MULTISOURCE REMOTE SENSING DATA CLASSIFICATION USING DEEP HIERARCHICAL RANDOM WALK NETWORKS

Xudong Zhao[1,2], *Ran Tao*[1,2,*] *and Wei Li*[3]

[1]School of Information and Electronics, Beijing Institute of Technology, Beijing, China
[2]Beijing Key Laboratory of Fractional Signals and Systems, Beijing Institute of Technology, Beijing, China
[3]College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China
[*]Email: rantao@bit.edu.cn

ABSTRACT

Collaborative classification of hyperspectral imagery (HSI) and light detection and ranging (LiDAR) data is investigated using effective hierarchical random walk networks, denoted as HRWN. The proposed HRWN jointly optimizes dual-tunnel CNN, pixelwise affinity and seeds map via a novel random walk layer, which enforces spatial consistency in the deepest layers of the network. In designed random walk layer, the predicted distribution of dual-tunnel CNN serves as global prior while pixelwise affinity reflects local similarity of pixel pairs, which preserves boundary localization and spatial consistency well. Experimental results validated with two real multisource remote sensing data demonstrate that the proposed HRWN can significantly outperform other state-of-art methods.

Index Terms— Hierarchical random walk, convolutional neural network (CNN), hyperspectral image (HSI), multisource remote sensing classification.

1. INTRODUCTION

Remotely-sensed hyperspectral images (HSI), with wealthy spectral information to uniquely discriminate various materials of interest, has been widely applied to the land-cover observation [1, 2]. Furthermore, as a supplementary, Light detection and ranging (LiDAR) data, which provide elevation information about the surveyed area, is valuable for better characterizating the same scene acquired solely by optical sensors, such as HSI [3]. Collaborative classification of HSI and LiDAR can help to integrate diverse information to further improve earth observation performance.

Numerous researches have stated that classification tasks can be modified by integrating HSI and LiDAR features [4, 5]. Deep learning-based methods have aroused wide attention for their capabilities of automatical extracting robust and high-level features, which are known to be generally invariant to most local changes of the input [6, 7]. For the sake of extracting high-level features in HSI, deep convolutional neural network (CNN) was exploited to extract useful information for HSI classification [6–9].

However, large receptive fields in the convolutional layers and the presence of pooling layers lead to low spatial resolution in the deepest CNN layers. Thus, the corresponding predicted classes tend to be spatially disjoint and lack fine object boundary details. These approaches often fail to accurately capture relationships between classes and lead to spatially fragmented classification. When employing CNN into HSI or LiDAR classification, this challenge always exists.

In this paper, we attempt to focus on the weak boundary and spatially fragmented classification issue. Inspired by random walk methods [10–12], a simple yet effective hierarchical random walk network (HRWN) is proposed. The proposed HRWN jointly optimizes dual-tunnel CNN, pixelwise affinity and initial seeds map via a novel random walk layer, which enforces spatial consistency in the deepest layers of the network. In the random walk layer, the predicted distribution of dual-tunnel CNN serves as global prior while pixelwise affinity reflects local similarity of pixel pairs. Finally, the classification map is obtained by computing the probability distribution of each unassigned pixel.

The main contributions are summarized as follows: (1) A novel dual-tunnel CNN predicting merged HSI and LiDAR date is applied to obtain prior distribution among pixels. (2) Pixel similarity of LiDAR image reflected by affinity matrix is considered for spatial consistency and boundary details. (3) A hierarchical random walk layer introduces spatial constraint and local seeds guidance into the deepest layer of CNN, which alleviates the problems of weak localization around the boundaries and spatially disjoint of classification map. Experimental results validated with real remote sensing datasets demonstrate the effectiveness of the proposed method.

This work was supported by the National Natural Science Foundation of China under Grants No.61331021, 61421001, 91638201 and U1833203 (corresponding author: Ran Tao).

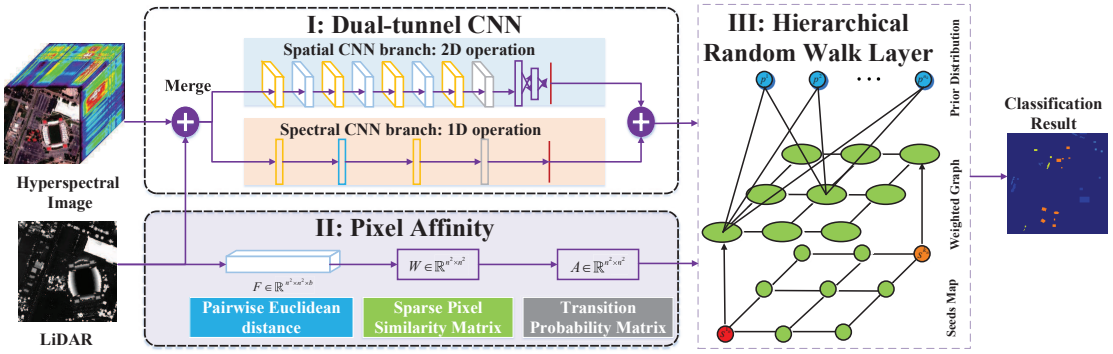


Fig. 1. The proposed feature extraction and classification framework of HRWN. Ellipse nodes denote the original nodes and circle nodes are auxiliary nodes. Green nodes are unseeded nodes, blue nodes are prior distribution and other color nodes are seeds. Only part of the transition edges are shown for simplification.

2. PROPOSED CLASSIFICATION FRAMEWORK

The framework of the proposed HRWN is illustrated in Fig. 1, which consists of three branches: (1) a dual-tunnel CNN branch that predicts classification potentials, (2) a pixel-level affinities branch predicting local similarities, and (3) initial seeds map indicating prior knowledge of training sets. A novel random walk layer merges the three branches to optimize classification jointly. Each of the components of the HRWN architecture will be described separately.

2.1. Dual-Tunnel CNN Branch

For dual-tunnel CNN branch as Part I shown in Fig. 1, HSI and LiDAR images are firstly merged by Gram-Schmidt pan-sharpening [13]. Then, as indicated in Fig. 2, a dual-tunnel CNN is designed for the merged image 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 radius r. The merged data patch $H_{ij}^{spa} \in \mathbb{R}^{ksize \times ksize}(ksize = 2 \times r + 1)$ is fed into the 2-D CNN tunnel, which includes simple operations such as 2-D convolution, activation, max-pooling and batch normalization layers. After these layers, the output spatial features are flattened. For the spectral tunnel, it concentrates on the center pixel H_{ij}^{spe} and consists of simple 1-D operations including convolution, activation, max-pooling and batch nomalization. The output spectral feature F_{ij}^{spe} is also flattened after max-pooling layer. Then the spatial and spectral features are concatenated and fed into the softmax classify layer to predict the probability distribution

$$p(i,k) = \frac{\exp(\theta_k | F_M)}{\sum_{k=1}^{C} \exp(\theta_k | F_M)}, \quad (1)$$

where C is the number of classes, i is the index of pixel, θ_k is the kth column of the weights in the prediction layer. The joint spatial-spectral feature F_M is expressed as $F_M = f(W \cdot (F_{ij}^{spe} \parallel F_{ij}^{spa}) + b)$, where W and b are the weight and bias of the full connection layer.

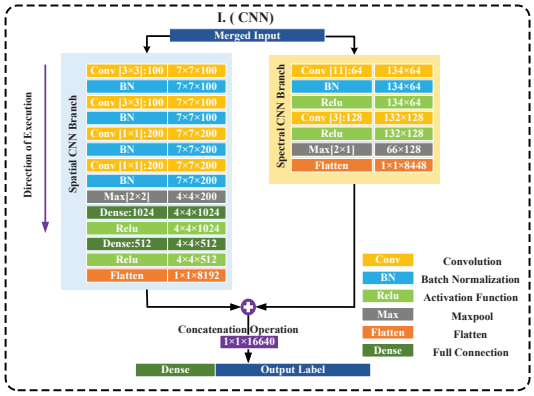


Fig. 2. Overall parameter configuration of the designed dual-tunnel CNN branch.

2.2. Pixel-level Affinity Branch

The dual-tunnel CNN can classify image using both spatial and spectral information, but pixel similarity of original image are not used. To learn the pixelwise affinities, a pixel-level affinity branch is employed as Part II in Fig. 1. This branch is connected with the input LiDAR image for its plentiful spatial information. The image is further considered as a weighted, undirected and connected graph $G = \{V, E, W\}$, where V is a finite set of vertices with $|V| = N$, E is a set of edges, and W is a weighted adjacency matrix indicating pixel similarity (see [14] for all definitions in this section). F is a sparse $n^2 \times n^2$ matrix storing Euclidean distance between each neighbor pixel pairs.

The non-normalized graph Laplacian is defined as $L := D - W$, where D is the diagonal degree matrix with the n-th element $d_n = \sum_{m \neq n} W_{mn}$ being the degree of vertex n. A weight $w_{ij} \in W$ 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 $w_{ij} = \exp(-\frac{||I_i - I_j||^2}{\sigma}) + \varepsilon$, where I_i and I_j are the pixel intensities at two nodes v_i and v_j, σ is the controlling parameter and ε is a small constant. Finally, the normalization matrix is applied to predict the ground truth pixel affinities A.

2.3. Hierarchical Random Walk Classification

Except of global distribution and pixel affinity, the training pixels also play a significant role in classification task. In the designed model, the training data are considered as initial seed map indicating the label of pixels in training set. These seeds provide local guidance for classification.

Then a hierarchical random walk layer is developed as Part III shown in Fig. 1, which incudes: (1) prior distribution $p(i,k) \in P$ obtained by dual-tunnel CNN as global guide, (2) affinity matrix w_{ij} reflecting pixel similarity for adjoint, and (3) seeds map $s(i,k) \in S$ for local guide to compute the probability of each not assigned pixel Δ. In the RW setting, the user marks some pixels in the image, then assumes that the random walker starts with each unlabeled pixel and calculates the probability that the random walker first arrives at the already tagged pixels. At each pixel, the tag with the maximum probability is selected as the final tag to obtain the final segmentation results. In the designed hierarchical random walk, the transition probability A on $V \cup \Delta \cup S \cup P$ is formulated as

$$A(i,j) = \begin{cases} c_i, & \text{if } i \in V, j \in \{\Delta\} \cup S \\ (1-c_i)\frac{\lambda p(i,k)}{d_i + \lambda g_i}, & \text{if } i \in V, j \in P \\ (1-c_i)\frac{w_{ij}}{d_i + \lambda g_i}, & \text{if } j \sim i \in V \\ 1, & \text{if } i = j \in \{\Delta\} \cup S \cup P \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

where $g_i = \sum_{k=1}^{K} p(i,k)$, λ is the weight parameter of prior distribution, c_i is the weight parameter of the ith seed.

Given the transition probability A on a graph with prior, the reaching probability r_i^k that a random walker from a node $v_i \in V$ reaching seeds $s(i,k)$ or prior node $p(i,k)$ is formulated as

$$r_i^k = \sum_{j \sim i \in V} (1-c_i)\frac{w_{ij} r_j^k}{d_i + \lambda g_i} + (1-c_i)\frac{p(i,k)}{d_i + \lambda g_i} + c_i s(i,k). \tag{3}$$

Then the classification map can be obtained as $R_i = \arg\max_k r_i^k$, where $k = 1, 2, \ldots, K$ is the label of pixel.

3. EXPERIMENTS AND ANALYSIS

In this section, two widely used remote sensing data sets are used to validate the effectiveness of the proposed method. The proposed HRWN is implemented using Python and Matlab language, and the networks are constructed using Tensorflow with the high-level API Keras.

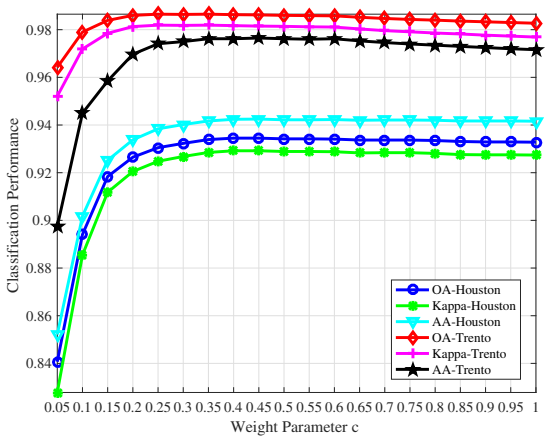


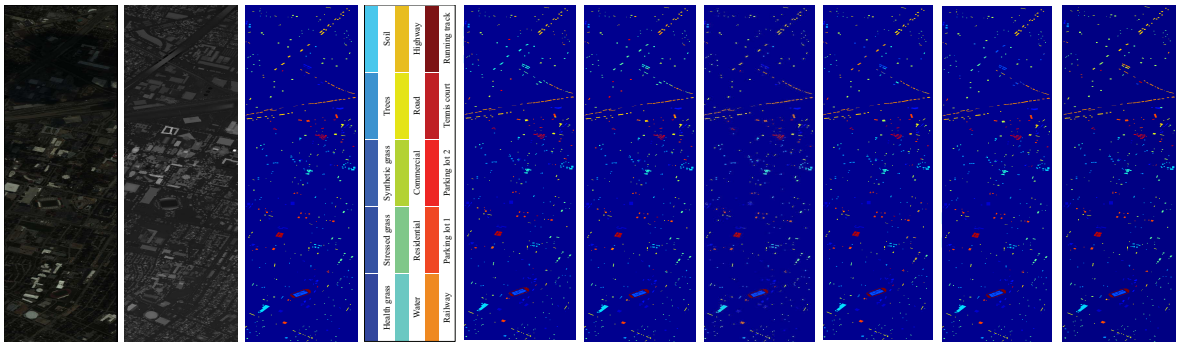
Fig. 3. The classification performance of HRWN with different weight parameter c.

Table 1. Comparison of the Classification Accuracy (%) Using the Houston Data.

No.	Class(Train/Test)	Performance					
		SVM	ELM	CNN-PPF	TB-CNN	C-CNN	HRWN
1	Health grass (198/1053)	82.43	83.10	83.57	83.10	84.89	85.77
2	Stressed grass (190/1064)	82.05	83.70	98.21	84.10	87.40	80.64
3	Synthetic grass (192/505)	99.80	100.00	98.42	100.00	99.86	99.14
4	Tress (188/1056)	92.80	91.86	97.73	93.09	93.49	92.52
5	Soil (186/1056)	98.48	98.86	96.50	100.00	100.00	100.00
6	Water (182/143)	95.10	95.10	97.20	99.30	98.77	98.15
7	Residential (196/1072)	75.47	80.04	85.82	92.82	82.81	95.82
8	Commercial (191/1036)	46.91	68.47	56.51	82.34	78.78	97.51
9	Road (193/1059)	77.53	84.80	71.20	84.70	82.51	87.62
10	Highway (191/1036)	60.04	49.13	57.12	65.44	59.41	85.74
11	Railway (181/1054)	81.02	80.27	80.55	88.24	83.24	98.95
12	Parking lot 1 (192/1041)	85.49	79.06	62.82	89.53	92.13	97.89
13	Parking lot 2 (184/285)	75.09	71.58	63.86	92.28	94.88	91.04
14	Tennis court (181/247)	100.00	99.60	100.00	96.76	99.77	100.00
15	Running track (187/473)	98.31	98.52	98.10	99.79	98.79	100.00
	OA	80.49	81.92	83.33	87.98	86.90	93.45
	AA	83.37	84.27	83.21	90.11	89.11	94.25
	Kappa	0.7898	0.8045	0.8188	0.8698	0.8589	0.9292

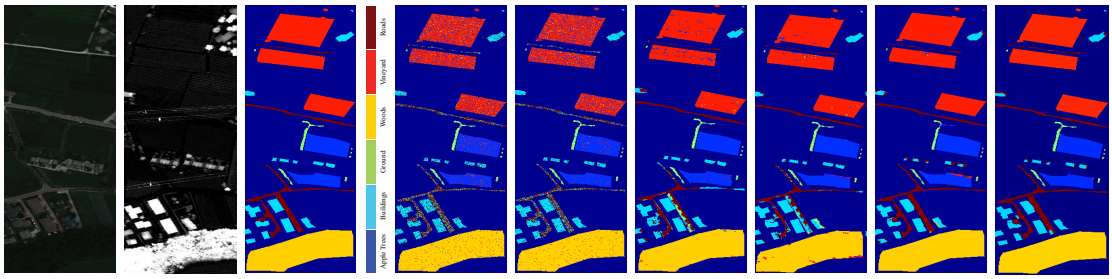
Houston Data: The scene was acquired over the area of University of Houston campus and neighbor area [15] in June 2012. The data consist of 349×1905 pixels covering 15 classes with a spatial resolution of 2.5m. Available training and testing samples are listed in Table 1. *Trento Data:* The second scene was acquired over a rural area in the south of the city of Trento, Italy. The data consist of 600×166 pixels covering 6 classes with a spatial resolution of 1m. Table 2 lists the available training and testing samples.

To validate the effectiveness, the proposed HRWN is compared with several classifiers, such as the standard SVM and ELM, recently proposed CNN-PPF [6], two-branch CNN (TB-CNN) [8] and the context CNN (C-CNN) [7]. Tables 1–2 list the overall accuracy (OA), averaged accuracy (AA), and Kappa Coefficient for two experimental datasets. The HSI and LiDAR data are concatenated together for classification. The proposed HRWN performs obviously better



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Fig. 4. Classification maps for the Houston data obtained with different methods: (a) Pseudo-color Image for HSI, (b) Gray image for LiDAR, (c) Ground truth Map, (d)Legend, (e) SVM (80.49%), (f) ELM (81.92%), (g) CNN-PPF (83.33%), (h) TB-CNN (87.98%), (i)C-CNN (86.90%) and (j) HRWN (93.45%).



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Fig. 5. Classification maps for the Trento data obtained with different methods including: (a) Pseudo-color Image for HSI, (b) Gray image for LiDAR, (c) Ground truth Map, (d)Legend, (e) SVM (92.77%), (f) ELM (91.32%), (g) CNN-PPF (94.76%), (h) C-CNN (96.11%), (i) TB-CNN (97.92%) and (j) HRWN (98.62%).

Table 2. Comparison of the Classification Accuracy (%) Using the Trento Data.

No.	Class(Train/Test)	Performance					
		SVM	ELM	CNN-PPF	TB-CNN	C-CNN	HRWN
1	Apple trees (129/3905)	88.62	95.81	90.11	98.07	99.26	99.78
2	Buildings (125/2778)	94.04	96.97	83.34	95.21	86.81	90.35
3	Ground (105/374)	93.53	96.66	71.13	93.32	97.91	99.79
4	Woods (154/8969)	98.90	99.39	99.04	99.93	97.31	100.00
5	Vineyard (184/10317)	88.96	82.24	99.37	98.78	99.82	100.00
6	Roads (122/3525)	91.75	86.52	89.73	89.98	84.63	95.97
	OA	92.77	91.32	94.76	97.92	96.11	98.62
	AA	92.63	92.93	88.97	96.19	94.29	97.65
	Kappa	0.9585	0.9042	0.9304	0.9681	0.9481	0.9815

than other methods. It is clear that due to more robust global and local features extraction, the hierarchical model for HSI, LiDAR and merged data can provide significant improvement in classification accuracy. For qualitative evaluation of the classification performance, visual maps are illustrated in Figs. 4-5. Also, the ground truth map and pseudo-color maps of entire image scenes are provided for clarity. The proposed method produces the most accurate and noiseless classification maps. Also, it can be concluded that the visual results are consistent with those in Tables 1–2. In order to validate the effects of weight between different layers, classification results using HRWN with different weight between prior distribution, pixel affinity and seeds map are shown in Fig. 3. An unique weight $c = 0.45$ can lead to best classification performance for both datasets, which indicates that both prior distribution and seed map contribute significantly.

4. CONCLUSION

In this paper, a hierarchical random walk network (HRWN) was proposed for the classification fusion of hyperspectral imagery (HSI) and light detection and ranging (LiDAR) data, which effectively mitigated the issue of boundary localization and spatial consistency. The proposed HRWN jointly optimized dual-tunnel CNN, pixelwise affinity and seeds map via a novel random walk layer, which enforced spatial consistency in the deepest layers of the network. Experimental results validated with real remote sensing datasets demonstrated that the proposed method can significantly outperform other state-of-art methods.

5. REFERENCES

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

[2] Xiangtao Zheng, Yuan Yuan, and Xiaoqiang Lu, "Dimensionality reduction by spatial-spectral preservation in selected bands," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 5185–5197, 2017.

[3] Jinha Jung, Edoardo Pasolli, Saurabh Prasad, James C. Tilton, and Melba M. Crawford, "A framework for land cover classification using discrete return LiDAR data: Adopting pseudo-waveform and hierarchical segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 2, pp. 491–502, Feb. 2014.

[4] Pedram Ghamisi, Bernhard Hofle, and Xiaoxiang Zhu, "Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 6, pp. 3011–3024, June 2017.

[5] Behnood Rasti, Pedram Ghamisi, and Richard Gloaguen, "Hyperspectral and LiDAR fusion using extinction profiles and total variation component analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3997–4007, 2017.

[6] Wei Li, Guodong Wu, Fan Zhang, and Qian Du, "Hyperspectral image classification using deep pixel-pair features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 844–853, 2017.

[7] H Lee and H Kwon, "Going deeper with contextual CNN for hyperspectral image classification.," *IEEE Transactions on Image Process*, vol. 26, no. 10, pp. 4843–4855, 2017.

[8] Xiaodong Xu, Wei Li, Qiong Ran, Qian Du, Lianru Gao, and Bing Zhang, "Multisource remote sensing data classification based on convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp. 1–13, 2018.

[9] 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 Transactions on Cybernetics*, pp. 1–12, 2018.

[10] Leo Grady, "Random walks for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768–1783, 2006.

[11] Leo Grady, "Multilabel random walker image segmentation using prior models," in *Computer Vision and Pattern Recognition*. IEEE, 2005, vol. 1, pp. 763–770.

[12] Xingping Dong, Jianbing Shen, Ling Shao, and Luc Van Gool, "Sub-markov random walk for image segmentation," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 516–527, 2016.

[13] Craig A Laben and Bernard V Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," 2000.

[14] R. K. Chung Fan, *Spectral Graph theory*, Published for the Conference Board of the mathematical sciences by the American Mathematical Society, 1997.

[15] Mahdi Khodadadzadeh, Jun Li, Saurabh Prasad, and Antonio Plaza, "Fusion of hyperspectral and LiDAR remote sensing data using multiple feature learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2971–2983, 2015.