Multiview Hierarchical Network for Hyperspectral and LiDAR Data Classification

Yishu Peng, *Member, IEEE*, Yuwen Zhang, *Student Member, IEEE*, Bing Tu, *Member, IEEE*,
Chengle Zhou, *Student Member, IEEE*, and Qianming Li, *Member, IEEE*

***Abstract*—In recent times, multisource remote sensing technology [e.g., hyperspectral image (HSI) and light detection and ranging (LiDAR) data] has been widely used in urban land-use recognition owing to its high classification effectiveness compared to using only single-source data. In this study, a multiview hierarchical network (MVHN) technique is developed for HSI and LiDAR data classification, which conducts the following execution procedures. First, based on the a preset band step length, the original HSI is sampled and divided into multiple groups with exactly the same number of bands to obtain spectral features. Then, principal components analysis is performed on the raw HSI to extract the first principal components (PCs) that meet the size of the LiDAR image. The Gabor filters are applied to the PCs and LiDAR to capture spatial details (i.e., textural features) of scenes. Specifically, a stacking mechanism is employed to generate fusion features once the above features are available. Next, a three-dimensional ResNet-like deep CNN is designed to extract the spectral–spatial information of the fusion feature. Finally, majority-voting is introduced into the classification results of the network trained using each fusion feature to achieve high-confidence final results. Experiments on three well-known HSI and LiDAR datasets (i.e., Houston, MUUFL, and Trento datasets) demonstrate the effectiveness of the proposed MVHN method compared to state-of-the-art comparable classification methods.**

***Index Terms*—Classification, Gabor feature, hyperspectral image (HSI), light detection and ranging (LiDAR), multisource remote sensing, residual network.**

Manuscript received September 1, 2021; revised December 5, 2021; accepted January 13, 2022. Date of publication January 21, 2022; date of current version February 7, 2022. 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 author: Bing Tu.)*

Yishu Peng, Yuwen Zhang, Chengle Zhou, and Qianming Li are with the School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang 414000, China (e-mail: lovepys@hnist.edu.cn; yuwen_zhang@vip.hnist.edu.cn; chengle_zhou@foxmail.com; lqm188@csu.edu.cn).

Bing Tu is with the School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang 414000, China, and also with the Guangxi Key Laboratory of Cryptography and Information Security, Guilin University of Electronic Technology, Guilin 541000, China (e-mail: tubing@hnist.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3144312

I. Introduction

URBAN land classification is an exciting topic in the remote sensing community. Accurate land classification not only improves the quality of urban land surveying and mapping but also helps managers to plan optimal urban layouts. With the rapid development of satellite sensor technology, multisource remote sensing data, like hyperspectral image (HSI) and light detection and ranging (LiDAR) images, are widely used in many practical applications. HSI is acquired by an imaging spectrometer, which provides a large amount of narrow-band spectral information from the visible spectrum to the infrared spectrum for each pixel to generate a complete and continuous spectrum curve [1], thus improving its discriminant ability of ground coverings compared with the RGB images. Thanks to the distinguishing advantages of HSI, several researchers have focused on various meaningful research directions in the field of hyperspectral image processing, such as classification [2]–[7], anomaly detection [8]–[12], and segmentation [13]–[16]. In fact, HSI aims to use the sensor to receive electromagnetic waves reflected from the ground to characterize the properties of the ground material. In contrast, LiDAR uses pulsed lasers to focus on the distance information of ground targets. Similarly, a large amount of relevant research work on the semantic segmentation, classification, and detection of LiDAR data can be found in previous studies [17]–[29]. The digital surface model (DSM) that is the image version obtained by prepossing of the LiDAR point cloud data, and as one of the LiDAR data has been successfully applied to complex scenes to obtain the elevation information of objects as it is not restricted by external conditions (e.g., weather and light). In particular, high-precision ground space information can still be obtained even under severe environmental conditions [30], [31]. Therefore, combining the respective advantages of HSI and LiDAR is a very suitable approach for fine classification of urban land.

In order to create complementary features between HSI and LiDAR data, researchers have devised many meaningful integration strategies such as feature-level and decision-level fusion. Feature-level fusion (such as feature stacking) is a commonly used fusion method, which aims to fuse the features of HSI and LiDAR images at the image feature level. In this work, a feature stacking-based feature-level fusion method is adopted to obtain the primary spectral–spatial advantage at the feature level. Even so, the following two aspects must be addressed: 1) The deep spatial–spectral information of HSI and LiDAR data cannot be captured intuitively [32]; 2) simply feature-stacking

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

inevitably results in information redundancy, which triggers the Hughes phenomenon, especially when the number of training samples is very limited [33]. Therefore, dimensionality reduction is unavoidable for a feature-level fusion method based on stacking. Principal component analysis (PCA) is a commonly used method of dimensionality reduction. PCA reduces the data dimensionality by extracting the principal components (PCs) of HSI [34], [35]. Furthermore, the kernel PCA (KPCA) [36] was applied to reduce the dimensionality of HSI and EP features. Compared with the traditional PCA algorithm, KPCA can effectively extract nonlinear PCs of HSI based on kernel tricks in high-dimension space. However, most dimension reduction methods suffer from the problem of information loss. Thus, choosing the best dimension to effectively replace the original data is still a challenge. Decision-level fusion is also a frequently used fusion method for HSI classification. The difference between the feature-level fusion method and the decision-level fusion method is that the former is often used before the classification task, while the latter is performed after the classification task. As an example of decision-level fusion, soft and hard voting [37] were used to fuse various classification results so as to obtain a more reliable final result. Zhang *et al.* [38] used various classifiers [i.e., K-nearest neighbor (KNN) [39], support vector machines (SVM) [40]–[42], and random forest (RF) [43]] to obtain their own individual classification results, followed by a combination strategy of majority- and weighting voting being performed on the classification results to achieve the final result. In addition, other decision fusion methods have been proposed to solve the imbalance between HSI features and LiDAR features [44], [45].

With the continuous development of the deep learning in the processing domain of remote sensing data, the classification accuracy of ground coverings including urban and agricultural land is being gradually improved [46], [47]. Xu *et al.* [48] proposed a two-branch CNN model (TBCNN), in which one branch was used to extract the spectral features of HSI and another branch was introduced into LiDAR data to obtain its spatial features. At the end of operations, a fully connected layer was employed to achieve the classification result after concatenating the features from both sources. A coupled CNN (CPCNN) model was proposed [49] based on the improvement of the TBCNN model. Specifically, by sharing the weights of the last two convolution layers, the model can not only reduce the training parameters of the network but also guide the two CNNs to learn from each other. In [50], a hierarchical random walk layer was designed under the CNN framework to effectively fuse HSI and LiDAR features, which significantly improved the classification accuracy of HSI and LiDAR data. Zhang *et al.* [51] proposed a multiscale patch-to-patch CNN model (PToP-CNN) to obtain the deep fusion features of HSI and LiDAR data. More recently, many three-dimensional (3-D) CNN models have been developed to fully extract the spatial–spectral features of HSI [52]–[54]. For instance, Swalpa *et al.* [55] proposed a hybrid spectral CNN (HybridSN) model by integrating two-dimensional (2-D) CNN and 3-D CNN into one operation line. The advantage of HybridSN is that the 3-D CNN can be used to extract the spatial–spectral features and the adoption of 2-D CNN can directly reduce the number of parameters. However, shallow CNN often cannot fully extract the features of HSI and LiDAR data, while deeper CNN are prone to network degeneration. In addition, CNN usually does not have the ability to accurately describe objects having changeable features. When the direction of the test object changes frequently, the classification accuracy will become unsatisfactory.

In this study, a multiview hierarchical network (MVHN) technique is developed for HSI and LiDAR data classification. MVHN involves the following execution procedures: First, the raw HSI is divided into various groups with exactly the same number of bands based on preset band step length to obtain the spectral features. Next, PCA is performed on the raw HSI to extract the first PCs that meet the size of the LiDAR image. The Gabor filters are applied to the PCs and LiDAR to capture spatial details (i.e., textural features) of the scene. Specifically, a stacking mechanism is employed to generate the fusion feature once the aforementioned features are available. Next, a 3-D ResNet-like deep CNN is designed to extract spatial–spectral information of the fusion feature. Finally, majority voting is introduced into the classification results of the network trained using each fusion feature to achieve high-confidence final results. The main contributions of our work can be summarized as follows.

1) A multiview strategy based on preset band step length is first introduced into HSI and LiDAR data classification. Its advantage lies in completely avoiding the loss of HSI information, while achieving dimensionality reduction, and maintains the physical meaning of the spectral features. Specifically, the information complementation of multiview strategy has been comprehensively utilized using a simple and effective majority-voting mechanism.
2) Gabor filters driven by HSI and LiDAR data are employed to extract Gabor features improving the CNN model. It is found that Gabor texture features can enhance the adaptability of the CNN model to changes of sample direction.
3) The 3-D ResNet-like deep CNN is designed by embedding residual units in a 3-D CNN. The 3-D residual network can capture spatial-spectral joint features having more semantic information compared with traditional 3-D CNN networks composed of shallow layers.

The rest of this article is organized as follows. Section II describes the proposed MVHN method. Section III presents an experimental analysis of the MVHN, including the setup and comparison with various state-of-the-art methods on the Houston, MUUFL, and Trento datasets. Finally, Section IV concludes this article.

II. Description of the Proposed Approach

In this section, we introduce the overall architecture of the proposed MVHN method, including the framework of the proposed method, multiview voting strategy, Gabor features extraction, and 3-D residual network for classification, which is described in detail as follows.

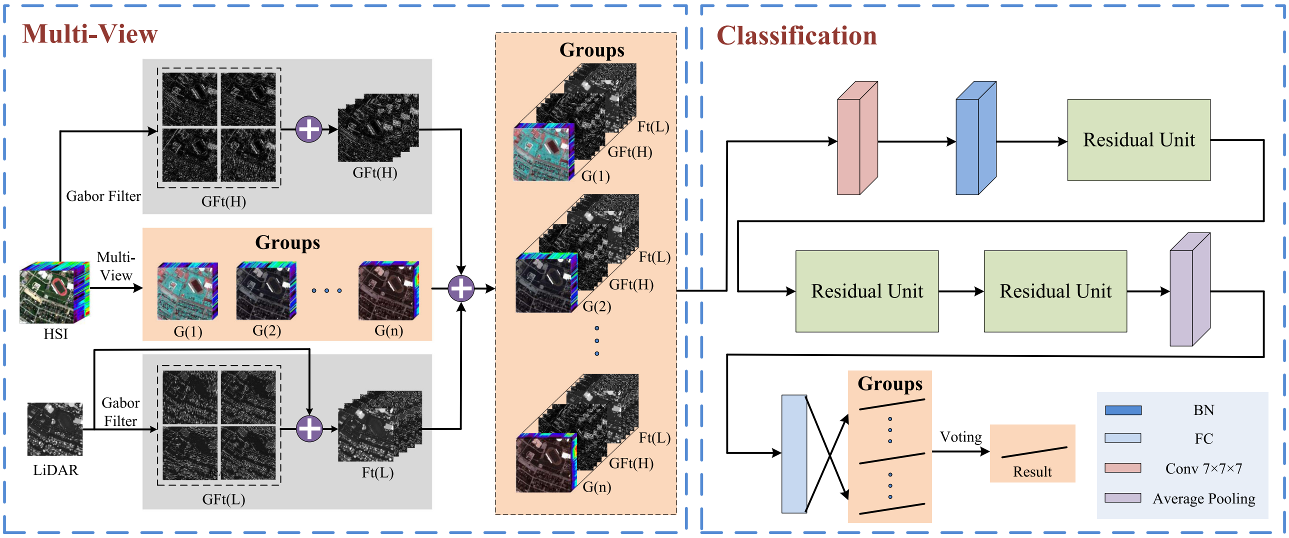


Fig. 1. Framework of proposed MVHN method, where the left subframework is the process of multiview strategy and feature level-based data fusion, and the CNN feature extraction and decision level-based results fusion for classification is presented in the right subframework. Note that the purple plus sign node in the left subframework refers to the data concatenation operation.

A. Framework of Proposed Method

As shown in Fig. 1, the specific framework of the proposed method is mainly composed of three parts. At the beginning, the original HSIs are partitioned-off multiple groups of local HSI with the same number of bands at intervals. Then, the PCA algorithm is applied to the original HSI to extract its first PCs. The Gabor filters are applied separately to the first PCs and LiDAR data to determine the Gabor features. Next, the multiview features of HSI are fused with the Gabor features as well as the LiDAR data in a stacked manner and are fed to the 3-D residual network to extract the corresponding spectral–spatial feature. Each feature in the various groups is classified by the full connection layer. Finally, to obtain a robust classification result, probability-based majority voting is applied to the classification results obtained from all group features. Specifically, we count the average probability of each prediction class for each test sample and take the label corresponding to the highest probability value as the final prediction result.

B. Multiview Voting Strategy

Owing to the tremendous amount of spectral information in the hyperspectral image, the data storage space requirements are large, and the network training time is increases considerably if all the HSI data are sent to the CNN together. Furthermore, the information between adjacent frequency bands is highly correlated and has a high information repetition rate, and, thus, there is serious redundancy between the data. For example, consider an HSI $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_N\} \in \mathbb{R}^{N\times H\times W}$. N is the number of bands. H and W stand for the height and width of the image, respectively. When $\mathbf{X}_1$ is used as the comparison object, there is a higher difference in spectral information between $\mathbf{X}_1$ and $\mathbf{X}_N$. The difference between $\mathbf{X}_1$ and $\mathbf{X}_2$ is extremely small, and there is duplication of information. Therefore, dimensionality reduction tools such as PCA and linear discriminant analysis have emerged in the field of hyperspectral data processing and are favored by scholars [56]. However, these dimensionality reduction tools generally suffer from information loss. To address this problem, as shown in Fig. 2, two dimensionality reduction schemes are designed in this work: 1) The HSI is divided into several groups in continuous bands (GHSICB); and 2) the HSI at equal intervals into multiple groups of local hyperspectral data (GHSIIB).

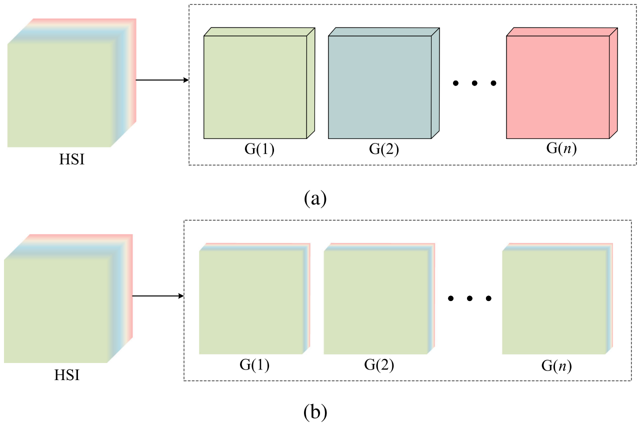


Fig. 2. Outline of two grouping strategies for HSI. (a) GHSICB. (b) GHSIIB.

As show in Fig. 2(a), for GHSICB method, assume that the HSI $\mathbf{X}$ is divided into several groups $\{\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_n\}$ in continuous bands. Here, each group shares the same number of bands. Then, each group can be expressed as follows:

$$\begin{aligned} &\mathbf{G}_1 = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m\} \\ &\mathbf{G}_2 = \{\mathbf{X}_{(m+1)}, \mathbf{X}_{(m+2)}, \ldots, \mathbf{X}_{2m}\} \\ &\qquad\vdots \\ &\mathbf{G}_n = \{\mathbf{X}_{(N-m+1)}, \mathbf{X}_{(N-m+2)}, \ldots, \mathbf{X}_N\} \end{aligned} \tag{1}$$

where *n* represents the number of groups, while *m* represents the number of bands in a group and also stands for the last band of the first group being in the band position of HSI. As one of the bands of HSI, although there is a certain correlation among the bands in different groups, the correlation between the bands within a group is stronger. Specifically, the GHSICB method makes the data within each group more relevant and reduces the correlation between the groups. Therefore, the amount of information contained in different groups is quite different, which has different degrees of impact on the classification results. It is worth noting that as some groups have less spectral information in the bands, the classification accuracies of these groups are significantly lower than that of other groups, thus inhibiting the performance of voting decision fusion.

As show in Fig. 2(b), the GHSIIB method aims at solving the high result difference of each group caused by the GHSICB method. Like the GHSICB, each group after grouping contains the same number of bands. Specifically, supposing the number of HSI bands are N and the grouping band interval is *n*, which is the same as the number of groups, then each group of partial images has N/n bands. Owing to the interval sampling, the bands in each group are as follows:

$$\begin{aligned}\mathbf{G_1} &= \{\mathbf{X_1}, \mathbf{X}_{n+\mathbf{1}}, \ldots, \mathbf{X}_{\mathbf{N}-n+\mathbf{1}}\}\\ \mathbf{G}_2 &= \{\mathbf{X}_2, \mathbf{X}_{n+2}, \ldots, \mathbf{X}_{N-n+2}\}\\ &\vdots\\ \mathbf{G}_n &= \{\mathbf{X}_n, \mathbf{X}_{2n}, \ldots, \mathbf{X}_N\}\end{aligned} \tag{2}$$

where $\mathbf{X}_i$, *i* is denoted the spectral position of the original HSI and $\mathbf{G}_j$, *j* represents the group label. In the GHSIIB grouping method, each group of local HSI not only eliminates data redundancy but also decreases correlation between the bands within each group. Specifically, the classification results coming from various groups will be output with small gaps, which prevents low classification accuracy in some groups. Finally, decision fusion is adopted to obtain the final result with strong robustness. Through the experimental comparison presented in Section III, it is found that the result of soft voting is better than that of hard voting. Therefore, soft voting is used in this study to obtain the final classification result.

Given a pixel $\mathbf{x}_{(i,j)} = \{\mathbf{x}^1_{(i,j)}, \mathbf{x}^2_{(i,j)}, ..., \mathbf{x}^K_{(i,j)}\} \in \mathbb{R}^{N\times 1\times 1}$, where (i, j) represents the pixel position at which the sample point is located in the image, K represents the number of groups, and N is the number of HSI bands, which is sent to the network first to get the output of the network and then transform it into a normalized probability output through the softmax layer. The expression is as follows:

$$\mathbf{p}^n_{(i.j)} = \mathrm{softmax}[\mathrm{Net}(\mathbf{x}^n_{(i,j)})](\mathbf{p}^n_{(i,j)} \in \mathbb{R}^{1\times C}) \tag{3}$$

where $\mathrm{Net}(\cdot)$ represents the CNN network, $\mathbf{x}^n_{(i,j)}$ is the $\mathbf{x}_{\ (i,j)}$ in n th group, and C refers to the number of class corresponding to the sample point. $\mathbf{p}^n_{(i,j)}$ is the possibility output corresponding to C class of the $\mathbf{x}_{\ (i,j)}$ in n th group. After obtaining all the possibility output of K groups $\{\mathbf{p}^1_{(i,j)}, \mathbf{p}^2_{(i,j)}, ..., \mathbf{p}^K_{(i,j)}\}$, the average value over each class is calculated. Then, the index corresponding to the value with the largest probability average can be determined as the final classification result of this pixel. The calculation process is as follows:

$$\mathbf{y}_{(i,j)} = \arg\max\left(\frac{\sum\limits_{n=1}^{K}\mathbf{p}^n_{(i,j)}}{K}\right) \tag{4}$$

where $\mathbf{y}_{(i,j)}$ is the final prediction label of the pixel, which has strong robustness, because it is the result of the decision fusion.

C. Gabor Features Extraction

As the CNN is very sensitive to the change of object position and geometry, the classification effect of CNN will be unsatisfactory when the direction of the object changes frequently. Furthermore, given that a single LiDAR-based DSM data that is the image version by preprocessing to the LiDAR point cloud data has only one band, simply using this data to fuse with HSI data has no obvious impact on the final result. Therefore, Gabor filters are used to extract the multidirectional features of the LiDAR data and the first PCs of HSI, which not only improves the adaptability of CNN to the rotation and deformation of features but also further expands the diversity of LiDAR data so that the LiDAR data can be used more fully. The essence of the Gabor transform is a short-time Fourier transform and its window function is a Gaussian kernel function, so the Gabor filter can extract features in various directions in the frequency domain. The definition of a 2-D Gabor function is as follows:

$$\begin{aligned}\mathrm{g}_{(\lambda,\theta,\psi,\sigma,\gamma)}(x, y) &= \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right)\\ x' &= \left(x - \frac{m+1}{2}\right)\cos\theta + \left(y - \frac{n+1}{2}\right)\sin\theta\\ y' &= \left(x - \frac{m+1}{2}\right)\sin\theta + \left(y - \frac{n+1}{2}\right)\cos\theta\end{aligned} \tag{5}$$

where m and n are the size of the Gabor filter, λ and θ, respectively, represent the wavelength of the sine function and the direction of the Gabor kernel function. ψ denotes the phase shift, while γ and σ are spatial aspect ratio and the standard deviation of the Gaussian function, respectively.

Regarding the specific feature extraction process, first, PCA is used to reduce the dimensionality of the original HSI data to obtain its PCs. Subsequently, in order to enable the CNN to have rotation invariance without adding too much redundant information, Gabor filters in 4 directions (i.e., 0, 45, 90, and 135 degrees) are designed in our approach. By applying the aforementioned Gabor filters to the PCs of HSI and LiDAR data, respectively, Gabor features in various directions are formed. Finally, concatenating the Gabor features of each direction, Gabor-HSI features and Gabor-LiDAR features are obtained. By using the Gabor filter, the classification accuracy of the network is significantly improved. The specific experimental results will be shown in Section III.

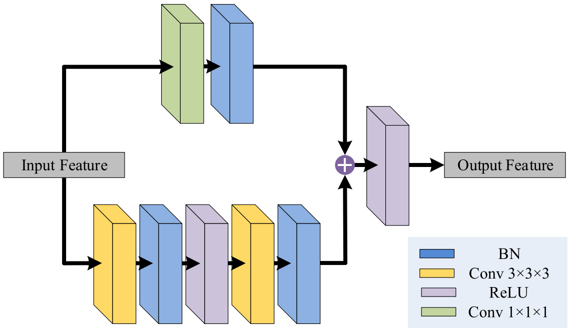


Fig. 3. Structure of residual uint in CNN.

D. 3-D Residual Network for Classification

In order to make full use of the spectral–spatial information of data, i.e., HSI and LiDAR, each pixel is expanded into a 3-D neighboring cube so as to take into account the spatial spectrum information of the feature. On the network side, a 3-D ResNet-like deep CNN is used to extract the spectral–spatial features of the input data. At the beginning of the network, we use a 7×7×7 convolutional layer to initially capture the spectral–spatial features of the input data and a batch normalization (BN) layer to accelerate the convergence of the network. Then, three residual units are introduced in sequence to extract features with more semantic information, and its structure is shown as Fig. 3. Specifically, each of them contains two 3×3×3 convolutional layers for three residual units. For the first residual unit, the stride of the two convolutional layers are both 1, and at the same time, the number of channels of the feature map is not changed. For the last two residual units, the stride of the first convolutional layer will be set to 2, which has the effect of downsampling, and the stride of the second convolutional layer is still 1. It is worth noting that when the size of the feature map is reduced by the downsampling size, the number of channels will be correspondingly expanded to twice the original size. After extracting the spectral–spatial features, a global average pooling layer is applied to transform the feature map into a block of 1×1×1, and then the block is flattened and sent to the fully connected layer (FC) (that includes the softmax layer) to complete the task of classification. Note that we inserted the dropout layer after the FC layer to prevent the overfitting phenomenon caused by less training data, and the parameter of dropout is set to 0.4 in this work.

III. Experimental and Analysis

In this section, three well-known hyperspectral and LiDAR datasets (i.e., Houston, MUUFL, and Trento datasets) are used to verify the effectiveness of the proposed MVHN method in terms of classification. Three commonly used evaluation metrics, i.e., average accuracy (AA), overall accuracy (OA), and kappa coefficient (Kappa), are applied to evaluate the performance between competitive methods and the proposed MVHN method objectively. All programming in the proposed MVHN method is completed on the Python 3.7 platform. The network in this work is built based on the PyTorch framework. PyTorch is an open-source machine learning framework that can not only achieve powerful GPU acceleration but also support dynamic neural networks. All our experiments are performed on a personal computer having a Windows 10 operation system, Intel Core i7-7800X CPU, 32-GB RAM, and a NVIDIA GeForce RTX 1080 Ti GPU.

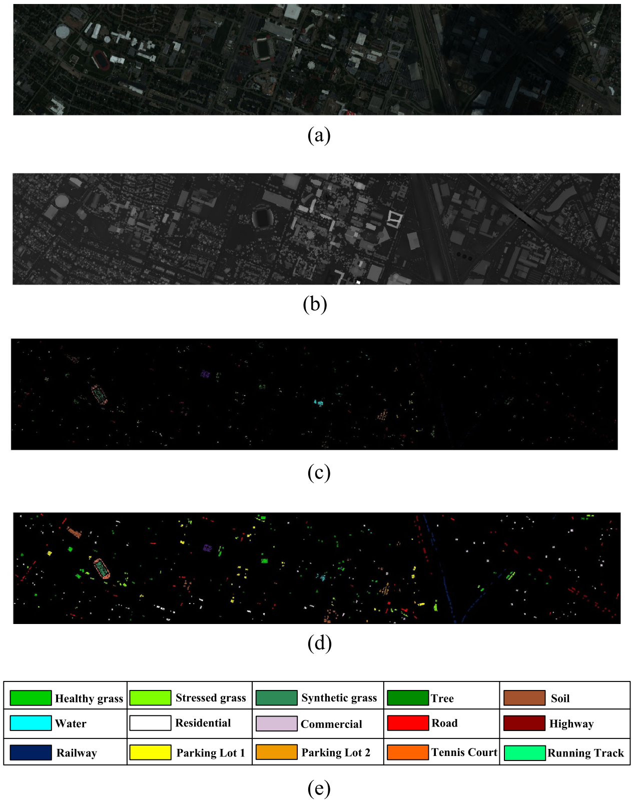
TABLE I
Number of Training, Validation, and Testing Samples for the Houston Dataset

Class		Number of Samples		
No	Name	Training	Validation	Testing
1	Health grass	198	158	895
2	Stressed grass	190	160	904
3	Synthetic grass	192	75	430
4	Trees	188	158	898
5	Soil	186	158	898
6	Water	182	21	122
7	Residential	196	160	912
8	Commercial	191	158	895
9	Road	193	159	900
10	Highway	191	155	881
11	Railway	181	158	896
12	Parking lot 1	192	156	885
13	Parking lot 2	184	43	242
14	Tennis Court	181	27	220
15	Running Track	187	28	445
	Total	2832	1774	10423

A. Datasets

1) Houston Dataset: The Houston dataset was created at the University of Houston campus and the neighboring urban area by the National Natural Science Foundation (NSF)-funded Center for Airborne Laser Mapping. This dataset is composed of hyperspectral data and LiDAR DSM data that are the result of preprocessing of LiDAR data, and the spatial resolutions of both these data sources are 2.5 m each. The hyperspectral data has 349×1905 pixels, including 144 bands from 380 to 1050 nm, and same with the hyperspectral data, LiDAR DSM data also has 349×1905 pixels. There are 15 classes included in the Houston dataset, and Table I lists the number of samples in the training set, validation set, and testing set applied in the experiment. It is worth noting that all the sets are randomly selected. Fig. 4 shows the hyperspectral pseudocolor image of the Houston dataset, the LiDAR DSM image, and the ground truth map of training and testing samples.

2) MUUFL Dataset: The MUUFL dataset was collected in November 2010 at the campus of the University of Southern Mississippi Gulfport by Optech, International. With the simultaneous use of Gemini LiDAR and CASI-1500 while flying in a single plane, the hyperspectral image data and LiDAR data were obtained together, both the HSI data and LiDAR-based DSM data have 325×220 pixels. The HSI has 64 bands and a spatial resolution of 1 m. There are 11 classes in the MUUFL dataset. Table II presents the number of samples for each class in the training set, validation set, and testing set used in our experiments. Fig. 5 presents the hyperspectral pseudocolor image and LiDAR-based DSM image of the MUUFL dataset,



(a)

(b)

(c)

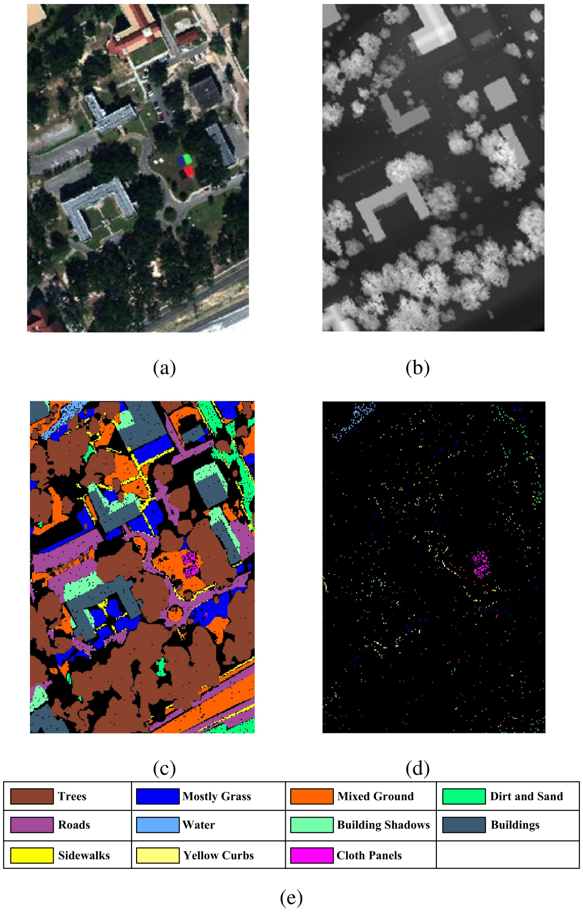
(d)

(e)

Fig. 4. Houston dataset. (a) Pseudocolor image for HSI. (b) LiDAR image. (c) Training set in ground truth. (d) Testing set in ground truth. (e) Color coding for each class.

TABLE II
NUMBER OF TRAINING, VALIDATION, AND TESTING SAMPLES FOR THE MUUFL DATASET

Class		Number of Samples		
No	Name	Training	Validation	Testing
1	Trees	100	3472	19674
2	Mostly Grass	100	625	3545
3	Mixed Ground	100	1017	5765
4	Dirt and Sand	100	259	1467
5	Roads	100	988	5599
6	Water	100	55	311
7	Building Shadows	100	320	1813
8	Buildings	100	921	5219
9	Sidewalks	100	193	1092
10	Yellow Curbs	100	12	71
11	Cloth Panels	100	25	144
	Total	1100	7887	44700



(a)

(b)

(c)

(d)

(e)

Fig. 5. MUUFL dataset. (a) Pseudocolor image for HSI. (b) LiDAR image. (c) Testing set in ground truth. (d) Training set in ground truth. (e) Color coding for each class.

TABLE III
NUMBER OF TRAINING, VALIDATION, AND TESTING SAMPLES FOR THE TRENTO DATASET

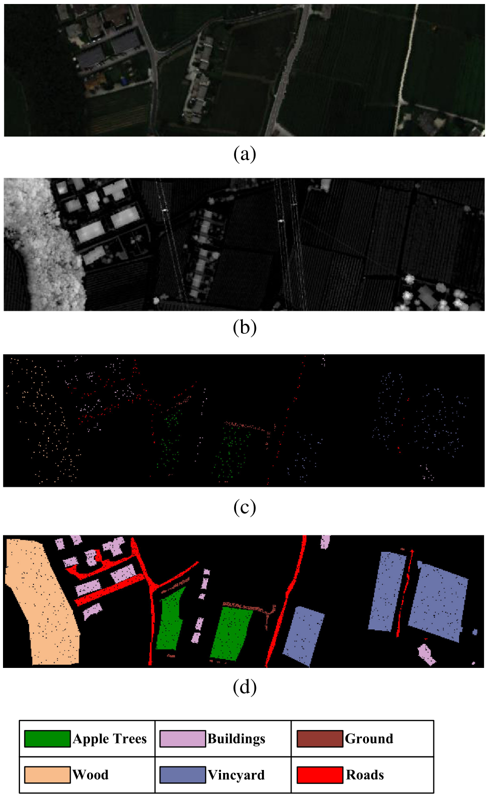
Class		Number of Samples		
No	Name	Training	Validation	Testing
1	Apple Trees	72	594	3368
2	Buildings	69	425	2409
3	Ground	58	63	358
4	Wood	86	1355	7682
5	Vineyard	102	1560	8839
6	Roads	68	466	2640
	Total	455	4463	25296

combined with the ground truth map of training samples and testing samples.

3) Trento Dataset: The Trento dataset was constructed in a rural area south of Trento, Italy. The HSI and LiDAR-based DSM data contained in the dataset consist of 166×600 pixels, in which HSI data contains 63 bands and LiDAR DSM data contains only one band. The Trento dataset has six object classes. Table III lists the number of samples in each class of the dataset, as well as the number of training set, validation set, and test set samples used in the experimental comparison in part D. Fig. 6 shows the hyperspectral pseudocolor map of the Trento dataset, the LiDAR data map, and the ground truth map of the training sample and the testing sample.

B. Analysis of Component on the Proposed Method

1) Analysis of Multiview Voting Strategy: With the application of multiview voting strategy, the robustness of final classification results has been considerably enhanced, especially in the case of small training samples. Table IV presents the classification accuracies of the proposed MVHN method using various grouping strategies (i.e., GHSICB and GHSIIB) on the Houston, MUUFL, and Trento datasets. It can be observed that the GHSIIB strategy


(a)

(b)

(c)

(d)

(e)

Fig. 6. Trento dataset. (a) Pseudocolor image for HSI. (b) LiDAR image. (c) Training set in ground truth. (d) Testing set in ground truth. (e) Color coding for each class.

TABLE IV
CLASSIFICATION PERFORMANCE OF THE PROPOSED MVHN METHOD USING DIFFERENT GROUPING STRATEGIES ON THE VARIOUS DATASETS

	Houston		MUUFL		Trento	
	GHSICB	GHSIIB	GHSICB	GHSIIB	GHSICB	GHSIIB
OA (%)	97.17	97.79	94.07	94.63	98.61	98.63
AA (%)	97.47	98.08	82.02	83.25	97.82	97.35
Kappa	96.94	97.62	92.12	92.87	98.15	98.18

TABLE V
OA (%) OF EACH GROUP OF THE PROPOSED MVHN METHOD USING DIFFERENT GROUPING STRATEGIES ON THE VARIOUS DATASETS

Group	Houston		MUUFL		Trento	
	GHSICB	GHSIIB	GHSICB	GHSIIB	GHSICB	GHSIIB
1	94.66	97.93	92.60	94.33	97.39	98.29
2	94.63	97.13	93.51	94.58	97.86	98.38
3	95.52	97.25	91.78	94.32	97.95	97.32
4	93.03	97.21	93.46	94.34	97.73	98.61
5	94.83	97.28	93.11	94.37	98.67	98.36
6	91.13	97.00	87.73	94.52	95.77	98.32
7	93.32	97.22	88.17	94.52	95.01	98.56
8	90.44	97.24	88.53	94.57	95.04	97.27
9					95.26	97.91

TABLE VI
CLASSIFICATION PERFORMANCE OF THE PROPOSED MULTIVIEW (MV) AND PCA IN VARIOUS DATASETS

	Houston		MUUFL		Trento	
	MV	PCA	MV	PCA	MV	PCA
OA (%)	99.74	99.81	90.15	88.55	99.48	99.47
AA (%)	99.78	99.85	89.78	89.17	99.24	99.22
Kappa	99.72	99.79	87.06	85.07	99.31	99.30

on the proposed MVHN method achieves higher classification performance than the GHSICB strategy on three datasets. This means that there is a high degree of correlation between adjacent bands of hyperspectral data. We also analyzed the classification accuracy of each group of the proposed MVHN method using various grouping strategies on the above datasets. Experimental reports are summarized in Table V. It can be seen that the classification results of each group based on the GHSICB strategy have obvious fluctuations, and it can even be observed that the classification accuracy of some groups is much lower than that of other groups. In comparison, the GHSIIB strategy can better overcome this phenomenon. Therefore, the GHSIIB strategy is selected as the default component of the proposed MVHN method in this work.

2) Analysis of Multiview Versus PCA: In order to verify the effectiveness of the grouping method proposed in this article, a comparative experiment of multiview and PCA methods is designed. Specifically, except for the method of extracting spectral features, the other processes and network framework remain unchanged. In addition, in order to ensure the fairness of the comparison experiment, the number of bands in each group are the same after the spectral feature extraction is performed through the multiview or PCA methods. The experiment was tested on three multisource remote sensing datasets, and the results are shown in Table VI. From the experimental results, it can be seen that for the Houston dataset and Trento dataset, the two methods are almost the same, but for the MUUFL dataset, the multiview method improves OA by 1.5% compared to the PCA method. This shows that the multiview grouping method proposed in this article can achieve better classification performance than the PCA method on the specific dataset, which proves the effectiveness of the method.

3) Analysis of Voting Methods: Two kinds of voting methods, i.e., soft voting and hard voting are compared in this section. The classification performance of soft and hard voting on three multiple-source datasets is summarized in Table VII. It can be seen that although the results of two voting methods are almost the same, in general, soft voting is better than hard voting. This is because hard voting is determined by the number of labels in the classification result, whereas soft voting is determined by the probability of the classification result, and, thus, avoids the situation where the original lower probability label is used as the final classification label. Therefore, the soft voting method is selected as a default setting on the proposed MVHN method based on its better classification accuracies.

4) Analysis of the Neural Network: The problem of the network degradation of the CNN model can be overcome by introducing a self-connected residual structure of the 3-D residual network (3-D ResNet). This means that the CNN model can build deeper hidden layers. Compared with the 3-D CNN, deeper semantic information can be obtained by a 3-D ResNet. In addition, the structure of a 3-D CNN can extract more detailed

TABLE VII
CLASSIFICATION PERFORMANCE OF THE PROPOSED METHOD USING DIFFERENT VOTING METHODS ON THREE MULTISOURCE REMOTE SENSING DATASETS

Method	Metrics	Houston					MUUFL					Trento				
		1%	5%	10%	15%	20%	1%	5%	10%	15%	20%	10	20	30	40	50
Soft Voting	OA(%)	88.17	97.79	98.23	99.60	99.69	88.91	94.60	96.19	97.16	97.67	98.67	99.28	99.37	99.41	99.39
	AA(%)	87.87	98.08	98.14	99.53	99.60	71.41	83.74	87.25	88.92	90.84	97.34	98.71	98.78	99.09	99.15
	Kappa	87.20	97.62	98.09	99.56	99.67	85.35	92.84	94.96	96.24	96.92	98.23	99.03	99.16	99.21	99.19
Hard Voting	OA(%)	84.72	96.73	98.37	99.36	99.53	88.85	94.58	96.11	97.13	97.70	97.74	99.13	99.41	99.36	99.46
	AA(%)	83.68	97.01	98.34	99.31	99.45	70.91	82.75	86.32	88.80	90.71	96.49	97.86	98.79	98.59	99.25
	Kappa	83.46	96.46	98.23	99.31	99.49	85.14	92.81	94.83	96.20	96.96	96.99	98.84	99.22	99.14	99.28

The number of training samples are selected per class from 1% to 20% of the total samples for Houston and MUUFL datasets and per class From 10 to 50 for Trento dataset.

TABLE VIII
OVERALL ACCURACY (%) OF PROPOSED METHOD USING DIFFERENT STRUCTURE OF CNN

	3D CNN	**2D ResNet**	**3D ResNet**
Houston	95.97	97.61	97.79
MUUFL	94.58	94.47	94.66
Trento	96.94	97.62	98.67

spatial–spectral features than the 2-D CNN. Table VIII presents the comparison report of the experimental result of various network structures. Note that the 3-D ResNet and the 2-D ResNet each use three residual units as mentioned in this article. It is difficult to design a 3-D CNN with that number of convolutional layers without the identity connection, and, thus, it contains three convolutional layers instead of the residual units of residual structure in this experiment. As summarized in Table VIII, the 3-D ResNet has achieved the best classification results on the three multisource remote-sensing datasets. This directly proves that the 3-D ResNet being selected as a component of the proposed MVHN method is credible in terms of classification.

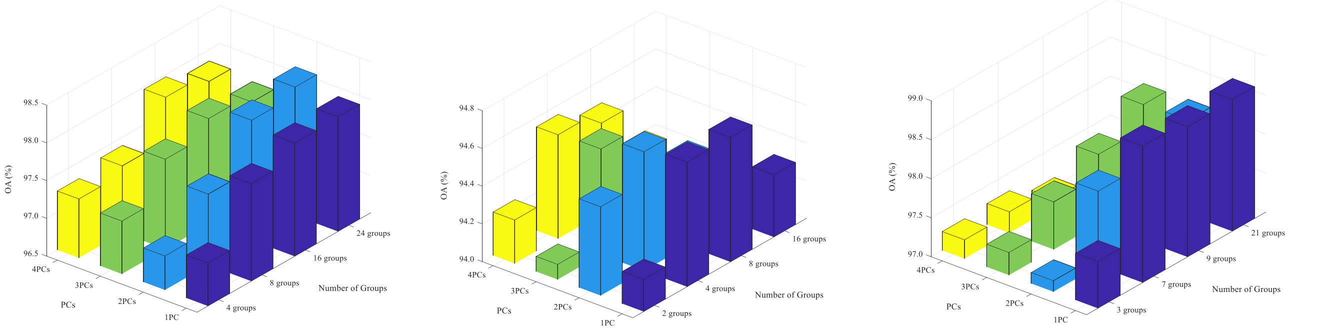
C. Experimental Setup

In this work, in addition to the parameters that have a deterministic effect on the performance of proposed MVHN method, we have set the default settings for some parameters based on experience, as follows: 1) The direction parameter of the Gabor filter on the proposed method is set to {0, 45, 90, 135} by default; 2) the network training epoch is 150 and the network training and testing batch sizes are 100 and 600, respectively; 3) the gradient optimization algorithm uses the stochastic gradient descent algorithm, where the momentum is 0.9 and the weight decay is 1e-4; 4) the learning rate decay decays to 0.1 times the initial value after 75 epochs, and again 0.1 times after 125 epochs; and 5) other parameters, including the number of PCs used for the Gabor filter, the number of groups, and the spatial size of the input 3-D block, are further determined by means of experimentation.

The number of PCs after the application of PCA and the groups of local HSI determine the number of bands for CNN input data. Excessively dense and large amounts of bands can easily cause information redundancy, increasing the occurrence of Hughes phenomenon. Conversely, a low number of bands will lead to insufficient feature extraction and inhibit classification performance. In our experiment, we used the first four PCs to analyze the classification effects of the proposed MVHN method under various grouping situations on the three datasets. Regarding the settings of the common parameters, the size of the input 3-D block size is set to 11×11 and the initial learning rate is set to 0.1. Next, 5% of the labeled data in the Houston and MUUFL datasets are randomly selected as the training samples to train the CNN, while the rest are used for validating and testing. For the Trento dataset, the training set selected has 10 samples per class.

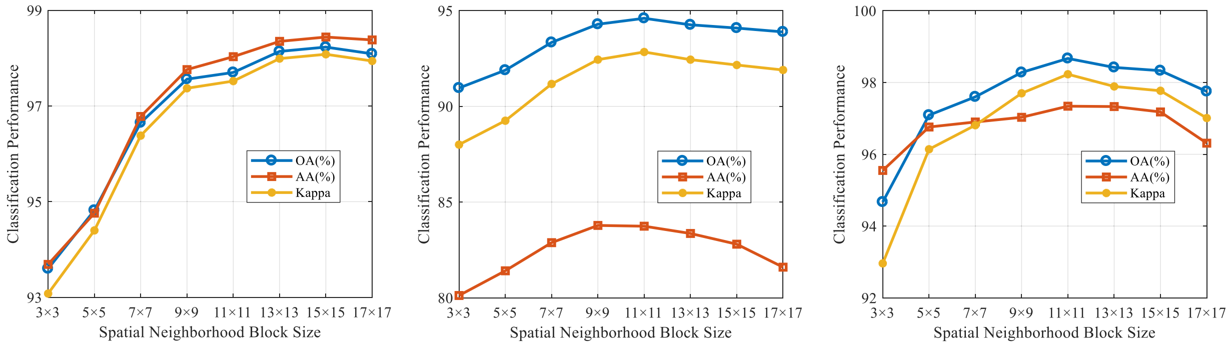
Fig. 7(a) shows the classification performance of the proposed MVHN method using various numbers of both groups and PCs. It can be observed that the combination of four groups and two PCs can maximize the performance of the proposed MVHN method on the Houston dataset. As shown in Fig. 7(b), the classification performance of the proposed MVHN method in the grouping manner of {2, 4, 8, 16} groups is analyzed on the MUUFL dataset. It can be seen that the MVHN method can achieve the best classification accuracy when the number of groups is 8 and the first PCs are selected for filtering. Similarly, on the Trento dataset, we tested the performance of the MVHN method when the grouping situation has $\{3, 7, 9, 21\}$ groups. As the result shows in Fig. 7(c), the best classification performance is obtained when the first PCs are used and 21 groups are grouped. Generally, when the number of PCs is selected as one and two, better classification results can be obtained by the MVHN. This shows that a small number of PCs is more conducive to reducing the redundancy of hyperspectral data. In addition, we emphasize that an appropriate increase in the number of groups can improve the performance expression of multiview decision fusion and effectively enhance the robustness of the MVHN method.

The input block size of the network also plays a decisive role on the classification effect of the proposed MVHN method. Specifically, a relatively large size of input data can obtain more Gabor spatial texture featurs and improve the expressive ability of features. However, a size that is too large will face interference from other samples, resulting in a decrease in classification accuracy. Therefore, this work tests the effect of the proposed method with the input size (from 3×3 to 17×17) on the above datasets, and the results are shown in Fig. 8. On the MUFFL dataset with relatively complex spatial distribution and relatively concentrated sample points, the proposed method using 11×11 input size can achieve the best classification performance (OA = 94.6%). For the Houston dataset, most of the testing sample points of different classes are scattered, making the sample less susceptible to interference from others compared with the MUUFL dataset. Therefore, the optimal input size of the



(a) (b) (c)

Fig. 7. Influence of different number of PCs and groups to proposed method on three remote sensing datasets. (a) Houston dataset. (b) MUUFL dataset. (c) Trento dataset.



(a) (b) (c)

Fig. 8. Influence of spatial neighborhood block size which as input of CNN. (a) Houston dataset. (b) MUUFL dataset. (c) Trento dataset.

proposed method is relatively larger, i.e., 15×15. The Trento dataset has the same test results as the other two datasets. According to experimental analysis, the optimal input size of Trento dataset is 11×11. Based on the above experimental results, the classification accuracies of the MVHN tend to improve when the input size increases and the accuracies will conversely decrease if the size is large enough.

The initial learning rate controls the convergence speed and simultaneously determines the convergence position of the network. A small learning rate will cause the network to converge too slowly, which is not conducive to jumping out of the local extreme point, making the network unable to converge to the global optimal solution. Furthermore, a large learning rate is similarly detrimental to the convergence of the network, even leading to divergence. The classification results of the proposed method are tested when the initial learning rate is set to {0.5, 0.1, 0.05, 0.01, 0.005, 0.001} on the Houston, MUUFL, and Trento datasets. For other parameter settings, the input size is 11×11, the number of PCs is 1, and the number of groups is 8. As presented in Table IX, on the Houston dataset, better classification accuracy of the MVHN is achieved when the learning rate is set between 0.05 and 0.1, and the OA reached more than 98%. Besides, similar results can be obtained on the MUUFL dataset and Trento dataset.

TABLE IX
CLASSIFICATION PERFORMANCE OF DIFFERENT INITIAL NETWORK LEARNING RATE FOR THE PROPOSED METHOD ON THREE MULTISOURCE REMOTE SENSING DATASETS

Dataset	Metrics	Initial learning rate for training					
		0.001	0.005	0.01	0.05	0.1	0.5
Houston	OA(%)	95.26	97.32	97.35	98.03	98.18	97.61
	AA(%)	94.40	97.48	97.58	98.20	98.40	97.86
	Kappa	94.87	97.10	97.14	97.87	98.03	97.42
MUUFL	OA(%)	93.32	93.80	94.10	94.47	94.66	94.50
	AA(%)	79.02	80.84	81.66	82.8	83.15	83.07
	Kappa	91.13	91.77	92.17	92.66	92.92	92.70
Trento	OA(%)	97.12	97.51	97.83	98.56	98.67	96.71
	AA(%)	93.51	95.02	95.83	97.17	97.34	93.79
	Kappa	96.16	96.67	97.11	98.08	95.23	97.17

D. *Experimental Comparison With Competitive Methods*

In order to verify the superiority of the proposed method compared to competitive classification methods of multisource remote-sensing data, experimental comparisons against a variety of methods on the Houston, MUUFL, and Trento datasets were carried out, including traditional methods such as SVM [57], extended multiattribute profiles (EMAP) [58], superpixel-wise PCA approach (SuperPCA) [59], deep encoder-decoder network (EndNet) [60] based on multilayer perceptron (MLP),

TABLE X
CLASSIFICATION PERFORMANCE OF THE SVM, SUPERPCA, EMAP, ENDNET, CPCNN, TBCNN, MVHN-A, AND MVHN-B CLASSIFICATION METHOD ON THE HOUSTON DATASET IN TERMS OF OA, AA, AND KAPPA

No	Class	**Classification performance of various methods**							
		SVM	SuperPCA	EMAP	EndNet	CPCNN	TBCNN	MVHN-A	MVHN-B
1	Healthy grass	98.02	95.53	**99.62**	98.58	90.69	**99.62**	98.39	98.67
2	Stressed grass	98.24	86.18	99.06	99.15	99.72	88.06	**100.0**	**100.0**
3	Synthetic grass	99.80	**100.0**	**100.0**	99.60	99.80	**100.0**	**100.0**	**100.0**
4	Trees	98.19	90.63	99.71	97.54	99.91	99.15	**100.0**	**100.0**
5	Soil	98.24	99.33	99.90	99.43	99.91	99.62	**100.0**	**100.0**
6	Water	99.79	84.52	**100.0**	98.60	**100.0**	**100.0**	**100.0**	**100.0**
7	Residential	95.01	96.66	99.34	93.10	98.13	99.16	**100.0**	**100.0**
8	Commercial	97.42	94.06	**99.62**	98.77	97.34	95.35	99.53	**99.62**
9	Roads	86.33	94.45	95.47	93.30	89.90	**98.96**	98.68	98.87
10	Highway	90.94	95.77	98.46	98.65	91.99	91.02	**100.0**	**100.0**
11	Railway	90.13	95.07	99.04	96.39	89.37	87.95	**100.0**	**100.0**
12	Parking Lot 1	93.66	98.93	97.74	94.72	93.85	88.28	99.71	**99.90**
13	Parking Lot 2	88.85	86.27	98.93	77.54	96.84	100.0	**100.0**	99.65
14	Tennis Court	97.22	**100.0**	97.62	99.60	99.19	99.60	**100.0**	**100.0**
15	Running Track	98.72	**100.0**	**100.0**	99.79	**100.0**	**100.0**	**100.0**	**100.0**
	OA (%)	94.94	94.68	98.88	96.79	95.63	95.44	99.68	**99.74**
	AA (%)	95.38	94.30	98.97	96.32	96.39	96.45	99.75	**99.78**
	Kappa	94.50	94.22	98.79	96.52	95.26	95.05	99.65	**99.72**

and recently emerging methods based on convolutional neural networks, such as TBCNN [48], and CPCNN [49].

1) The SVM method directly classifies the concatenated data of HSI and LiDAR by applying the SVM classifier. The code of SVM is implemented using the LIBSVM toolbox of MATLAB, and fivefold cross-validation is used to train the model with the Gaussian RBF kernel function.
2) EMAP is a method based on morphology algorithm to obtain EMAPs, in order to make full use of the spatial information of fused remote-sensing data. In the experiment, the number of bands with HSI are reduced to 3 under the method of PCA, and then the PCs are extended to 60-band profiles with the use of EMAPs. Similarly, the LiDAR data are extended to 15-band profiles.
3) SuperPCA is a PCA dimensionality-reduction method based on super pixel blocks, which aims to fully consider the spatial information of HSI in the process of dimensionality reduction. Regarding the design of experimental parameters, the number of PCs after PCA dimensionality reduction is 30. We set the number of super pixel blocks to 500 on the Houston dataset, 100 on the MUUFL dataset, and 100 on the Trento dataset.
4) EndNet is a deep learning network based on MLP. Its objective is to avoid the occurrence of information loss like CNN. The application of the encoder–decoder structure can more compactly integrate HSI and LiDAR data. In the experimental design, owing to the lack of invariant attribute profile codes used in the original article, the number of bands was not expanded for LiDAR data, but even so, it still achieved superior classification results.
5) TBCNN is a two-branch structure of CNN. One branch is used to extract the spatial spectrum features of HSI, while the other is used to obtain the elevation information of LiDAR data. In the experimental design, in order to reduce the cost of computation, the PCA was used to reduce the number of original HSI bands to 30. In addition, the input data sizes of HSI and LiDAR are both set to 11×11.
6) CPCNN is an optimization model based on a two-branch network. The model realizes the mutual learning of HSI and LiDAR features by sharing the convolution weights of the latter two layers, while reducing training parameters and speeding up training time. The input data size of HSI and LiDAR are both set to 11×11 in the experiment.
7) In order to analyze whether the PCA method should be used before grouping or after grouping, this article tests the two cases of MVHN-B and MVHN-A. MVHN-B is the application of PCA to the original HSI data before grouping. In MVHN-A, PCA is applied to each local HSI after grouping. Both of them are adopted with the same hyperparameters in experiment. Specifically, the number of groups is 8, the initial learning rate is 0.1, and the first PCs are used as input for Gabor feature extraction.

Note that all the traditional methods are implemented on MATLAB, and all the methods based on deep learning are coded using Python 3.7.

Tables X–XII summarize the classification accuracies (i.e., OA (%), AA (%), and Kappa) for each classification method on three well-known multisource remote-sensing datasets. In order to ensure the fairness of the experiment, each method uses the same number of training samples for training on the same device. After each experimental method is repeatedly tested 10 times, the average value of each test result is taken as the final classification result. As the results presented in the tables indicate, the performances of the proposed methods (including MVHN-A and MVHN-B) are significantly better than other competitive methods on the three datasets.

1) *Houston Dataset:* Table X presents the classification performance of each method on the Houston dataset. The traditional SVM method, which does not use the spatial information of HSI and LiDAR data, only obtained 94.94% of OA. For the SuperPCA method, the PCA dimensionality reduction is conducted on each superpixel block to enhance the consistency of classes and expand the differences between classes. However, owing to the

TABLE XI
CLASSIFICATION PERFORMANCE OF THE SVM, SUPERPCA, EMAP, ENDNET, CPCNN, TBCNN, MVHN-A, AND MVHN-B CLASSIFICATION METHOD ON THE MUUFL DATASET IN TERMS OF OA, AA, AND KAPPA

No	Class	**Classification performance of various methods**							
		SVM	SuperPCA	EMAP	EndNet	CPCNN	TBCNN	MVHN-A	MVHN-B
1	Trees	97.93	72.22	**98.34**	82.81	92.84	83.44	93.70	93.85
2	Mostly Grass	53.02	81.63	79.71	81.34	43.02	**85.64**	77.53	79.66
3	Mixed Ground	79.84	71.70	**87.83**	72.07	82.03	71.35	79.83	79.78
4	Dirt and Sand	78.10	83.49	88.90	88.18	**98.73**	88.24	94.79	94.61
5	Roads	86.78	85.12	88.24	89.51	77.24	**91.74**	89.04	87.92
6	Water	60.89	90.71	72.31	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
7	Building Shadows	50.83	83.26	57.21	91.00	88.84	90.76	**96.72**	95.78
8	Buildings	94.82	84.04	97.93	94.38	**98.32**	90.68	96.21	96.74
9	Sidewalks	50.44	70.12	59.98	75.49	37.35	74.16	76.81	**76.89**
10	Yellow Curbs	35.26	78.31	21.17	**96.39**	43.37	78.31	78.31	86.75
11	Cloth Panels	72.80	98.22	82.26	96.45	98.22	96.45	**99.41**	**99.41**
OA (%)		82.25	76.00	88.91	84.02	84.85	84.32	90.12	**90.24**
AA (%)		69.19	82.55	75.81	87.97	78.18	86.43	89.30	**90.13**
Kappa		77.16	70.99	85.57	79.50	80.04	79.90	87.03	**87.19**

TABLE XII
CLASSIFICATION PERFORMANCE OF THE SVM, SUPERPCA, EMAP, ENDNET, CPCNN, TBCNN, MVHN-A, AND MVHN-B CLASSIFICATION METHOD ON THE TRENTO DATASET IN TERMS OF OA, AA, AND KAPPA

No	Class	**Classification performance of various methods**							
		SVM	SuperPCA	EMAP	EndNet	CPCNN	TBCNN	MVHN-A	MVHN-B
1	Apple Trees	70.12	99.39	96.27	89.40	99.20	**99.72**	98.49	98.72
2	Buildings	96.40	84.91	96.73	97.18	97.60	91.67	99.42	**99.63**
3	Ground	76.40	45.49	99.04	95.49	99.33	91.92	**100.0**	**100.0**
4	Wood	99.76	99.32	99.81	99.25	99.99	95.32	**100.0**	**100.0**
5	Vineyard	94.19	99.65	99.57	88.52	99.89	97.07	**99.99**	99.98
6	Roads	94.24	89.47	97.31	90.86	97.04	98.55	**97.42**	97.08
OA (%)		91.71	95.46	98.68	93.06	99.30	96.46	99.47	**99.48**
AA (%)		88.52	86.37	98.12	93.45	98.84	95.71	99.22	**99.24**
Kappa		89.02	93.95	98.24	90.79	99.07	95.28	99.29	**99.31**

complexity of the Houston dataset landscape, the single-scale superpixel segmentation method finds it difficult to ensure that the superpixel blocks are all consistent classes, so its accuracy is lower than the SVM method. The EMAP method based on the morphological algorithm fully extracts the morphological texture features of HSI and LiDAR data and its OA is increased by 3.94% compared with the SVM method. However, for the above multisource remote sensing fusion methods based on traditional algorithms, there is no feature extraction on the spectral dimension information of the fusion data, which means that the information of multisource data is not fully utilized, and the performance of the classifier is suppressed. The MVHN method based on 3-D ResNet-like deep CNN can extract the spatial features and fully consider the feature extraction of the spectral dimension of the fusion data. Therefore, the OA indicator of the MVHN method increases by 0.86% compared with the EMAP method. EndNet based on the encoder–decoder structure realizes the information fusion of HSI and LiDAR data in the spectral dimension but fails to use the multisource data spatial structure information and ignores the influence of the correlation between the query pixels and its spatial domain pixel due to its fully connected layer network structure. Although the CPCNN and TBCNN methods realize the spatial–spectral feature extraction of HSI and LiDAR data, they have not yet overcome the shortcoming that CNN is sensitive to shape changes in objects. In contrast, the proposed MVHN method in this article uses Gabor filters to extract multidirectional Gabor features for HSI and LiDAR data, thus providing rotation invariance for the CNN. Furthermore, the proposed MVHN method not only uses feature-level fusion, it also utilizes multiview-based decision fusion to achieve the robust classification results of multisource data. Fig. 9 shows the full-pixel classification results of each experimental method on the Houston dataset, and also shows an enlarged view of some scenes in the classification map. It can be seen from the figure that the CNN method based on a block as the input obtains a smoother classification result map than the traditional methods. Thus, the proposed MVHN method is superior in HSI and LiDAR data classification compared with competitive methods.

2) *MUUFL Dataset:* As Table XI presents, the proposed MVHN-A method obtains an OA of 90.12%, and the proposed MVHN-B method achieves an OA of 90.24%. By contrast, the OAs of the competitive methods are all less than 90%, establishing the effectiveness of the proposed methods (including MVHN-A and MVHN-B) for HSI and LiDAR data classification compared to the traditional and advanced classification methods. Fig. 10 shows the full-pixel classification maps of all classification methods on the MUUFL dataset. It can be observed that the proposed MVHN method obtains a clearer boundary and achieves

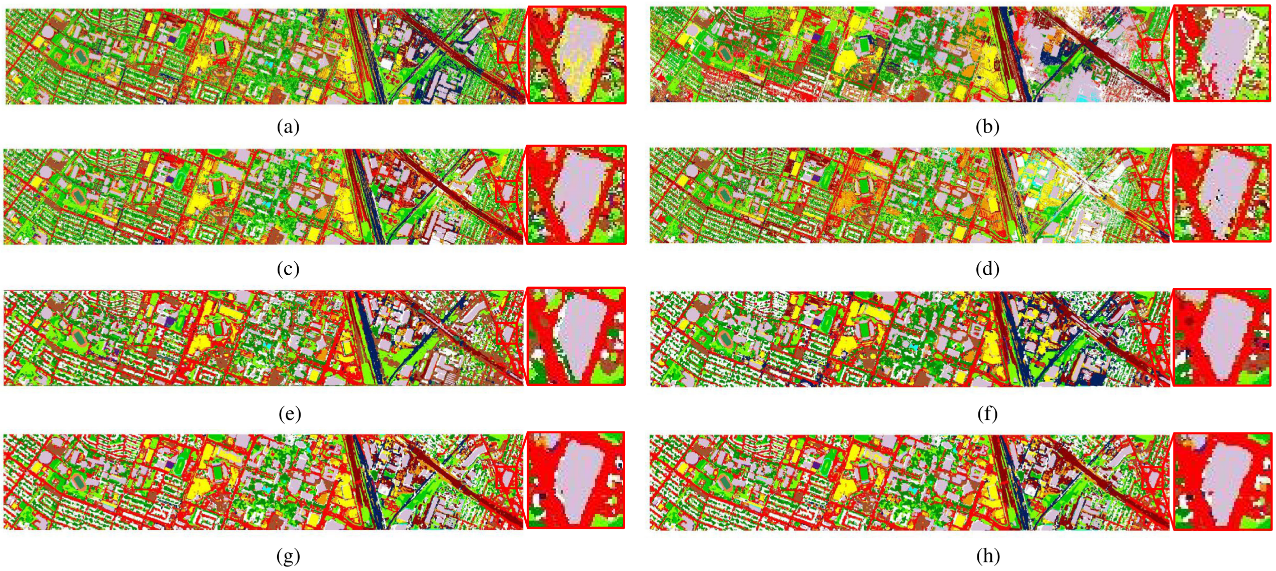


Fig. 9. Classification result maps for different comparison methods on the Houston dataset. (a) SVM (OA = 94.94%). (b) SuperPCA (OA = 95.01%). (c) EMAP (OA = 94.68%). (d) EndNet (OA = 98.88%). (e) CPCNN (OA = 96.79%). (f) TBCNN (OA = 95.63%). (g) MVHN-A (OA = 99.68%). (h) MVHN-B (OA = 99.74%).

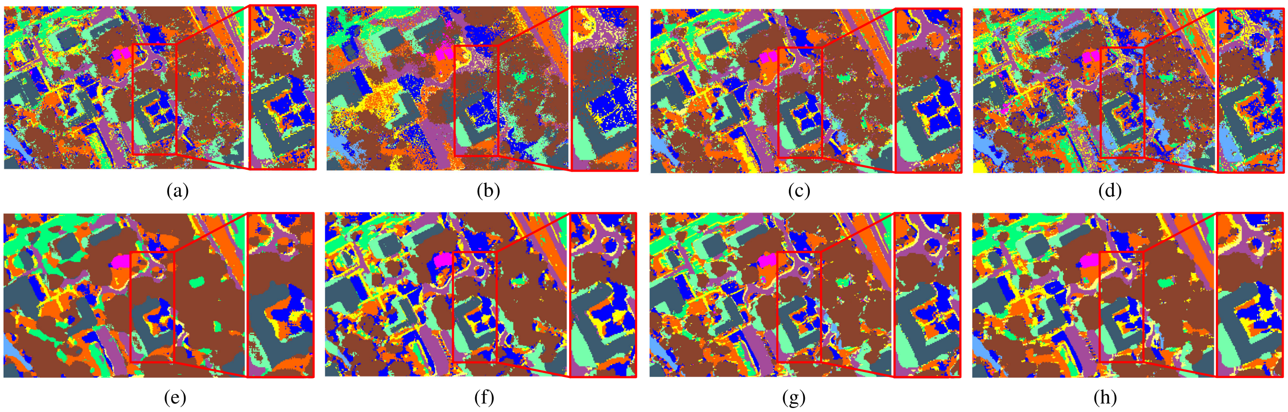


Fig. 10. Classification result maps for different comparison methods on the MUUFL dataset. (a) SVM (OA = 82.25%). (b) SuperPCA (OA = 76.00%). (c) EMAP (OA = 88.91%). (d) EndNet (OA = 84.02%). (e) CPCNN (OA = 84.85%). (f) TBCNN (OA = 84.32%). (g) MVHN-A (OA = 90.12%). (h) MVHN-B (OA = 90.24%).

fewer misclassifications. The classification effectiveness of the proposed MVHN method on the MUUFL dataset can also be further verified.

3) *Trento Dataset:* The Trento test image is a relatively regular and single remote-sensing scene of pixel distribution. Specifically, the pixels in the image are mostly distributed in blocks. Therefore, from Table XII, the classification accuracies of the spatial classification methods i.e., SuperPCA and EMAP, are significantly better than the pixelwise SVM classifier owing to the spatial information of pixels being taken into consideration. The CNN-based TBCNN and CPCNN are also significantly better than the EndNet method that does not consider the spatial information. However, the MVHN method combines Gabor filter and 3-D ResNet-like deep CNN to effectively account for the spatial–spectral features of HSI and LiDAR data, thus achieving better classification accuracy. As the classification results in Table XII indicate, the MVHN-B method obtained the highest classification result with an OA of 99.48%, which proves that this method still has obvious advantages over other methods on the Trento dataset. The same conclusion can also be drawn from the classification map in Fig. 11, where the MVHN method obtains a more accurate classification result map. Furthermore, based on the results presented in Tables X–XII, the classification performance of MVHN-B is slightly better than that of MVHN-A. This is because applying PCA to the original HSI before the multiview strategy allows the Gabor filter to capture more abundant spatial spectrum information.

To test the classification performance of each method, we conduct another experiment on the Houston, MUUFL, and Trento datasets. Fig. 12 presents the classification performance

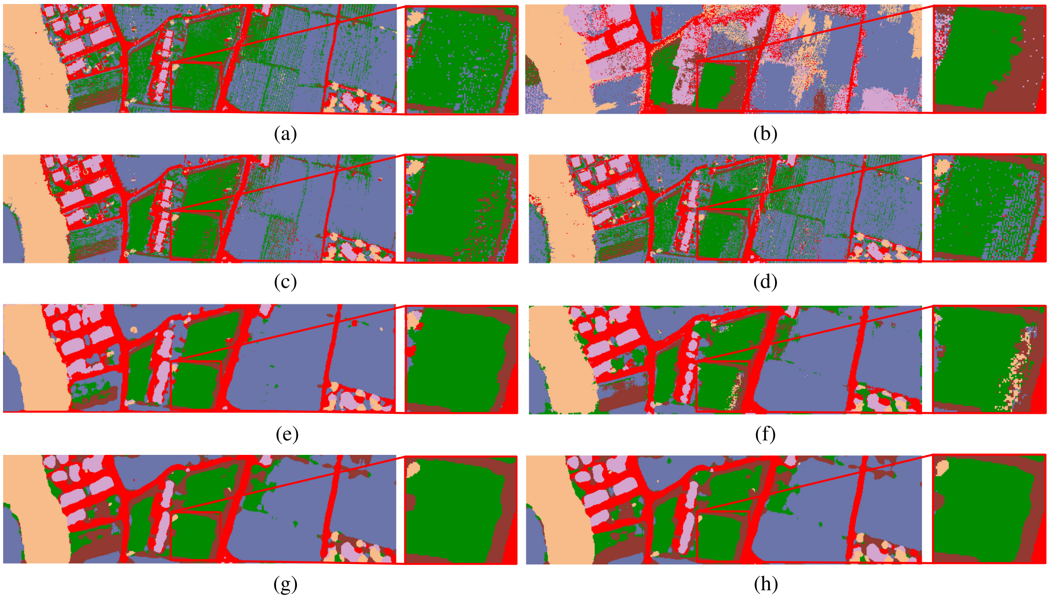


Fig. 11. Classification result maps for different comparison methods on the Trento dataset. (a) SVM (OA = 91.71%). (b) SuperPCA (OA = 95.46%). (c) EMAP (OA = 98.68%). (d) EndNet (OA = 93.06%). (e) CPCNN (OA = 99.30%). (f) TBCNN (OA = 96.46%). (g) MVHN-A (OA = 99.47%). (h) MVHN-B (OA = 99.48%).

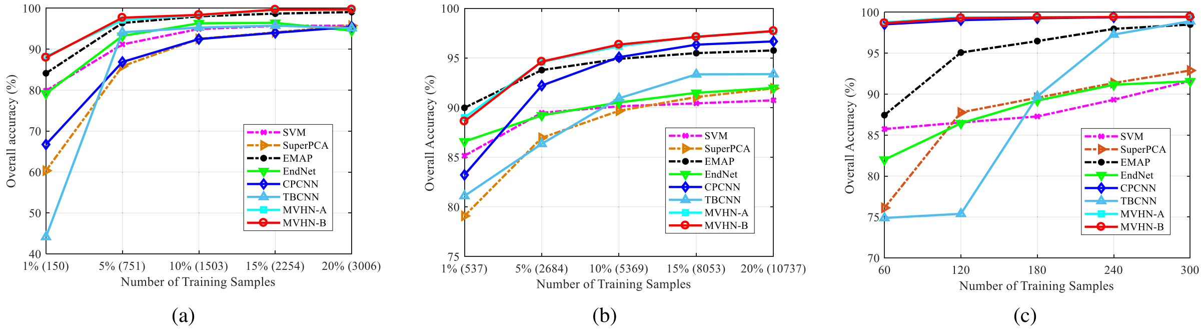


Fig. 12. Classification performance of the proposed MVHN method with different number of training samples on the various datasets. (a) Houston dataset. (b) MUUFL dataset. (c) Trento dataset.

of each method under various numbers of training samples. The experiment is used to evaluate the sensitivity of each classification method to different numbers of training samples. For the Houston and the MUUFL datasets, the proportion of each class of training sample to the labeled sample changed from 1% to 20%. For the Trento dataset, the number of training samples are selected from 10 to 50 per class. In Fig. 12, the methods that take the neural network as the main body are marked by a solid line and the methods take traditional algorithms are marked by a dashed line. It can be seen that the proposed MVHN method is better than other classification methods for almost any number of training samples. Especially for a smaller number of samples, the use of multiple view strategies significantly improves the accuracy of the classification results and enhances the robustness of the classification process. In addition, as the number of samples increases, the proposed MVHN method also demonstrates advantages in the classification task.

E. Ablation Experiments for the Proposed Method

The components to be ablated include HSI, LiDAR, Gabor filter, and multiview. The training samples account for 5% of labeled data of the Houston as well as MUUFL datasets. For the Trento dataset, ten labeled data per class are selected to construct the training set used by the proposed MVHN method for training. Table XIII summarizes the classification accuracies of the proposed MVHN method under ablation conditions on the three datasets. The results in Table XIII indicate that the introduction of LiDAR data enhances the classification performance of the MVHN and, consequently, the effectiveness of the fusion data of HSI and LiDAR for the MVHN in the classification task. In Table XIII, when the Gabor filter is ablated from the MVHN method, the classification accuracy of the MVHN method is degraded by about 1–2% on the Houston and Trento datasets. This is because, for these two datasets, in which for a certain category of samples, the spatial distribution of the sample blocks

TABLE XIII
CLASSIFICATION ACCURACY OF THE PROPOSED MVHN METHOD USING DIFFERENT ABLATION STRATEGIES ON THE THREE MULTISOURCE DATASETS

HSI	LiDAR	Gabor Filter	Multi-View	Houston			MUUFL			Trento		
				OA (%)	AA (%)	Kappa	OA (%)	AA (%)	Kappa	OA (%)	AA (%)	Kappa
√	×	√	√	96.83	97.10	96.58	94.49	83.07	92.70	96.26	93.67	95.01
×	√	√	√	63.27	61.41	60.27	77.56	52.02	69.88	75.99	77.54	69.35
√	√	×	√	97.47	97.58	97.27	94.84	84.02	93.16	96.68	95.23	95.58
√	√	√	×	98.23	98.45	98.08	93.08	80.65	90.82	97.54	95.51	96.71
√	√	√	√	98.28	98.49	98.15	94.64	84.10	92.89	98.67	97.34	98.23

TABLE XIV
COMPARISON OF RUNNING TIME OF EACH METHOD ON DIFFERENT DATASETS

		Running time of various methods							
		SVM	SuperPCA	EMAP	EndNet	CPCNN	TBCNN	MVHN-A	MVHN-B
Houston	Train(s)	69.04	36.21	52.81	9.42	60.67	15.74	61.82	63.07
	Test(s)	1.82	1.05	0.74	0.17	0.28	1.81	4.15	4.37
MUUFL	Train(s)	8.85	7.50	11.83	5.45	31.57	13.52	42.32	41.73
	Test(s)	1.99	2.23	2.15	0.19	0.99	4.51	10.30	10.01
Trento	Train(s)	1.85	3.96	2.23	3.42	14.03	11.55	39.21	36.52
	Test(s)	0.39	0.47	0.40	0.03	0.54	3.05	6.83	6.90

mostly has different directions, and Gabor filter can enhance the adaptability of features to direction changes during network training. Therefore, introducing the Gabor filter into these two datasets can effectively improve the generalization ability and classification performance of the network. In addition, as indicated in Table XIII, the multiview strategy has a significant promotional effect on the performance of the MVHN method. Thus, the use of the multiview method helps the MVHN to overcome the information loss in the process of dimensionality reduction, preserving the integrity of the data, and improves the robustness of the network. For instance, the classification accuracies of the multiview method increased by approximately 2–3% (in terms of OA, AA, and Kappa indicators) upon application of the multiview strategy on the Trento dataset with a very limited number of training samples. This also shows that in the classification task for a small number of training samples, the multiview method can significantly improve the classification accuracy. Thus, the experimental results indicate that the integration of HSI, LiDAR, Gabor filter, and multiview has a polar contribution to the classification performance of the MVHN method. It also proves that the MVHN method is extremely practical in HSI and LiDAR data classification tasks.

F. Comparison of Running Time

The calculation time experiment calculated the running time of the proposed method MVHN and other comparison methods, including training time and test time. During this experiment, all time calculation experiments are run on the same computer. In addition, the SVM, SpuerPCA, and EMAP codes run on the MATLAB 2018a platform, while the rest of the codes run on Python 3.7. The dataset and the corresponding training and test samples are given in Section III-A. During the training process, the training process of all data sets is 20 epochs. The experimental results are shown in Table XIV.

It can be seen from the results in the Table that the MVHN method proposed in this article is not superior in terms of computational efficiency. The main reason is that the feature extraction process of the 3-D residual structure and the classification process of the fully connected layer are expensive in time. It is worth noting that the training time for MVHN is for one group. If the number of groups increases, the calculation time will increase to varying degrees.

IV. CONCLUSION

A multiview hierarchical network is proposed for HSI and LiDAR data classification. Experiments on the Houston, MUUFL, and Trento real datasets compared several well-known traditional classification methods and state-of-the-art deep learning networks, proving that the proposed MVHN method can obtain the best classification accuracy on the aforementioned datasets compared to other state-of-the-art classification methods. We emphasize that the proposed multiview strategy based on a preset band step length is indeed effective for HSI and LiDAR data classification, and the 3-D residual network designed in this work can capture the spectral–spatial semantic features of HSI as well as LiDAR data effectually. In the future, research on the collaborative paradigm of deep learning and active learning for HSI and LiDAR data classification tasks will be our focus. While the performance of the deep network has a strong dependence on training samples, active learning compensates for this defect of the deep network. Therefore, the design of a collaborative model of deep learning and active learning will be of immense importance.

REFERENCES

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

[2] B. Liu, A. Yu, X. Yu, R. Wang, and W. Guo, "Deep multiview learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7758–7772, Sep. 2021.

[3] Y. Xu, Q. Du, W. Li, and N. H. Younan, "Efficient probabilistic collaborative representation-based classifier for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 11, pp. 1746–1750, Nov. 2019.

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

[5] 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, 2021, Art. no. 5006916.

[6] C. Zhou, B. Tu, Q. Ren, and S. Chen, “Spatial peak-aware collaborative representation for hyperspectral imagery classification,” *IEEE Geosci. Remote Sens. Lett.*, vol. 70, 2021, Art. no. 5506805.

[7] N. Audebert, B. Le Saux, and S. Lefevre, “Deep learning for classification of hyperspectral data: A comparative review,” *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 159–173, Jun. 2019.

[8] S. Song, H. Zhou, Y. Yang, and J. Song, “Hyperspectral anomaly detection via convolutional neural network and low rank with density-based clustering,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3637–3649, Sep. 2019.

[9] Q. Wang, Z. Yuan, Q. Du, and X. Li, “GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2019.

[10] R. Tao, X. Zhao, W. Li, H.-C. Li, and Q. Du, “Hyperspectral anomaly detection by fractional Fourier entropy,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4920–4929, Dec. 2019.

[11] W. Xie, B. Liu, Y. Li, J. Lei, C.-I. Chang, and G. He, “Spectral adversarial feature learning for anomaly detection in hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2352–2365, Apr. 2020.

[12] 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.

[13] P. Fu, X. Sun, and Q. Sun, “Hyperspectral image segmentation via frequency-based similarity for mixed noise estimation,” *Remote Sens.*, vol. 9, no. 12, 2017, Art. no. 1237.

[14] A. M. Saranathan and M. Parente, “Uniformity-based superpixel segmentation of hyperspectral images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1419–1430, Mar. 2016.

[15] X. Zhang, J. Zhang, C. Li, C. Cheng, L. Jiao, and H. Zhou, “Hybrid unmixing based on adaptive region segmentation for hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3861–3875, Jul. 2018.

[16] Q. Leng, H. Yang, J. Jiang, and Q. Tian, “Adaptive multiscale segmentations for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5847–5860, Aug. 2020.

[17] H. Aytaylan and S. E. Yuksel, “Fully-connected semantic segmentation of hyperspectral and lidar data,” *IET Comput. Vis.*, 2019.

[18] X. Zhang, H. Fu, and B. Dai, “Lidar-based object classification with explicit occlusion modeling,” in *Proc. Int. Conf. Intell. Human-Mach. Syst. Cybern.*, vol. 2, 2019, pp. 298–303.

[19] S. Morchhale, V. P. Pauca, R. J. Plemmons, and T. C. Torgersen, “Classification of pixel-level fused hyperspectral and lidar data using deep convolutional neural networks,” in *Proc. Workshop Hyperspectral Image Signal Process. Evol. Remote Sens.*, 2016, pp. 1–5.

[20] Z. Wen, B. Hu, L. Jing, M. E. Woods, and P. Courville, “Automatic forest species classification using combined LiDAR data and optical imagery,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2008, pp. III-134–III-137.

[21] S. Morsy, A. Shaker, and A. El-Rabbany, “Multivariate Gaussian decomposition for multispectral airborne LiDAR data classification,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 8741–8744.

[22] S. McCrae and A. Zakhor, “3D object detection for autonomous driving using temporal lidar data,” in *Proc. IEEE Int. Conf. Image. Process.*, 2020, pp. 2661–2665.

[23] M. Dalponte, L. Bruzzone, and D. Gianelle, “Fusion of hyperspectral and lidar remote sensing data for classification of complex forest areas,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1416–1427, May 2008.

[24] H. Gao, C. Bo, J. Wang, K. Li, and D. Li, “Object classification using CNN-based fusion of vision and LiDAR in autonomous vehicle environment,” *IEEE Trans. Ind. Inf.*, vol. 14, no. 9, pp. 4224–4231, Sep. 2018.

[25] Y. Gu, Q. Wang, X. Jia, and J. A. Benediktsson, “A novel MKL model of integrating lidar data and MSI for urban area classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5312–5326, Oct. 2015.

[26] T. Matsuki, N. Yokoya, and A. Iwasaki, “Hyperspectral tree species classification of Japanese complex mixed forest with the aid of lidar data,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens*. vol. 8, no. 5, pp. 2177–2187, May 2015.

[27] J. Vauhkonen *et al.*, “Classification of spruce and pine trees using active hyperspectral LiDAR,” *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1138–1141, Sep. 2013.

[28] J. Schlosser, C. K. Chow, and Z. Kira, “Fusing LIDAR and images for pedestrian detection using convolutional neural networks,” in *Proc. IEEE Int. Conf. Rob. Autom.*, 2016, pp. 2198–2205.

[29] S. Matteoli, L. Zotta, M. Diani, and G. Corsini, “POSEIDON: An analytical end-to-end performance prediction model for submerged object detection and recognition by LiDAR fluorosensors in the marine environment,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 11, pp. 5110–5133, Nov. 2017.

[30] C. Weitkamp, *LiDAR: Range-Resolved Optical Remote Sensing of the Atmosphere*. Singapore: Springer, 2005.

[31] C. Mallet and F. Bretar, “Full-waveform topographic LiDAR: State of the art,” *ISPRS J. Photogramm. Remote Sens.*, vol. 64, no. 1, pp. 1–16, 2009.

[32] P. Ghamisi, B. Hfle, and X. X. Zhu, “Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 3011–3024, Jun. 2017.

[33] Q. Cao, Y. Zhong, A. Ma, and L. Zhang, “Urban land use/land cover classification based on feature fusion fusing hyperspectral image and LiDAR data,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 8869–8872.

[34] P. Ghamisi, J. A. Benediktsson, and S. Phinn, “Land-cover classification using both hyperspectral and LiDAR data,” *Int. J. Image Data Fusion*, vol. 6, no. 3, pp. 189–215, 2015.

[35] K. G. Toker and S. E. Yksel, “Deep canonical correlation analysis for hyperspectral image classification,” *Remote Sens. Ocean, Sea Ice, Coastal Waters, Large Water Regions*, vol. 11150, 2019, Art. no. 1115009.

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

[37] Y. Zhang, H. L. Yang, S. Prasad, E. Pasolli, J. Jung, and M. Crawford, “Ensemble multiple Kernel active learning for classification of multisource remote sensing data,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 2, pp. 845–858, Feb. 2015.

[38] C. Zhang, M. Smith, and C. Fang, “Evaluation of Goddard’s LiDAR, hyperspectral, and thermal data products for mapping urban land-cover types,” *GISci. Remote Sens.*, vol. 55, no. 1, pp. 90–109, 2017.

[39] B. Tu, C. Zhou, D. He, S. Huang, and A. Plaza, “Hyperspectral classification with noisy label detection via superpixel-to-pixel weighting distance,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4116–4131, Jun. 2020.

[40] F. Melgani and L. Bruzzone, “Classification of hyperspectral remote sensing images with support vector machines,” *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[41] 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.

[42] B. Tu, C. Zhou, X. Liao, G. Zhang, and Y. Peng, “Spectral-spatial hyperspectral classification via structural-kernel collaborative representation,” *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 861–865, May 2021.

[43] M. Dalponte, H. O. Orka, T. Gobakken, D. Gianelle, and E. Naesset, “Tree species classification in Boreal forests with hyperspectral data,” *IEEE Trans. Image Process.*, vol. 51, no. 5, pp. 2632–2645, May 2012.

[44] W. Liao, R. Bellens, A. Piurica, S. Gautama, and W. Philips, “Combining feature fusion and decision fusion for classification of hyperspectral and LiDAR data,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2014, pp. 1241–1244.

[45] Y. Zhong, Q. Cao, Z. Ji, A. Ma, Z. Bei, and L. Zhang, “Optimal decision fusion for urban land-use/land-cover classification based on adaptive differential evolution using hyperspectral and lidar data,” *Remote Sens.*, vol. 9, no. 8, 2017, Art. no. 868.

[46] Y. Zhu, W. Li, M. Zhang, Y. Pang, and Q. Du, “Joint feature extraction for multi-source data using similar double-concentrated network,” *Neurocomputing*, vol. 450, pp. 70–79, 2021.

[47] X. Zhao, R. Tao, W. Li, W. Philips, and W. Liao, “Fractional Gabor convolutional network for multisource remote sensing data classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5503818.

[48] X. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, “Multisource remote sensing data classification based on convolutional neural network,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 937–949, Feb. 2018.

[49] R. Hang, Z. Li, P. Ghamisi, D. Hong, G. Xia, and Q. Liu, “Classification of hyperspectral and LiDAR data using coupled CNNs,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4939–4950, Jul. 2020.

[50] X. Zhao *et al.*, "Joint classification of hyperspectral and LiDAR data using hierarchical random walk and deep CNN architecture," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7355–7370, Oct. 2020.

[51] M. Zhang, W. Li, Q. Du, L. Gao, and B. Zhang, "Feature extraction for classification of hyperspectral and LiDAR data using patch-to-patch CNN," *IEEE Trans. Cybern.*, vol. 50, no. 10, pp. 100–111, Oct. 2020.

[52] M. Kanthi, T. H. Sarma, and C. S. Bindu, "A 3D-deep CNN based feature extraction and hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 229–232.

[53] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, "Hyperspectral image super-resolution based on multi-scale wavelet 3D convolutional neural network," in. *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 2770–2773.

[54] S. Mei, J. Ji, Y. Geng, Z. Zhang, X. Li, and Q. Du, "Unsupervised spatial-spectral feature learning by 3D convolutional autoencoder for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6808–6820, Sep. 2019.

[55] 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.

[56] F. Ye, Z. Shi, and Z. Shi, "A comparative study of PCA, LDA and kernel LDA for image classification," in *Proc. Int. Symp. Ubiquitous Virtual Reality*, 2009, pp. 51–54.

[57] C. C. Chang and C. J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2007, Art. no. 27.

[58] K. Wang, H. Rui, and S. Qian, "Spectral-spatial hyperspectral image classification using extended multi attribute profiles and guided bilateral filter," in *Proc. Int. Conf. Comput. Sci. Mech. Autom.*, 2015, pp. 235–239.

[59] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, "SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4581–4593, Aug. 2018.

[60] D. Hong, L. Gao, R. Hang, B. Zhang, and J. Chanussot, "Deep encoder–decoder networks for classification of hyperspectral and LiDAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2020, Art. no. 5500205.

Yishu Peng received the B.Sc., M.Sc., and the Ph.D. degrees in mechanical design and theory from Northeastern University, Shenyang, China, in 2009, 2011, and 2017 respectively.

From 2017 to 2019, he was with the School of Mechanical and Engineering, Hunan Institute of Science and Technology, Yueyang, China, and since 2019, he has been with the School of Information Science and Technology. His research interests include the image processing, object detection, and target tracing.

Yuwen Zhang (Student Member, IEEE) received the B.S. degree in electrical engineering and automation in 2020 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 image processing, classification of multisource remote sensing data, and object detection.

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.

Chengle Zhou (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering and automation, 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 image processing, pattern recognition, hyperspectral image classification, anomaly detection, and noisy label detection.

Mr. Zhou was the recipient of the Excellent Student Award at the 2021 IEEE Geoscience and Remote Sensing Society Summer School (GR4S) on Modeling in Microwave and Optical Remote Sensing and the National Scholarship for Postgraduate granted by the Chinese Ministry of Education in 2020 and 2021.

Qianming Li (Member, IEEE) received the M.S. degree in urban planning and the Ph.D. degree in civil architecture and planning design from Central South University, Changsha, China, in 2016 and 2021, respectively.

Since 2021, she has been a Lecturer with the School of Information and Engineering, Hunan Institute of Science and Technology, Yueyang, China. Her research interests include image processing, regional spatial structure, and urban analysis in remote sensing.