Joint Classification of Hyperspectral and LiDAR Data Using Hierarchical Random Walk and Deep CNN Architecture

Xudong Zhao, *Student Member, IEEE*, Ran Tao, *Senior Member, IEEE*, Wei Li, *Senior Member, IEEE*, Heng-Chao Li, *Senior Member, IEEE*, Qian Du, *Fellow, IEEE*, Wenzhi Liao, *Senior Member, IEEE*, and Wilfried Philips, *Senior Member, IEEE*

***Abstract*— Earth observation using multisensor data is drawing increasing attention. Fusing remotely sensed hyperspectral imagery and light detection and ranging (LiDAR) data helps to increase application performance. In this article, joint classification of hyperspectral imagery and LiDAR data is investigated using an effective hierarchical random walk network (HRWN). In the proposed HRWN, a dual-tunnel convolutional neural network (CNN) architecture is first developed to capture spectral and spatial features. A pixelwise affinity branch is proposed to capture the relationships between classes with different elevation information from LiDAR data and confirm the spatial contrast of classification. Then in the designed hierarchical random walk layer, the predicted distribution of dual-tunnel CNN serves as global prior while pixelwise affinity reflects the local similarity of pixel pairs, which enforce spatial consistency in the deeper layers of networks. Finally, a classification map is obtained by calculating the probability distribution. Experimental results validated with three real multisensor remote sensing data demonstrate that the proposed HRWN significantly outperforms other state-of-the-art methods. For example, the two branches CNN classifier achieves an accuracy of 88.91% on the University of Houston campus data set, while the proposed HRWN classifier obtains an accuracy of 93.61%, resulting in an improvement of approximately 5%.**

***Index Terms*— Convolutional neural network (CNN), hyperspectral image (HSI), multisensor data fusion, hierarchical random walk.**

Manuscript received May 28, 2019; revised September 11, 2019, November 6, 2019, January 7, 2020, and February 9, 2020; accepted March 16, 2020. Date of publication April 6, 2020; date of current version September 25, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61421001, Grant 91638201, Grant 61922013, Grant 61571033, and Grant U1833203. The work of Xudong Zhao was supported by the China Scholarship Council. *(Corresponding author: Ran Tao.)*

Xudong Zhao is with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China, and also with Image Processing and Interpretation, IMEC Research Group, Ghent University, 9000 Ghent, Belgium (e-mail: zhaoxudong@bit.edu.cn).

Ran Tao and Wei Li are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: rantao@bit.edu.cn; liwei089@ieee.org).

Heng-Chao Li is with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China (e-mail: lihengchao_78@163.com).

Qian Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA (e-mail: du@ece.msstate.edu).

Wenzhi Liao is with the Sustainable Materials Management, Flemish Institute for Technological Research (VITO), 2400 Mol, Belgium, and also with the Image Processing and Interpretation, IMEC Research Group, Ghent University, 9000 Ghent, Belgium (e-mail: wenzhi.liao@vito.be).

Wilfried Philips is with Image Processing and Interpretation, IMEC Research Group, Ghent University, 9000 Ghent, Belgium (e-mail: wilfried.philips@ugent.be).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2020.2982064

I. Introduction

REMOTELY SENSED hyperspectral imagery can provide detailed spectral information for potential material identification [1], [2]. Unlike the hyperspectral image (HSI) data, light detection and ranging (LiDAR) data provide elevation information about the area under investigation, which can be acquired at any time of the day and under adverse weather conditions [3]–[5]. Several techniques fusing data collected from different sensors have been investigated for better scene characterization [6]–[11]. Joint classification of HSI and LiDAR further improves the classification performance in remote sensing [8], [12].

Numerous effective classification and machine learning methods using HSI data have been developed [13]–[16]. Support vector machines (SVMs) [17] with nonlinear kernel functions have been widely used, especially when the number of training data is small. A simple classification method called extreme learning machine (ELM) [18] has been introduced with comparable performance to SVM. Deep learning-based methods have attracted wide attention for their capabilities of automatically extracting robust and high-level features, which are known to be generally invariant to most local changes of input [19], [20]. Specifically, convolution neural network (CNN) has been extensively studied in remote sensing, and it outperforms classical machine learning methods. For example, a CNN-based pixel-pair feature framework [19] and a contextual CNN [20] were recently proposed for HSI classification. A convolutional recurrent neural network (CRNN) was proposed to learn more discriminative features for hyperspectral data classification [21]. A CNN architecture based on spectral–spatial capsule networks (CapsNets) is proposed to achieve accurate classification of HSI while significantly reducing the network design complexity [22]. Residual network (ResNet)

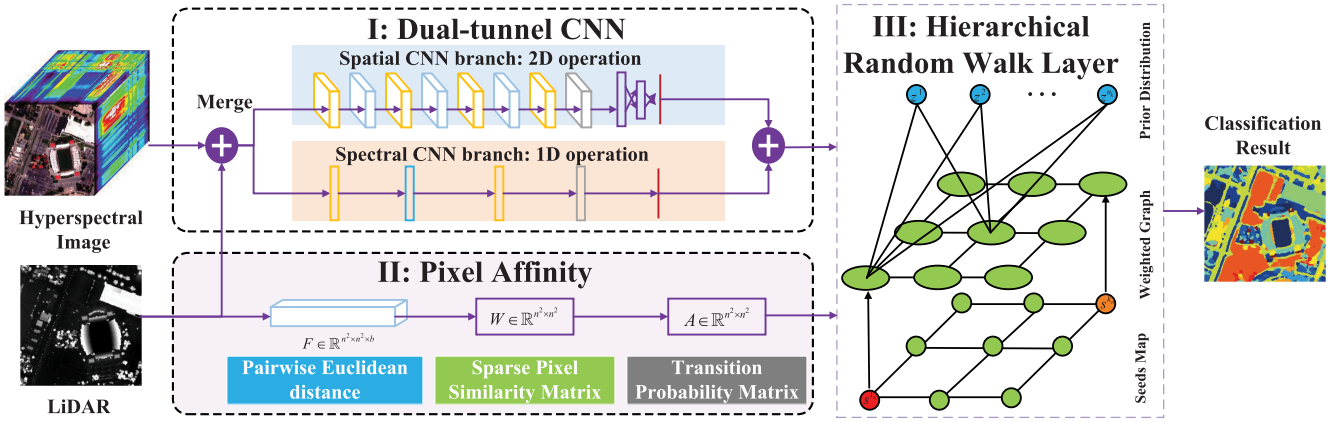


Fig. 1. Proposed feature extraction and classification framework of HRWN. In the random walk layer, ellipse nodes denote the graph nodes, and circle nodes are auxiliary nodes. Green nodes are unseeded nodes, blue nodes are prior distribution, and other color nodes are seeds. Only part of the transition edges are shown for simplification.

[23] and extended model based on pyramidal bottleneck residual units (Pyramidal ResNet) [24] enable feature reuse. Dense convolutional networks (DenseNets) [25] enable new feature exploration. Both are important for extracting discriminative features of HSI. A dual-path network method was proposed to combine ResNet and DenseNet so as to reuse the previous features by skipping connections between adjacent layers, and explore new features containing more details by connecting each layer to other layers in a feedforward fashion [26].

With the increased availability of HSI and LiDAR data sets of the same area, studies have indicated that classification performance can be improved by fusing HSI and LiDAR data [8], [12]. In [9], morphological extinction-profiles were exploited for joint feature extraction of both HSI and LiDAR-based DSM. In [27], topological information and morphological operations were included in the classification task. However, simple concatenation or stacking of features may be limited in individual feature extraction [28]. In particular, to solve the problem caused by an unbalance between different features, decision-fusion methods for HSI and LiDAR classification have been presented [10], [11]. CNN-based methods combined HSI and LiDAR-based DSM using multibranch architecture [29] or end-to-end hierarchical fusion module [30] for efficient classification. Although these features and decision-fusion-based approaches have shown excellent performance for classification tasks, the large receptive fields in the convolutional layers and the presence of pooling layers lead to low spatial resolution in the deeper CNN layers [31]. Thus, the corresponding predicted classes tend to be spatially fragmented and lack fine object boundary details.

To address spatial continuity, Markov random field (MRF) was integrated into the HSI classification framework as a post-processing step [32], [33]. However, these models typically involve a large number of parameters and complex loss functions requiring specialized model training. In [35] and [36], MRF or conditional random field (CRF) was integrated into CNN by employing graphs with a fixed grid structure, which made training and testing inflexible and complicated. Thus, simple but flexible random walk methods were introduced for the efficiency of the focusing task [36]–[39]. It was originally developed for image segmentation and has been successfully applied in HSI classification [40]–[42].

In this article, focusing on the weak boundary and spatially fragmented classification issue, a simple yet effective hierarchical random walk network (HRWN) is proposed. The proposed HRWN jointly optimizes dual-tunnel CNN and pixelwise affinity via a novel random walk layer, which enforces spatial consistency in the deeper layers of the network. Because the LiDAR-based DSM (digital surface model) possesses clear boundaries between objects with different elevations and homogeneous regions inside real objects with the same elevations, a pixel affinity branch using LiDAR-based DSM is employed to capture the relationships between classes with different elevation information and confirm spatial contrast. Then, in the hierarchical random walk layer, the predicted distribution of dual-tunnel CNN serves as a global *a priori* while the affinity of pixels reflects the local similarity and spatial contrast of pixel pairs. Finally, the classification map is obtained by computing the probability distribution of each pixel to be classified. Experimental results with real remote sensing data sets demonstrate the effectiveness of the proposed method.

The main contributions are summarized as follows.

1) A hierarchical random walk layer exploits spatial constraint and local seed guidance into the deeper layer of CNN. The random walk layer alleviates the problems of weak localization around boundaries and spatial fragmentation of the classification map.
2) Considering detailed spectral signatures of HSI and elevation information of LiDAR, the proposed dual-tunnel CNN can utilize complementary features and analyze the joint features from the fused data spatially and spectrally.
3) For spatial consistency and contrast of LiDAR-based DSM, a new pixel affinity matrix is defined, which captures the elevation similarity and confirms the spatial contrast between classes.

The remainder of this article is organized as follows. The proposed framework is introduced in Section II. In Section III,

experimental results and analysis are presented. Finally, Section IV summarizes with some concluding remarks.

II. Proposed Classification Framework

The HRWN framework is designed to classify pixels by fusing multisource remote sensing images in both feature extraction and decision prediction aspects. First of all, the random walk process is integrated into the deep architecture to encourage coherent classification among pixels that are similar to each other. By doing so, poor localization around boundaries and spatial fragmented classification problems are alleviated. The framework of the proposed HRWN is illustrated in Fig. 1, which consists of three parts: 1) a dual-tunnel CNN branch that exploits classification potentials (part I); 2) the other pixel-level affinity branch that captures local similarities (part II); and 3) a novel random walk layer that merges the two branches to obtain classification map (part III).

A. Prior Distribution Prediction by Dual-Tunnel CNN

As illustrated in Fig. 1, part I reflects dual-tunnel CNN branch, which is elaborated on more details in Fig. 2. For the input of the dual-tunnel CNN branch, both spectral and spatial features are needed, but the original HSI images are limited for low spatial resolution and cloud interference. Therefore, the HSI and LiDAR-based DSM images with geographical registration are first fused to combine the advantages of both. Classical multisensor information fusion methods include color normalized transformation, principal component spectral sharpening, and Gramm–Schmidt methods. Specifically, Gramm–Schmidt method is used for the fusion of HSI and LiDAR-based DSM after geographical registration [43]. Then, a novel dual-tunnel CNN is designed for the fused image $\mathbf{H}$, which consists of a spectral tunnel and a spatial tunnel.

For the spatial tunnel, the input is a patch centered at the pixel p_{ij} with a radius r. The fused data patch $\mathbf{H}_{ij}^{\text{spa}} \in \mathbb{R}^{\text{ksize}\times\text{ksize}}(\text{ksize} = 2\times r+1)$ is fed into the 2-D CNN tunnel, which includes simple operations such as 2-D convolution, activation, max-pooling, and batch normalization layers [44], [45]. The spatial feature $\mathbf{F}_{ij}^{\text{spa}}$ around the central target pixel p_{ij} is first extracted by the 2-D convolution and batch normalization operations. The convolution operations are executed with valid padding, and the value of convolution stride stride is set to 2, then the range of samples is $[(\text{Width}*\text{Height})/\text{stride}^2]$, where $i = 1,\ldots,\text{Width}/\text{stride}$, $j = 1,\ldots,$ Height/stride, Width and Height are the size of the original image. Each convolution process involves certain operations, including 2-D CNN (conv2D) and batch normalization. The kernel size of the convolutional layer shown in Fig. 2 is set to 3×3 and then 1×1 for detailed spatial information while the number of filters changes from 100 to 200. The activation function is a rectified linear unit (ReLu) while dropout is used to prevent overfitting. Since the dimensionality of feature maps from the last layer is different from that of the spectral tunnel, the output spatial features $\mathbf{F}_{ij}^{\text{spa}}$ are vectorized as a D-dimensional vector by the fully connected layer.

For the spectral tunnel, it concentrates on the center pixel $\mathbf{H}_{ij}^{\text{spe}}$ and consists of simple 1-D operations including convolution, activation, max-pooling, and batch normalization.

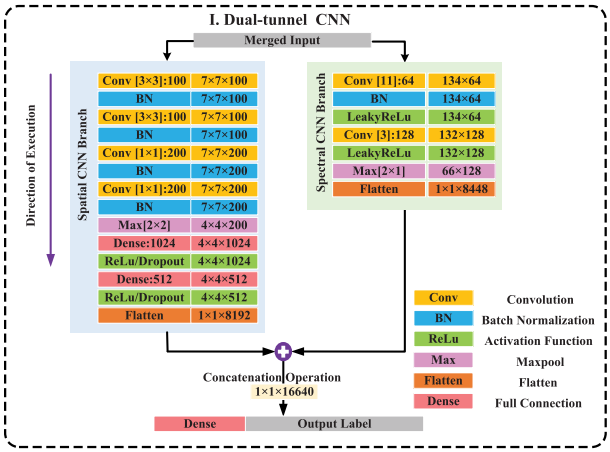


Fig. 2. Overall parameter configuration of the designed dual-tunnel CNN branch.

1-D CNN (conv1D) is applied to extract spectral features, where the value of convolution stride stride is set to 2 for sufficient training samples. Filter length of convolutional layers shown in Fig. 2 are set to 11 and 3 while the number of filters changes from 64 to 128 for detailed spectral feature extraction. The output spectral feature $\mathbf{F}_{ij}^{\text{spe}}$ is also vectorized as a D-dimensional vector by a fully connected layer after the max-pooling layer. Then the spatial and spectral features are concatenated as

$$\mathbf{F}_M = f\left(\mathbf{W}\cdot\left(\mathbf{F}_{ij}^{\text{spe}} \parallel \mathbf{F}_{ij}^{\text{spa}}\right)+\mathbf{b}\right) \tag{1}$$

where $\mathbf{W}$ and $\mathbf{b}$ are the weight and bias of the fully connected layer, respectively, and $\parallel$ means concatenating the spectral and spatial features.

Then the joint spatial–spectral feature $\mathbf{F}_M$ is fed into the softmax classify layer to predict the probability distribution

$$p_i^k = \frac{\exp(\theta_k|\mathbf{F}_M)}{\sum_{k=1}^{K}\exp(\theta_k|\mathbf{F}_M)} \tag{2}$$

where K is the number of classes and k indicates the kth class, i is the index of pixel, and θ_k is the kth column of the weights in the prediction layer. The probability distribution indicated by p_i^k is used as prior distribution in the hierarchical random walk layer (part III).

B. Spatial Similarity by Pixel-Level Affinity Branch

The dual-tunnel CNN utilizes both spatial and spectral information, but the large receptive fields in the convolutional layers and the presence of pooling layers lead to low spatial resolution in the deeper CNN layers. To capture the relationship between classes with different elevation information and confirm spatial contrast, a pixel-level affinity branch is employed with its own learning objective as part II in Fig. 1.

This branch is connected with the LiDAR-based DSM image with N pixels. The input image is modeled as a weighted, undirected, and connected graph $G = \{V, E\}$, where V is a finite set of vertices with $|V| = N$ and E is a set of edges. Then $\mathbf{Eu}$ is a sparse $N\times N$ matrix storing Euclidean

distance between each pixel and all its neighbors

$$\mathbf{Eu}(i, j) = ||\mathbf{DSM}(i) - \mathbf{DSM}(j)||_2 \tag{3}$$

where $i = 1, \ldots, N$ and $j = 1, \ldots, N$ with N being the number of pixels. The resulting **Eu** is used as an input feature of the affinity branch, as shown in Fig. 1. Note that **Eu** is a sparse matrix, as only a small fraction of all the entries are populated with nonzero values.

Then the Euclidean distance layer is optimized to predict the pixel similarity matrix **W**, where a weight $w_{ij} \in \mathbf{W}$ of edge e_{ij} measures the probability that a random walker will cross this edge. As many other algorithms based on graph expressed, a weight w_{ij} is formulated as

$$w_{ij} = \exp\left(-\frac{\mathbf{Eu}(i, j)^2}{\sigma}\right) + \varepsilon \tag{4}$$

where $\mathbf{Eu}(i, j)$ is the Euclidean distance between two nodes v_i and v_j, σ is the controlling parameter and ε is a small constant. The nonnormalized graph Laplacian is defined as $\mathbf{L} := \mathbf{D} - \mathbf{W}$, where **D** is the diagonal degree matrix with the nth element $d_n = \sum_{m \neq n} \mathbf{W}_{mn}$ being the degree of vertex n. Finally, normalization is applied to predict the ground-truth pixel affinities by matrix operation $\mathbf{A} = \mathbf{D}^{-1}\mathbf{W}$, which can be expressed as

$$\mathbf{A}_{ij} = \begin{cases} w_{ij}/d_i, & \text{if } i \sim j \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

The pixel affinities matrix **A** is used as a spatial constraint in the hierarchical random walk layer (part III) to build up an extended graph.

C. Hierarchical Random Walk Classification

To integrate the dual-tunnel CNN prediction and pixel affinities, a hierarchical random walk layer (part III) is designed, as shown in Fig. 1. The hierarchical random walk layer is connected to the two branches, including: 1) prior distribution $p_i^k \in \mathbf{P}$ obtained by dual-tunnel CNN as a global guide and 2) affinity matrix **A** reflecting pixel similarity for adjoint and seeds map $s_i^k \in \mathbf{S}$ for the local guide.

With the prior information, a graph with priori G_e is constructed. Assume a random walker starts at each unlabeled pixel, the probabilities r_i^k is computed such that random walkers first reach the labeled pixels. For the standard random walker, the probability r_i^k can be computed by minimizing the energy function

$$E_{\text{spatial}}^k(\mathbf{r}^k) = (\mathbf{r}^k)^T \mathbf{L}\mathbf{r}^k. \tag{6}$$

The analytical solution of the energy function can be computed by solving a system of linear equations [36].

However, the random walker layer will not work if **S** is empty. To incorporate global prior distribution and local prior information, another nonspatial energy function is given as

$$\begin{aligned} E_{\text{non-spatial}}^k(\mathbf{r}^k) &= E_{\text{seed}}^k(\mathbf{r}^k) + E_{\text{prior}}^k(\mathbf{r}^k) \\ &= (\mathbf{r}^k - \mathbf{s}^k)^T \mathbf{C}(\mathbf{r}^k - \mathbf{s}^k) + \cdots + (\mathbf{r}^k - \mathbf{I})^T \mathbf{P}^k(\mathbf{r}^k - \mathbf{I}) \\ &\quad + \sum_{q=1, q\neq k}^{N} (\mathbf{r}^q)^T \mathbf{P}^q \mathbf{r}^q \end{aligned} \tag{7}$$

where $\mathbf{P}^k$ is a diagonal matrix representing the prior probability with the ith element p_i^k being the probability for pixel v_i, $\mathbf{s}^k$ is the seed matrix, and **C** is the weight matrix controlling the trade-off of seeds. As (7) indicated, three parts of energy constitute the nonspatial energy. $(\mathbf{r}^k - \mathbf{s}^k)^T \mathbf{C}(\mathbf{r}^k - \mathbf{s}^k)$ is considered as the energy of random walker reaching the training seeds, and mininizing this term can gather nodes to the seed nodes in their class. $(\mathbf{r}^k - \mathbf{I})^T \mathbf{P}^k(\mathbf{r}^k - \mathbf{I})$ and $\sum_{q=1, q\neq k}^{N} (\mathbf{r}^q)^T \mathbf{P}^q \mathbf{r}^q$, respectively, indicate reaching probabilities to nodes in the prior distribution of current class labels and labels for other classes.

Thus, the spatial and nonspatial energy functions are combined as

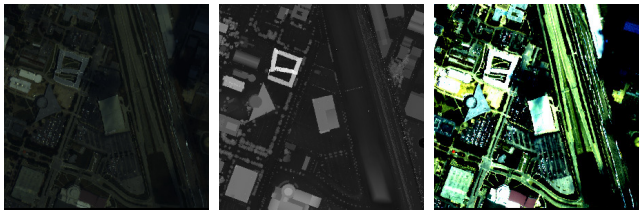
$$E^k(\mathbf{r}^k) = E_{\text{spatial}}^k(\mathbf{r}^k) + \lambda E_{\text{non-spatial}}^k(\mathbf{r}^k) \tag{8}$$

where λ is the weight number controlling the tradeoff of spatial and nonspatial energy. Each prior distribution node is connected with all nodes in V, and the weight of an edge between a prior node h_k and node v_i is proportional to probability p_i^k.

Then, in the designed hierarchical random walk layer, the transition probability matrix **Q** on $\mathbf{V} \cup \Delta \cup \mathbf{S} \cup \mathbf{P}$ is

Algorithm 1 HRWN Model

Require: Fused image **H**, LiDAR-based DSM **DSM**, training seeds **S**, training epochs *epochs*.
Ensure: Classification map **R**.
1: Initialize all weights
2: Step1: **Prior Distribution Prediction**
3: **for** *epoch* < *epochs* **do**
4: Extract spatial features $\mathbf{H}_{ij}^{spa}$ and spectral feature $\mathbf{H}_{ij}^{spe}$ from the fused image **H**.
5: Train the Dual-tunnel CNN shown in Fig. 2
6: Obtain probability distribution p_i^k by Eq. (2).
7: **end for**
8: Step2: **Pixel-Level Affinity**
9: Build graph G from LiDAR-based DSM image **DSM**.
10: Compute weight matrix **W** using Euclidean distance **Eu** and compute pixel affinity matrix **A** by Eq. (5).
11: Step3: **Hierarchical Random Walk**
12: Compute the transition probability matrix **Q** and the reaching probability r_i^k by Eq. (9) and Eq. (10).
13: Optimize random walk path and obtain classification map **R** by Eq. (11).



(a) (b) (c)

Fig. 3. Part of HSI and LiDAR-based DSM data with their fused image for joint feature extraction. (a) HSI. (b) LiDAR-based DSM. (c) Fused image.

formulated as

$$\mathbf{Q}(i,j)=\begin{cases} c, & \text{if} i\in\mathbf{V}, j\in\mathbf{S} \\ \dfrac{\lambda p_i^k}{d_i+\lambda g_i}, & \text{if} i\in\mathbf{V}, j\in\mathbf{P} \\ \dfrac{w_{ij}}{d_i+\lambda g_i}, & \text{if} j\sim i\in\mathbf{V} \\ 1, & \text{if} i=j\in\{\Delta\}\cup\mathbf{S}\cup\mathbf{P} \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where $g_i=\sum_{k=1}^{K}p_i^k$, λ is the weight parameter of prior distribution, c is the weight parameter of seeds, and Δ is the set of unlabeled pixels.

Given the transition probability $\mathbf{Q}$ on a graph with prior, the reaching probability r_i^k that a random walker from a node $v_i\in V$ reaching seed s_i^k or prior node p_i^k is formulated as

$$r_i^k=\sum_{j\sim i\in V}\frac{w_{ij}r_j^k}{d_i+\lambda g_i}+\frac{\lambda p_i^k}{d_i+\lambda g_i}+cs_i^k \tag{10}$$

where the weight parameter λ and c balance two effects: 1) propagating the classification information across the nodes using random walk transition matrix and 2) not deviating too much from the initial classification.

Then the final labeling result can be obtained as

$$\mathbf{R}(i)=\arg\max_k r_i^k \tag{11}$$

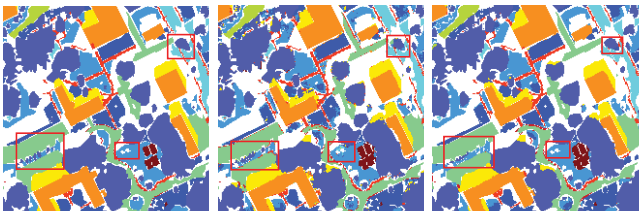
where $\mathbf{R}$ is the final label that generates the greatest probability. With some deduction, the solution of (11) is equivalent to the optimal solution of minimizing objective function in (8). These algorithm steps are shown in Algorithm 1.

D. Motivation of the Proposed HRWN

For the joint classification of multisensor data, the detailed spectral signatures of ground covers using HSI and elevation information of the same area using LiDAR-based DSM need to be extracted jointly. Taking the HSI and LiDAR-based DSM data sets in Table I as an example, the motivations of each step are analyzed as follows. In the proposed dual-tunnel CNN, the joint utilization of 2-D and 1-D CNN integrates spatial and spectral information. Due to the high spectral resolution, narrow bandwidth and a large amount of information, HSI can be used to distinguish and detect ground targets with powerful diagnostic ability. However, as shown in Fig. 3(a), its poor spatial contrast between objects limits its performance in spatial feature presentation. The LiDAR-based DSM shown in Fig. 3(b) provides more accurate elevation information and useful spatial contrast. Thus, fusing HSI and LiDAR-based DSM is helpful to synthesize diverse information on the investigated areas. However, as shown in the top right corner of Fig. 3(c), the occluded part still has no obvious contrast in the fused image. Thus, the pixel affinity branch is further developed using LiDAR-based DSM to confirm the spatial contrast between classes with different elevations.

(a) (b)

Fig. 4. Spatial comparison of pixel similarity with boundary region specifically. (a) HSI. (b) LiDAR-based DSM. (Detailed specifications of the HSI and LiDAR-based DSM are depicted in Table I.)



(a) (b) (c)

Fig. 5. Classification maps comparing the usefulness of hierarchical random walk. (a) Ground truth. (b) Dual-tunnel CNN classification. (c) Improved classification by hierarchical random walk.

The reason to choose LiDAR-based DSM rather than HSI or the fused image for pixel affinity is that LiDAR-based DSM possesses clearer boundary regions and homogeneous regions inside the real objects as shown in Fig. 4. With the elevation information of LiDAR-based DSM, objects composed of the same material but different elevations are distinguishable locally, such as highway, roofs, and roads. Even for the occluded part on the top right corner, the local pixel similarity is extracted accurately because LiDAR works well in all light conditions. In the proposed classification framework, adjacent pixels are most likely to own the same label. This spatial smooth prior defined in (4) encourages neighboring pixels to be assigned with the same labels. The pairwise interaction terms obtain a small distance when neighboring pixels are similar. In this way, this smoothness prior can reinforce piecewise smooth classification in homogeneous regions. On the contrary, obvious dissimilarity is encouraged with a larger w in the boundary regions. The exploitation of this intuitive prior distribution of LiDAR-based DSM with proper characteristics tends to dramatically improve classification accuracy.

TABLE I
HSI AND LiDAR-BASED DSM DATA SETS USED FOR EVALUATION

Dataset	Houston [47]		Trento [8]		MUUFL [48]	
Location	Houston, Texas, USA		Trento, Italy		Long Beach, Mississippi, USA	
Sensor Type	HSI	LiDAR-based DSM	HSI	LiDAR-based DSM	HSI	LiDAR-based DSM
Image Size	349*1905	349*1905	600*166	600*166	325*220	325*220
Spatial Resolution	2.5m	2.5m	1m	1m	0.54 m *1.0m	0.60 m *0.78 m
Number of Bands	144	1	63	1	64	2
Wavelength Range	$0.38-1.05\mu m$		$0.42-0.99\mu m$		$0.38-1.05\mu m$	$1.06\mu m$
Sensor Name	CASI-1500		AISA Eagle	Optech ALTM 3100EA	CASI-1500	Gemini ALTM LIDAR

The probability maps estimated by the dual-tunnel CNN are usually noisy and not aligned with real boundaries of objects, which leads to low classification accuracy as shown in Fig. 5(b) (the MUUFL Gulfport data set used in Fig. 5 for illustration will be introduced in Section III-A). This is because of piecewise spatial constraint and pixel similarity are not considered. With the pixel affinity branch confirming spatial contrast using only LiDAR-based DSM image, the classifier performance on classes with similar elevations is degraded. To combine the advantages of spatial–spectral and elevation information, a hierarchical random walk layer is designed to integrate the dual-tunnel CNN prediction and pixel affinities. As shown in Fig. 5(c), this layer not only improves classification accuracy in homogeneous regions but also ensures that real object boundaries are well aligned with the probability map.

III. EXPERIMENTS AND ANALYSIS

In this section, three multisensor remote sensing data sets are used to validate the effectiveness of the proposed method. All the programs are implemented using Python 3.6 and MATLAB R2018a [47], and the networks are constructed using Tensorflow[1] with the high-level API Keras.[2] Tensorflow is an open-source software library for numerical computation using data flow graphs, and Keras can be seen as a simplified interface to Tensorflow. All experiments are conducted using a personal computer equipped with Windows 7 and NVIDIA Quadro K2200.

A. Experimental Data

To evaluate the performance of the proposed HRWN, three multisensor data sets are tested for quantitative and qualitative evaluation. The first scene was acquired by the NSF-funded Center in June 2012 over the University of Houston campus and neighboring areas [28]. The data set is composed of HSI and LiDAR-based DSM with detailed information shown in Table I. Available training and testing samples are listed in Table II. The second scene was acquired over a rural area in the south of the city of Trento, Italy, with detailed information depicted in Table I [8]. Table IV lists the available training and testing samples. The third scene was collected on November 8, 2010, over the University of Southern Mississippi Gulfport Campus [46]. Detailed data descriptions can be found in Table I. Table VI lists the available training and testing samples.

[1]http://tensorflow.org/
[2]https://github.com/fchollet/keras

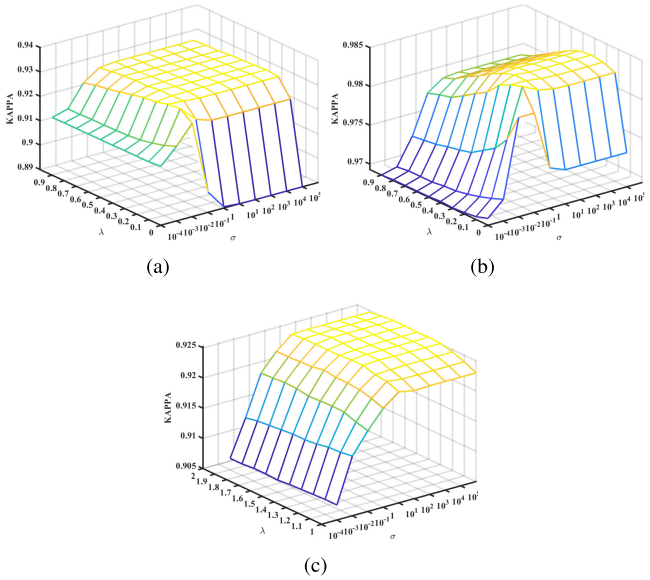


Fig. 6. Influence of the parameters on classification accuracies of the proposed HRWN with different parameters, i.e., λ and σ. (a) Houston. (b) Trento. (c) MUUFL.

B. Parameter Tuning

The classification performance is closely related to the designed architecture of the deep network. In order to validate the proposed HRWN, the classification results using different weight parameters and patch sizes are compared.

1) Weight Parameter Comparison: For the proposed HRWN, (4) and (10) can be used to calculate the parameters λ and σ. In order to validate the effects of weights between different layers, classification tasks are tested using different weights between prior distribution, pixel affinity and seed map. The Kappa coefficient (Kappa) of the proposed method as the two varying parameters is shown in Fig. 6. It is indicated that when σ is lower than 10^{-1}, Kappa of the proposed method may decrease significantly, because the value of the weighted graph is too large when σ is quite small. In this case, λ needs to be a relatively large value to ensure the balance between spatial and nonspatial information of the energy function. Similarly, if the value of λ is large, the resulting classification map looks

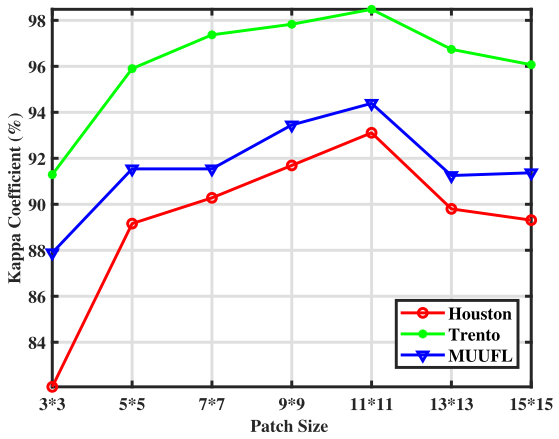


Fig. 7. Classification performance of the proposed HRWN with different patch scales.

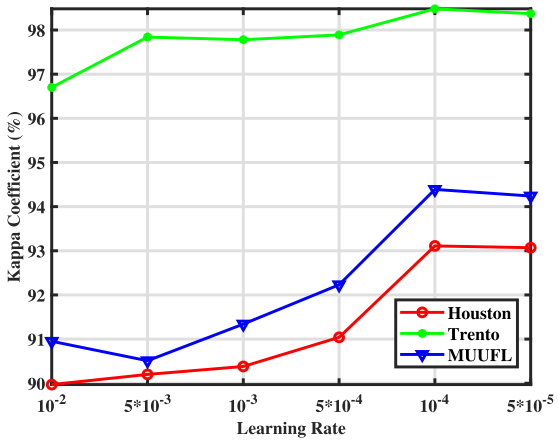


Fig. 9. Classification performance of the proposed HRWN with different learning rates.

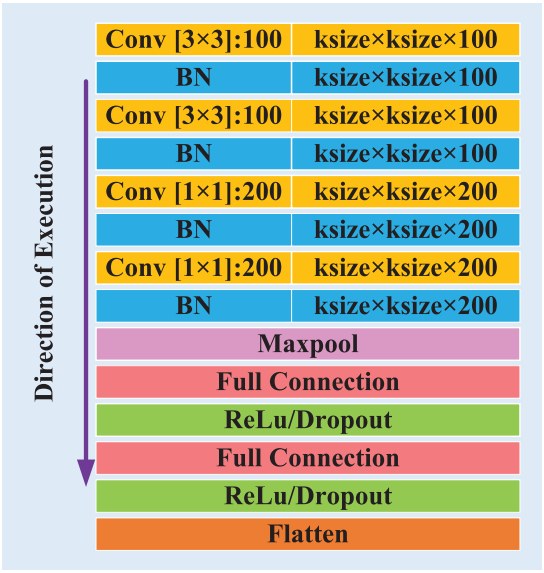


Fig. 8. Structure of the spatial-tunnel of HRWN (ksize denotes the size of neighbor patches).

relatively similar to that produced by the dual-tunnel CNN. When λ is too small, an overfitting issue may occur.

2) Multiscale Comparison: The effect of several different sizes of the neighbor patch is further studied. The performance on different window sizes of 7×7, 11×11, and 15×15 is demonstrated in Fig. 7. The experimental results indicate that features extracted by various patch sizes yield different classification performance. The experimental results indicate that features extracted by various patch sizes yield different classification performance. The parameter fitting on training and validation samples during the training process is shown in Fig. 8. The experimental results indicate that the patch size influences the classification performance. Fig. 7 shows that the optimal patch size is 11×11. It also shows that the result for 13×13 is worse than for 9×9, which is because a larger patch containing more classes in one patch is difficult to classify.

3) Learning Rate Comparison: The learning rate controls the step of gradient descent in the training process, and is one of the factors determining the convergence rate. The parameter is set with an initial value according to Adam [48]. In the training process of HRWN, different learning rates are tested, and the corresponding spatial–spectral features are utilized. Based on our empirical study, classification performance may not be improved with a higher learning rate. As listed in Fig. 9, the best learning rate is 0.0001 for all the three experimental data sets. Since the Trento data set contains fewer classes and more uniform regions, it results in relatively higher classification accuracy, which relatively improves with increasing learning rates.

C. Classification Performance and Analysis

To validate the effectiveness, the proposed HRWN is compared with several other classifiers, such as the standard SVM [17], ELM [18], recently proposed CNN-PPF [19], two-branch CNN [29], context CNN [20], CRNN [21], and CNN-MRF [35]. Note that SVM is implemented using the LIBSVM toolbox,[3] and trained using a Gaussian radial basis function (with gamma $=$ 0.5). We optimized the parameters of all the methods used in the comparison. Furthermore, several classification frameworks including SVM, ELM, CNN-PPF, two-branch CNN, two-branch CNN(Merge), context CNN, CRNN, and CNN-MRF, are discussed. In two-branch CNN (Merge), the input of two-branch CNN [29] is changed to the fused data. Besides, for a fair comparison, all the training and testing samples are exactly the same. In the data preprocessing procedure of the proposed HRWN and other competitive classifiers [21], [29], data augmentation is used to increase the diversity of training samples, improve the robustness of the model, and avoid overfitting. The used augmentation method applies orientation flipping, injection of noise randomly to the data, which reduces the model's dependence on certain attributes and thereby improving the generalization performance.

Tables II, IV and VI list the overall accuracy (OA), average accuracy (AA), and Kappa coefficient (Kappa) for three experimental data sets. The proposed HRWN performs

[3]http://www.csie.ntu.edu.tw/ cjlin/libsvm/

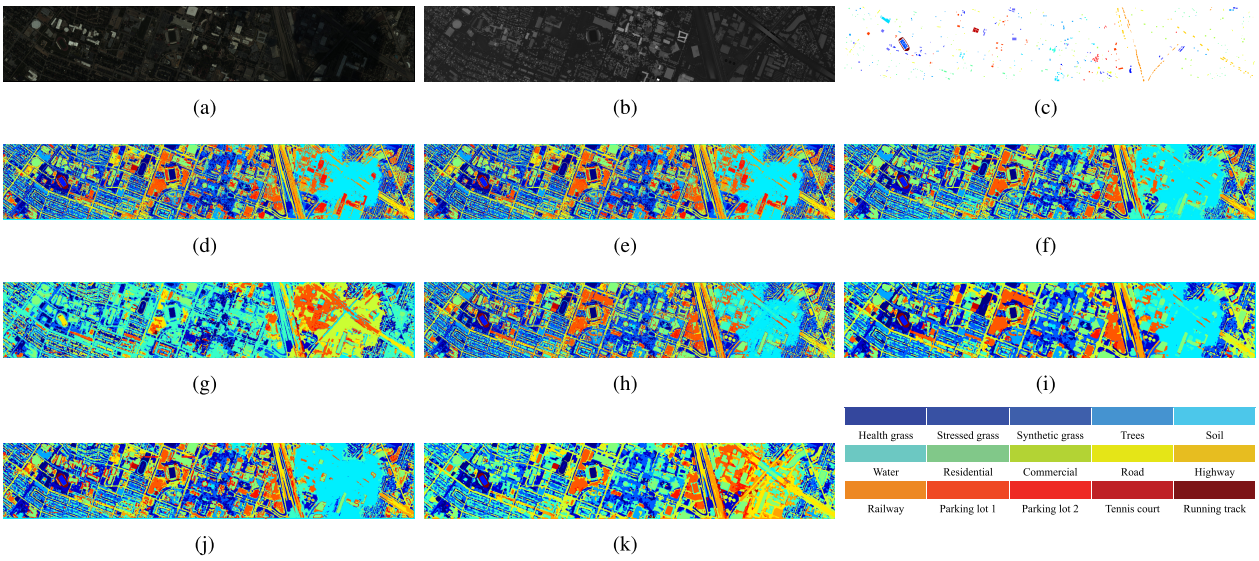


Fig. 10. Classification maps for the Houston data obtained with different methods. (a) Pseudocolor image for HSI. (b) LiDAR-based DSM. (c) Ground-truth map. (d) SVM (80.49%). (e) ELM (81.92%). (f) CNN-PPF (83.33%). (g) Two-Branch CNN (88.91%). (h) Context CNN (86.90%). (i) CNN-MRF (90.61%). (j) CRNN (88.55%). (k) HRWN (93.61%).

TABLE II
COMPARISON OF THE CLASSIFICATION ACCURACY (%) USING THE HOUSTON DATA

No.	Class(Train/Test)	Performance								
		SVM	ELM	CNN-PPF	Two-Branch CNN	Two-Branch CNN (Merge)	Context CNN	CNN-MRF	CRNN	HRWN
1	Health grass (198/1053)	82.43± 6.95	83.10±0.13	83.57±0.42	83.10±0.22	83.10±0.24	84.89±1.12	**85.77±0.31**	83.00±0.23	85.61±0.09
2	Stressed grass (190/1064)	82.05±1.31	83.70±4.03	**98.21±0.15**	84.10±1.85	81.20±1.49	87.40±1.40	86.28±0.45	79.41±0.98	85.17±1.25
3	Synthetic grass (192/505)	99.80±0.04	**100.00±0.00**	98.42±0.44	**100.00±0.00**	**100.00±0.00**	99.86±0.46	99.00±0.02	99.8±0.12	99.57±0.18
4	Tress (188/1056)	92.80±3.84	91.86±0.92	**97.73±0.27**	93.09±0.08	92.90±0.10	93.49±3.42	92.85±0.87	90.15±0.78	92.20±0.62
5	Soil (186/1056)	98.48±0.97	98.86±0.38	96.50±1.03	**100.00±0.00**	99.81±0.30	**100.00±0.00**	**100.00±0.00**	99.71±0.12	**100.00±0.00**
6	Water (182/143)	95.10±2.06	95.10±1.20	97.20±2.29	99.30±0.08	**100.00±0.00**	98.77±0.52	98.15±0.25	83.21±5.44	98.15±0.86
7	Residential (196/1072)	75.47±0.75	80.04±1.69	85.82±0.59	92.82±3.16	92.54±2.20	82.81±4.54	91.64±0.38	88.06±1.76	**95.98±0.89**
8	Commercial (191/1036)	46.91±8.43	68.47±5.58	56.51±0.58	82.34±4.29	94.87±1.10	78.78±5.30	80.79±0.56	88.61±2.37	**97.59±0.52**
9	Road (193/1059)	77.53±2.76	84.80±2.92	71.20±1.04	84.70±5.72	83.85±2.65	82.51±4.41	**91.37±2.08**	66.01±9.32	88.66±0.63
10	Highway (191/1036)	60.04±1.05	49.13±5.27	57.12±1.82	65.44±3.47	69.89±5.52	59.41±4.83	73.35±0.84	52.22±7.84	**86.23±0.34**
11	Railway (181/1054)	81.02±2.26	80.27±7.03	80.55±0.36	88.24±5.01	86.15±3.97	83.24±7.65	**98.87±1.33**	81.97±3.02	97.98±3.06
12	Parking lot 1 (192/1041)	85.49±2.99	79.06±2.84	62.82±2.64	89.53±1.77	92.60±1.29	92.13±0.67	89.38±2.71	69.83±7.64	**97.40±0.44**
13	Parking lot 2 (184/285)	75.09±0.97	71.58±1.20	63.86±0.77	92.28±0.37	79.30±3.42	**94.88±0.31**	92.75±1.30	79.64±5.13	91.47±1.25
14	Tennis court (181/247)	**100.00±0.00**	99.60±0.05	**100.00±0.00**	96.76±1.10	**100.00±0.00**	99.77±0.03	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**
15	Running track (187/473)	98.31±0.14	98.52±0.38	98.10±0.39	99.79±0.10	**100.00±0.00**	98.79±0.47	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**
	OA	80.49±0.38	81.29±1.01	83.33±0.70	87.98±0.24	88.91±0.57	86.90±1.14	90.61±0.54	88.55±0.71	**93.61±0.75**
	AA	83.37±1.10	84.27±0.34	83.21±0.33	90.11±0.52	90.42±0.86	89.11±1.45	92.01±0.48	90.30±0.64	**94.40±0.63**
	Kappa	78.98±0.45	80.45±0.77	81.88±0.77	86.98±0.27	87.96±0.62	85.89±2.53	89.87±0.58	87.56±0.77	**93.09±0.81**

obviously better than other methods. Taking the Houston data, for example, the Kappa of the proposed HRWN is 93.09%, which is the best among all the classifiers. Due to the more robust extraction of global and local features, the hierarchical model for HSI, LiDAR-based DSM, and fused data can provide significant improvement. One of the main challenges in classification is the problem of misclassification. Specifically, it is common to misclassify soil as shrub or grass as crops in natural settings. In order to evaluate the proposed HRWN in terms of misclassification, the confusion matrices for three data sets are shown in Tables III, V and VII. The conditional Kappa (with 5% confidence interval) for the classes are listed as a reliable indicator of classwise accuracy. The conditional Kappa is a statistic that is used to measure interrater reliability for qualitative items. The HSI data should not be used to differentiate targets consisting of the same material such as road and highway in Houston data, grass and ground in MUUFL data. However, the efficient usage of LiDAR-based DSM enables the proposed HRWN to distinguish objects with different elevations as shown in Tables III, V and VII. On the other hand, the spatial–spectral information obtained by the dual-tunnel CNN branch can enable the classifier to distinguish targets with similar elevations. Also, the pixel affinity branch using only LiDAR-based DSM may degrade the performance. Thus, the trade-off between spectral–spatial similarity and elevation contrast affected by λ is demonstrated in Fig. 6.

For qualitative evaluation of the classification performance, visual maps are illustrated in Figs. 10–12. Also, the ground-truth map and pseudocolor maps of entire image scenes are provided for clarity. The proposed method produces the most accurate classification maps. The use of a hierarchical random walk layer enables the proposed HRWN method to produce

TABLE III
CONFUSION MATRIX OF THE HOUSTON DATA CLASSIFICATION

Classified Data	Reference Data															
	Health grass	Stressed grass	Synthetic grass	Trees	Soil	Water	Residential	Commercial	Road	Highway	Railway	Parking lot 1	Parking lot 2	Tennis court	Running track	Total
Health grass	1071	0	0	0	0	2	24	0	100	0	0	0	0	54	0	1251
Stressed grass	0	1068	32	0	0	0	28	0	11	0	115	0	0	0	0	1254
Synthetic grass	0	0	694	0	0	0	0	0	0	0	0	0	0	0	3	697
Trees	0	0	0	1147	0	0	88	0	0	0	0	0	9	0	0	1244
Soil	0	0	0	0	1242	0	0	0	0	0	0	0	0	0	0	1242
Water	0	0	0	0	0	319	6	0	0	0	0	0	0	0	0	325
Residential	0	0	0	0	0	0	1217	0	18	6	27	0	0	0	0	1268
Commercial	0	0	0	0	0	0	1	1214	0	0	27	2	0	0	0	1244
Road	0	0	0	0	24	0	0	0	1110	105	12	1	0	0	0	1252
Highway	0	0	0	0	0	0	0	139	0	1058	0	30	0	0	0	1227
Railway	0	0	0	0	0	0	0	0	15	10	1210	0	0	0	0	1235
Parking lot 1	0	0	0	0	0	0	0	0	0	32	0	1201	0	0	0	1233
Parking lot 2	0	0	0	0	0	0	0	0	0	0	0	40	429	0	0	469
Tennis court	0	0	0	0	0	0	0	0	0	0	0	0	0	428	0	428
Running track	0	0	0	0	0	0	0	0	0	0	0	0	0	0	660	660
Total	1071	1068	726	1147	1266	321	1364	1353	1254	1211	1391	1274	438	482	663	15029
95% Conditional Kappa(%)	91.06	91.32	97.42	95.59	98.96	98.73	91.76	92.88	87.55	85.63	91.41	95.43	94.43	93.88	99.76	

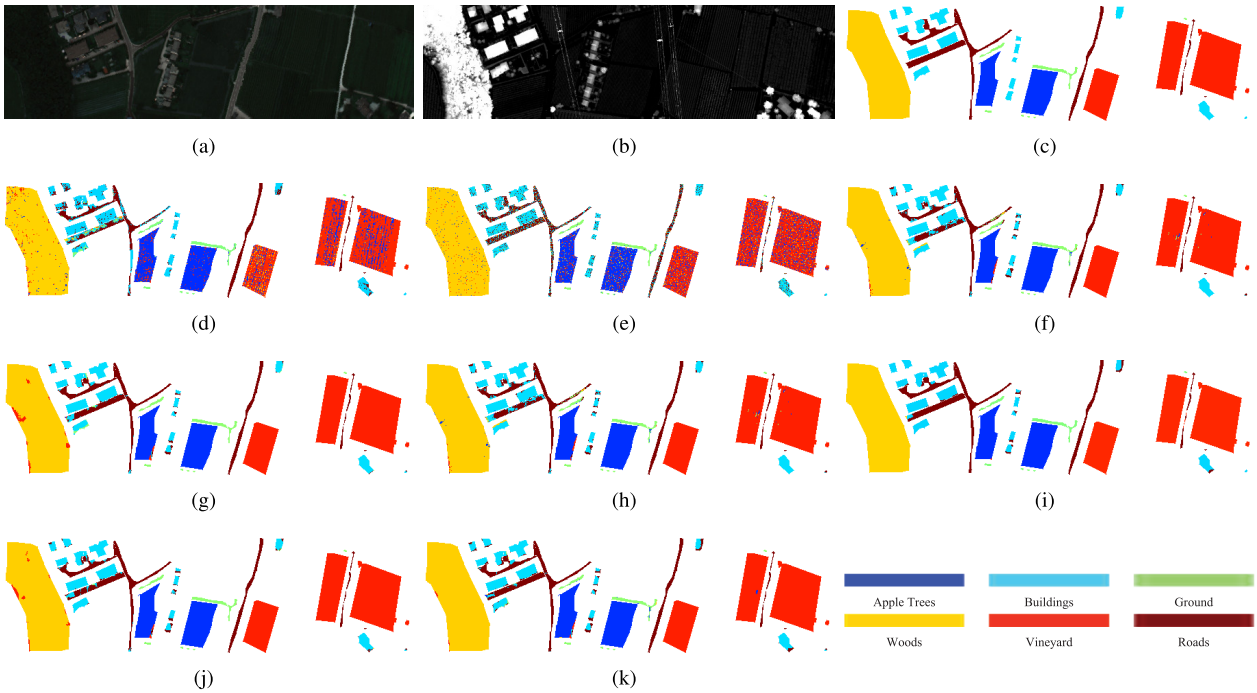


Fig. 11. Classification maps for the Trento data obtained with different methods including (a) Pseudocolor image for HSI, (b) LiDAR-based DSM, (c) ground-truth map, (d) SVM (92.77%), (e) ELM (91.32%), (f) CNN-PPF (94.76%), (g) two-branch CNN (97.92%), (h) context CNN (96.11%), (i) CNN-MRF(98.40%), (j) CRNN (97.22%), and (k) HRWN (98.86%).

spatially smooth results with less false boundaries. It can be concluded that the visual results are consistent with those in Tables II, IV and VI.

For the sensitivity of the training-samples size, classification results with different numbers of training samples are shown in Fig. 13. The percentage of training samples varies from 20% to 100% (100% percentage means using all the training samples of original data sets listed in Tables II, IV and VI). As for various percentages, training samples are selected randomly and averaged results are reported. Obviously, the proposed method consistently outperforms other classifiers. Note that even for a small training data size, such as 20% of the training data, the proposed network still provides excellent classification performance. In Fig. 13(c), the Kappa of the proposed HRWN is approximately 85%, whereas those of other methods are all below 75%. This obvious improvement gap verifies the effectiveness of the HRWN method.

To test the applicability of HRWN for medium or low-resolution data, the experimental data sets used are resampled to lower resolution. To test HRWN with respect to the quality of the input data, a Gaussian downsampling operation is applied to decrease the spatial resolution of both the HSI and LiDAR-based DSM. Then, the data set is simply interpolated to match the size of the original ones. For the sensitivity of all

TABLE IV

COMPARISON OF THE CLASSIFICATION ACCURACY (%) USING THE TRENTO DATA

No.	Class(Train/Test)	Performance								
		SVM	ELM	CNN-PPF	Two-Branch CNN	Two-Branch CNN (Merge)	Context CNN	CNN-MRF	CRNN	HRWN
1	Apple trees (129/3905)	88.62±0.30	92.17±0.28	90.11±2.70	98.07±1.73	98.04±0.24	93.53±1.04	**99.95±0.05**	98.39±0.12	99.75±0.06
2	Buildings (125/2778)	94.04±0.16	90.91±0.47	83.34±0.88	**95.21±0.50**	94.56±1.48	91.97±1.88	89.97±2.96	90.46±1.56	94.32±0.94
3	Ground (105/374)	93.53±0.91	94.36±0.48	71.13±3.30	93.32±1.04	94.36±0.82	98.33±0.54	98.33±0.97	**99.79±0.08**	98.75±0.07
4	Woods (154/8969)	98.90±0.70	97.12±0.15	99.04±0.24	99.93±0.02	96.57±1.04	96.50±0.39	**100.00±0.00**	96.96±0.51	**100.00±0.00**
5	Vineyard (184/10317)	88.96±0.36	78.58±3.27	99.37±0.09	98.78±0.27	99.62±0.07	98.49±0.47	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**
6	Roads (122/3525)	91.75±0.60	63.23±1.11	89.73±4.10	89.98±1.70	74.47±1.42	71.24±2.77	**97.86±0.50**	81.63±0.38	94.90±1.14
	OA	92.77±0.13	85.81±0.78	94.76±0.72	97.92±0.26	96.44±0.22	93.73±0.42	98.40±1.19	97.30±0.00	**98.86±0.38**
	AA	92.63±0.14	86.06±0.92	88.97±1.23	96.19±0.60	92.94±0.24	94.23±0.24	97.04±0.63	94.54±1.04	**97.95±1.11**
	Kappa	95.85±0.17	81.36±1.07	93.04±0.97	96.81±0.34	95.26±0.29	93.70±0.57	97.86±0.74	96.39±0.70	**98.48±0.51**

TABLE V

CONFUSION MATRIX OF THE TRENTO DATA CLASSIFICATION

Classified Data	Reference Data						
	Apple trees	Buildings	Ground	Woods	Vineyard	Roads	Total
Apple trees	4024	0	10	0	0	0	4034
Buildings	0	2738	0	0	0	165	2903
Ground	0	0	473	0	0	6	479
Woods	0	0	0	9123	0	0	9123
Vineyard	0	0	0	0	10501	0	10501
Roads	0	101	0	5	56	3012	3647
Total	4024	2839	483	9128	10557	3183	30214
95% Conditional Kappa(%)	99.86	94.88	98.31	99.96	99.59	86.79	

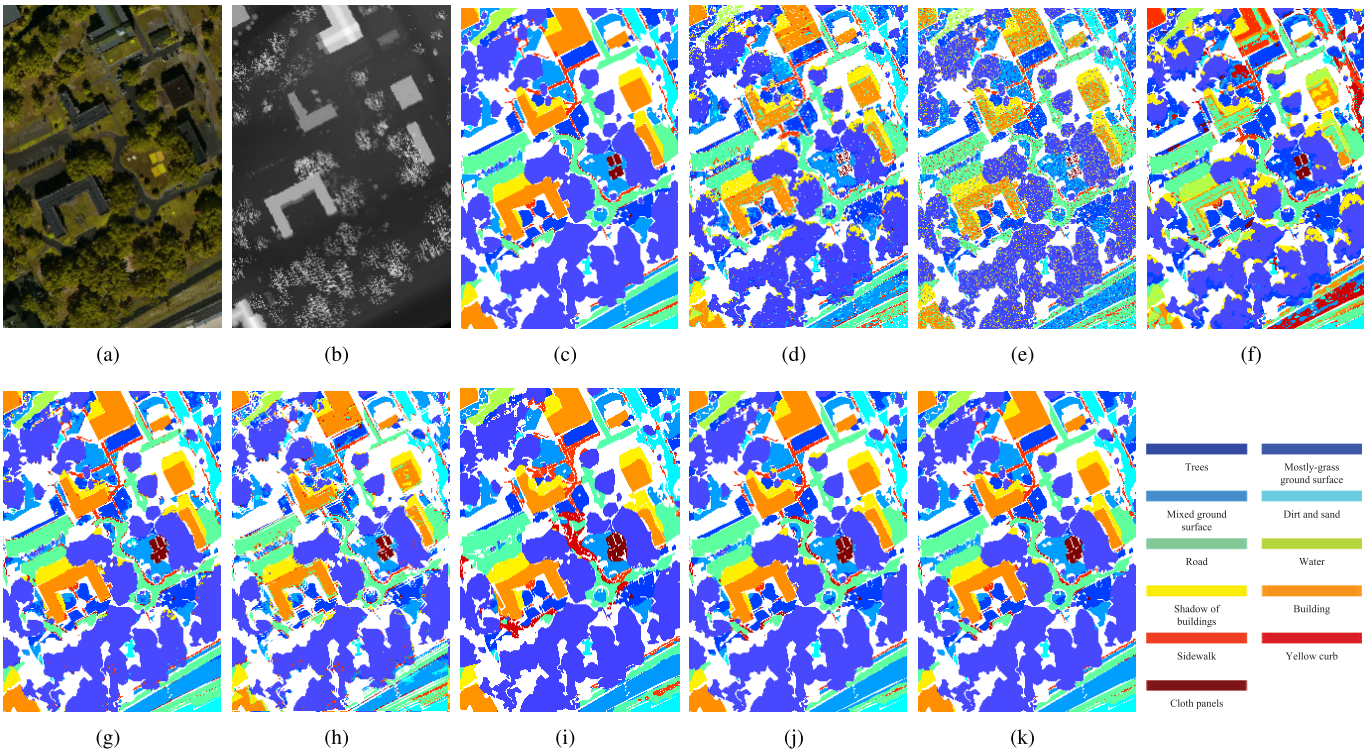


Fig. 12. Classification maps for the MUUFL Gulfport data obtained with different methods including (a) pseudocolor image for HSI, (b) LiDAR-based DSM, (c) ground truth map, (d) SVM (82.74%), (e) ELM (80.41%), (f) CNN-PPF (90.97%), (g) two-branch CNN (90.35%), (h) context CNN (86.07%), (i) CNN-MRF(88.94%), (j) CRNN (91.38%), and (k) HRWN (94.31%).

the methods to the spatial resolutions of data sets, classification performance with different resolutions of data sets are listed in Fig. 14. The resampled scales of HSI and LiDAR-based DSM are changed from 0.2 to 1 (1 means using the original data sets listed in Table I). From Fig. 14, the performance of all the methods is affected by resolution changes, and the HRWN consistently outperforms other classifiers. Note that even for the low-resolution data, the proposed network still provides better classification performance. In Fig. 14(a), the Kappa of the proposed HRWN performs approximately 3% better than

TABLE VI

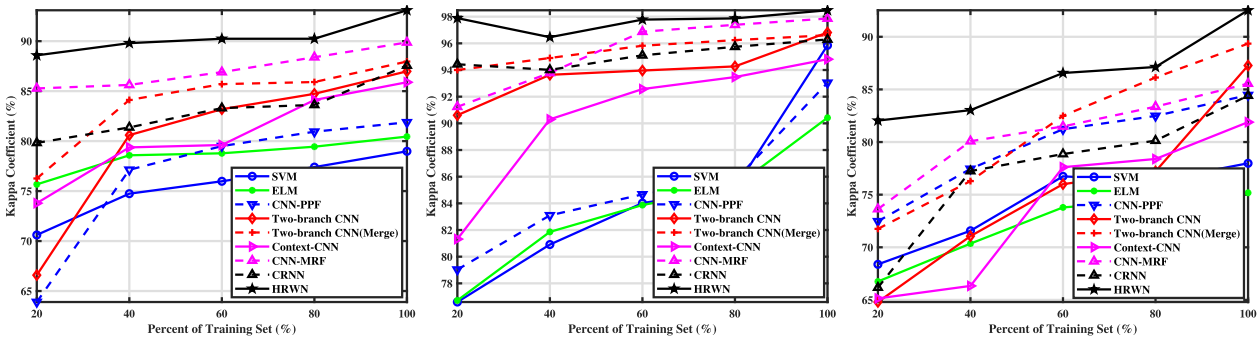
COMPARISON OF THE CLASSIFICATION ACCURACY (%) USING THE MUUFL GULFPORT DATA

No.	Class(Train/Test)	Performance								
		SVM	ELM	CNN-PPF	Two-Branch CNN	Two-Branch CNN (Merge)	Context CNN	CNN-MRF	CRNN	HRWN
1	Trees (150/23246)	83.22±0.21	81.02±0.97	89.07±7.28	92.35±0.64	94.28±0.75	87.47±4.25	93.04±0.87	90.40±1.57	**96.08±0.50**
2	Mostly grass (150/4270)	82.34±0.12	82.01±1.13	85.71±0.71	59.30±6.23	82.84±1.72	73.23±7.79	60.17±10.58	85.08±4.01	**90.59±0.99**
3	Mixed ground surface (150/6882)	72.87±0.11	69.78±3.87	80.15±1.40	**94.47±1.48**	84.58±0.51	77.13±7.41	90.60±2.74	79.29±5.65	90.38±1.33
4	Dirt and sand (150/1826)	89.27±0.18	87.95±0.76	93.10±0.62	93.74±3.31	97.32±0.39	90.80±0.62	97.20±0.86	97.32±1.80	**98.47±0.35**
5	Road (150/6687)	88.07±0.31	88.60±0.60	88.98±2.84	92.76±1.21	92.96±0.34	89.46±1.97	92.00±1.22	91.63±0.78	**93.36±0.90**
6	Water (150/466)	94.64±0.06	95.71±0.92	98.93±0.14	98.42±0.44	99.05±0.26	98.28±1.54	99.68±0.13	**100.00±0.00**	**100.00±0.00**
7	Building shadow (150/2233)	88.67±0.15	87.82±2.81	89.07±0.61	**95.68±0.70**	93.28±0.91	92.92±1.07	95.39±0.66	95.66±1.24	92.75±0.81
8	Building (150/6240)	81.67±0.39	68.94±3.46	92.15±1.85	94.01±0.98	95.42±0.46	82.93±10.05	94.71±0.87	95.90±0.23	**96.49±0.63**
9	Sidewalk (150/1385)	76.25±0.54	74.66±2.54	75.45±0.39	**86.64±4.30**	84.13±0.63	83.18±2.60	30.53±15.91	82.53±3.96	84.19±5.03
10	Yellow curb (150/183)	98.91±0.72	98.91±0.39	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	98.91±0.26	36.36±23.82	**100.00±0.00**	96.72±1.62
11	Cloth panels (150/269)	**100.00±0.00**	97.77±0.39	**100.00±0.00**	96.64±1.05	97.48±0.69	99.63±0.04	95.80±1.53	98.51±0.70	99.26±0.15
	OA	82.74±0.10	79.82±0.65	90.97±0.56	90.35±0.73	91.95±0.47	85.15±2.45	88.94±0.77	89.43±0.65	**94.31±0.18**
	AA	86.90±0.11	84.83±1.17	90.24±1.03	91.27±1.46	92.58±0.06	88.54±1.45	85.02±2.03	92.39±0.69	**94.39±0.50**
	Kappa	77.98±0.11	74.40±0.87	84.46±0.64	87.27±1.44	89.35±0.58	86.02±2.02	85.55±0.99	86.16±0.84	**92.52±0.75**

TABLE VII

CONFUSION MATRIX OF THE MUUFL GULFPORT DATA CLASSIFICATION

Classified Data	Reference Data											
	Trees	Mostly grass	Mixed ground surface	Dirt and sand	Road	Water	Building shadow	Building	Sidewalk	Yellow curb	Cloth panels	Total
Trees	22334	194	374	14	49	36	215	12	11	4	3	23246
Mostly grass	71	3868	259	0	2	20	9	9	5	26	1	4270
Mixed ground surface	87	410	6220	3	28	6	2	2	26	19	79	6882
Dirt and sand	3	0	0	1798	4	0	0	0	21	0	0	1826
Road	122	2	66	100	6243	0	25	1	47	81	0	6687
Water	0	0	0	0	0	466	0	0	0	0	0	466
Building shadow	70	12	32	0	23	0	2071	18	7	0	0	2233
Building	36	18	5	0	0	0	123	6021	37	0	0	6240
Sidewalk	33	21	12	43	85	0	2	18	1166	5	0	1385
Yellow curb	1	0	0	0	5	0	0	0	0	177	0	183
Cloth Panels	0	1	1	0	0	0	0	0	0	0	267	269
Total	22757	4526	6969	1958	6439	528	2447	6081	1320	312	350	53687
95% Conditional Kappa(%)	94.92	86.87	88.30	94.85	94.45	93.70	92.13	97.44	85.85	71.39	86.05	



(a) (b) (c)

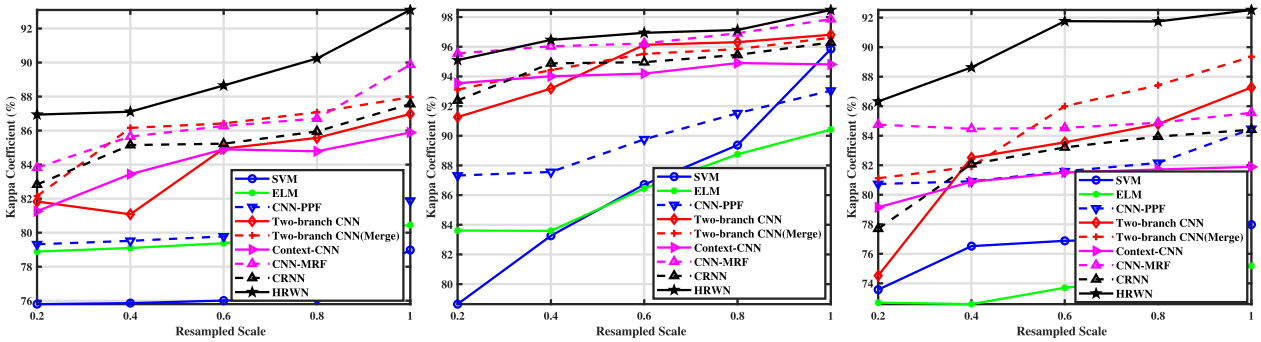
Fig. 13. Classification performance with different sizes of training samples using (a) Houston, (b) Trento, and (c) MUUFL.

other classifiers. This discussion shows the effectiveness of the HRWN method for medium resolution.

Actual HSI data inevitably contain considerable noise. To deal with feature noise and label noise, several noise-robust approaches have been proposed [49]–[51]. To test the applicability of HRWN for noisy data, Gaussian white noise with different degrees of zero mean is added to the fused data during simulation. It is assumed that the effect of noise on the density map satisfies the following model:

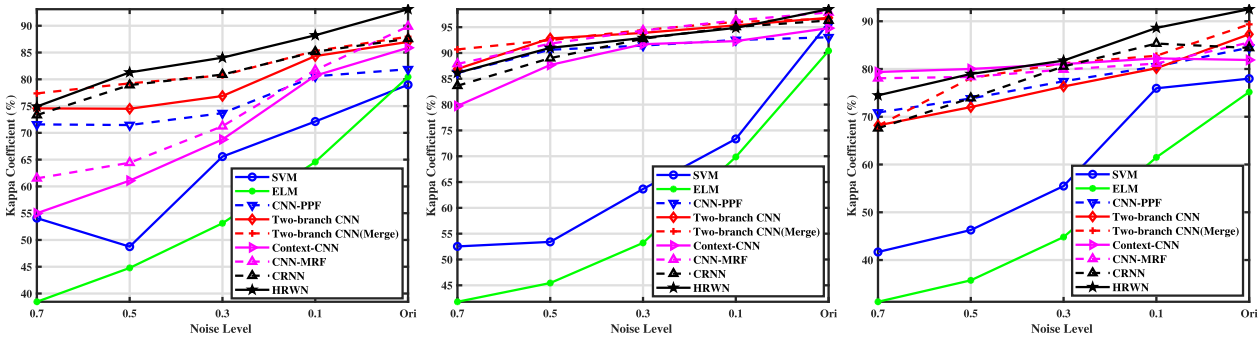
$$E' = E(1 + \sigma G) \tag{12}$$

where E and E' are, respectively, the original and noisy HSI and LiDAR-based DSM data sets, the parameter σ is the noise weight, G denotes Gaussian white noise. Fig. 15 shows the noise sensitivity of the classification performance of all methods. σ varies from 0.1 to 0.7 (ori means using the original data sets listed in Table I). Fig. 15 shows that the performance of HRWN is affected by the noise level. The proposed HRWN may classify real objects into neighbor regions embedded in noise. The hierarchical random walk may aggravate this effect and further lower the accuracy. Though the HRWN method is still effective under noisy conditions, its performance is not



(a) (b) (c)

Fig. 14. Classification performance with different resolutions of data sets using (a) Houston, (b) Trento, and (c) MUUFL.



(a) (b) (c)

Fig. 15. Classification performance under different noise level using (a) Houston, (b) Trento, and (c) MUUFL.

robust in the presence of considerable noise, which means further efforts are needed to improve its robustness.

To demonstrate the statistical significance of improved performance, the standardized McNemar's test [52], [53] is employed, and the multiclassifier z-test results are listed in Table VIII. The Z values of McNemar's test larger than 1.96 and 2.58 mean that two results are statistically different at the 95% and 99% confidence levels, respectively. The sign of Z indicates whether classifier 1 outperforms classifier 2 ($Z > 0$) or vice versa. In the experiment, we run the comparison between HRWN and other classifiers. As shown in Table VIII, all the values larger than 2.58 mean that the HRWN classifier outperforms other classifiers with a 99% confidence level. The standardized McNemar's test confirms that the proposed HRWN is highly discriminative.

Table IX summarizes the training and testing computational cost of the proposed method. All the experiments of computational time are implemented in the same configuration of hardware and software. The training process takes longer, and the test of the whole scene is relatively faster. The HRWN is not time-consuming because the iteration epochs for the HRWN are less than the competitive deep learning methods. The prior probability maps estimated by the dual-tunnel CNN using fewer epochs are usually noisy and not aligned with real boundaries of objects, which are improved by the hierarchical random walk layer.

TABLE VIII

STATISTICAL SIGNIFICANCE FROM THE STANDARDIZED MCNEMAR'S TEST ABOUT THE DIFFERENCE BETWEEN METHODS

the Houston	the Trento	the MUUFL
mean Z/significant ?		
HRWN *vs* SVM		
9.01/yes	33.35/yes	10.21/yes
HRWN *vs* ELM		
24.41/yes	27.74/yes	24.34/yes
HRWN *vs* CNN-PPF		
4.71/yes	27.68/yes	7.56/yes
HRWN *vs* Two-Branch CNN		
3.77/yes	7.25/yes	6.22/yes
HRWN *vs* Two-Branch CNN (Merge)		
4.49/yes	5.59/yes	8.02/yes
HRWN *vs* Context CNN		
3.77/yes	7.18/yes	9.43/yes
HRWN *vs* CNN-MRF		
11.09/yes	7.12/yes	5.27/yes
HRWN *vs* CRNN		
4.90/yes	4.00/yes	11.94/yes

D. Ablation Studies

In addition, we investigate and analyze the performance improvement of our HRWN method by comparing using different components, since the HRWN is involved with multiple steps, that is, data augmentations in preprocessing and hierarchical random layer. Successively implementing these steps of the proposed HRWN leads to a higher classification result for the HSI and LiDAR-based DSM. This also indicates the reasonableness and advancement of the proposed HRWN classification framework.

1) With/Without Hierarchical Random Walk Comparison: The hierarchical random walk layer plays a crucial role in achieving better classification performance. Considering the

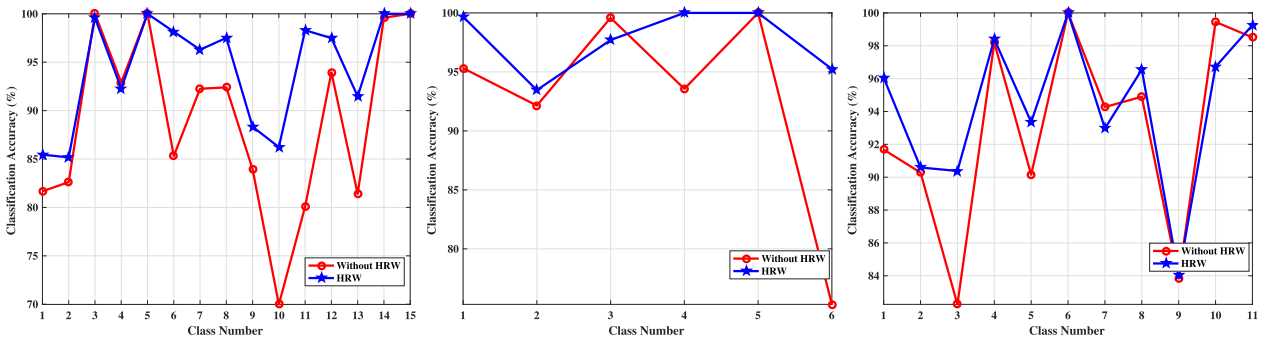


Fig. 16. Influence of with/without hierarchical random walk layer on classification accuracies of HRWN. (a) Houston. (b) Trento. (c) MUUFL.

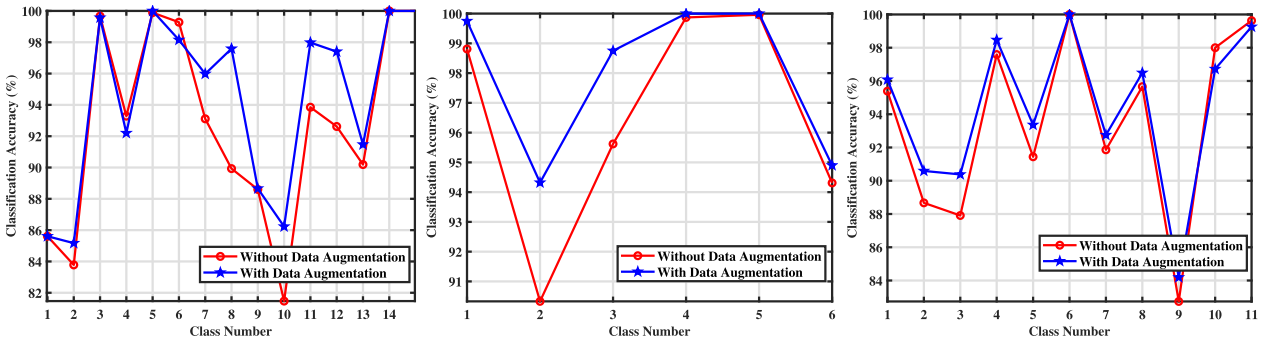


Fig. 17. Influence of with/without data augmentations on classification accuracies of HRWN. (a) Houston. (b) Trento. (c) MUUFL.

TABLE IX

ELAPSED TIME (*h*: HOURS, *s*: SECOND) OF TRAINING AND TESTING TIME FOR THE PROPOSED METHOD USING THE EXPERIMENTAL DATA SETS

		Houston	*Trento*	*MUUFL*
SVM	Training(s)	22.94	3.63	23.12
	Testing(s)	10.26	1.13	4.49
ELM	Training(s)	0.15	0.08	0.07
	Testing(s)	0.36	0.26	1.05
CNN-PPF	Training(h)	10.92	8.60	10.21
	Testing(s)	8.34	7.46	11.27
Two-Branch CNN	Training(h)	0.20	0.15	0.18
	Testing(s)	15.24	17.21	28.45
Context CNN	Training(h)	0.61	0.54	0.52
	Testing(s)	9.02	8.70	20.15
CNN-MRF	Training(h)	0.72	0.63	0.67
	Testing(s)	12.01	10.78	11.95
CRNN	Training(h)	0.23	0.19	0.21
	Testing(s)	12.36	14.52	12.68
HRWN	Training(h)	0.10	0.03	0.05
	Testing(s)	20.86	25.95	43.01

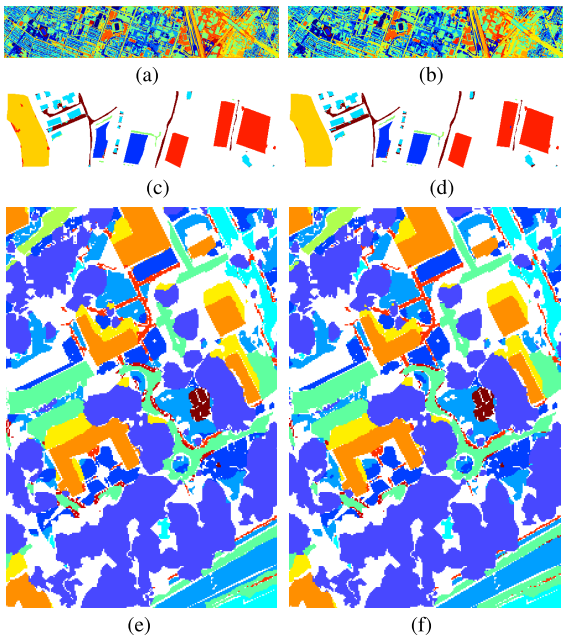


Fig. 18. Classification maps obtained by HRWN with/without data augmentations. (a) Houston data set without data augmentations (OA = 91.56% and Kappa = 90.88%). (b) Houston data set with data augmentations (OA = 93.61% and Kappa = 93.09%). (c) Trento data set without data augmentations (OA = 98.19% and Kappa = 97.59%). (d) Trento data set with data augmentations (OA = 98.86% and Kappa = 98.48%). (e) MUUFL data set without data augmentations (OA = 93.11% and Kappa = 90.96%). (f) MUUFL data set without data augmentations (OA = 94.31% and Kappa = 92.52%).

experimental analysis of a random walk layer, the dual-tunnel CNN branch is trained first, and the random walk strategy is adopted based on the pretrained layers. Fig. 16 illustrates the classification performance with and without the hierarchical random walk layer. It is observed that the random walk layer achieves better performance than that without a random walk for most of the classes. However, for some classes with spatial dispersion and small areas, the hierarchical random walk may misclassify them into large neighboring areas. In addition, because the random walk layer combines the LiDAR-based DSM information, the performance of the classifier in classes with similar elevations may be degraded. For example, ground

class in the Trento data set may be misclassified as roads, building shadows and yellow curb classes in the MUUFL data set are more likely to be classified as adjacent areas. However, the random walk layer can effectively improve the classification performance in general.

2) With/Without Data Augmentations: In the data preprocessing procedure of the proposed HRWN and other competitive classifiers, data augmentation is used to increase the diversity of training samples, improve the robustness of the model, and avoid overfitting. To validate the improvement of performance by data augmentation, both augmented and not augmented results are provided in Fig. 17. Classification accuracy is reduced when disabling data augmentation because of the insufficient training. From the training aspect, data augmentation is useful for increasing the diversity of training samples, improve the robustness of the model. Take the training and evaluating of the University of Houston data set as an example, the used augmentation method applies orientation flipping, injection of noise randomly to the data, which reduces the model's dependence on certain attributes and thereby improving the generalization performance. From the quantitative and qualitative comparisons in Fig. 18, using data augmentation is useful for sufficient training and offers a $1\% - 2\%$ improvement in the Kappa coefficient.

IV. Conclusion

In this article, an HRWN was proposed to explore deep features for the joint classification of HSI and LiDAR-based DSM, which effectively mitigated the issue of boundary localization and spatial inconsistency. The proposed HRWN jointly optimized dual-tunnel CNN and pixelwise affinity via a novel random walk layer, which enforced spatial consistency in the deeper layers of the network. Experimental results validated with real multisensor remote sensing data sets demonstrated that the proposed method can significantly outperform other state-of-the-art methods. Compared with other joint classification methods, the major advantage of the proposed method is that it combines spectral–spatial classification, graph-based elevation feature and hierarchical random walk in a joint learning framework. With the proposed framework, the problems of weak localization around boundaries and spatial fragmentation of the classification map are reduced. As a result, the proposed HRWN is able to obtain a satisfactory classification accuracy. However, the proposed method still has some of the common limitations of HSI classification methods. For example, the performance declines sharply on a noisy data set. As actual HSI data inevitably contain considerable noise, preprocessing is needed to improve the robustness of the proposed HRWN. Furthermore, extending the proposed HRWN for large-scene HSIs and LiDAR-based DSMs (with higher spatial resolution) also requires more research. Lastly, lowering the computational burden of the proposed HRWN can further promote its practical application. Our code and all the results are available at https://github.com/xudongzhao461/HRWN.

Acknowledgment

The authors would like to thank the Editor-in-Chief, the anonymous Associate Editor, and the reviewers for their rigorous and valuable suggestions, which greatly helped them to improve the technical quality and presentation of this article.

References

[1] W. Li, E. W. Tramel, S. Prasad, and J. E. Fowler, "Nearest regularized subspace for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 477–489, Jan. 2014.

[2] X. Zheng, Y. Yuan, and X. Lu, "Dimensionality reduction by spatial–spectral preservation in selected bands," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5185–5197, Sep. 2017.

[3] J. Jung, E. Pasolli, S. Prasad, J. C. Tilton, and M. M. Crawford, "A framework for land cover classification using discrete return LiDAR data: Adopting pseudo-waveform and hierarchical segmentation," *IEEE J. Sel. Topics Appl. Earth Observat., Remote Sens.*, vol. 7, no. 2, pp. 491–502, Feb. 2014.

[4] C. Weitkamp, *LiDAR: Range-Resolved Optical Remote Sensing of the Atmosphere*, vol. 102. Springer, 2006.

[5] 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, Jan. 2009.

[6] M. Khodadadzadeh, A. Cuartero, J. Li, A. Felicisimo, and A. Plaza, "Fusion of hyperspectral and lidar data using generalized composite kernels: A case study in Extremadura, Spain," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 61–64.

[7] M. Zhang and W. Li, "Collaborative classification of hyperspectral and visible images with convolutional neural network," *J. Appl. Remote Sens.*, vol. 11, no. 4, Sep. 2017, 042607.

[8] 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.

[9] W. Liao, R. Bellens, A. Pizurica, S. Gautama, and W. Philips, "Graph-based feature fusion of hyperspectral and LiDAR remote sensing data using morphological features," in *Proc. IGARSS*, Melbourne, VIC, Australia, Jul. 2013, pp. 4942–4945.

[10] W. Liao, R. Bellens, A. Pizurica, S. Gautama, and W. Philips, "Combining feature fusion and decision fusion for classification of hyperspectral and LiDAR data," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Montreal, QC, Canada, Jul. 2014, pp. 1241–1244.

[11] C. Zhao, X. Gao, Y. Wang, and J. Li, "Efficient multiple-feature learning-based hyperspectral image classification with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4052–4062, Jul. 2016.

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

[13] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[14] W. Li, C. Chen, M. Zhang, H. Li, and Q. Du, "Data augmentation for hyperspectral image classification with deep CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 593–597, Apr. 2019.

[15] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.

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

[17] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul. 1998.

[18] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

[19] W. Li, G. Wu, F. Zhang, and Q. Du, "Hyperspectral image classification using deep pixel-pair features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 844–853, Feb. 2017.

[20] H. Lee and H. Kwon, "Going deeper with contextual CNN for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4843–4855, Oct. 2017.
[21] H. Wu and S. Prasad, "Convolutional recurrent neural networks forHyperspectral data classification," *Remote Sens.*, vol. 9, no. 3, p. 298, 2017.
[22] M. E. Paoletti *et al.*, "Capsule networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2145–2160, Apr. 2019.
[23] Z. Zhong, J. Li, L. Ma, H. Jiang, and H. Zhao, "Deep residual networks for hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 1824–1827.
[24] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, "Deep pyramidal residual networks for spectral–spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 740–754, Feb. 2018.
[25] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
[26] X. Kang, B. Zhuo, and P. Duan, "Dual-path network-based hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 447–451, Mar. 2019.
[27] A. Merentitis, C. Debes, R. Heremans, and N. Frangiadakis, "Automatic fusion and classification of hyperspectral and LiDAR data using random forests," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 1245–1248.
[28] M. Khodadadzadeh, J. Li, S. Prasad, and A. Plaza, "Fusion of hyperspectral and LiDAR remote sensing data using multiple feature learning," *IEEE J. Sel. Topics Appl. Earth Observat., Remote Sens.*, vol. 8, no. 6, pp. 2971–2983, Jun. 2015.
[29] 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.
[30] 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. 1, pp. 100–111, Jan. 2020.
[31] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.
[32] W. Li, S. Prasad, and J. E. Fowler, "Hyperspectral image classification using Gaussian mixture models and Markov random fields," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 153–157, Jan. 2014.
[33] 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.
[34] F. I. Alam, J. Zhou, A. W.-C. Liew, X. Jia, J. Chanussot, and Y. Gao, "Conditional random field and deep feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1612–1628, Mar. 2019.
[35] X. Cao, F. Zhou, L. Xu, D. Meng, Z. Xu, and J. Paisley, "Hyperspectral image classification with Markov random fields and a convolutional neural network," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2354–2367, May 2018.
[36] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.
[37] L. Grady, "Multilabel random walker image segmentation using prior models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 763–770.
[38] X. Dong, J. Shen, L. Shao, and L. Van Gool, "Sub-Markov random walk for image segmentation," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 516–527, Feb. 2016.
[39] G. Bertasius, L. Torresani, S. X. Yu, and J. Shi, "Convolutional random walk networks for semantic image segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 858–866.
[40] X. Kang, S. Li, L. Fang, M. Li, and J. A. Benediktsson, "Extended random walker-based classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 144–153, Jan. 2015.
[41] B. Sun, X. Kang, S. Li, and J. A. Benediktsson, "Random-Walker-Based collaborative learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 212–222, Jan. 2017.
[42] B. Cui, X. Xie, X. Ma, G. Ren, and Y. Ma, "Superpixel-based extended random walker for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3233–3243, Jun. 2018.
[43] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent 6 011 875, 2000.
[44] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, 2013, p. 3.
[45] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167
[46] G. Paul, Z. Alina, C. Ryan, A. Jen, and T. Grady, "MUUFL gulfport hyperspectral and LiDAR airborne data set," Univ. Florida, Gainesvill, FL, USA, Tech. Rep. REP-2013-570, 2013.
[47] *The Mathworks*, MATLAB MathWorks, Natick, MA, USA, 1992.
[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[49] J. Jiang, J. Ma, Z. Wang, C. Chen, and X. Liu, "Hyperspectral image classification in the presence of noisy labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 851–865, Feb. 2019.
[50] D. Varade, A. K. Maurya, and O. Dikshit, "Unsupervised hyperspectral band selection using ranking based on a denoising error matching approach," *Int. J. Remote Sens.*, vol. 40, no. 20, pp. 8031–8053, Oct. 2019.
[51] C. Pelletier, S. Valero, J. Inglada, N. Champion, C. Marais Sicre, and G. Dedieu, "Effect of training class label noise on classification performances for land cover mapping with satellite image time series," *Remote Sens.*, vol. 9, no. 2, p. 173, 2017.
[52] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865–4876, Dec. 2011.
[53] G. Foody, "Assessing the accuracy of remotely sensed data: Principles and practices," *Photogramm. Rec.*, vol. 25, no. 130, pp. 204–205, 2010.

Xudong Zhao (Student Member, IEEE) received the B.S. degree from the Science and Technology of Electronic Information Department, Beijing Institute of Technology, Beijing, China, in 2016, where he is pursuing the Ph.D. degree with the Information and Communication Engineering Department. He is also pursuing the Ph.D. degree in computer science engineering with Ghent University, Ghent, Belgium.

His research interests include fractional signal processing and remote sensing image process.

Ran Tao (Senior Member, IEEE) received the B.S. degree from the Electronic Engineering Institute of PLA, Hefei, China, in 1985, and the M.S. and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in 1990 and 1993, respectively.

He has been a Senior Visiting Scholar with the University of Michigan, Ann Arbor, MI, USA, in 2001, and the University of Delaware, Newark, DE, USA, in 2016. He is a Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. He has 3 books and more than 100 peer-reviewed journal articles. His research interests include fractional fourier transform and its applications, theory, and technology for radar and communication systems.

Dr. Tao is a fellow of the Institute of Engineering and Technology and the Chinese Institute of Electronics. He was a recipient of the National Science Foundation of China for Distinguished Young Scholars in 2006, and a Distinguished Professor of Changjiang Scholars Program in 2009. He was a recipient of the First Prize of Science and Technology Progress in 2006 and 2007, and the First Prize of Natural Science in 2013, both awarded by the Ministry of Education. He has been a Chief Professor of the Creative Research Group with the National Natural Science Foundation of China since 2014, and he was a Chief Professor of the Program for Changjiang Scholars and Innovative Research Team in University from 2010 to 2012. He is the Vice Chair of the IEEE China Council and the International Union of Radio Science (URSI) China Council and a member of the Wireless Communication and Signal Processing Commission, URSI.

Wei Li (Senior Member, IEEE) received the B.E. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2007, the M.S. degree in information science and technology from Sun Yat-sen University, Guangzhou, China, in 2009, and the Ph.D. degree in electrical and computer engineering from Mississippi State University, Starkville, MS, USA, in 2012.

Subsequently, he spent one year as a Post-Doctoral Researcher with the University of California at Davis, Davis, CA, USA. He is a Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. His research interests include hyperspectral image (HSI) analysis, pattern recognition, and data compression.

Dr. Li received the 2015 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society (GRSS) for his service for the IEEE JSTARS. He is serving as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. He served as a Guest Editor for special issue of the *Journal of Real-Time Image Processing*, *Remote Sensing*, and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS).

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

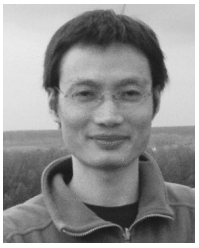
From 2013 to 2014, he was a Visiting Scholar with the University of Colorado, Boulder, CO, USA. He is a Professor with the Sichuan Provincial Key Laboratory of Information Coding and Transmission, Southwest Jiaotong University. His research interests include the statistical analysis of synthetic aperture radar images, remote sensing image processing, and signal processing in communications.

Dr. Li was a recipient of several scholarships or awards, especially including the Special Grade of the Financial Support from the China Postdoctoral Science Foundation in 2009 and the New Century Excellent Talents in University from the Ministry of Education of China in 2011. He serves as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS).

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

She is the Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA, and also an Adjunct Professor with the College of Surveying and Geo-informatics, Tongji University, Shanghai, China. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is a Fellow of the SPIC International Society for Optics and Photonics. She was a recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society (GRSS). She was the Chair of the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She served as the Co-Chair for the Data Fusion Technical Committee of the IEEE GRSS from 2009 to 2013. She was the General Chair for the fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing held at Shanghai, in 2012. She served as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS), the *Journal of Applied Remote Sensing*, and the IEEE SIGNAL PROCESSING LETTERS. Since 2016, she has been the Editor-in-Chief of the IEEE JSTARS.



Wenzhi Liao (Senior Member, IEEE) received the B.Sc. degree in mathematics from Hainan Normal University, Haikou, China, in 2006, the Ph.D. degree in engineering from the South China University of Technology, Guangzhou, China, in 2012, and the Ph.D. degree in computer science engineering from Ghent University, Ghent, Belgium, in 2012.

Since 2012, he has been working as a Post-Doctoral Research Fellow first with Ghent University and then with the Research Foundation Flanders (FWO), Vlaanderen, Belgium. Since 2020, he has been with the Sustainable Materials Management, Flemish Institute for Technological Research (VITO), Mol, Belgium. His research interests include pattern recognition, remote sensing, and image processing. In particular, his interests include mathematical morphology, multitask feature learning, multisensor data fusion, and hyperspectral image (HSI) restoration.

Dr. Liao was a recipient of the Best Paper Challenge Awards in both the 2013 IEEE GRSS Data Fusion Contest and the 2014 IEEE GRSS Data Fusion Contest. He serves as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS) and the *IET Image Processing*.

Wilfried Philips (Senior Member, IEEE) was born in Aalst, Belgium, in 1966. He received the Diploma degree in electrical engineering and the Ph.D. degree in applied sciences from Ghent University, Ghent, Belgium, in 1989 and 1993, respectively.

He is a Senior Full Professor with the Department of Telecommunications and Information Processing, Ghent University, where he heads the Image Processing and Interpretation Research Group. He also leads the activities in image processing and sensor fusion with the research institute IMEC, Leuven, Belgium. His main research interests include image and video quality improvement and estimation, real-time computer vision, and sensor data processing.

Dr. Philips is also a Co-Founder of the Senso2Me company, which provides the Internet of Things solutions for elderly care.