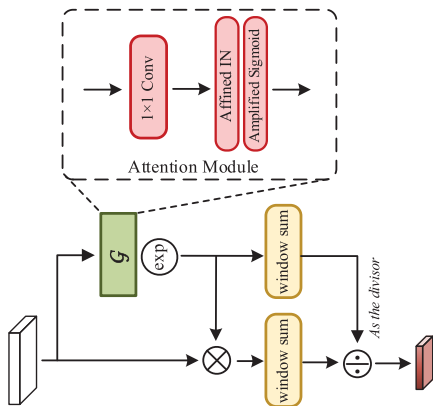


Fig. 5. Flowchart of the proposed GL-CNN.

in the spatial subnetwork to solve these problems. Next, we will give a detailed description of the proposed global-local-pooling fusion-based CNN (GL-CNN). Fig. 5 shows the classification framework of GL-CNN. First, we perform PCA dimensionality reduction processing on the original spectrum to achieve the purpose of reducing computational consumption and then construct a patch centered on the target pixel as the input of CNN. After each convolution, we consider three pooling strategies (LA pooling, max pooling, and average pooling), by constructing GL-PF module to obtain downsampled output features.

We describe the construction of LA-pooling and global-local pooling fusion below.

1) *Local Attention Pooling*: From the perspective of local importance, for the feature map I after convolution, Ω is the kernel index set corresponding to the relative sampling position $(\Delta x, \Delta y)$ in the sliding window, and the starting position of the sliding window (i.e., the top left of the feature map I) is set to (x, y) , the input position is set to (x', y') , the pooling process can be expressed as

$$O_{x',y'} = \frac{\sum_{(\Delta x, \Delta y) \in \Omega} F(I)_{x+\Delta x, y+\Delta y} I_{x+\Delta x, y+\Delta y}}{\sum_{(\Delta x, \Delta y) \in \Omega} F(I)_{x+\Delta x, y+\Delta y}} \quad (18)$$

where $F(I)$ represents the importance map, and the size is the same as the input feature map I . We learn the importance map through a learnable attention module $\mathcal{G}(I)$ and determine the downsampling features through the attention map. As shown in Fig. 6, $\mathcal{G}(I)$ is composed of a 1×1 convolution, which is a tiny fully convolutional network used to capture important maps. AffinedIN denote affine instance normalization, which makes each channel of each feature map obey the normal distribution, and then use the amplified sigmoid function to assist it to adjust the range. The attention module $\mathcal{G}(I)$ can be written by the following form:

$$\mathcal{G}(I) = \sigma(IN(I * W_1 + b_1)) \quad (19)$$

where W_1 and b_1 , respectively, denote convolution kernel weights and bias parameters, IN denotes instance normalization, and σ is the sigmoid function. In order to keep the importance weight positive and propagate backward more readily, $\exp(\cdot)$ is added after the attention module $\mathcal{G}(I)$. The importance map can be described as

$$F(I) = \exp(\mathcal{G}(I)). \quad (20)$$

Based on the importance map obtained, the feature output of location (x, y) using LA pooling can then be described as

$$O_{x',y'} = \frac{\sum_{(\Delta x, \Delta y) \in \Omega} I_{x+\Delta x, y+\Delta y} \exp(\mathcal{G}(I))_{x+\Delta x, y+\Delta y}}{\sum_{(\Delta x, \Delta y) \in \Omega} \exp(\mathcal{G}(I))_{x+\Delta x, y+\Delta y}}. \quad (21)$$

2) *Global-Local-Pooling Fusion Module*: Average pooling can reduce the increase in the variance of the estimated value caused by the limited size of the neighborhood, and max pooling can reduce the estimated mean deviation caused by the convolutional layer parameter error theory [41]. Average pooling takes the mean value of the overall data as the output of pooling, which is focused on the background information of the image

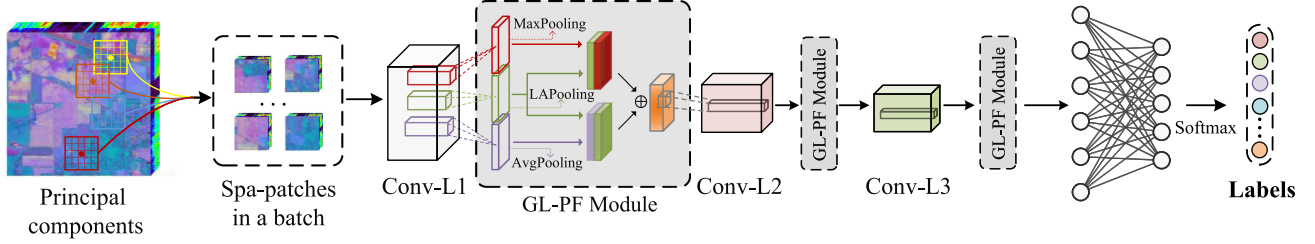
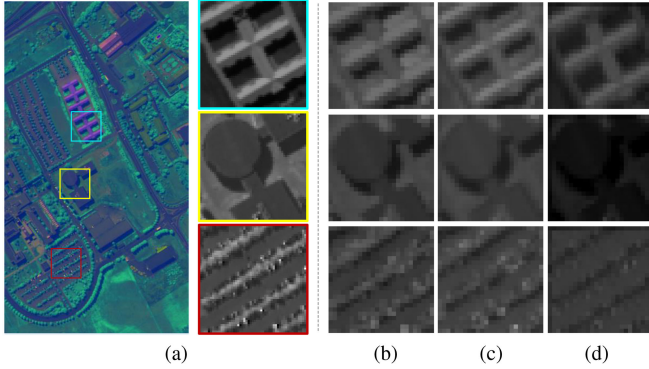


Fig. 6. Process of LA pooling.

Fig. 7. Illustration of feature maps obtained by different pooling layers in the three scenes of Pavia University. (a) Input patches; (b)–(d) show the downsampled feature maps of max pooling, average pooling, and LA pooling with $1 \times$ pooling operation, respectively.

and reflects the global features, whereas max pooling takes the maximum value as the pooling output, which focuses more on the texture information of the image and reflects the local features. Fig. 7 illustrates the Pavia University (see Section IV-A for details) as an example, and selects three scenes to illustrate the feature maps obtained by performing a pooling operation with three different pooling strategies. To fully explore the global and local features in the pooling layer, we consider the strong complementarity and correlation between the most discriminative activation of the pixels obtained by LA-pooling downsampling and the background features and texture features, by cascading the features obtained by LA pooling and the features obtained by max pooling and average pooling, respectively, and finally adopt elementwise addition and fusion to obtain the final downsampling feature vector. The processes are described as follows:

$$\begin{aligned}
 s_i^{(M-L)} &= [\text{MaxPooling}(I_i), \text{LAPooling}(I_i)] \\
 s_i^{(A-L)} &= [\text{LAPooling}(I_i), \text{AvePooling}(I_i)] \\
 \mathcal{S}_i &= s_i^{(M-L)} \oplus s_i^{(A-L)}
 \end{aligned} \quad (22)$$

where $I_i (i = 1, 2, 3)$ represents the features after different convolutional layers, $s_i^{((M/A)-L)} (i = 1, 2, 3)$ represents the features of different downsampling layers of LAP-pooling operation and Max/Ave-pooling operation that are cascading, and $\mathcal{S}_i (i = 1, 2, 3)$ represents the fusion features after different

downsampling layers. $[\cdot, \cdot]$ denotes concatenation and \oplus denotes elementwise addition. After the last downsampling operation, we add an FC layer to obtain the feature vector, and finally, determine the label of the target pixel through the softmax layer.

C. Spectral–Spatial: Hierarchical Weighting Fusion

The fusion of spectral and spatial features is of great importance for HSI classification. Previous studies adopted linear fusion strategy [49], which only considered the first-order spectral and spatial statistical features. The proposed GLH-WFN makes full use of the second-order spectral and spatial statistical features to obtain a more discriminative global representation. Furthermore, considering that the importance of spectral and spatial features varies with objects and scenes, we design a weighting matrix to improve the spectral–spatial second-order statistics of the fusion layer. It can be trained in the network along with other parameters. In this section, we describe the proposed GLH-WFN in detail.

For the features $\mathcal{F}_{\text{spe}} \in \mathbb{R}^{1 \times M}$ and $\mathcal{F}_{\text{spa}} \in \mathbb{R}^{1 \times M}$ of the pixel x extracted by the two branches, first perform the outer product learning high-order statistical information

$$\mathbf{Q} = \mathcal{F}_{\text{spe}}^T \mathcal{F}_{\text{spa}} \in \mathbb{R}^{M \times M} \quad (23)$$

where $\mathbf{Q} \in \mathbb{R}^{M \times M}$ is the second-order-pooling matrix of spectral–spatial characteristics. The outer product can consider the paired interaction of spatial–spectral features to constrain their output, similar to the feature expansion in the secondary kernel [46]. Given that the spectral dimension and the spatial dimension have different constraints to the learned new features, we first calculate the difference values of different nodes on the spectrum and spatial feature vectors

$$\mathcal{D}_{(i,j)} = |f_{\text{spe}}^i - f_{\text{spa}}^j| \quad (24)$$

where f_{spe}^i and $f_{\text{spa}}^j (i, j \in \{1, 2, \dots, M\})$ denote the i th and the j th node of the spatial and spectral feature vectors, respectively. With the above obtained $\mathcal{D}_{(i,j)}$, then by mapping it to adjust the range to get the weight $w_{(i,j)}$

$$w_{(i,j)} = e^{-\mathcal{D}_{(i,j)}} \quad (25)$$

the weight matrix $W \in \mathbb{R}^{M \times M}$ can be constructed by $w_{(i,j)}$. Then, the weighted second-order-pooling matrix can be expressed as

$$\mathbf{Q}_w = W \cdot \mathbf{Q} \in \mathbb{R}^{M \times M}. \quad (26)$$

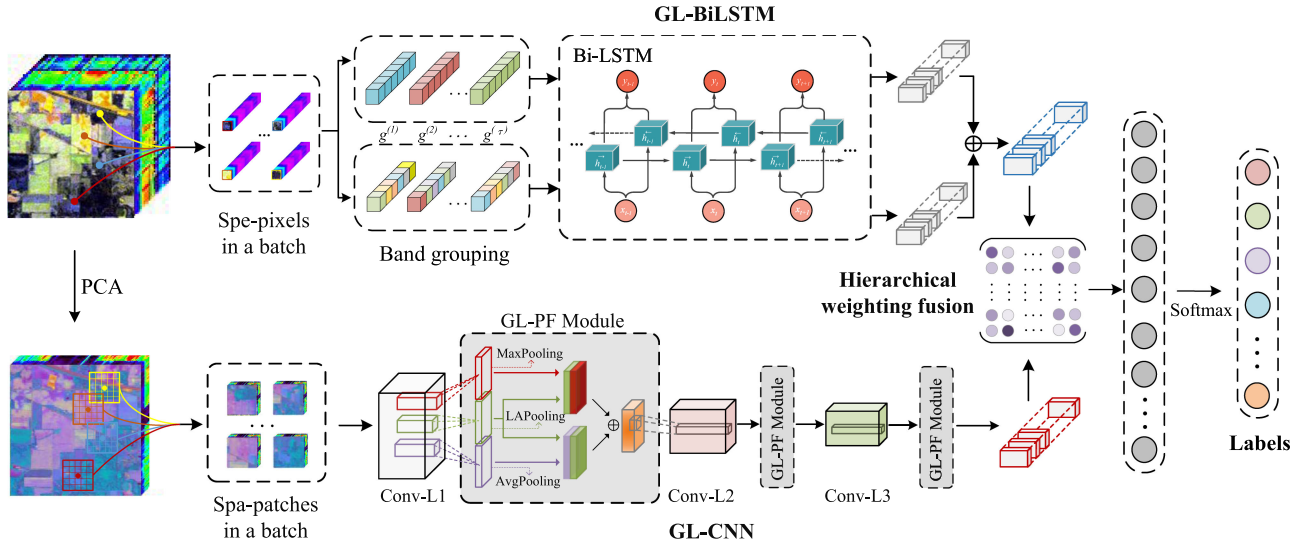


Fig. 8. Flowchart of the proposed GLH-WFN.

Finally, perform matrix normalization and L2 norm normalization on the weighting matrix Q_w , and flatten it into a vector. The specific description is as follows:

$$X = \text{sign}(Q_w) \sqrt{|Q_w|} \in \mathbb{R}^{M \times M} \quad (27)$$

$$Y = X / \|X\|_2 \in \mathbb{R}^{M \times M} \quad (28)$$

$$Z = \text{vec}(Y) \in \mathbb{R}^{MM \times 1}. \quad (29)$$

Fig. 8 illustrates the proposed GLH-WFN. After obtaining the fused vector Z , we add an FC layer to obtain the joint feature vector, and finally use the softmax layer to determine the probability of each category to complete the classification.

IV. EXPERIMENTAL RESULTS

The organizational structure of the experiment is as follows. First, four benchmark datasets for experimental analysis are introduced. Second, the impact of three parameters in the model on classification accuracy is analyzed. Third, the ablation experiments on the spectral and spatial branches are analyzed. Fourth, the effects of different feature fusion strategies on the model effect are compared. Fifth, we compare the proposed GLH-WFN with different classification methods on four hyperspectral datasets and analyze their corresponding time complexity. Finally, the practicability of GLH-WFN is analyzed using the hyperspectral dataset of Dongting Lake watershed obtained by the GF-5 satellite.

This article uses overall accuracy (OA), average accuracy (AA), category accuracy (CA), and the kappa coefficient (Kappa) as evaluation indicators. All experiments are repeated 10 times to improve the scientific credibility of the experimental results, and the average value and standard deviation are used as the final results of the experiment. The experiment is implemented under Keras 2.2.4 with TensorFlow backend, The batch size is set to 128, training epoch to 500, and learning rate to $1e-4$ in the experiment. Adam optimizer is adopted to update

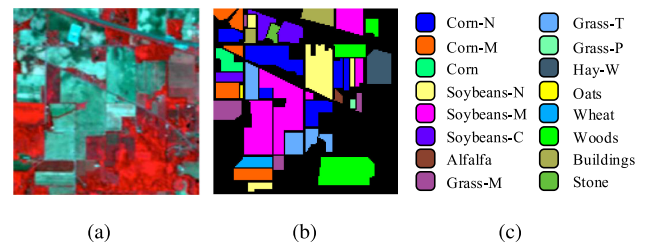


Fig. 9. Indian Pines dataset. (a) False-color composite map. (b) Groundtruth. (c) Color coding of the label.

network parameters. All experiments are performed on a PC with an Intel(R) Core(TM) i7-7800X CPU 3.50 GHz, NVIDIA GTX 1080 Ti GPU, and 32 GB RAM.

A. Datasets Description

Experiments were conducted on four benchmark hyperspectral datasets (Indian Pines, Pavia University, Salinas, and Washington DC) to evaluate the effectiveness of the proposed GLH-WFN. The corresponding false-color composite image and Groundtruth map are shown in Figs. 9–12, and corresponding specific categories information and training samples are shown in Table I. The specific introduction of the datasets is as follows.

- 1) The Indian Pine dataset was acquired by the AVIRIS hyperspectral sensor in the Indian Pine forest test area in northwestern Indiana, USA. Its spatial dimensions are 145×145 , and its spatial and spectral resolutions are 20 m and 10 nm, respectively. 20 wavebands absorbed by atmospheric water were removed from the original bands, leaving 200 bands for classification, and the image contains 10 249 labeled samples and 16 categories.
- 2) The Pavia University dataset was taken by the ROSIS-3 sensor at the University of Pavia in Italy. Its spatial size

TABLE I
NUMBER OF SAMPLES IN EACH CATEGORY USED FOR TRAINING AND TESTING IN THE INDIAN PINE, PAVIA UNIVERSITY, SALINAS, AND WASHINGTON DC DATASETS

Indian Pines			
Number	Class Name	Training	Test
1	Alfalfa	5	41
2	Corn-notill	143	1285
3	Corn-min	120	710
4	Corn	40	197
5	Grass/Pasture	60	423
6	Grass/Trees	90	640
7	Grass/Pasture-mowed	3	25
8	Hay-windrowed	48	430
9	Oats	4	16
10	Soybeans-notill	98	874
11	Soybeans-min	200	2255
12	Soybeans-clean	80	513
13	Wheat	21	184
14	Woods	126	1139
15	Building-Grass-Trees-Drives	60	326
16	Stone-steel Towers	20	73
Total		1118	9131
Pavia University			
Number	Class Name	Training	Test
1	Asphalt	198	6433
2	Meadows	559	18090
3	Gravel	146	1953
4	Tress	153	2911
5	Metal sheets	40	1305
6	Bare soil	150	4879
7	Bitumen	66	1264
8	Bricks	110	3572
9	Shadows	47	900
Total		1469	41307
Salinas			
Number	Class Name	Training	Test
1	Weeds_1	47	1962
2	Weeds_2	75	3651
3	Fallow	40	1936
4	Fallow_plow	28	1366
5	Fallow_smooth	54	2624
6	Stubble	80	3879
7	Celery	72	3507
8	Grapes	226	11045
9	Soil	125	6078
10	Corn	66	3212
11	Lettuce 4wk	22	1046
12	Lettuce 5wk	39	1888
13	Lettuce 6wk	19	897
14	Lettuce 7wk	30	1040
15	Vinyard untrained	146	7122
16	Vinyard trellis	37	1770
Total		1106	53023
Washington DC			
Number	Class Name	Training	Test
1	Roofs	50	3079
2	Road	50	1740
3	Grass	50	1352
4	Trail	50	1214
5	Trees	50	1144
6	Shadow	50	1070
Total		300	9599

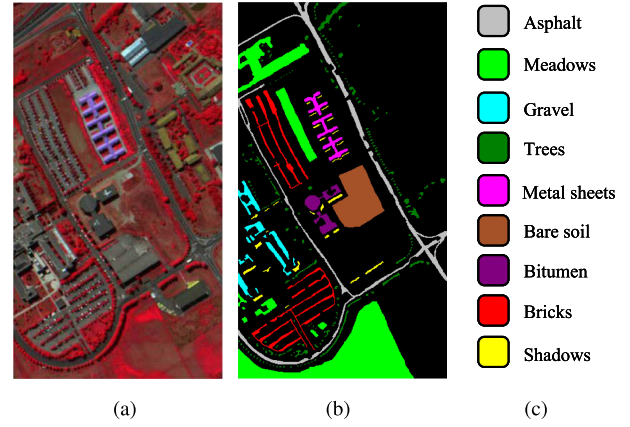


Fig. 10. Pavia University dataset. (a) False-color composite map. (b) Groundtruth. (c) Color coding of the label.

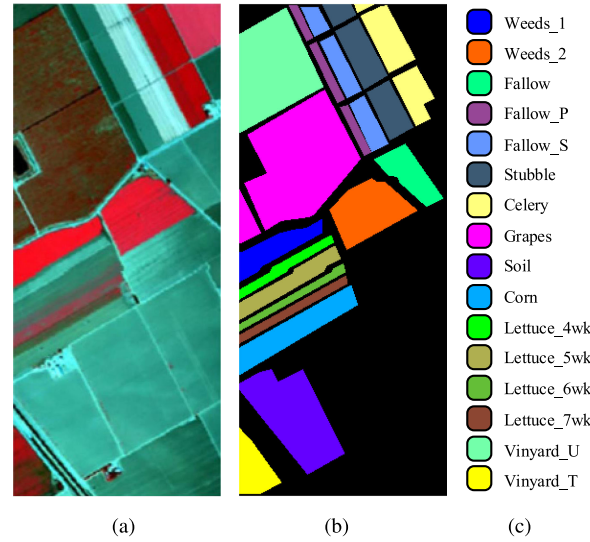


Fig. 11. Salinas dataset. (a) False-color composite map. (b) Groundtruth. (c) Color coding of the label.

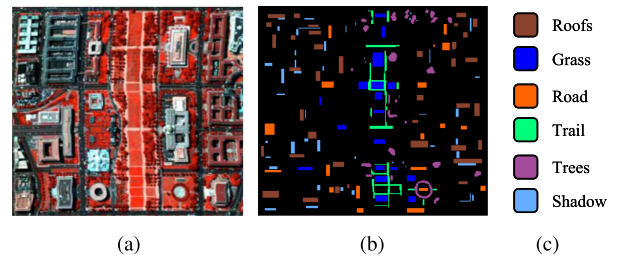


Fig. 12. Washington DC dataset: (a) False-color composite map. (b) Groundtruth. (c) Color coding of the label.

is 610×340 , the spatial resolution is 1.3 m, and the wavelength range is $0.4\text{--}0.86 \mu\text{m}$. Remove 12 noise bands based on the original band, leaving 103 bands. The HSI data contains 41 176 labeled pixels and 9 categories.

- 3) The Salinas dataset was taken by the AVIRIS hyperspectral sensor in the Salinas Valley of California, USA.

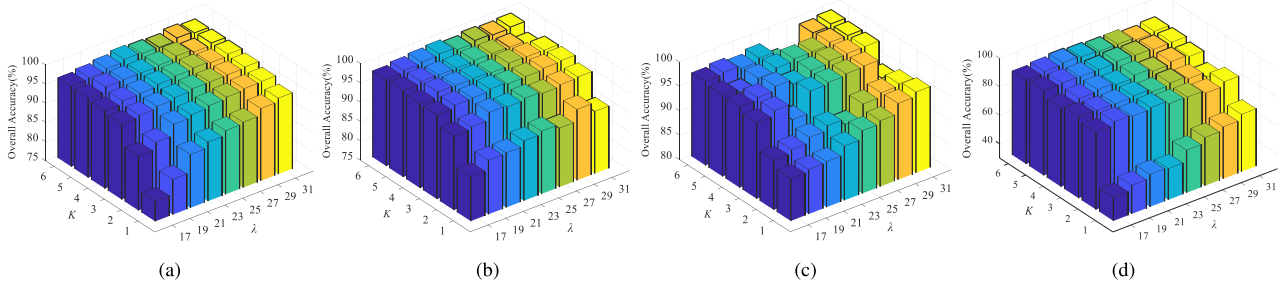


Fig. 13. Effect of principal component K and spatial neighborhood patch λ on classification performance. (a) Indian Pine dataset. (b) Pavia University dataset. (c) Salinas dataset. (d) Washington DC dataset.

The original data contains 224 bands, and 204 bands remain after removing 20 bands absorbed by atmospheric water. The image size is $512 \times 217 \times 204$. It contains 54 129 labeled pixels and 16 categories.

- 4) The Washington DC dataset was collected by the HYDCIE hyperspectral sensor over the Washington DC mall. Its space size is 280×307 , contains 210 ($0.4\text{--}2.4 \mu\text{m}$) continuous bands, and 19 water absorption bands ($0.9\text{--}1.4 \mu\text{m}$) are removed, leaving 191 bands for analysis. The data contains 9899 labeled pixels and 6 categories.

B. Parameter Tuning

For the proposed GLH-WFN model, in the spatial subnetwork, the PCA-based dimensionality reduction method is first adopted to reduce the original HSIs. Then the spatial neighborhood patch centered on the target pixels is selected as the input of the CNN. In addition to the above two critical parameters, in the spectral subnetwork, the step size of the band-grouping also is essential to the classification of the model. This section will explore the influence of the number of principal components, spatial neighborhood block size and band-grouping step on classification accuracy.

1) *Number of Principal Components (PCs)*: Dimensionality reduction of the original HSI can reduce the training cost while retaining most of the valuable information. We set the range of PCs K to [1–6] to analyze the impact on classification. As shown in Fig. 13, when the number of principal components K increases, OA values shows a trend of first increasing and then leveling off. To balance the effects of training cost and classification accuracy, we set K as 5 for Indian Pine, Pavia University, and Salinas, and 6 for Washington DC in subsequent experiments.

2) *Size of Spatial Neighborhood Patch*: We evaluate the impact of the spatial neighborhood patch size on the classification accuracy in detail by setting the spatial neighborhood patch λ size range of [17–31] with a step size of 2. As shown in Fig. 13, the OA values obtained by different spatial neighborhoods differ. Generally speaking, a larger spatial neighborhood patch λ can represent a larger homogeneous area. However, it will inevitably introduce unnecessary information. In the subsequent experiment, We set the spatial patch size to $[27 \times 27]$ for Indian Pine, Pavia University, and Salinas, and $[21 \times 21]$ for Washington DC.

3) *Spectral Grouping Step*: As shown in Fig. 14, when the timestep changes, it will lead to different classification performance. In general, the OA values decrease with increasing timestep, which indicates that the shallower network is conducive to improving the performance of the model. Moreover, when the timestep is small, OA fluctuates less in Indian Pine, Pavia University, and Salinas, which means that the model is more stable under this setting. In the following experiment, the timestep of all datasets is set to 2.

C. Analysis of Spectral and Spatial Model Branches

In the third experiment, the effects of different grouping strategies in the spectral branch and pooling methods in the spatial branch on classification accuracy are explored.

1) *Spectral Branch Grouping Strategy*: The band-grouping strategy is essential for extracting the rich spectral information of the HSI. In this experiment, we analyzed the impact of different grouping strategies on classification performance and compared them with the proposed grouping strategy. The results of the experiment are shown in Table II. In the band-by-band-grouping method, each band corresponds to the time step of the BiLSTM. This grouping method is often accompanied by a deeper network that produces high training time costs. The “all-in-one” grouping strategy means dividing all bands in HSI into one time step as the input of BiLSTM. It can capture the global context information of the entire band and retain the continuity of the original frequency band, while ignoring the local context information in the local adjacent frequency bands. Adjacent grouping and interval grouping indicate that the input sequences are divided into n frequency band groups in an adjacent and spaced manner with n time steps, corresponding to (15) and (16).

Based on the results shown in the table, the proposed global and local grouping strategy achieved the highest accuracy in four datasets. Compared to Band-by-band, Indian Pine, Pavia University, Salinas, and Washington DC increased by 16.37%, 10.9%, 6.8%, and 9.29%, respectively in OA, due to GL-BiLSTM which integrates global information and local information between bands. Furthermore, the time step set in this article is 2, which has a higher advantage in training efficiency.

2) *Spatial Branch Ablation Analysis*: We explore the impact of pooling combination methods on the accuracy of spatial branch

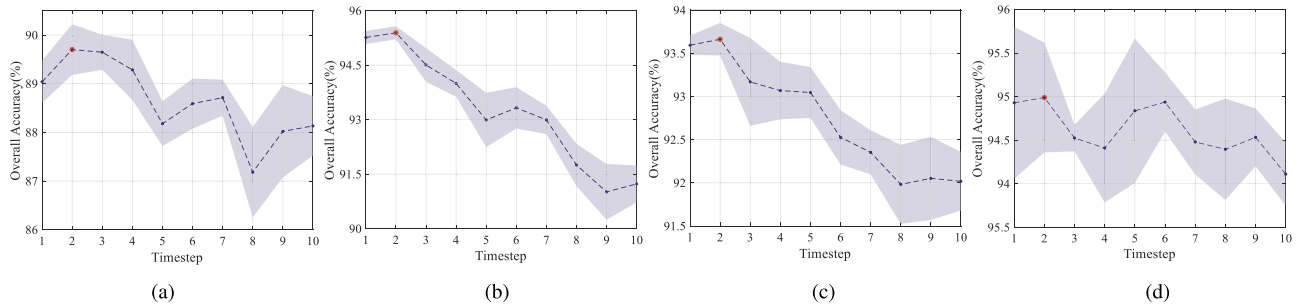


Fig. 14. Effect of grouping timestep on classification performance. The shaded area represents the standard deviation. (a) Indian Pines dataset. (b) Pavia University dataset. (c) Salinas dataset. (d) Washington dc dataset.

TABLE II
PERFORM CLASSIFICATION PERFORMANCE ANALYSIS ON THE DIFFERENT BAND GROUPING METHODS OF THE SPECTRAL NETWORK BRANCH IN THE FOUR BENCHMARK DATASETS

Method	Indian Pines			Pavia University			Salinas			Washington DC		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
Band-by-band	73.06	64.63	69.22	84.18	83.72	78.56	86.28	91.17	84.72	85.34	88.89	81.99
All-in-one	81.77	78.62	79.21	91.55	90.22	88.75	90.58	94.16	89.50	90.79	93.25	88.64
Adjacent grouping	81.82	77.76	79.26	91.29	90.20	88.41	90.24	93.79	89.12	90.63	93.31	88.46
Interval grouping	83.47	80.22	81.14	91.42	90.08	88.57	90.85	94.27	89.80	91.23	93.47	89.17
GL-BiLSTM	89.43	86.43	87.92	95.08	94.25	93.45	93.08	96.14	92.29	94.63	96.25	93.35

The best accuracy value for each column is highlighted in bold.

TABLE III
CLASSIFICATION PERFORMANCE OF DIFFERENT POOLING COMBINATIONS OF SPATIAL NETWORK BRANCHES IN FOUR BENCHMARK DATASETS

Method	Indian Pines		Pavia University		Salinas		Washington DC	
	OA	Param#(M)	OA	Param#(M)	OA	Param#(M)	OA	Param#(M)
LAP	97.12	1.013	97.65	1.009	98.09	1.013	84.11	0.695
LAP+Max	97.80	1.013	98.20	1.009	98.95	1.013	88.86	0.695
LAP+Ave	97.66	1.013	97.93	1.009	98.96	1.013	88.37	0.695
Max+Ave	97.87	0.930	98.24	0.926	98.65	0.930	89.53	0.612
Max+Ave+LAP	98.09	1.013	98.16	1.009	98.89	1.013	87.78	0.695
[LAP,Max]	97.32	1.927	97.39	1.923	98.24	1.927	87.68	1.297
[LAP,Ave]	97.40	1.927	97.36	1.923	98.52	1.927	87.61	1.297
[Max,Ave]	97.42	1.844	98.01	1.840	98.61	1.844	87.88	1.214
[Max,Ave,LAP]	97.85	2.841	98.27	2.837	98.80	2.841	87.94	1.898
GL-PF	98.12	1.927	98.34	1.923	99.04	1.927	90.36	1.297

The best accuracy value for each column is highlighted in bold.

classification by comparing the performance of different pooling combination methods on four different datasets. The relevant results are shown in Table III. We used common fusion methods to integrate features, where “+” indicates elementwise addition and “[, ·]” denotes connection along the channel axis. Compared to elementwise addition, connection increases the number of channels for feature mapping, which leads to an increase in network parameters. As can be seen from the table, fusion of pooling layers is effective in improving the classification accuracy compared to single pooling in the four benchmark datasets. Furthermore, elementwise addition-based fusion can obtain better classification performance than connection-based fusion. However, as the number of pooling layers fused increases, elementwise addition-based fusion generates some redundant

information, leading to a decrease in classification accuracy, which is more evident in the Washington DC dataset. In contrast, the proposed GL-PF module strengthens the most important features of the central pixel in the patch and simultaneously integrating the complementary nature of other pooling layers, and achieves the best results on all four benchmark datasets.

D. Comparison of Different Feature Fusion Strategies

The fusion of spectral and spatial features is an important process in the proposed model. An effective fusion strategy can improve the correlation and complementarity of HSI spectral-spatial information. In this section, we compare different fusion strategies (i.e., addition, concatenation, multiplication, and

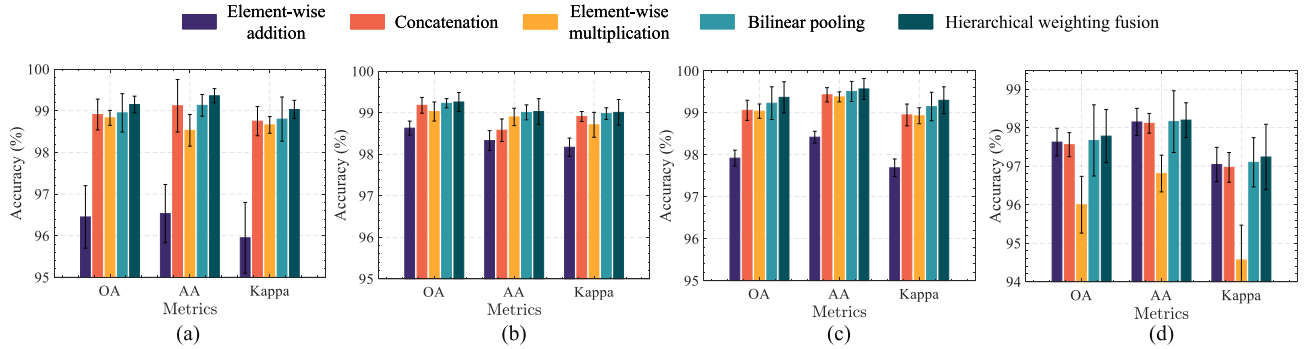


Fig. 15. Performance analysis of the proposed method under different fusion strategies. Error bar represents the standard deviation. (a) Indian Pines dataset. (b) Pavia University dataset. (c) Salinas dataset. (d) Washington DC dataset.

bilinear pooling) to verify the effectiveness of the proposed fusion strategy. Fig. 15 shows the classification indexes (OA, AA, Kappa) and corresponding standard deviation results obtained using different fusion strategies in four datasets. The bilinear pooling-based fusion method is superior to the other three simple linear fusion methods in three different classification indexes. Hierarchical weighting fusion can further improve the correlation of spectral–spatial information and obtain more discriminative spectral–spatial features. In addition, hierarchical weighting fusion has a smaller standard deviation than bilinear pooling, which indicates that the model under the fusion method is more stable. This result is more evident for the Indian Pines and Pavia University.

E. Comparisons With Other Approaches

In this section, we quantitatively analyze and compare the classification results obtained by the proposed GL-BiLSTM, GL-CNN, and GLH-WFN methods with several advanced classification methods. These methods include extended morphological profiles (EMP) [50] represented by traditional methods, and classification methods based on DL such as LDCR [51], 2DCNN, 3DCNN [31], HybridSN [52], spectral–spatial attention networks (SSAN) [36], and spectral–spatial residual network (SSRN) [53].

These comparison methods can be divided into spectral-based, spatial-based, and spectral-spatial-based methods depending on the feature information used for classification. For the classification methods based on spectral level, including LDCR, and our proposed GL-BiLSTM. The LDCR proposed a multitask DL model to learn the compact representation of HSI through end-to-end strategy, and its parameter settings, see [51]. In the classification method based on the spatial level, we compared EMP, 2DCNN, and the proposed GL-CNN. The EMP uses morphology-based methods to extract spatial structure features, and the support vector machines (SVM) based on RBF kernel were used to classify the extracted features. 2DCNN and GL-CNN have a similar architecture, consisting of three convolutional blocks and two FCs. The last FC layer is the softmax layer. Each convolutional block contains a convolutional layer, a

ReLU activation layer, and a max-pooling layer. The specific receptive field size is consistent with GL-CNN. Table VIII presents the architecture and specific parameters of GL-CNN, in addition to the FC layer specific parameters. 3DCNN, HybridSN, SSAN, SSRN, and our proposed GLH-WFN were divided into classification methods based on spectral–spatial level. The parameters in 3DCNN, HybridSN, SSAN, and SSRN set the default values of their references.

The first comparative experiment is to analyze the Indian Pines dataset. Fig. 16 illustrates the results obtained by different classification methods. Due to the lack of spatial information, LDCR and GL-BiLSTM inevitably produces some noise effects. In contrast, other methods that use spatial information achieve superior results in eliminating noise and can obtain more accurate classification maps. Table IV presents the accuracy values obtained by different classification methods. The OA value of our proposed GL-BiLSTM reaches 89.43%. Although it is competitive in the spectral-based classification method, it still has a particular gap with the spatial-based comparison method. In the comparison method based on spatial information, GL-CNN has 4.29% and 1.19% higher OA values than EMP and 2DCNN, respectively. Among the classification methods based on spectral-spatial level, GLH-WFN is the best in terms of OA, AA, and Kappa, and has a lower standard deviation, which indicates that the method is more stable.

The second and third experiments were conducted on the Pavia University and Salinas. Figs. 17 and 18 show the results of different comparison methods of experiments on the two databases. Compared with other comparison methods, the proposed GLH-WFN can more accurately identify the edge contours of the target features and can remove the interference pixels of small samples and obtain a more accurate classification result map on the two datasets. Furthermore, as shown in Tables V and VI, the OA values of the proposed method on Pavia University and Salinas reached 99.26% and 99.37%, which were increased by 0.62% and 1.90% compared with SSRN, respectively. The proposed method has obvious advantages over other classification methods.

The last experiment was conducted on the Washington DC dataset. Table VII shows the quantitative results of the different methods in Washington DC. The proposed GLH-WFN has

TABLE IV
CLASSIFICATION ACCURACY AND RUNNING TIME OF DIFFERENT ALGORITHMS ON THE INDIAN PINES DATASET. THE DISTRIBUTION OF TRAINING SET AND TEST SET ARE SHOWN IN TABLE I

Metrics	Spectral		Spatial			Spectral-Spatial				
	LDCR	GL-BiLSTM	EMP	2DCNN	GL-CNN	3DCNN	HybridSN	SSAN	SSRN	GLH-WFN
OA(%)	85.93 (1.01)	89.43 (0.73)	93.83 (0.28)	96.93 (0.40)	98.12 (0.27)	90.80 (0.78)	97.67 (0.72)	97.22 (0.85)	97.99 (0.80)	99.15 (0.20)
AA(%)	85.31 (1.64)	86.43 (1.31)	90.1 (2.25)	96.73 (0.88)	98.05 (0.58)	87.89 (1.21)	96.59 (1.77)	95.31 (0.77)	98.43 (0.70)	99.36 (0.17)
Kappa(%)	83.94 (1.16)	87.92 (0.85)	92.94 (0.32)	96.49 (0.47)	97.85 (0.31)	89.49 (0.89)	97.33 (0.83)	96.81 (0.98)	97.70 (2.10)	99.03 (0.22)
1	75.12	42.28	77.64	98.78	95.43	52.03	87.80	91.23	100.00	99.73
2	80.39	86.36	87.39	95.58	97.60	87.39	96.98	94.65	94.69	97.35
3	83.10	83.85	94.58	97.61	98.80	90.47	98.40	94.99	96.83	99.64
4	74.21	86.13	90.69	98.98	99.49	82.06	99.38	99.44	99.40	99.49
5	91.63	95.90	95.15	96.87	98.23	92.75	96.85	93.56	97.70	99.37
6	95.78	99.11	97.92	98.09	99.04	97.24	99.15	99.45	98.98	99.93
7	80.80	77.33	66.00	94.00	96.50	64.00	89.78	91.17	100.00	99.11
8	97.67	99.30	99.73	99.42	99.62	99.69	98.94	100.00	98.96	100.00
9	80.00	79.17	66.67	90.63	96.88	95.83	95.14	72.86	100.00	100.00
10	86.70	89.40	87.72	92.96	95.01	87.64	97.62	95.45	98.68	97.55
11	82.66	86.05	95.04	96.69	98.11	89.30	97.02	98.79	98.50	99.31
12	86.39	83.89	92.11	96.69	98.73	89.15	96.21	95.61	98.48	99.35
13	96.52	100.00	98.55	96.20	98.85	98.01	97.10	99.62	100.00	99.52
14	93.03	97.34	98.90	99.52	98.92	96.34	99.26	98.63	99.51	99.80
15	68.40	79.45	97.24	98.47	98.50	86.61	97.72	99.44	98.95	99.83
16	92.60	97.26	96.35	97.26	99.14	97.72	98.02	100.00	94.20	99.39
Training(s)	1259.63	72.38	34.94	45.83	110.32	68.40	403.88	2484.27	412.85	249.70
Testing(s)	0.65	1.23	1.69	0.54	0.89	0.74	3.48	10.52	5.16	2.43

The value in brackets represents the corresponding standard deviation. The best accuracy values base on different levels of methods are highlighted in bold.

TABLE V
CLASSIFICATION ACCURACY AND RUNNING TIME OF DIFFERENT ALGORITHMS ON THE PAVIA UNIVERSITY DATASET, THE DISTRIBUTION OF TRAINING SET, AND TEST SET ARE SHOWN IN TABLE I

Metrics	Spectral		Spatial			Spectral-Spatial				
	LDCR	GL-BiLSTM	EMP	2DCNN	GL-CNN	3DCNN	HybridSN	SSAN	SSRN	GLH-WFN
OA(%)	92.36 (0.13)	95.08 (0.36)	96.35 (0.31)	97.66 (0.35)	98.34 (0.32)	95.28 (0.40)	99.08 (0.17)	97.50 (0.40)	98.64 (0.13)	99.26 (0.23)
AA(%)	91.42 (0.22)	94.25 (0.51)	96.38 (0.45)	96.45 (0.69)	97.56 (0.61)	94.33 (0.05)	98.24 (0.34)	96.76 (0.36)	98.10 (0.30)	99.03 (0.31)
Kappa(%)	89.48 (0.32)	93.45 (0.48)	95.13 (0.42)	96.89 (0.46)	97.79 (0.43)	93.73 (0.52)	98.77 (0.22)	96.68 (0.53)	98.20 (0.76)	99.01 (0.31)
1	91.29	93.21	97.17	96.05	96.68	93.36	99.33	95.77	99.73	98.43
2	96.04	98.33	97.86	99.48	99.48	98.24	99.93	99.29	99.55	99.86
3	87.99	90.63	97.59	95.51	97.66	93.12	98.97	95.75	93.32	98.63
4	93.56	96.66	97.73	96.55	97.35	96.42	96.92	97.70	99.72	99.23
5	99.10	99.46	99.69	95.46	99.48	97.83	99.85	99.16	100.00	100.00
6	86.65	89.77	86.06	97.97	98.95	90.89	99.82	95.67	97.70	98.77
7	90.35	92.01	93.91	95.53	95.12	90.51	99.95	96.23	97.57	97.86
8	77.95	88.66	98.49	94.81	97.08	89.91	95.88	95.01	95.27	98.82
9	99.87	99.53	98.89	96.69	96.19	98.74	93.47	96.34	100.00	99.71
Training(s)	829.16	75.91	27.75	52.32	142.23	155.57	216.06	1381.05	375.28	279.47
Testing(s)	2.52	4.76	7.43	1.98	3.23	1.97	6.93	25.63	9.13	9.90

The value in brackets represents the corresponding standard deviation.

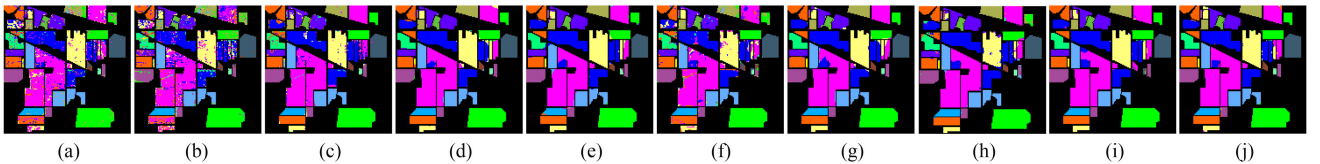


Fig. 16. Classification map obtained by (a) LDCR (85.93%), (b) GL-BiLSTM (89.43%), (c) EMP (93.83%), (d) 2DCNN (96.93%), (e) GL-CNN (98.12%), (f) 3DCNN (90.80%), (g) HybridSN(97.67%), (h) SSAN (97.22%), (i) SSRN (97.99%), and (j) **GLH-WFN(99.15%)** in Indian Pines. Number in brackets represents the OA (in%).

TABLE VI
CLASSIFICATION ACCURACY AND RUNNING TIME OF DIFFERENT ALGORITHMS ON THE SALINAS DATASET, THE DISTRIBUTION OF TRAINING SET AND TEST SET ARE SHOWN IN TABLE I

Metrics	Spectral		Spatial			Spectral-Spatial				
	LDCR	GL-BiLSTM	EMP	2DCNN	GL-CNN	3DCNN	HybridSN	SSAN	SSRN	GLH-WFN
OA(%)	91.47 (0.52)	93.08 (0.43)	97.36 (0.25)	98.19 (0.37)	99.04 (0.59)	92.51 (0.78)	99.03 (0.20)	97.12 (0.85)	97.47 (0.82)	99.37 (0.47)
AA(%)	95.00 (0.55)	96.14 (0.32)	98.05 (0.21)	98.47 (0.35)	99.23 (0.43)	96.08 (1.19)	99.18 (0.39)	98.09 (0.53)	98.83 (0.15)	99.57 (0.35)
Kappa(%)	92.85 (0.44)	92.29 (0.47)	97.06 (0.28)	97.99 (0.42)	98.93 (0.32)	91.65 (0.86)	98.92 (0.23)	96.79 (0.94)	97.18 (0.92)	99.30 (0.52)
1	99.32	98.90	99.83	98.04	99.93	99.10	99.99	93.11	100.00	100.00
2	99.10	99.89	99.39	99.60	97.93	99.85	99.95	99.10	99.99	99.99
3	97.25	98.81	98.83	98.59	99.71	98.80	99.79	97.95	99.47	99.97
4	99.55	99.63	99.36	99.38	99.94	99.11	97.88	99.22	99.05	99.96
5	98.26	98.50	96.78	99.61	99.63	98.26	99.55	99.11	99.96	99.86
6	99.68	99.91	98.81	99.28	99.85	99.75	99.97	99.98	99.99	100.00
7	99.25	99.67	99.48	99.44	99.00	99.61	99.95	99.45	99.98	99.86
8	82.16	86.57	95.93	97.18	98.19	82.16	98.55	93.69	93.07	98.64
9	98.99	99.76	99.51	99.83	99.61	99.69	99.94	99.04	99.78	99.98
10	92.04	95.02	95.87	98.58	99.24	94.25	98.60	99.59	99.32	99.38
11	91.99	93.76	97.30	97.30	97.44	95.74	99.67	98.64	96.78	98.34
12	99.80	99.88	99.60	97.44	99.61	99.37	99.78	99.87	99.84	100.00
13	97.39	98.41	98.42	98.97	99.02	98.58	97.89	98.06	100.00	99.84
14	95.94	96.38	97.37	99.21	99.83	97.24	98.79	99.03	99.83	99.81
15	74.76	75.53	93.12	95.73	98.87	81.38	97.09	93.65	94.24	98.37
16	94.55	97.57	99.15	97.30	99.95	94.34	99.55	99.89	99.97	99.11
Training(s)	748.60	60.68	25.11	40.01	101.05	361.66	83.35	2581.11	459.36	205.67
Testing(s)	3.12	6.75	4.66	2.53	3.82	5.39	5.22	63.41	28.69	11.50

The value in brackets represents the corresponding standard deviation. The best accuracy values base on different levels of methods are highlighted in bold.

TABLE VII
CLASSIFICATION ACCURACY AND RUNNING TIME OF DIFFERENT ALGORITHMS ON THE WASHINGTON DC DATASET. THE DISTRIBUTION OF TRAINING SET AND TEST SET ARE SHOWN IN TABLE I

Metrics	Spectral		Spatial			Spectral-Spatial				
	LDCR	GL-BiLSTM	EMP	2DCNN	GL-CNN	3DCNN	HybridSN	SSAN	SSRN	GLH-WFN
OA(%)	90.76 (0.60)	94.63 (1.15)	92.56 (1.47)	85.31 (1.24)	90.36 (0.50)	92.35 (0.64)	88.32 (2.43)	88.47 (0.67)	92.62 (0.70)	97.78 (0.69)
AA(%)	93.02 (0.53)	96.25 (0.67)	93.47 (1.14)	85.48 (0.84)	90.13 (0.71)	93.78 (0.41)	88.90 (1.34)	90.09 (1.01)	93.06 (0.18)	98.20 (0.45)
Kappa(%)	88.60 (0.73)	93.35 (1.41)	90.74 (1.82)	81.86 (1.47)	88.03 (0.59)	90.52 (0.78)	85.57 (2.90)	85.65 (0.69)	90.71 (0.80)	97.24 (0.85)
1	80.75	87.65	88.47	84.59	90.61	86.68	84.39	82.00	91.86	95.91
2	96.86	98.81	98.74	85.24	92.59	95.86	91.69	95.43	99.26	99.37
3	91.72	94.64	83.43	84.93	87.43	88.91	90.89	78.04	96.80	97.01
4	96.08	98.52	96.38	85.30	90.48	94.89	94.41	94.55	78.70	98.65
5	96.35	98.40	95.98	91.80	95.07	98.95	92.20	91.92	96.07	99.55
6	96.39	99.49	97.85	81.05	84.60	97.38	79.79	98.62	96.05	98.71
Training(s)	1236.44	25.82	3.74	11.79	30.36	63.17	67.86	1146.13	172.3	83.05
Testing(s)	1.63	1.25	1.55	0.57	0.75	0.76	1.68	10.12	4.88	4.36

The value in brackets represents the corresponding standard deviation. The best accuracy values base on different levels of methods are highlighted in bold.

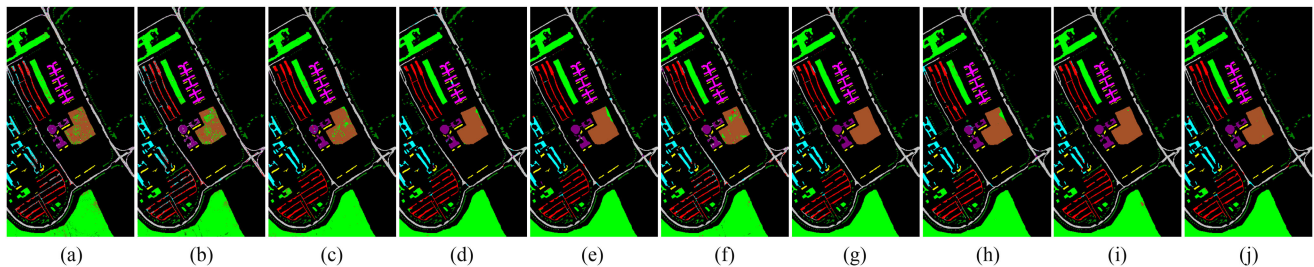


Fig. 17. Classification map obtained by (a) LDCR (92.36%), (b) GL-BiLSTM (95.08%), (c) EMP (96.35%), (d) 2DCNN (97.66%), (e) GL-CNN (98.34%), (f) 3DCNN (92.51%), (g) HybridSN(99.08%), (h) SSAN (97.50%), (i) SSRN (98.64%), and (j) **GLH-WFN (99.26%)** in Pavia University. Number in brackets represents the OA (in%).

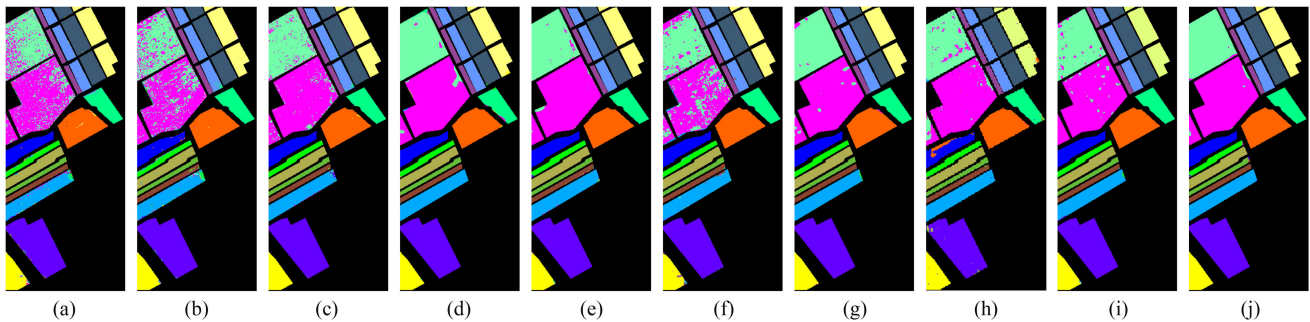


Fig. 18. Classification map obtained by (a) LDCR (91.47%), (b) GL-BiLSTM (93.08%), (c) EMP (97.36%), (d) 2DCNN (98.19%), (e) GL-CNN (99.04%), (f) 3DCNN (92.51%), (g) HybridSN (99.03%), (h) SSAN (97.06%), (i) SSRN (97.47%), and (j) **GLH-WFN (99.37%)** in Salinas. Number in brackets represents the OA (in%).

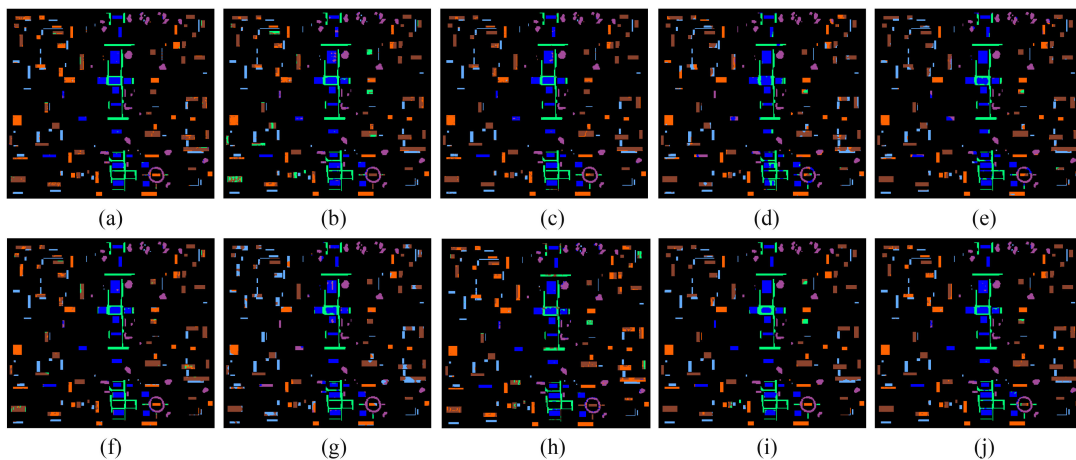


Fig. 19. Classification map obtained by (a) LDCR (90.76%), (b) GL-BiLSTM (94.63%), (c) EMP (92.56%), (d) 2DCNN (85.31%), (e) GL-CNN (90.36%), (f) 3DCNN (92.35%), (g) HybridSN (88.32%), (h) SSAN (88.47%), (i) SSRN (92.62%), and (j) **GLH-WFN (97.78%)** in Washington DC. Number in brackets represents the OA (in%).

TABLE VIII
NETWORK SETTINGS OF GLH-WFN, SPATIAL BRANCHES AND SPECTRAL
BRANCHES, WHERE N IS THE CLASS OF DATASETS

Component	Layer	Size	Activation	Strides	Padding	
Spe	FC	128	ReLU	-	0	
	Conv	L1	$3 \times 3 \times 32$	ReLU	(1,1)	1
		L2	$3 \times 3 \times 64$			
L3		$3 \times 3 \times 128$				
Spa	Max-pooling	2×2	-	(2,2)	0	
	Ave-pooling		-			
	LA-pooling		Sigmoid			
Merge	FC	128	ReLU	-	0	
	FC1	1024	ReLU	-	-	
	FC2	N	Softmax	-	0	

significant advantages over other methods. Moreover, due to the dense distribution of objects in this scene, the spatial-based method classification performs poorly in terms of accuracy compared to the spectral-based method. Excellent algorithms like HybridSN and SSRN have not achieved exciting performance in this scene. Fig. 19 shows the classification maps of the different methods on the Washington DC.

F. Computing Cost

In this experiment, we analyzed the computational cost of different comparison methods. As presented in Tables IV to VII, the training time is positively correlated with the spatial size of the data samples and the spectral depth. The computational cost of SSAN is the highest among all methods due to the SSAN having a deeper spectral branch network, which produces huge training costs. In contrast, our proposed spectral classification model with a grouping strategy greatly reduced the training cost. Compared with SSRN and HybridSN, the computational cost of our method has certain advantages. Moreover, although a large amount of time is consumed in the training stage, the time required in the test stage is greatly reduced.

V. PRACTICAL APPLICATION OF ANALYSIS

In this section, we further verify the practicability of the proposed GLH-WFN method, using the HSI data collected by the GF-5 satellite for the next experiment. The GF-5 is the world's first full-spectrum hyperspectral satellite to achieve comprehensive observations of the atmosphere and land. GF-5 is loaded with two payloads for land observation and four payloads

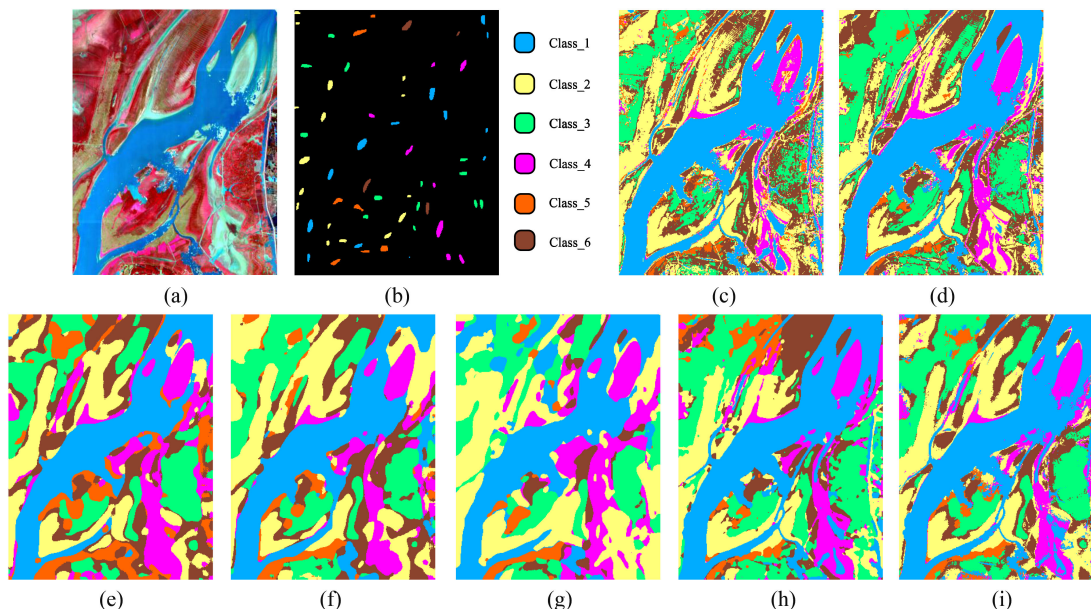


Fig. 20. Classification map for GF-5 dataset. (a) False-color composite map. (b) Groundtruth. (c) GL-BiLSTM (95.42%). (d) EMP (95.57%). (e) 2DCNN (96.03%). (f) GL-CNN (98.02%). (g) HybridSN (97.52%). (h) SSRN (97.90%). (i) **GLH-WFN (98.41%)**. Number in brackets represents the OA (in%).

for atmospheric observation. As the world's first satellite-borne hyperspectral camera that takes into account broad coverage and a wide spectrum, the visible shortwave infrared hyperspectral camera can obtain a wealth of information on ground features [54]–[56]. This article uses the HSI of the DongTing Lake basin acquired by the payload on January 22, 2019, and uses one of the scenes as an example for experiment.

The raw data have been preprocessed, including atmospheric correction and radiation correction. The spectral coverage of this scene is 400–2500 nm, the spatial resolution is 30 m, 20 bad bands are eliminated, 310 spectral bands are reserved for classification, and the image window size is 456×352 pixels. Fig. 20 illustrates the false-color composite map and the ground truth map. The scene is marked with 6 classes and a total of 4816 labeled samples.

In the experiment classifying the GF-5 hyperspectral dataset, we randomly selected 15 label samples per annotated class in the ground truth map as the training set and the remaining samples of each annotated class as the test set. Fig. 20 shows the qualitative classification maps corresponding to different classification methods. From the perspective of OA, our proposed GL-BiLSTM can achieve 95.42%. However, the spectral-based classification method inevitably produces some salt-and-pepper noise that affects the classification result. This phenomenon is evident from the classification map. DL-based spatial extraction methods can significantly improve this phenomenon. Compared with other spatial classification methods, GL-CNN can achieve excellent classification results, but it also inevitably produces oversmoothing phenomenon. In contrast, classification methods based on the fusion of spectral and spatial information have obvious advantages. Compared with the state-of-the-art classification method SSRN and HybridSN, the proposed GLH-WFN can more accurately classify the pixels in the edge area, providing more similar results to the ground truth. In this experiment, our

proposed GLH-WFN is superior to other comparison methods in OA value.

VI. CONCLUSION

This article proposes a GLH-WFN end-to-end framework for hyperspectral classification. The framework consists of GL-BiLSTM for spectral features and GL-CNN for spatial feature extraction. In GL-BiLSTM, the entire spectral bands are redefined in groups from global and local perspectives in order to obtain more discriminative and robust spectral characteristics and to overcome the training difficulties caused by band redundancy. In GL-CNN, we proposed the GL-PF module that considers the correlation and complementarity between different pooling strategies to overcome the influence of interfering information in the patch. Finally, the hierarchical weighting fusion is used to model the high-order information of the extracted spectral and spatial features, and generate a more discriminative global representation. The experimental results on four real datasets and a GF-5 satellite dataset illustrate that the GLH-WFN proposed in this article has better classification performance than other advanced classification methods.

In the future, our work will focus on optimizing the high-dimensional information generated by high-level information modeling in feature fusion. Moreover, due to the limited labeling pixels in HSI, we will try to combine transfer learning and deep networks to solve the HSI classification problem with few or no labeled pixels.

REFERENCES

- [1] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, "Deep pyramidal residual networks for spectral-spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.* vol. 57, no. 2, pp. 740–754, Feb. 2019.

- [2] K. Liu, H. Su, and X. Li, "Estimating high-resolution urban surface temperature using a hyperspectral thermal mixing (HTM) approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 804–815, Feb. 2016.
- [3] J. Guan, H. Shen, X. Li, W. Gan, and L. Zhang, "Climate control on net primary productivity in the complicated mountainous area: A case study of Yunnan, China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4637–4648, Dec. 2018.
- [4] Y. Yu *et al.*, "Accuracy and stability improvement in detecting WuChang rice adulteration by piece-wise multiplicative scatter correction in the hyperspectral imaging system," *Analytical Methods*, vol. 10, pp. 3224–3231, May 2018.
- [5] P. Wu *et al.*, "Spatially continuous and high-resolution land surface temperature product generation: A review of reconstruction and spatiotemporal fusion techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 3, pp. 112–137, Sep. 2021.
- [6] B. Tu, C. Zhou, J. Peng, G. Zhang, and Y. Peng, "Feature extraction via joint adaptive structure density for hyperspectral imagery classification," *IEEE Trans. Instrum. Meas.*, vol. 70, Jan. 2021, Art. no. 5006916.
- [7] L. Fang, C. Wang, S. Li, and J. A. Benediktsson, "Hyperspectral image classification via multiple-feature-based adaptive sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1646–1657, Jul. 2017.
- [8] J. M. Haut, S. Bernabé, M. E. Paoletti, R. Fernandez-Beltran, A. Plaza, and J. Plaza, "Low-high-power consumption architectures for deep-learning models applied to hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 776–780, May 2019.
- [9] B. Tu, C. Zhou, X. Liao, Q. Li, and Y. Peng, "Feature extraction via 3-D block characteristics sharing for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10503–10518, Dec. 2020.
- [10] A. W. Bitar, L. Cheong, and J. Ovarlez, "Sparse and low-rank matrix decomposition for automatic target detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5239–5251, Aug. 2019.
- [11] B. Tu, X. Yang, C. Zhou, D. He, and A. Plaza, "Hyperspectral anomaly detection using dual window density," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8503–8517, Dec. 2020.
- [12] Y. Zhang, K. Wu, B. Du, and X. Hu, "Multitask learning-based reliability analysis for hyperspectral target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2135–2147, Jul. 2019.
- [13] L. Zhang, W. Wei, C. Bai, Y. Gao, and Y. Zhang, "Exploiting clustering manifold structure for hyperspectral imagery super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5969–5982, Dec. 2018.
- [14] W. Xie, X. Jia, Y. Li, and J. Lei, "Hyperspectral image super-resolution using deep feature matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6055–6067, Aug. 2019.
- [15] J. Li, Q. Yuan, H. Shen, X. Meng, and L. Zhang, "Hyperspectral image super-resolution by spectral mixture analysis and spatial-spectral group sparsity," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1250–1254, Sep. 2016.
- [16] P. Duan, X. Kang, S. Li, P. Ghamisi, and J. A. Benediktsson, "Fusion of multiple edge-preserving operations for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10336–10349, Dec. 2019.
- [17] C. Zhou, B. Tu, N. Li, W. He, and A. J. Plaza, "Structure-aware multi-kernel learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9837–9854, Sep. 2021.
- [18] Y. Xu, L. Zhang, B. Du, and F. Zhang, "Spectral-spatial unified networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5893–5909, Oct. 2018.
- [19] L. Zhang, M. Lan, J. Zhang, and D. Tao, "Stagewise unsupervised domain adaptation with adversarial self-training for road segmentation of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2021.3104032](https://doi.org/10.1109/TGRS.2021.3104032).
- [20] C. Zhou, B. Tu, Q. Ren, and S. Chen, "Spatial peak-aware collaborative representation for hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: [10.1109/LGRS.2021.3083416](https://doi.org/10.1109/LGRS.2021.3083416).
- [21] G. Cheng, X. Xie, J. Han, L. Guo, and G. S. Xia, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3735–3756, Jun. 2020.
- [22] X. Kang, C. Li, S. Li, and H. Lin, "Classification of hyperspectral images by Gabor filtering based deep network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1166–1178, Apr. 2018.
- [23] 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.
- [24] P. Zhong, Z. Gong, S. Li, and C. Schönlieb, "Learning to diversify deep belief networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3516–3530, Jun. 2017.
- [25] H. Zhang, Z. Wang, and D. Liu, "A comprehensive review of stability analysis of continuous-time recurrent neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 7, pp. 1229–1262, Jul. 2014.
- [26] K. Greff, R.K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space Odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
- [27] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.
- [28] H. Lee and H. Kwon, "Going deeper with contextual CNN for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4843–4855, Oct. 2017.
- [29] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, pp. 1–12, Jul. 2015.
- [30] J. Xie, N. He, L. Fang, and P. Ghamisi, "Multiscale densely-connected fusion networks for hyperspectral images classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 1, pp. 246–259, Jan. 2021.
- [31] Y. Li, H. Zhang, and Q. Shen, "Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, May 2017, Art. no. 67.
- [32] S. Ghaderizadeh, D. Abbasi-Moghadam, A. Sharifi, N. Zhao, and A. Tariq, "Hyperspectral image classification using a hybrid 3D-2D convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7570–7588, Jul. 2021.
- [33] J. Feng *et al.*, "Deep reinforcement learning for semisupervised hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–19, 2022, Art no. 5501719, doi: [10.1109/TGRS.2021.3049372](https://doi.org/10.1109/TGRS.2021.3049372).
- [34] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, "Learning and transferring deep joint spectral-spatial features for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4729–4742, Aug. 2017.
- [35] 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.
- [36] X. Mei *et al.*, "Spectral-spatial attention networks for hyperspectral image classification," *Remote Sens.*, vol. 11, no. 8, Apr. 2019, Art. no. 963.
- [37] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5384–5394, Aug. 2019.
- [38] J. Carreira, R. Caseiro, J. Batista, and C. Sminchisescu, "Free-form region description with second-order pooling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1177–1189, Jun. 2015.
- [39] P. Li, J. Xie, Q. Wang, and W. Zuo, "Is second-order information helpful for large-scale visual recognition?" in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2089–2097.
- [40] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," 2015, *arXiv:1508.01991*. [Online]. Available: <https://arxiv.org/abs/1508.01991>
- [41] M. D. Zeiler and R. Fergus, "Stochastic pooling for regularization of deep convolutional neural networks," in *Proc. Int. Conf. Learn. Representations*, 2013, pp. 1–9.
- [42] D. Yu, H. Wang, P. Chen, and Z. Wei, "Mixed pooling for convolutional neural networks," in *Proc. Rough Sets Knowl. Technol.*, 2014, pp. 364–375.
- [43] F. Saeedan, N. Weber, M. Goesele, and S. Roth, "Detail-preserving pooling in deep networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 9108–9116.
- [44] Q. Hou, L. Zhang, M.-M. Cheng, and J. Feng, "Strip pooling: Rethinking spatial pooling for scene parsing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 4002–4011.
- [45] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
- [46] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN models for fine-grained visual recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1449–1457.

- [47] Z. Xue, M. Zhang, Y. Liu, and P. Du, "Attention-based second-order pooling network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9600–9615, Nov. 2021.
- [48] Z. Gao, L. Wang, and G. Wu, "LIP: Local importance-based pooling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3354–3363.
- [49] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.
- [50] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
- [51] L. Zhang, J. Zhang, W. Wei, and Y. Zhang, "Learning discriminative compact representation for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8276–8289, Oct. 2019.
- [52] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 277–281, Feb. 2020.
- [53] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.
- [54] H. Ren, X. Ye, R. Liu, J. Dong, and Q. Qin, "Improving land surface temperature and emissivity retrieval from the Chinese Gaofen-5 satellite using a hybrid algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 1080–1090, Feb. 2018.
- [55] B. Tang, "Nonlinear split-window algorithms for estimating land and sea surface temperatures from simulated Chinese Gaofen-5 satellite data," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6280–6289, Nov. 2018.
- [56] X. Ye, H. Ren, R. Liu, Q. Qin, Y. Liu, and J. Dong, "Land surface temperature estimate from Chinese Gaofen-5 satellite data using split-window algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5877–5888, Oct. 2017.



Bing Tu (Member, IEEE) received the M.S. degree in control science and engineering from the Guilin University of Technology, Guilin, China, in 2009, and the Ph.D. degree in mechatronic engineering from the Beijing University of Technology, Beijing, China, in 2013.

From 2015 to 2016, he was a Visiting Researcher with the Department of Computer Science and Engineering, University of Nevada, Reno, NV, USA, which is supported by the China Scholarship Council. Since 2018, he has been an Associate Professor

with the School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang, China. His research interests include sparse representation, pattern recognition, and analysis in remote sensing.



Wangquan He (Student Member, IEEE) received the B.S. degree in communication engineering, in 2019, from the Hunan Institute of Science and Technology, Yueyang, China, where he is currently working toward the M.S. degree in information and communication engineering.

His research interests include pattern recognition, hyperspectral image classification, remote sensing image scene classification, and noisy label detection.



Wei He (Member, IEEE) received the B.S. degree in computer science and technology from Hunan Institute of Science and Technology, Yueyang, China, the M.S. degree in computer software and theory from Changsha University of Science and Technology, Changsha, China, and the Ph.D. degree in information and communication Engineering from Hoseo University, Asan, South Korea.

He was a Visiting Researcher with the Computer Vision Lab, University of Nevada, Reno, NV, USA, from August 2015 to 2016. He is currently an Associate Professor with the School of Information Science and Engineering, Hunan Institute of Science and Technology. His research interests include computer vision, pattern recognition, and video surveillance applications.



XianFeng Ou (Member, IEEE) received the B.S. degree in electronic information science and technology and the M.S. degree in communication and information system from Xinjiang University, Urumchi, China, in 2006 and 2009, respectively, and the Ph.D. degree in communication and information system from Sichuan University, Chengdu, China, in 2015.

He was a Visiting Researcher with the Internet Media Group, Polytechnic di Torino, Turin, Italy, from January to April 2014, working on distributed video coding and transmission. His research interests

include image and video coding process technologies, object detection, and hyperspectral image classification.



Antonio Plaza (Fellow, IEEE) received the M.Sc. and the Ph.D. degrees in computer engineering from the Hyperspectral Computing Laboratory, the 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. He has authored more than 600 publications, including over 200 JCR journal papers (over

160 in IEEE journals), 23 book chapters, and around 300 peer-reviewed conference proceeding papers. He has reviewed more than 500 manuscripts for over 50 different journals. He has guest-edited ten special issues on hyperspectral remote sensing for different journals. His research interests include hyperspectral data processing and parallel computing of remote sensing data.

Dr. Plaza was the recipient of the Best Reviewer of IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, in 2009, and the Best Reviewer of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, in 2010. He was also the recipient of the Best Column Award of *IEEE Signal Processing Magazine*, in 2015, the 2013 Best Paper Award for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING journal, and the Most Highly Cited Paper, from 2005 to 2010, in *Journal of Parallel and Distributed Computing*, the best paper awards at the IEEE International Conference on Space Technology, the IEEE Symposium on Signal Processing and Information Technology, and the recognition as an Outstanding Associate Editor for IEEE ACCESS, in 2017. He has served as an Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, from 2007 to 2012. He has served as the Editor-in-Chief of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, from 2013 to 2017. He is also an Associate Editor for the IEEE ACCESS. He was a member of the Editorial Board of IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER, from 2011 to 2012, and *IEEE Geoscience and Remote Sensing Magazine*, in 2013. He was also a member of the Steering Committee of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. He has served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS), from 2011 to 2012, and as the President of the Spanish Chapter of IEEE GRSS, from 2012 to 2016.