DF2Net: Differential Feature Fusion Network for Hyperspectral Image Classification

Qingwang Wang, *Member, IEEE*, Jiangbo Huang, *Graduate Student Member, IEEE*, Yuanqin Meng, and Tao Shen, *Member, IEEE*

***Abstract*—Recently, hybrid networks, combining graph convolutional networks (GCNs) and convolutional neural networks into a unified framework, have garnered significant attention in hyperspectral image (HSI) classification. However, existing hybrid networks have the following limitations. 1) Existing methods primarily utilize simple fusion strategies such as concatenation or direct addition, resulting in the ineffective utilization of advantageous features. 2) Traditional GCNs only consider the relationship between pairs of vertices, limiting their ability to capture complex high-order and long-range correlations. In this work, a novel differential feature fusion network (DF2Net) is proposed for HSI classification. Specifically, DF2Net utilizes two subnetworks to learn features at different abstraction levels: 1) the spectral–spatial hypergraph convolutional network for capturing complex high-order and long-range correlations, and the spectral–spatial convolution network for pixel-level local information extraction. Subsequently, we introduce an advantageous feature differential enhancement fusion module, in which mutual enhancement of advantageous features from different network structures is performed, thereby improving the classification robustness of different regions in HSI. The experiments on four HSI benchmark datasets demonstrate that our DF2Net exhibits superior advantages over state-of-the-art models, particularly when the training samples are limited.**

***Index Terms*—Advantageous feature differential enhancement fusion (AFDEF), convolutional neural networks (CNNs), hypergraph convolutional network, hyperspectral image (HSI) classification.**

I. Introduction

HYPERSPECTRAL images (HSIs) generally consist of dozens or even hundreds of contiguous spectral bands and can provide rich high-resolution spectral information on land covers. With this distinctive advantage, HSIs have found extensive applications in various fields, such as land management [1], agriculture assessment [2], and environmental protection [3], [4].

Manuscript received 29 March 2024; revised 9 May 2024; accepted 13 May 2024. Date of publication 21 May 2024; date of current version 5 June 2024. This work was supported in part by the Yunnan Fundamental Research Projects under Grant 202101BE070001-008, Grant 202401AW070019, and Grant 202301AV070003, in part by the Youth Project of the National Natural Science Foundation of China under Grant 62201237, in part by the Major Science and Technology Projects in Yunnan Province under Grant 202202AD080013 and Grant 202302AG050009, in part by the Youth Project of the Xingdian Talent Support Plan of Yunnan Province under Grant KKRD202203068. *(Corresponding author: Tao Shen.)*

The authors are with the Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650093, China (e-mail: wangqingwang@kust.edu.cn; jiangbohuang@stu.kust.edu.cn; mengyuanqin@stu.kust.edu.cn; shentao@kust.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3403863

HSI classification aims to classify each pixel with specific labels by analyzing spectral–spatial information of land covers. At first, the combination of spectral feature extraction and machine learning algorithms was primarily utilized for HSI classification. Commonly employed spectral feature extraction techniques encompass principal component analysis [5], linear discriminant analysis [6], wavelet transform [7], and multiple kernel learning [8]. These techniques reduce the data dimension and extract the most representative spectral information within the HSI data. Subsequently, extracted features are subjected to training and classification using machine learning methods, such as support vector machines (SVM) [9], [10] and random forest (RF) [11]. However, these methods require manual selection and design of feature extraction methods. In addition, the richness of spectral information also makes it difficult for these methods to obtain accurate results.

In recent years, many deep learning methods [12] have been proposed for HSI classification, such as convolutional neural networks (CNNs) [13], [14], [15], graph convolutional networks (GCNs) [16], [17], and recurrent neural networks (RNNs) [18]. They allow extracting deeper and more discriminative features from HSI through a series of hierarchical layers. Among them, CNNs are able to efficiently extract pixel-level local information by applying learnable filters. For example, in [13], a simple 1-D CNN, consisting of only one convolutional layer, achieves superior classification performance compared to traditional methods, such as SVM. Subsequently, many CNN variants, including 2DCNN [19], 3DCNN [20], [21], and hybrid spectral CNN (HybridSN) [22], have been proposed for HSI classification, which has more powerful feature extraction capability. CNN-based models have unique advantages in extracting pixel-level local information, exhibiting superior classification performance in fine areas. However, its receptive field is influenced by the size of convolutional kernels, resulting in the long-range relations cannot be adequately captured. In addition, labeled pixels in HSI are typically rare and expensive to obtain. CNN-based methods are prone to overfitting when the training sample is limited.

Graph CNNs (GCNs) [23], [24] offer a novel approach to extract deep features in HSI. On the one hand, GCNs combine the graph structure and node features to achieve feature aggregation in non-Euclidean space, which makes GCN naturally suitable for extracting long-range features. On the other hand, GCNs can fully utilize unlabeled samples through the graph structure, effectively alleviating the burden on labeled samples. Based on the success of GCN, many variants of GCN have

been proposed for HSI classification. Qin et al. [25] proposed the spectral–spatial GCN (S2GCN), which integrates spectral information and local spatial window information and is the first GCN model designed for HSI classification. Since considering each pixel point of an HSI as a node of the graph leads to a large computational load, limiting its applicability. Wan et al. [26] proposed the multiscale dynamic GCN (MDGCN), which uses superpixel nodes instead of pixel nodes, and proposed to utilize a dynamic graph structure representing the relationships between superpixel nodes. In addition, superpixels can reflect features such as the shape and patterns of land covers, giving GCNs a significant advantage in classifying smooth regions. However, due to the loss of pixel-level spectral–spatial information during the superpixel segmentation process, superpixel-based GCNs often exhibit poorer classification performance in fine regions.

CNNs and GCNs both demonstrate powerful feature extraction capabilities. However, it can be observed that CNNs and GCNs perform well in different regions: CNNs exhibit better classification performance in fine regions while GCNs perform better in smooth regions. Single-network structure, whether CNN or GCN, captures limited information. Therefore, to achieve robust classification of different regions in HSI, it is necessary to design a new hybrid network, which can extract and integrate the advantageous features of different network structures.

In recent years, many hybrid networks [27], [28], [29], [30], [31] have been proposed and garnered significant attention in HSI classification. Among them, the CNN-enhanced GCN (CEGCN) [27] is the first to achieve feature interaction and fusion between CNN and GCN structure. Subsequently, the weighted feature fusion network (WFCG) [28] uses GAT to enhance the aggregation of features and combines the channel attention mechanism and the spatial attention mechanism to design the branches of the CNN network. Recently, the attention multihop graph and multiscale convolutional fusion network (AMGCFN) [29] uses multiscale fully CNN and multihop GCN to extract the multilevel information of HSI. However, these methods have the following limitations. 1) Existing methods, such as CEGCN and WFCG, primarily utilize simple fusion strategies such as concatenation or direct addition, resulting in the ineffective utilization of advantageous features. 2) Traditional GCNs only consider the relationship between pairs of vertices, limiting their ability to capture complex high-order correlations and long-range correlations. Deep GCNs enhance the representation capability of complex high-order and long-range correlations by increasing the network depth. However, they may suffer from the issue of feature oversmoothing. Yu et al. [31] proposed a synergetic interaction network combined with convolution and transformer (HyperSINet), which replaces GCN with a transformer to enhance modeling capabilities for long-range correlations. However, the transformer requires significant computational resources and lacks interpretability.

To cope with the previous problems, a novel differential feature fusion network (DF2Net) is proposed for HSI classification. The DF2Net consists of three subnetworks: 1) the spectral–spatial hypergraph convolutional network (S2HGCN), 2) the spectral–spatial convolution network (S2CN), and 3) an advantageous feature differential enhancement fusion (AFDEF) module. Specifically, the S2HGCN subnetwork utilizes spectral and spatial hypergraphs to capture complex high-order and long-range correlations contained in HSI. Compared to simple graphs, hypergraphs have stronger representation capabilities. The S2CN subnetwork utilizes spectral–spatial convolution (SSConv) for pixel-level local information extraction. Finally, through the AFDEF module, mutual enhancement of advantageous features from different network structures is performed, thereby improving the classification robustness of different regions in HSI. The main contributions of this article are as follows.

1) We systematically analyze the performance of different network structures, including CNN, GCN, and HGCN structures, for HSI classification. The experimental results indicate that different network structures can extract complementary advantageous features.
2) Aiming at improving the ability to capture complex high-order and long-range correlations contained in HSI, we extend the simple graph to the hypergraph and design the spectral–spatial fusion hypergraph convolutional network S2HGCN as a nonlocal feature extraction subnetwork.
3) A novel advantageous feature enhancement hybrid network DF2Net is proposed. DF2Net utilizes the AFDEF module to achieve mutual enhancement of advantageous features derived from different network structures, thereby improving the classification robustness of different regions in HSI.

The rest of the article is organized as follows. In Section II, we systematically analyze the performance of different network structures for HSI classification. Section III provides a detailed introduction to our proposed DF2Net. The experiments are illustrated in Section IV. The discussion and analyses are performed in Section V. Finally, Section VI concludes this article.

II. Different Network Structures for HSI Classification

In this section, we compare the performance of different network structures for HSI classification. Specifically, two sets of comparative experiments are performed: 1) comparing the CNN (based on regular convolutional) with GCN (based on graph convolutional) structures, and 2) comparing the GCN (based on the simple graph) with HGCN (based on the hypergraph) structures. The dataset used for the experiments is the University of Pavia, which will be introduced in detail in Section IV.

A. Comparing the CNN With GCN Structures for HSI Classification.

We initially designed a classification network based on CNN, which consists of two layers of regular convolution. The kernel size for the first layer is set to 5×5 and the second layer is set to 3×3. Similarly, we design a classification network based on GCN, incorporating two layers of graph convolution. The dimension of the hidden layers in both CNN and GCN is set to 128. The overall accuracy (OA) under different training label ratios is shown in Fig. 1. Furthermore, we visualize the

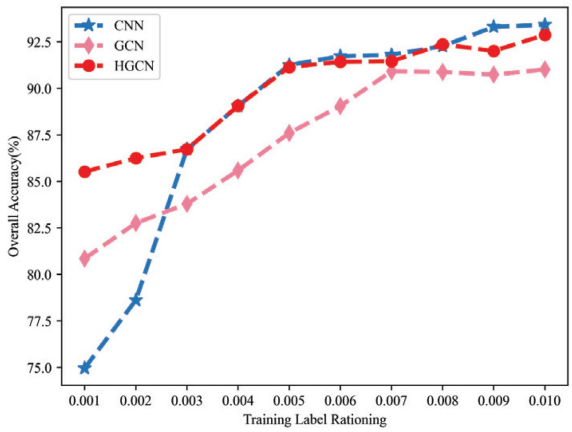

Fig. 1. OA of different network structures in the University of Pavia dataset classification.

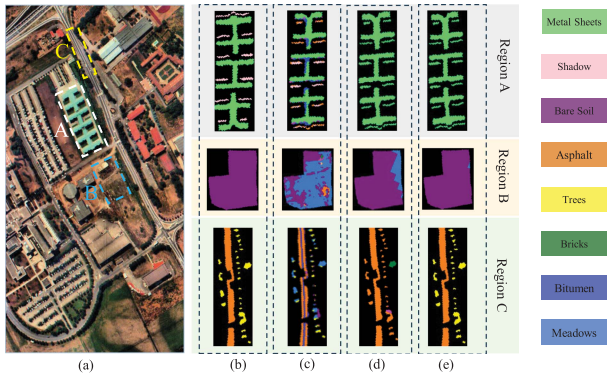

Fig. 2. Classification result maps of different network structures for three representative regions in the University of Pavia dataset classification. The CNN structure achieves better classification results in fine regions while GCN and HGCN structures perform better in smooth regions. (a) HSI. (b) GT. (c) CNN. (d) GCN. (e) HGCN.

classification maps of some representative regions in Fig. 2, where region A covers fine areas such as shadows, and region B covers smooth areas such as bare soil.

From the earlier data, we can observe that GCN achieves 80.86% OA when the number of training label ratio is only 0.1%. However, the CNN has only 74.97% OA under the constraint of limited training samples. Subsequently, with the increasing number of training samples, the OA of CNN grows rapidly while GCN remained relatively steady. When the training label ratio reaches 0.1%, CNN achieves 93.14% OA, improved by 2% over GCN.

The results above validate that the CNN network structure exhibits a strong dependence on training samples. With sufficient training samples, the CNN structure model can have a good classification performance, but its performance deteriorates significantly with a reduction in training samples. However, the scarcity of training samples is a common and inevitable challenge in HSI classification. In contrast, GCN exhibits good classification performance even with a limited number of samples. This is likely attributed to the fact that GCNs can utilize unlabeled samples through the graph structure.

From Fig. 2, we can observe that the CNN structure performs better in the classification of shadows in region A. This is

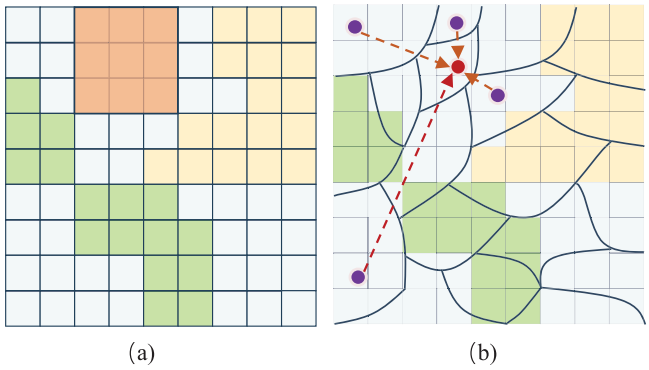

Fig. 3. Illustrations of the working principles of different network structures. (a) CNN can efficiently extract pixel-level local information by applying learnable filters (orange square). (b) GCN utilizes the graph structure based on superpixel node and graph convolution to learn superpixel-level high-order features. (a) CNN: Regular convolution. (b) GCN: Graph convolution.

attributed to the CNN structure can extract pixel-level spectral–spatial information through local convolution, enabling better classification performance in fine areas, as shown in Fig. 3(a). For Region B, we observe that GCN exhibits better classification performance in smooth areas than CNN. This is primarily due to the strong representation capability of the graph structure for nonlocal features, giving the GCN structure a significant advantage in classifying smooth regions, as shown in Fig. 3(b).

B. *Comparing the GCN With HGCN Structures for HSI Classification*

In this section, we compare the classification performance of GCN and HGCN structures on the University of Pavia dataset. They are based on simple graphs and hypergraphs, respectively, aiming to validate whether the hypergraph provides a more powerful representation than the simple graph in specific HSI classification. The implementation details for GCN are consistent with the previous. Simultaneously, we design a simple classification network HGCN based on the hypergraph. HGCN contains two layers of hypergraph convolution and the dimension of the hidden layer is set to 128 as in GCN.

As shown in Fig. 1, HGCN has better classification performance than GCN under different number of training samples. Notably, the improvement is more pronounced when the training samples are limited. For instance, when the training label ratio is 1%, HGCN exhibits an improvement of approximately 1% over GCN. When the training label ratio is 0.1%, HGCN has an improvement of approximately 4%. This is primarily because the hypergraph has a stronger representation capability than the graph and better utilizes unlabeled sample data. As shown in Fig. 4, the simple edge in the simple graph can only connect two nodes while the hyperedge in the hypergraph can connect any number of nodes, providing a more natural way to model complex correlations. From Fig. 2, we can also observe that HGCN has better classification performance in the classification of regions A and C than GCN. This phenomenon validates the more powerful representation of the hypergraph over the simple graph in capturing long-range and complex high-order correlations contained in HSI.

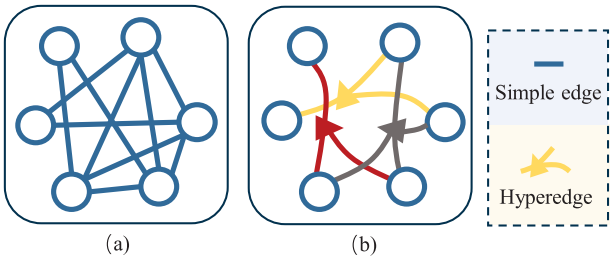

Fig. 4. Illustrations of the simple graph and the hypergraph. (a) Simple graphs tend to focus on relationships between pairs of vertices. (b) Hypergraph is the extension of the simple graph that can represent relationships between any number of nodes. (a) Simple graph. (b) Hypergraph.

In summary, models with different structures exhibit distinct performances in the classification of different regions. The CNN structure (based on regular convolution) shows better classification performance in fine regions, whereas the GCN (based on graph convolution) structure performs better in the classification of smooth areas. Therefore, it is essential to integrate models with different structures to enhance the robustness of the model across different regions. Indeed, many hybrid models [27], [28], [29], [30], [31] have recently been proposed for HSI classification and achieve commendable classification performance. However, the effective fusion of complementary advantageous information has been overlooked. Existing methods primarily employ simple fusion strategies, such as linear combination or concatenation, resulting in the extracted complementary features not being fully exploited and utilized. Therefore, it is necessary to propose a more effective fusion strategy, which can efficiently leverage the complementary advantages of features. In addition, conventional graphs, focusing on pairwise vertex relationships, have limitations in modeling intricate land cover distributions. From the previous experimental results, it is also necessary to extend GCN into HGCN and construct a hybrid CNN and HGCN network for HSI classification.

III. Proposed DF2Net

The overall structure of our proposed DF2Net is shown in Fig. 5. We first perform feature extraction and transformation (FET) of HSI to fit the inputs of different structure models. Subsequently, the extracted features are processed through two subnetworks, 1) S2HGCN and 2) S2CN, to extract complementary advantages features. Finally, the AFDEF module is utilized to enhance advantageous features, achieving robust classification across different regions. A detailed description of the main modules of the DF2Net is provided ahead.

A. Feature Extraction and Transformation

We denote spectral features of HSI as $X_{\text{spe}} \in \mathbb{R}^{M\times L\times B}$, where M, L, and B denote the height, width, and number of spectral bands, respectively. Then, we extract spatial features $X_{\text{spa}} \in \mathbb{R}^{M\times L\times 2}$ through pixel coordinates, denoted as

$$X_{\text{spa}}(i,j) = [i,j] \tag{1}$$

where i and j represent the horizontal and vertical coordinates, respectively. The spectral features and spatial features are normalized separately and then concatenated together to obtain the spectral–spatial feature $X \in \mathbb{R}^{M\times L\times (B+2)}$, denoted as

$$X = X_{\text{spe}} \parallel X_{\text{spa}} \tag{2}$$

the spectral–spatial feature X is directly passed to the S2CN subnetwork for pixel-level local information extraction.

Considering the large number of pixels in HSI, treating each pixel as a node in a hypergraph results in a considerable computational burden. Therefore, we employ the simple linear iterative clustering [27] algorithm to perform superpixel segmentation on the HSI data, replacing pixel nodes with superpixel nodes. Specifically, we obtain a transformation matrix $Q \in \mathbb{R}^{ML\times N}$ where N denotes the number of superpixels that can map the associations between pixels and superpixels [27], denoted as

$$Q_{i,j} = \begin{cases} 1, & \text{if } \widetilde{X}_i \in S_j \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where $\widetilde{X} = \text{Flatten(X)}$, represents the flattened HSI by the spatial dimension, $\widetilde{X}_i$ is the ith pixel in $\widetilde{X}$ and S_j is the jth superpixel. Therefore, $Q_{i,j}$ can reflect the map of the associations between pixel $\widetilde{X}_i$ and superpixel S_j. Finally, we can achieve pixel-level to superpixel-level feature conversion using the following formula:

$$V = \text{Encoder(X;}Q) = \widehat{Q}^T \widetilde{X} \tag{4}$$

where $\widehat{Q}$ denotes the normalized Q by column. V is the superpixel node, and its feature is the average value of all pixel points. V serves as the input to the S2HGCN subnetwork. After the features V undergo hypergraph convolution, we can similarly achieve the feature transformation from superpixels to pixels using the association matrix Q, denoted as

$$X^* = \text{Decoder(V;}Q) = QV. \tag{5}$$

B. Spectral–Spatial Hypergraph Convolutional Network

1) Hypergraph construction for HSI: The hypergraph is a generalization of the graph, enabling a more intricate representation of relationships. In previous works [32], [33], [34], [35], [36], [37], the hypergraph is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$, where $\mathcal{V}$ is the set of vertices, the $\mathcal{E}$ is the set of hyperedge, and $\mathcal{W}$ is the hyperedge weight matrix. For a hypergraph $\mathcal{G}$, it is typically described by an incidence matrix $\mathcal{H}$. Specifically, when given a vertex $v \in \mathcal{V}$ and a hyperedge $e \in \mathcal{E}$, the incidence matrix $\mathcal{H}$ is defined as

$$\mathcal{H}(v,e) = \begin{cases} 0, & \text{if } v \in e \\ 1, & \text{otherwise} \end{cases} \tag{6}$$

furthermore, the vertex degree is defined as $D_v(v) = \sum_{e\in\varepsilon} W(e)\mathcal{H}(v,e)$ and edge degree is defined as $D_e(e) = \sum_{v\in\mathcal{V}} \mathcal{H}(v,e)$. For superpixel node v_i, we find its K nearest neighbors $N_{\text{spe}}(v_i)$ and $N_{\text{spa}}(v_i)$ form the spectral hyperedge and spatial hyperedge, respectively,

$$\begin{aligned} e_{\text{spe}}(v_i) &= v_i \cup N_{\text{spe}}(v_i) \\ e_{\text{spa}}(v_i) &= v_i \cup N_{\text{spa}}(v_i) \end{aligned} \tag{7}$$

then, all spectral and spatial hyperedges form the spectral and spatial hypergraph, respectively,

$$\mathcal{H}_{\text{spe}} = e_{\text{spe}}(v_1)||\cdots||e_{\text{spe}}(v_N)$$

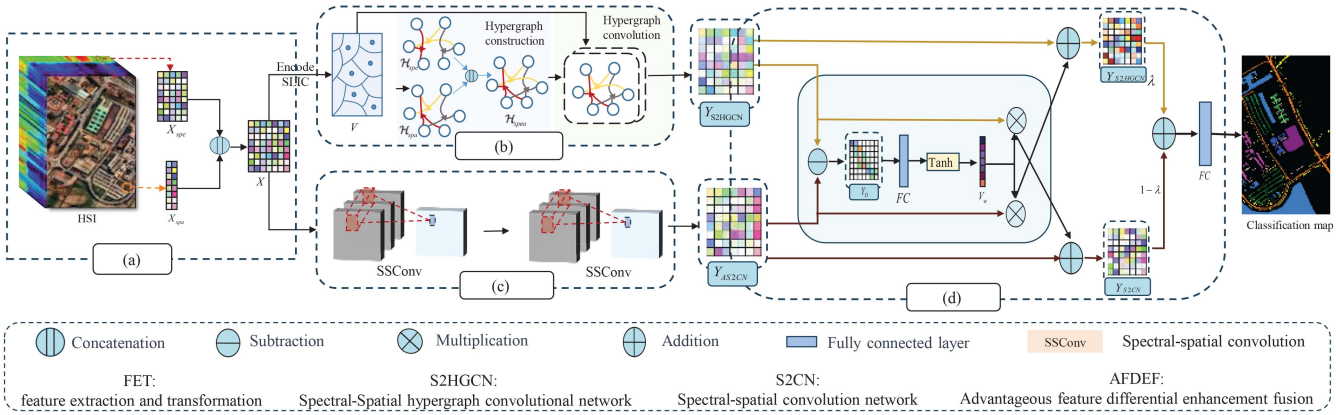

Fig. 5. Flowchart of the proposed DF2Net. DF2Net mainly includes four parts. (a) Feature extraction and transformation module FET. (b) S2HGCN for capturing long-range correlations and complex high-order correlations. (c) S2CN for pixel-level local information extraction. (d) Differential feature enhancement fusion module AFDEF for mutual enhancement of advantageous features derived from different network structures. (a) FET. (b) S2HGCN. (c) S2CN. (d) AFDEF.

$$\mathcal{H}_{\text{spa}} = e_{\text{spa}}(v_1)||\cdots||e_{\text{spa}}(v_N) \tag{8}$$

where $||$ represents concatenation operation. Finally, the spectral and spatial hypergraph are fused to obtain the hypergraph representation of HSI $\mathcal{H}_{\text{spea}}$

$$\mathcal{H}_{\text{spea}} = \mathcal{H}_{\text{spe}}||\mathcal{H}_{\text{spa}} \tag{9}$$

2) Hypergraph Convolution: convolution operation on the hypergraph is similar to graph convolution [32]. After obtaining the superpixel node features V and the hypergraph representation of HSI $\mathcal{H}_{\text{spea}}$, the convolution operation on the hypergraph can be formulated by

$$Y = D_v^{-\frac{1}{2}}\mathcal{H}_{\text{spea}}WD_e^{-1}{\mathcal{H}_{\text{spea}}}^T D_v^{-\frac{1}{2}}V\Theta \tag{10}$$

where D_v and D_e denote, respectively, the diagonal matrix of the vertex degrees and edge degrees of $\mathcal{H}_{\text{spea}}$, W is the hyperedge weight matrix, and Θ is the parameter to be learned. In this study, we use two-layer hypergraph convolution to realize the mining of higher-order correlations contained in HSI, formulated by

$$\widehat{Y} = \sigma\left(T\sigma\left(TV\Theta^{(1)}\right)\Theta^{(2)}\right) \tag{11}$$

where $T = D_v^{-\frac{1}{2}}\mathcal{H}_{\text{spea}}WD_e^{-1}{\mathcal{H}_{\text{spea}}}^T D_v^{-\frac{1}{2}}$, σ is nonlinear activation function, and $\Theta^{(1)}$ with $\Theta^{(2)}$ are the learnable weight parameters. Y is the output signal. To enhance the stability of the hypergraph convolution, Y is processed through a fully connected layer

$$\tilde{Y}_{\text{S2HGCN}} = \sigma\left(W_{\text{fc}}\widehat{Y} + b_{\text{fc}}\right) \tag{12}$$

where W_{fc} is the weight matrix and b_{fc} is the bias term. Finally, we can similarly achieve the feature transformation from superpixels to pixels using the association matrix Q, formulated by

$$Y_{\text{S2HGCN}} = \text{Decoder}\left(\widetilde{\text{Y}}_{\text{S2HGCN}};Q\right) = Q\widetilde{\text{Y}}_{\text{S2HGCN}}. \tag{13}$$

C. Spectral–Spatial Convolution Network

In the S2CN network, we utilize SSConv [27], [38] for pixel-level local information extraction. Compared to the traditional 2-D CNN structure, SSConv can greatly reduce the parameters and enhance the robustness to overfitting. Specifically, SSConv consists of two independent convolutional layers: 1) spectral dimension convolution and 2) spatial dimension convolution. In a spectral dimension convolutional layer, we utilize a 1×1 convolutional kernel to do a nonlinear transformation of each image element in spectral–spatial feature X, formulated as

$$Y_{\text{spe}} = \sigma(W_{\text{spe}} * \text{BN}(X) + b_{\text{spe}}) \tag{14}$$

where σ is activation function, W_{spe} is the weight of the spectral dimension convolution, b_{spe} is the bias of the spectral dimension convolution, * denotes the convolution operation. Through the convolutional transformation of HSI using a 1×1 convolutional kernel, deep spectral features can be extracted using fewer parameters. Then, the spatial dimensional convolutional layer performs a 2-D convolution using a 2-D convolutional kernel on Y_{spe}, denoted as

$$Y_{\text{SSConv}} = \sigma(W_{\text{spa}} * Y_{\text{spe}} + b_{\text{spa}}) \tag{15}$$

where W_{spa} is the weight of the spatial dimension convolution and b_{spa} is the bias of the spatial dimension convolution. SSconv can utilize all channels of the image element in the feature extraction process at the same time, and its spectral dimensional receptive field is not limited, so it can extract the spatial–spectral features more effectively. In this study, we utilize a two-layer SSConv to construct the S2CN subnetwork, aiming to explore pixel-level local information. The final output of the S2CN subnetwork is denoted as Y_{S2CN}.

D. Advantageous Feature Differential Enhancement Fusion

In Section II, we have demonstrated that different structure models extract distinct advantageous features. Specifically, the CNN structure exhibits superior classification performance in fine regions, whereas the GCN/HGCN structure exhibits superior classification performance in smooth regions. Therefore, to enhance the classification robustness of the model, in this section, we introduce AFDEF to achieve mutual enhancement between different structure models with distinct advantageous

features, thereby achieving good classification performance in different regions. The principle of the AFDEF module [39] is analogous to differential amplifier circuits in which the common features are suppressed and the advantageous features are amplified, which can be formulated as

$$\begin{aligned} Y_{\text{S2HGCN}} &= \frac{Y_{\text{S2HGCN}} + Y_{\text{S2HGCN}}}{2} + \frac{Y_{\text{S2CN}} - Y_{\text{S2CN}}}{2} \\ &= \frac{Y_{\text{S2HGCN}} + Y_{\text{S2CN}}}{2} + \frac{Y_{\text{S2HGCN}} - Y_{\text{S2CN}}}{2} \end{aligned} \tag{16}$$

$$\begin{aligned} Y_{\text{S2CN}} &= \frac{Y_{\text{S2CN}} + Y_{\text{S2CN}}}{2} - \frac{Y_{\text{S2HGCN}} - Y_{\text{S2HGCN}}}{2} \\ &= \frac{Y_{\text{S2CN}} + Y_{\text{S2HGCN}}}{2} + \frac{Y_{\text{S2CN}} - Y_{\text{S2HGCN}}}{2} \end{aligned} \tag{17}$$

where the accumulation part $Y_{\text{S2HGCN}} + Y_{\text{S2CN}}$ reflects common features, and the differential part $Y_{\text{S2HGCN}} - Y_{\text{S2CN}}$ and $Y_{\text{S2CN}} - Y_{\text{S2HGCN}}$ reflects advantageous features captured by S2HGCN and S2CN structure. Therefore, The core of AFDEF is to preserve the original features and compensate according to the differential features. The workflow of the AFDEF module is illustrated in Fig. 5(d). We first obtain the differential feature Y_D by direct subtraction of two subnetworks, denoted as

$$Y_D = Y_{\text{S2HGCN}} - Y_{\text{S2CN}}. \tag{18}$$

Then, we obtain the derivation of the fusion weight vector V_w formulated as

$$V_w = \text{Tanh}(\text{FC}(Y_D)) \tag{19}$$

where FC represents a fully connected layer, and Tanh is the activation function, which normalizes the fusion weights to the range of -1 to 1. Subsequently, the features Y_{S2HGCN} and Y_{S2CN} from the two subnetworks are recalibrated by the fusion weight vector V_w through channelwise multiplication, resulting in the advantageous complementary features Y_{AHG} and Y_{ACN} extracted by S2HGCN and S2CN, respectively,

$$\begin{aligned} Y_{\text{AHG}} &= V_w Y_{\text{S2HGCN}} \\ Y_{ACN} &= V_w Y_{\text{S2CN}} \end{aligned} \tag{20}$$

finally, the complementary features Y_{AHG} and Y_{ACN} are added to the original feature as complementary information, resulting in a more discriminative feature representation

$$\begin{aligned} \widetilde{Y}_{\text{S2HGCN}} &= Y_{\text{S2HGCN}} + \eta_{\text{ACN}} Y_{\text{ACN}} \\ \widetilde{Y}_{S2CN} &= Y_{\text{S2CN}} + \eta_{\text{AHG}} Y_{\text{AHG}} \end{aligned} \tag{21}$$

compared to the work in [39], we introduce an enhancement factor η to amplify advantageous features. The enhanced features from the two subnetworks $\widetilde{Y}_{\text{S2HGCN}}$ and $\widetilde{Y}_{\text{S2CN}}$ are then fused and through a fully connected layer to output the classification result

$$Y = \text{FC}\left(\gamma \widetilde{Y}_{\text{S2HGCN}} + (1-\gamma)\widetilde{Y}_{\text{S2CN}}\right) \tag{22}$$

where γ represents the fusion weight.

TABLE I
CATEGORY INFORMATION OF THE INDIAN PINES DATASET

Class id	Color	Class name	Training	Validation	Test
1		Alfalfa	1	1	14
2		Corn Notill	15	15	1398
3		Corn Mintill	9	9	812
4		Corn	3	3	231
5		Grass Pasture	5	5	473
6		Grass Trees	8	8	714
7		Grass Pasture Mowed	1	1	26
8		Hay Windrowed	5	5	468
9		Oats	1	1	18
10		Soybean Notill	10	10	952
11		Soybean Mintill	25	25	2405
12		Soybean Clean	6	6	581
13		Wheat	3	3	199
14		Woods	13	13	1239
15		Buildings Grass Trees Drives	4	4	378
16		Stone Steel Towers	1	1	91
		Total	110	110	9619

The cross-entropy is employed as the loss function for DF2Net

$$\mathcal{L}oss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}\log(p_{ij}) \tag{23}$$

where N is the number of samples, C is the number of classes, y represents the true labels, and p represents the predicted probabilities.

IV. EXPERIMENTAL

A. Dataset Description

In this section, we evaluate the proposed DF2Net on four HSI benchmark datasets, i.e., Indian Pines, University of Pavia, Xuzhou, and LaoYu River dataset.

1) Indian Pines Dataset: The Indian Pines dataset was collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over Northwestern Indiana, USA, in 1992 [40]. The dataset comprises 145×145 pixels with a spatial resolution of 20 m and a spectral resolution of 10 nm. After filtering out the noisy bands, the remaining 200 bands from the original dataset, containing 10 366 samples, labeled into 9 different classes are used for further analysis and experimentation. Since the Indian Pines dataset suffers from severe class imbalance issues. For instance, the Oats comprise only 20 pixels, posing challenges for accurate classification. Therefore, 1%, 1%, and 98% of labeled samples per class are chosen as the training, validation, and testing sets, respectively. The detailed category information is shown in Table I.

2) University of Pavia Dataset: The University of Pavia dataset was collected by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor over the University of Pavia, Italy, in 2001 [41]. The dataset comprises 610×340 pixels with a spatial resolution of 1.3 m. After removing the disturbing bands, the remaining 103 bands, containing 207 400 samples, labeled into 9 different classes are used for further analysis and experimentation. Since the University of Pavia dataset has a higher proportion of labeled samples (over 20%), Therefore, 0.1%, 0.1%, and 99.8% of labeled samples per class are chosen

TABLE II
CATEGORY INFORMATION OF THE UNIVERSITY OF PAVIA DATASET

Class id	Color	Class name	Training	Validation	Test
1		Asphalt	7	7	6617
2		Meadows	19	19	18 611
3		Gravel	3	3	2093
4		Trees	4	4	3056
5		Metal Sheets	2	2	1341
6		Bare Soil	6	6	5017
7		Bitumen	2	2	1326
8		Bricks	4	4	3674
9		Shadows	1	1	945
	Total		48	48	42 680

TABLE III
CATEGORY INFORMATION OF THE XUZHOU DATASET

Class id	Color	Class name	Training	Validation	Test
1		Bareland1	27	27	26 342
2		Lakes	5	5	4017
3		Coals	3	3	2777
4		Cement	6	6	5202
5		Crops-1	14	14	13 156
6		Trees	3	3	2430
7		Bareland2	7	7	6976
8		Crops	5	5	4767
9		Red-title	4	4	3062
	Total		74	74	68 729

as the training, validation, and testing sets, respectively. The detailed category information is shown in Table II.

3) Xuzhou Dataset: The Xuzhou dataset was collected by an airborne HYSPEX hyperspectral camera over the Xuzhou peri-urban site in November 2014 [42]. This dataset consists of 500 × 260 pixels, with a very high spatial resolution of 0.73 m/pixel. After removing the disturbing bands, the remaining 436 bands are used for further analysis and experimentation. About 0.1%, 0.1%, and 99.8% of labeled samples per class are chosen as the training, validation, and testing sets, respectively. The detailed category information is shown in Table III.

4) LaoYu River Dataset: The LaoYu River dataset was collected by OHS satellite over Dianchi Lake and its adjacent LaoYu River wetland parks. In this area, a large number of Metasequoias have been planted along Dianchi Lake, which plays a crucial role in purifying the water quality of Dianchi Lake and protecting the ecological environment. The dataset consists of 391 × 591 pixels with the spatial resolution of 10 m. The wavelength range of the LaoYu River dataset is 0.4 to 1 μm. After removing the disturbing bands, the remaining 32 spectral bands in the dataset are labeled into 8 different classes for further analysis and experimentation. Since, the LaoYu River dataset has a low spatial resolution, 0.3%, 0.3%, and 99.4% of labeled samples per class are chosen as the training, validation, and testing sets, respectively. The detailed category information is shown in Table IV.

B. Implementation Details

1) Comparison With State-of-the-Art Models: Eight state-of-the-art deep learning models are selected for comparison: 1) 2DCNN [19], 2) multiscale 3-D deep CNN (M3DCNN) [20], 3) hybrid spectral network (HybridSN) [22], 4) RNN [18], 5) GCN [24], 6) the feature fusion hypergraph convolution neural network (F2HGNNss) [35], 7) CNN-enhanced GCN (CEGCN) [27], and 8) attention multihop graph and multiscale convolutional fusion network (AMGCFN) [29]. In these methods, M3DCNN and HybridSN are based on the CNN structure, GCN and F2HGNNss are based on the graph/hypergraph structure, and CEGCN and AMGCFN are based on the hybrid network structure.

TABLE IV
CATEGORY INFORMATION OF THE LAOYU RIVER DATASET

Class id	Color	Class name	Training	Validation	Test
1		Metasequoia	17	17	5473
2		Other Tree Species	9	9	2864
3		Greenhouse Farmland	20	20	6626
4		Bare Land	7	7	2142
5		Water Bodies	24	24	7654
6		Buildings	8	8	2440
7		Asphalt	21	21	6761
8		Pitches	1	1	166
	Total		107	107	34 126

2) Evaluation Metrics: To comprehensively evaluate the performance of the proposed DF2Net and other state-of-the-art models, four evaluation metrics are adopted, namely 1) OA, 2) average accuracy (AA), 3) kappa coefficient (KC), and 4) per class classification accuracy. In addition, all results presented in this article are the mean values of 10 experiments for consistency.

3) Configuration of Hyperparameters: The number of layers for both S2CN and S2HGCN subnetworks is set to 2. For the S2CN subnetwork, the size of convolutional kernels is set to 5×5, and the number of channels in the first and the second layers are set to 128 and 64, respectively. For the S2HGCN subnetwork, the superpixel density for the University of Pavia and Xuzhou datasets during the SLIC process is set to 100. For the Indian Pines and LaoYu River datasets, the superpixel density is set to 50 as they have a low spatial resolution. The number of channels in the first and the second layers of the S2HGCN subnetwork are set to 128 and 64. The number of neighbor nodes for superpixel nodes K is set to 10. The fusion weight γ is set to 0.95 and the enhancement factor η is set to 0.2. We utilize the Adam optimizer to train DF2Net with a weight decay set of 0.0005. The initial learning rate is set to 0.001, and DF2Net is trained for 600 epochs. All experiments are conducted on the PyTorch platform, utilizing the NVIDIA GeForce RTX 3060 GPU for hardware acceleration.

C. Comparison of Classification Performance

In this section, we compare and analyze the results of all methods on each dataset separately. The classification accuracies of different methods are detailed in Tables V–VIII and the classification result maps generated by these methods are shown in Figs. 6–9,

1) Indian Pines Dataset: As shown in Table V, M3DCNN performs poorly on the Indian Pines dataset, with OA of 61.36% and KC of 55.43%. In addition, in Fig. 6(c), the classification result maps of M3DCNN exhibit a lot of pepper-salt noise.

TABLE V
CLASSIFICATION RESULTS (%) ON THE INDIAN PINES DATASET

Class ID	M3D-CNN	HybridSN	RNN	GCN	F2HGNNss	CEGCN	AMGCFN	DF2Net (ours)
1	6.09±2.83	73.92±19.61	27.54±24.44	**83.18±20.00**	72.27±20.99	18.89±15.24	81.85±17.78	29.54±9.85
2	29.26±4.02	64.31±3.49	39.87±11.63	73.61±5.34	79.17±1.77	81.67±3.72	**82.36±6.54**	79.02±6.10
3	29.09±3.06	67.44±5.8	37.75±5.98	67.41±7.51	73.37±1.90	44.76±10.52	75.77±10.43	**83.40±2.90**
4	32.66±8.1	82.6±15.22	9.99±1.95	77.83±9.11	**92.64±6.70**	27.03±11.06	79.94±13.86	80.95±4.36
5	57.48±1.52	76.30±1.19	44.24±12.13	76.32±6.32	87.31±0.51	62.59±11.86	74.24±9.74	**93.99±0.94**
6	96.49±1.64	89.08±6.41	88.17±2.61	90.81±0.88	94.53±1.02	98.24±1.43	90.83±4.19	**99.86±0.12**
7	28.57±9.78	75.89±17.83	41.66±13.52	**100.00±0.00**	**100.00±0.00**	32.79±17.83	93.70±9.94	98.46±1.88
8	99.04±0.98	99.11±1.04	60.39±24.4	95.08±1.76	98.67±1.61	96.33±3.56	99.08±0.89	**99.65±0.58**
9	45.00±9.35	96.25±7.50	28.33±14.43	28.89±24.44	52.22±14.31	19.06±8.43	**97.30±4.88**	88.89±12.17
10	51.63±2.93	74.77±7.81	44.55±4.65	80.41±1.83	84.30±1.81	73.12±7.13	81.18±4.84	**84.53±3.89**
11	75.93±4.26	81.01±4.00	58.76±4.70	81.34±2.23	84.03±1.65	89.23±5.22	**90.19±4.04**	87.63±2.32
12	28.06±2.66	57.89±10.62	26.7±10.24	51.36±7.33	**82.09±2.31**	46.78±15.95	67.55±10.26	53.25±8.31
13	85.37±2.95	93.41±2.47	92.2±6.09	91.26±3.87	94.37±1.21	96.01±3.80	96.25±7.44	**98.89±0.86**
14	94.02±1.32	97.97±1.44	88.22±3.08	95.64±1.25	93.96±1.03	99.90±0.09	92.39±7.35	**99.98±0.03**
15	33.99±8.03	68.2±20.04	13.99±4.94	67.46±5.67	87.67±1.25	41.57±18.08	80.49±11.42	**94.34±2.38**
16	53.33±9.51	88.98±15.81	59.5±28.07	99.34±1.31	**99.12±1.28**	45.77±27.26	82.50±15.40	87.91±18.41
OA	61.36±0.78	78.8±1.09	53.96±2.34	79.87±1.11	85.83±0.66	77.58±0.33	85.18±2.19	**87.06±1.10**
AA	52.87±0.64	80.44±1.82	47.62±2.59	78.74±2.51	85.98±0.22	60.89±3.23	**85.35±2.64**	84.95±1.46
KC	55.43±0.77	75.87±1.30	47.28±2.70	77.14±1.25	83.87±0.75	74.02±0.39	83.16±2.40	**85.25±1.23**

The bold values indicate the highest value in the corresponding evaluation metric.

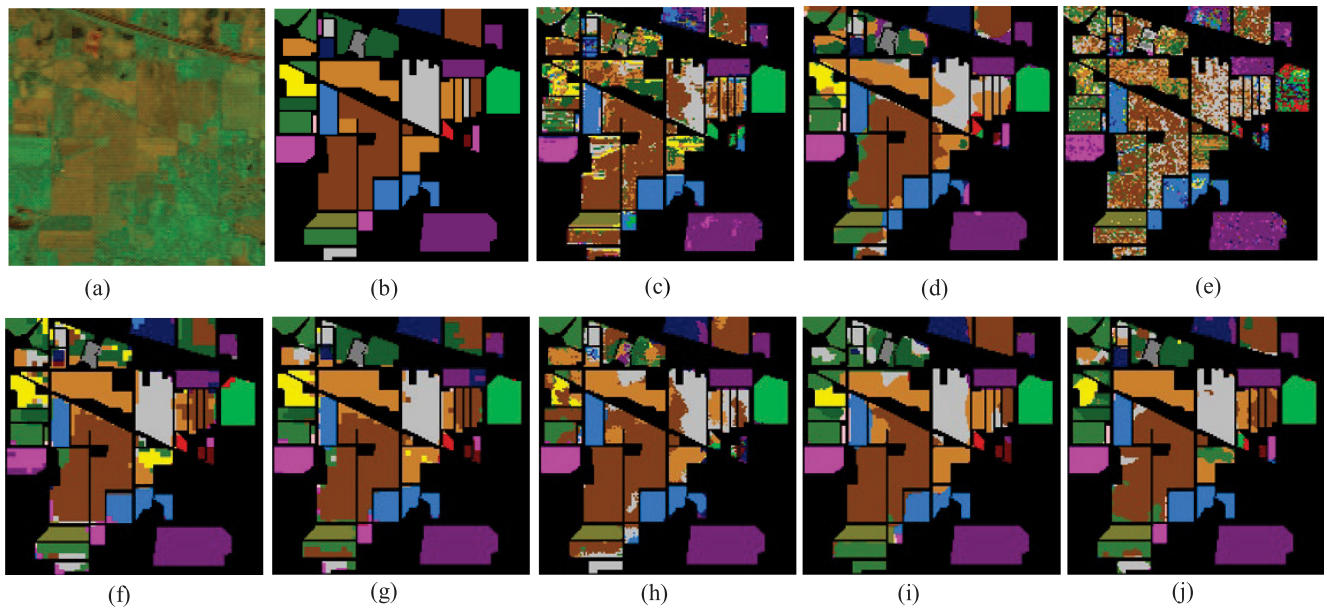

Fig. 6. Classification maps on the Indian Pins dataset. (a) False-color image. (b) Ground truth. (c) M3DCNN. (d) HybidSN. (e) RNN. (f) GCN. (g) F2HGNNss. (h) CEGCN. (i) AMGCFN. (j) DF2Net.

This is mainly due to the high dependence of M3DCNN on training samples, resulting in poor classification performance when facing with limited and imbalanced samples. HybridSN integrates convolutions from different dimensions, enhancing its representative capacity. However, its classification performance remains unsatisfactory. We also compared the classification performance of RNN when there is a lack of samples, and the results were similarly poor. Simple GCN surpasses HybridSN by 1% in OA, benefiting from its effective modeling of nonlocal features and the utilization of unlabeled samples through the graph structure. Compared to GCN, F2HGNNss utilizes the hypergraph to capture high-order correlations, significantly enhancing the classification performance of HSI. For instance, F2HGNNss achieves smoother and more accurate classification for class 7, as illustrated in Fig. 6. But both GCN and F2HGNN are based on superpixel nodes, which results in the loss of pixel-level spectral information, leading to poorer classification performance in fine regions.

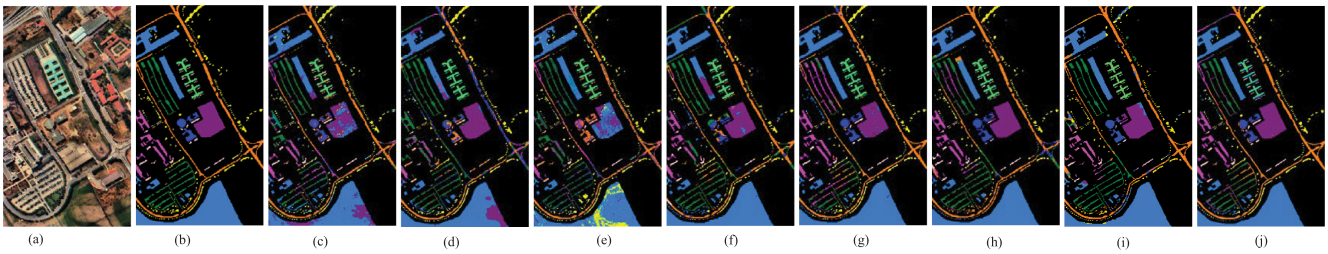

Fig. 7. Classification maps on the University of Pavia dataset. (a) False-color image. (b) Ground truth. (c) M3DCNN. (d) HybridSN. (e) RNN. (f) GCN. (g) F2HGNNss. (h) CEGCN. (i) AMGCFN. (j) DF2Net.

TABLE VI
CLASSIFICATION RESULTS (%) ON THE UNIVERSITY OF PAVIA DATASET

Class ID	M3D-CNN	HybridSN	RNN	GCN	F2HGNNss	CEGCN	AMGCFN	DF2Net (ours)
1	83.88±1.31	49.87±7.65	74.18±3.71	81.66±5.55	88.03±5.08	**89.26±10.35**	90.21±4.47	87.79±2.19
2	85.48±6.34	93.23±4.24	81.65±6.98	93.69±3.40	97.29±2.39	96.74±1.96	97.90±1.32	**99.58±0.32**
3	42.15±13.34	44.26±5.29	27.36±11.52	69.48±5.12	64.79±26.09	86.95±6.80	74.15±15.60	**90.85±3.41**
4	85.91±7.32	84.95±2.05	81.54±9.22	65.01±10.32	80.29±6.57	75.12±11.72	68.01±8.86	**87.02±1.68**
5	93.35±7.11	99.38±1.24	94.54±8.88	98.64±1.65	99.18±1.44	**100.00±0.00**	99.55±0.92	90.03±6.51
6	68.61±3.02	94.34±2.13	49.67±18.31	**95.23±3.84**	90.43±8.26	93.38±10.22	94.77±4.71	94.45±1.87
7	76.36±2.89	30.50±11.33	22.11±9.40	73.35±23.12	84.33±11.04	85.42±22.55	73.38±19.46	**95.98±2.71**
8	76.81±9.45	54.83±18.13	78.75±11.88	84.60±8.04	72.26±19.89	78.14±20.01	81.51±7.750	**87.66±2.77**
9	84.75±15.01	14.74±10.48	75.84±22.31	44.42±10.29	72.83±2.88	58.88±23.74	50.85±18.93	**90.85±7.99**
OA	80.36±3.39	76.84±1.43	72.23±3.05	86.44±1.82	89.20±1.81	90.47±3.50	89.87±2.25	**94.19±0.44**
AA	77.48±2.87	62.90±3.08	65.07±1.70	78.48±3.54	83.27±2.88	84.88±6.52	81.15±4.42	**91.58±1.21**
KC	74.20±4.22	69.54±1.77	63.47±3.60	82.14±2.29	85.65±2.39	87.33±4.75	86.46±3.05	**92.28±0.58**

The bold values indicate the highest value in the corresponding evaluation metric.

For hybrid networks, CEGCN combines CNN and GCN into a hybrid network, which was expected to have good classification performance but yielded poor experimental results. This is mainly because it employs a simple concatenation strategy to fuse features from different structures, failing to effectively utilize advantageous complementary information. Building upon CEGCN, AMGCFN introduces an attention multihop graph and various attention fusion mechanisms to enhance the extraction and utilization of deep features, achieving OA of 85.18%. However, improving the representation capability of GCN subnetworks for high-order correlations through multihop graphs is not a favorable direction, as it leads to feature oversmoothing and consumes a considerable amount of memory resources. Our proposed DF2Net achieved the best classification performance with an OA of 87.06%. DF2Net combines hypergraph convolution with convolution and utilizes the AFDEF module to achieve mutual enhancement of advantageous features derived from different network structures, thereby improving the classification robustness of different regions in HSI, as shown in Fig. 6.

2) University of Pavia Dataset: Compared to the Indian Pines dataset, the University of Pavia dataset has more labeled samples and is more balanced. Thus, the OA of both M3DCNN and RNN is improved. For GCN, its OA reaches 86.44%, but its AA is only 77.48%, indicating an imbalance issue in its classification performance. For instance, in classes 9 and 4 with fewer samples, its accuracy is only 44.42% and 65.01%, respectively. In addition, from Fig. 7, it can be observed that GCN performs poorly in classifying asphalt, particularly at road intersections, indicating its weak representation capability for the complex land cover distributions contained in HSI. In contrast, F2HGNNss exhibits superior classification performance at road intersections, as depicted in Fig. 7. This validates the stronger representational power of hypergraphs over simple graphs for capturing complex correlations. The hybrid networks CEGCN and AMGCFN achieve high OA of 90.47% and 89.87%, respectively. However, the classification accuracy for class 9 is below 60% and indicates that there is also an imbalance issue in classification performance. As shown in Table VI, our proposed DF2Net exhibits good classification performance across all classes. Particularly, it achieves an accuracy of 90.85% even in class 9, where other methods perform poorly. In addition, all the comparative methods exhibited notable variability in results across multiple runs, indicating instability in their performance. The proposed DF2Net demonstrates relatively stable performance on the University of Pavia dataset.

3) Xuzhou Dataset: As shown in Table VII, hybrid structure networks exhibit superior classification performance compared to single-structure networks. For example, CEGCN achieves an OA of 94.52% while GCN achieves an OA of 91.56%. This is mainly due to their ability to extract complementary information at different levels. As shown in Fig. 8, M3DCNN and RNN exhibit a significant amount of pepper-salt noise. In all classification methods, our proposed DF2Net achieved the best classification performance on the Xuzhou dataset, with OA of 95.49% and KC of 94.29%. Compared to the single hypergraph convolutional network structure F2HGNNss, DF2Net exhibits

TABLE VII
CLASSIFICATION RESULTS (%) ON THE XUZHOU DATASET

Class ID	M3D-CNN	HybridSN	RNN	GCN	F2HGNNss	CEGCN	AMGCFN	DF2Net (ours)
1	95.13±1.26	96.36±0.26	92.41±0.45	94.14±1.87	95.53±1.97	94.81±1.21	94.80±1.52	**96.43±0.54**
2	96.25±0.58	89.30±2.45	96.65±1.16	95.32±2.33	97.97±2.17	99.72±0.34	98.45±0.89	**99.86±0.03**
3	71.02±6.47	94.70±1.70	74.09±9.27	**94.13±8.50**	92.52±5.63	93.01±10.22	93.78±9.55	87.53±3.20
4	82.38±6.92	86.26±5.56	84.35±2.79	68.57±3.39	78.45±7.73	86.16±6.42	83.48±13.18	**97.52±0.82**
5	93.68±2.32	95.14±2.42	84.70±2.67	95.85±1.29	96.99±1.77	97.79±1.18	97.16±2.56	**98.44±0.61**
6	18.05±7.39	31.35±4.46	10.27±2.66	**91.33±6.80**	67.66±22.60	74.86±14.12	87.15±10.95	72.08±1.11
7	72.17±6.83	92.27±2.58	55.88±3.39	94.58±2.24	94.19±3.69	94.97±3.40	97.95±1.36	**98.32±1.14**
8	93.37±5.51	98.32±0.69	95.82±2.74	93.70±5.39	95.76±2.16	**100.00±0.00**	99.32±0.64	99.40±0.18
9	87.90±4.95	79.47±14.07	89.21±6.21	72.72±13.41	79.70±5.47	**93.02±3.10**	86.69±8.26	78.93±6.89
OA	87.48±1.33	91.93±1.39	83.31±1.15	91.56±0.38	92.83±1.30	94.52±0.64	94.57±0.99	**95.49±0.26**
AA	78.88±2.16	84.91±2.49	75.93±2.06	88.93±1.01	88.87±3.24	92.70±1.09	**93.20±1.87**	92.05±0.67
KC	83.96±1.72	89.72±1.76	78.73±1.46	89.31±0.45	90.90±1.65	93.06±0.82	93.13±1.24	**94.29±0.32**

The bold values indicate the highest value in the corresponding evaluation metric.

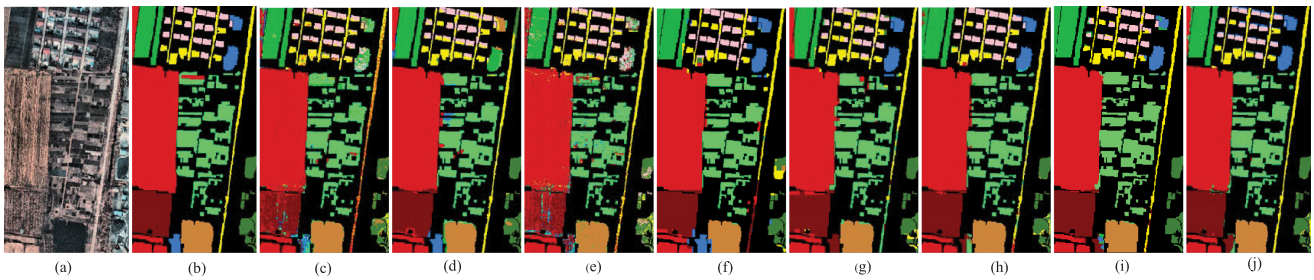

Fig. 8. Classification maps on the Xuzhou dataset. (a) False-color image. (b) Ground truth. (c) M3DCNN. (d) HybidSN. (e) RNN. (i) AMGCFN. (j) DF2Net.

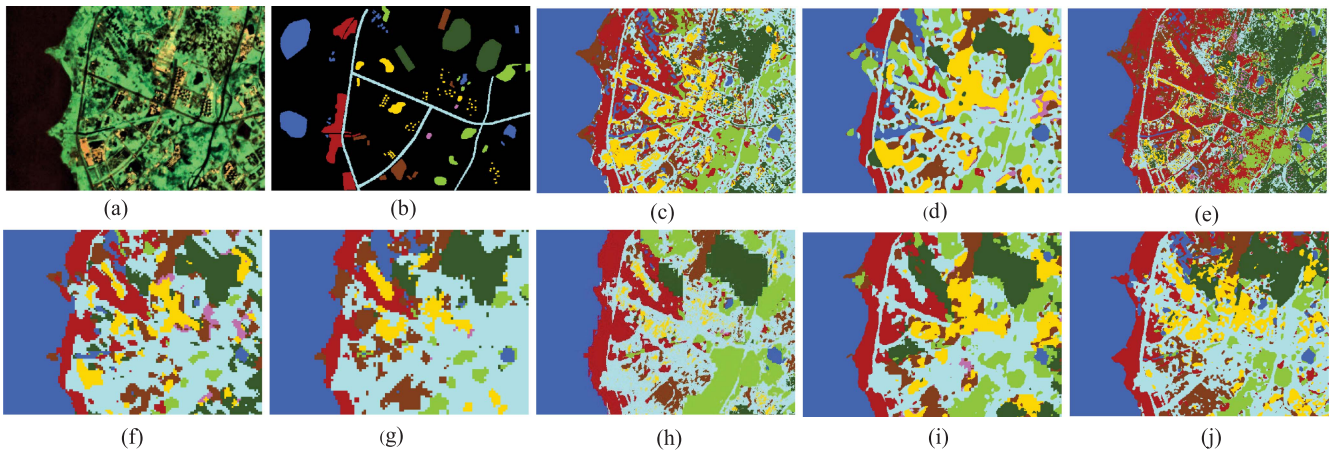

Fig. 9. Classification maps on the LaoYu River dataset. (a) False-color image. (b) Ground truth. (c) M3DCNN. (d) HybridSN. (e) RNN. (f) GCN. (g) F2HGNNss. (h) CEGCN. (i) AMGCFN. (j) DF2Net.

a 3% improvement in OA. This is mainly attributed to the loss of pixel-level local information in F2HGNNss. DF2Net combines S2HGCN and S2CN subnetworks to mine information at different abstraction levels, effectively enhancing classification performance. In addition, it is noteworthy that DF2Net surpasses CEGCN and AMGCFN by 1% in OA, further validating its capability in extracting and integrating advantageous complementary features.

4) LaoYu River Dataset: Three important observations need to be emphasized again. First, graph-based methods are better at capturing global features while regular convolution-based methods are better at capturing local features. As shown in Fig. 9(c), M3DCNN can effectively extract textures of land covers and classify the fine areas better. In addition, RNN also extracts texture details of the land covers well, as shown in Fig. 9(e). However, they exhibit errors in large-area classification, for instance,

TABLE VIII
CLASSIFICATION RESULTS (%) ON THE LAOYU RIVER DATASET

Class ID	M3D-CNN	HybridSN	RNN	GCN	F2HGNNss	CEGCN	AMGCFN	DF2Net (ours)
1	91.47±1.53	86.56±2.06	93.50±1.67	87.62±1.89	89.58±0.83	**96.51±1.46**	93.44±4.22	95.93±0.83
2	87.51±2.43	89.79±3.25	81.74±4.06	58.98±6.38	69.47±3.21	95.86±1.67	**97.65±2.23**	76.03±2.71
3	81.68±1.55	91.08±3.42	72.21±5.46	95.28±1.17	98.41±0.68	**99.48±0.20**	98.56±1.42	97.14±1.90
4	54.96±10.54	70.14±4.87	52.15±8.12	72.92±6.48	81.85±7.28	74.31±11.21	72.16±7.57	**81.91±7.14**
5	97.44±1.41	96.65±2.07	97.46±0.38	95.30±1.55	94.92±0.69	94.58±2.09	94.51±2.45	**97.98±0.82**
6	59.76±6.10	68.70±12.17	26.59±3.59	71.48±6.48	**72.03±3.16**	51.80±8.80	58.61±7.48	70.46±5.71
7	70.64±3.67	82.59±4.31	50.94±4.80	76.59±6.11	82.85±4.54	83.80±8.16	86.06±6.93	**86.74±4.04**
8	5.51±1.02	21.88±5.80	3.13±1.78	16.62±8.24	22.29±13.49	4.82±3.19	35.77±28.23	**48.65±23.27**
OA	81.47±1.39	86.56±1.02	73.01±1.07	83.82±1.15	87.46±1.53	88.20±1.48	89.46±0.76	**90.15±0.88**
AA	68.62±1.56	75.92±1.88	59.71±1.44	71.85±1.95	76.55±3.01	74.27±1.05	79.60±2.60	**81.74±2.94**
KC	77.68±1.65	83.80±1.21	67.44±1.31	80.43±1.38	84.85±1.85	85.94±1.77	87.28±0.87	**88.10±1.08**

The bold values indicate the highest value in the corresponding evaluation metric.

TABLE IX
CLASSIFICATION RESULTS (%) OF DIFFERENT SUBNETWORKS ON EACH DATASET

Method	Indian Pines			University of Pavia			Xuzhou			LaoYu River		
	OA	AA	KC	OA	AA	KC	OA	AA	KC	OA	AA	KC
S2CN	75.01±3.74	65.85±4.77	71.15±4.33	90.01±2.36	85.37±4.19	86.59±3.22	93.22±0.26	89.63±0.43	91.68±0.32	86.54±1.14	70.06±1.87	83.65±1.44
S2HGCN	85.74±0.53	82.44±1.68	83.76±0.59	88.97±0.09	81.37±1.31	85.41±0.13	93.08±0.27	92.13±0.25	91.55±0.34	87.33±0.37	78.37±3.65	84.72±0.44
DF2Net	**87.06±1.10**	**84.95±1.46**	**85.25±1.23**	**94.19±0.44**	**91.58±1.21**	**92.28±0.58**	**95.49±0.26**	**92.05±0.67**	**94.29±0.32**	**90.15±0.88**	**81.74±2.94**	**88.10±1.08**

The bold values indicate the highest value in the corresponding evaluation metric.

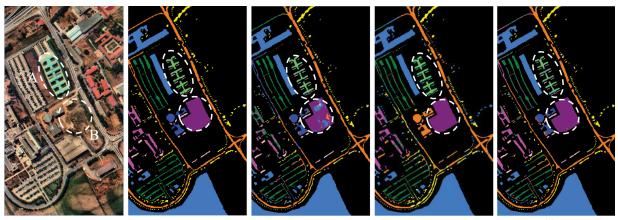

Fig. 10. Visualization of different subnetworks on the University of Pavia dataset. (a) False-color image. (b) Ground truth. (c) S2CN. (d) S2HGCN. (e) DF2Net.

misclassifying many other tree species as metasequoia. In contrast, GCN and F2HGNNss demonstrate better performance in large-area classification, as shown in Fig. 9(f) and (g). However, since they are superpixel-based methods, a large amount of pixel-level information is lost. In addition, due to oversmoothing issues, nodes within the same connected component tend to converge to the same value (the class with the highest number of nodes). Therefore, in the results in Fig. 9(e) and (g), many regions are incorrectly classified as asphalt. Second, hypergraphs demonstrate better performance than graphs. As shown in Table VIII, the OA and KC of F2HGNNss increased by about 4% compared to GCN. Third, hybrid networks demonstrate more robust classification performance compared to single-structure networks. For example, CEGCN achieves an OA of 88.20% and AMGCFN achieves an OA of 89.46%. They are approximately 5% higher than GCN. Among all hybrid structure methods, the proposed DF2Net achieved the best classification results, with an OA of 90.15%, AA of 81.74%, and KC of 88.10%. This once again validates the effectiveness of extending simple graphs to hypergraphs in DF2Net, as well as the impactful role of the introduced AFDEF module.

V. DISCUSSION

A. Comparison of Subnetworks

In this section, we design and conduct various ablation experiments on four datasets to evaluate the effectiveness of each module. Specifically, we first compare the performance of two subnetworks, S2HGCN and S2CN, along with DF2Net. The experimental results are shown in Table IX. Since, the University of Pavia dataset covers typical fine and smooth regions, we provide the visualization of different subnetworks on the University of Pavia dataset, as is shown in Fig. 10.

From Table IX, it can be observed that the S2CN subnetwork exhibits better classification performance on the University of Pavia dataset compared to S2HGCN. Conversely, S2HGCN demonstrates better classification performance on the Indian Pines dataset. The reasons could be as follows: First, the University of Pavia and Xuzhou datasets have larger image dimensions, leading to a higher density of superpixels. Consequently, a substantial amount of local spectral–spatial information remains uncaptured by the S2HGCN network based on superpixel nodes. Second, the University of Pavia dataset contains more fine regions; hence, the convolution-based S2CN has more advantages on this dataset. From Fig. 10(c), we observe that S2CN exhibits better classification performance in fine regions such as region A, but its performance is poorer in region B. In Fig. 10(d), S2HGCN demonstrates better classification performance in smooth region B, yet it performs poorly in identifying shadows in region A due to pixel-level information loss. In Fig. 10(e), our proposed DF2Net achieves accurate classification performance across different types of regions by integrating two subnetworks, demonstrating robust classification performance.

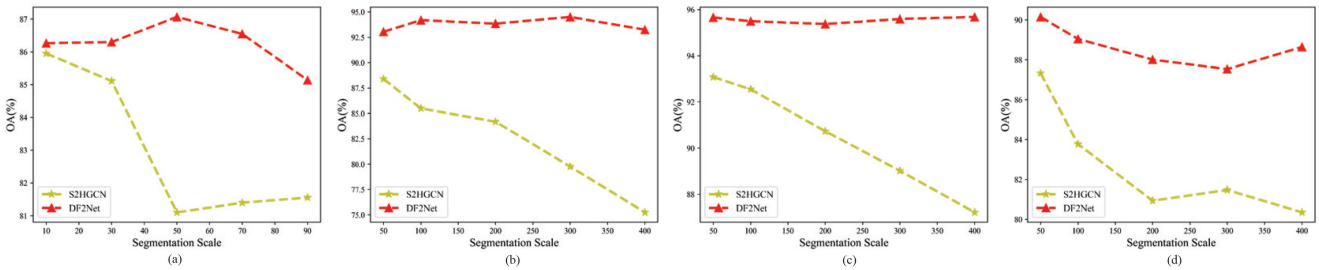

Fig. 11. Classification accuracies of DF2Net with the S2HGCN subnetwork under different segmentation scales on each dataset. (a) Indian Pines. (b) University of Pavia. (c) Xuzhou. (d) LaoYu River.

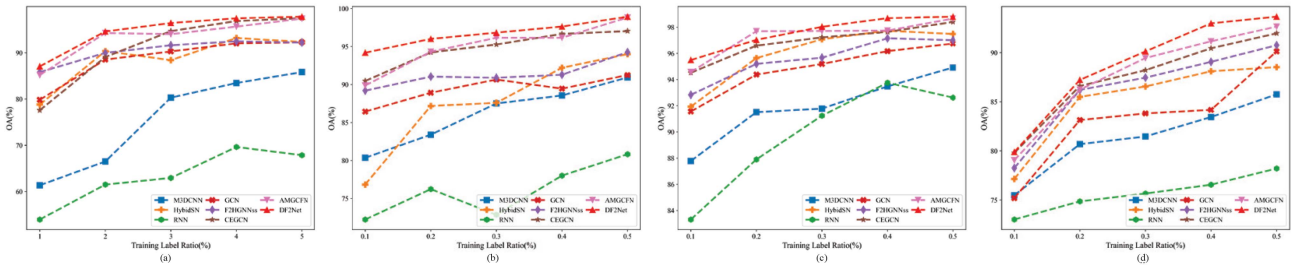

Fig. 12. Classification accuracies under different training label ratios on each dataset. (a) Indian Pines. (b) University of Pavia. (c) Xuzhou. (d) LaoYu River.

B. Influence of Segmentation Scale

The segmentation scale λ determines the size of superpixels. A larger λ means that more information about large objects can be provided while a smaller λ can reduce the loss of pixel-level spectral–spatial information. Therefore, the segmentation scale λ has an impact on the classification performance of DF2Net. As shown in Fig. 11, we tested the classification performance of DF2Net and S2HGCN subnetworks on four datasets under different segmentation scales λ. Specifically, for the Indian Pines dataset with a smaller size, segmentation scale λ was set to 10, 30, 50, 70, and 90. For the University of Pavia, Xuzhou, and LaoYu River datasets with larger sizes, segmentation scale λ was set to 50, 100, 200, 300, and 400. We can observe that as the segmentation scale λ increases, the performance of the S2HGCN subnetwork rapidly decreases due to the increasing loss of pixel-level information. However, DF2Net achieves good classification performance across different segmentation scale λ sizes by combining the S2HGCN and S2CN subnetworks. This phenomenon indicates that the influence of segmentation scale λ on the hybrid network diminishes.

C. Comparison of Different Training Sample Ratios

As we all know, labeled pixels in HSI are usually rare and expensive to acquire. Therefore, in the previous experiments, we evaluated the HSI classification performance of our proposed DF2Net using a small sample. To further validate the robustness of DF2Net, we tested its classification performance on four datasets under different ratios of labeled samples. Specifically, for the Indian Pines dataset, as its low sample annotation rate and validation imbalance, we selected training proportions of 1%, 2%, 3%, 4%, and 5%. For the University of Pavia, Xuzhou, and LaoYu River datasets, given their abundance of labeled samples, we chose training proportions of 0.1%, 0.2%, 0.3%, 0.4%, and 0.5%. As shown in Fig. 12, it is observed that the performance of all methods improves with an increase in the number of training samples. Overall, hybrid networks outperform single-structure networks. Our proposed DF2Net demonstrates significant advantages in scenarios with limited labeled samples, which may be attributed to the utilization of more unlabeled samples facilitated by the hypergraph.

D. Fusion Strategy Analysis

In this section, we compared different fusion strategies, including simple concatenation, direct addition (where each branch is multiplied by 0.5 before addition), and AFDEF. The experimental results are presented in Table X. It can be observed that different fusion strategies have a significant impact on the final classification performance, especially on the Indian Pines dataset. Using the concatenation method only achieves an OA of 75.72% while using AFDEF can achieve an OA of 87.06%. In addition, the fusion strategy of direct addition achieves an OA of 79.43%, which is 4% higher than the concatenation. This could be attributed to the poor classification performance of the S2CN subnetwork on the Indian Pines dataset when faced with limited and imbalanced samples. Employing fusion strategies such as simple concatenation, and direct addition, especially simple concatenation, at this time may lead the fusion network to be influenced by the S2CN subnetwork, resulting in the underutilization of advantageous features. It is noteworthy that the hybrid network, which employs the AFDEF fusion strategy, demonstrates strong classification robustness across all four datasets. As is shown in Fig. 13, we provide the visualization of

TABLE X
CLASSIFICATION RESULTS (%) OF DIFFERENT FUSION STRATEGIES ON EACH DATASET

Fusion Strategy	Indian Pines			University of Pavia			Xuzhou			LaoYu River		
	OA	AA	KC	OA	AA	KC	OA	AA	KC	OA	AA	KC
Concatenation	75.72±1.64	67.46±2.56	71.96±1.92	90.71±1.76	86.16±3.02	87.55±2.41	94.76±0.49	91.37±0.80	93.37±0.62	87.48±0.73	70.48±1.31	84.55±0.92
Direct addition	79.43±2.37	68.10±2.76	76.26±2.78	92.40±0.73	89.83±1.20	89.82±1.01	95.01±0.39	91.62±0.94	93.68±0.50	88.34±1.04	72.79±1.16	85.85±1.37
AFDEF	**87.06±1.10**	**84.95±1.46**	**85.25±1.23**	**94.19±0.44**	**91.58±1.21**	**92.28±0.58**	**95.49±0.26**	**92.05±0.67**	**94.29±0.32**	**90.15±0.88**	**81.74±2.94**	**88.10±1.08**

The bold values indicate the highest value in the corresponding evaluation metric.

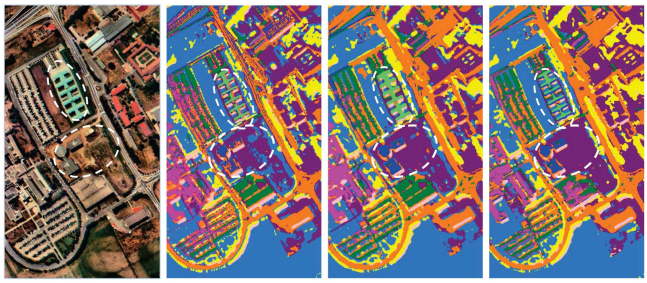

Fig. 13. Visualization of different feature fusion strategies on the University of Pavia dataset. (a) False-color image. (b) Concatenation. (c) Direct addition. (d) AFDEF.

different feature fusion strategies for all areas on the University of Pavia dataset, as the University of Pavia dataset covers typical fine and smooth regions. The classification maps demonstrate that the AFDEF module can achieve mutual enhancement of advantageous features derived from different network structures, thereby enabling robust classification of different regions in HSI.

VI. CONCLUSION

In this article, we systematically compared the performance of different network structures (i.e., CNN, GCN, and HGCN) in HSI classification. The experimental results demonstrate that the CNN structure achieves better classification results in fine regions while GCN and HGCN structures perform better in smooth regions. In addition, the HGCN structure exhibits better representation ability for complex spectral–spatial information compared to GCN. However, features extracted by single-network structures are always partial. To achieve robust classification across different regions, we propose an advantageous feature enhancement hybrid network, named DF2Net, for HSI classification. DF2Net consists of two subnetworks: S2HGCN is employed for capturing long-range correlations and complex high-order correlations and S2CN for pixel-level local information extraction, along with an AFDEF module. Through the AFDEF module, DF2Net achieves mutual enhancement of advantageous features derived from different network structures, thereby improving the classification robustness of HSI.

REFERENCES

[1] J. Feng, J. Zhang, T. Li, and Y. Zhang, "Spectral–spatial joint feature extraction for hyperspectral image based on high-reliable neighborhood structure," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9609–9623, Sep. 2021.

[2] B. Lu, P. D. Dao, J. Liu, Y. He, and J. Shang, "Recent advances of hyperspectral imaging technology and applications in agriculture," *Remote Sens.*, vol. 12, no. 16, 2020, Art. no. 2659.

[3] P. Duan, Z. Xie, X. Kang, and S. Li, "Self-supervised learning-based oil spill detection of hyperspectral images," *Sci. China Technol. Sci.*, vol. 65, no. 4, pp. 793–801, 2022.

[4] P. Duan, X. Kang, P. Ghamisi, and S. Li, "Hyperspectral remote sensing benchmark database for oil spill detection with an isolation forest-guided unsupervised detector," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Apr. 2023, Art. no. 5509711.

[5] W. Sun, G. Yang, J. Peng, and Q. Du, "Lateral-slice sparse tensor robust principal component analysis for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 1, pp. 107–111, Jan. 2019.

[6] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.

[7] Y. Y. Tang, Y. Lu, and H. Yuan, "Hyperspectral image classification based on three-dimensional scattering wavelet transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2467–2480, May 2015.

[8] Q. Wang, Y. Gu, and D. Tuia, "Discriminative multiple kernel learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 3912–3927, Jul. 2016.

[9] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, Oct. 2010.

[10] P. Duan, P. Ghamisi, X. Kang, B. Rasti, S. Li, and R. Gloaguen, "Fusion of dual spatial information for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7726–7738, Sep. 2021.

[11] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.

[12] 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.

[13] 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, 2015.

[14] X. Lu, D. Yang, F. Jia, Y. Yang, and L. Zhang, "Hyperspectral image classification based on multilevel joint feature extraction network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10977–10989, Oct. 2021.

[15] M. Esmaeili, D. Abbasi-Moghadam, A. Sharifi, A. Tariq, and Q. Li, "ResMorCNN model: Hyperspectral images classification using residual-injection morphological features and 3DCNN layers," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 219–243, Oct. 2024.

[16] Y. Ding, X. Zhao, Z. Zhang, W. Cai, and N. Yang, "Multiscale graph sample and aggregate network with context-aware learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4561–4572, Apr. 2021.

[17] H. Hu, M. Yao, F. He, and F. Zhang, "Graph neural network via edge convolution for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Sep. 2022, Art. no. 5508905.

[18] 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.

[19] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 4959–4962.

[20] M. He, B. Li, and H. Chen, "Multi-scale 3D deep convolutional neural network for hyperspectral image classification," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 3904–3908.

[21] S. M. M. Nejad, D. Abbasi, A. Sharifi, N. Farmonov, K. Amankulova, and M. Laszlz, "Multispectral crop yield prediction using 3D-convolutional neural networks and attention convolutional LSTM approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 254–266, Nov. 2023.
[22] 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.
[23] Y. Ding, Y. Guo, Y. Chong, S. Pan, and J. Feng, "Global consistent graph convolutional network for hyperspectral image classification," *IEEE Trans. Instrum. Meas.*, vol. 70, Feb. 2021, Art. no. 5501516.
[24] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Represent.*, 2017, pp. 1–14.
[25] 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.
[26] 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.
[27] 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.
[28] Y. Dong, Q. Liu, B. Du, and L. Zhang, "Weighted feature fusion of convolutional neural network and graph attention network for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 31, pp. 1559–1572, Jan. 2022.
[29] H. Zhou, F. Luo, H. Zhuang, Z. Weng, X. Gong, and Z. Lin, "Attention multihop graph and multiscale convolutional fusion network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Apr. 2023, Art. no. 5508614.
[30] M. Liu, Z. Chai, H. Deng, and R. Liu, "A CNN-transformer network with multiscale context aggregation for fine-grained cropland change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4297–4306, May 2022.
[31] Q. Yu, W. Wei, D. Li, Z. Pan, C. Li, and D. Hong, "HyperSINet: A synergetic interaction network combined with convolution and transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Feb. 2024, Art. no. 5508118.
[32] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao, "Hypergraph neural networks," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 3558–3565.
[33] Y. Gao, Y. Feng, S. Ji, and R. Ji, "HGNN$^+$: General hypergraph neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3181–3199, Mar. 2023.
[34] Y. Tian et al., "Fully-weighted HGNN: Learning efficient non-local relations with hypergraph in aerial imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 191, pp. 263–276, 2022.
[35] Z. Ma, Z. Jiang, and H. Zhang, "Hyperspectral image classification using feature fusion hypergraph convolution neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 2022, Art. no. 5517314.
[36] Q. Wang, J. Huang, T. Shen, and Y. Gu, "EHGNN: Enhanced hypergraph neural network for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, Mar. 2024, Art. no. 5504405.
[37] Y. Yan et al., "Learning multi-granular hypergraphs for video-based person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2899–2908.
[38] Q. Liu, L. Xiao, F. Liu, and J. Huan, "SSCDenseNet: A spectral-spatial convolutional dense network for hyperspectral image classification," *Acta Electronica Sinica*, vol. 48, no. 4, pp. 751–762, 2020.
[39] K. Zhou, L. Chen, and X. Cao, "Improving multispectral pedestrian detection by addressing modality imbalance problems," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 787–803.
[40] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.
[41] X. Lu, J. Zhang, D. Yang, L. Xu, and F. Jia, "Cascaded convolutional neural network-based hyperspectral image resolution enhancement via an auxiliary panchromatic image," *IEEE Trans. Image Process.*, vol. 30, pp. 6815–6828, Jul. 2021.
[42] K. Tan, F. Wu, Q. Du, P. Du, and Y. Chen, "A parallel Gaussian–Bernoulli restricted Boltzmann machine for mining area classification with hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 2, pp. 627–636, Feb. 2019.

Qingwang Wang (Member, IEEE) received the B.E. and Ph.D. degrees in electronics and information engineering, and information and communication engineering from the Harbin Institute of Technology, Harbin, China, in 2014 and 2020, respectively.

From 2020 to 2021, he was a Senior Engineer with Huawei Technology Company, Ltd., in the area of autonomous driving. Since 2022, he has been with the Kunming University of Science and Technology, Kunming, China. His research interests include machine learning and its application to remote sensing data analysis, autonomous driving, and edge calculation, focusing on developments of deep learning, broad learning, and graph convolutional-neural-network-based methods for RGB-T images, hyperspectral images, LiDAR data, and multispectral LiDAR point clouds.

Jiangbo Huang(Graduate Student Member, IEEE) received the B.S. degree in information and computing science from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2022. He is currently working toward the M.S. degree in computer systems organization with the School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China.

His research interests include machine learning, graph neural networks, and hypergraph neural networks.

Yuanqin Meng received the double B.E. degrees in mechatronics from Tongji University, Shanghai, China, and Zittau/Görlitz University of Applied Science, Zittau, Germany, in 2013, and the M.S. degree in electrical engineering and computer science from the Technical University of Berlin, Berlin, Germany, in 2016. She is currently working toward the Ph.D. degree in computer science and technology with the Kunming University of Science and Technology, Kunming, China.

From 2017 to 2021, she was a Software Developer with ALTRAN Deutschland S.A.S & CO.KG, Wolfsburg, Germany. Her research interests include deep learning, medical data processing, image segmentation, and their applications in clinical practice.

Tao Shen (Member, IEEE) received the Ph.D. degree in electrical engineering from the Illinois Institute of Technology, Chicago, IL, USA, in 2013.

He is currently a Professor with the Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China. His research interests include intelligent perception and computation, artificial intelligence, blockchain, and industrial Internet.