



Collaborative Classification for Woodland Data Using Similar Multi-concentrated Network

Yixuan Zhu¹, Mengmeng Zhang², Wei Li^{2(✉)}, Ran Tao², and Qiong Ran¹

¹ Beijing University of Chemical Technology, Beijing 100029, China

² Beijing Institute of Technology, Beijing 100081, China

liwei089@ieee.org

Abstract. With the increasing of the forest area and complexity of tree species, collaborative classification using multi-source remote sensing data has been drawn increasing attention. Fusion of hyperspectral and LiDAR data can improve to acquire a comprehensive information which is conducive to the forest land classification. In this work, a similar multi-concentrate network focusing on the fine classification of tree species, denoted as SMCN, is proposed for woodland data. More specific, a preprocessing stage named pixel screening for data intensity critical control is firstly designed. Then, a similar multi-concentrate network is developed to capture spectral and spatial features from hyperspectral and LiDAR data and make specific connections, respectively. Experimental results validated on Belgian data have favorably demonstrated that the proposed SMCN outperforms other state-of-the-art methods.

Keywords: Multi-source remote sensing data · Collaborative classification · Convolutional neural network · Woodland classification

1 Introduction

With the development of geospatial science and sensor technology, classification technologies of remote sensing image faced to forest land information have made great progress [13]. Collaborative classification of hyperspectral image (HSI) and light detection and ranging (LiDAR) data takes advantages of the complementary information from multi-source data [4]. For example, hyperspectral image provides abundant biophysical and chemical canopy properties information which is convenient to discriminate various materials of interest target [4, 10]. And LiDAR data provides elevation information which can be acquired free from the limit of time and weather conditions, it is more suited to assess the horizontal and vertical canopy structure of forest area [4].

Many studies conclude that combining multi-sensor data could achieve better classification accuracy than using either data set individually. Collaborative

This work was supported by the National Natural Science Foundation of China under Grants NSFC-91638201, 61922013.

classification is beneficial to synthesize diverse forest information to more accurate forest data classification performance [8,12]. Liao et al. proposed a new deep fusion framework to integrate the complementary information from multi-sensor data [6]. Recently-proposed dense Convolutional Network [1] and UNET network [7] demonstrated that they can be used as an effective method for tree species classification. However, these deep learning architectures might not perform better for tree species mapping in complex and closed forest canopies.

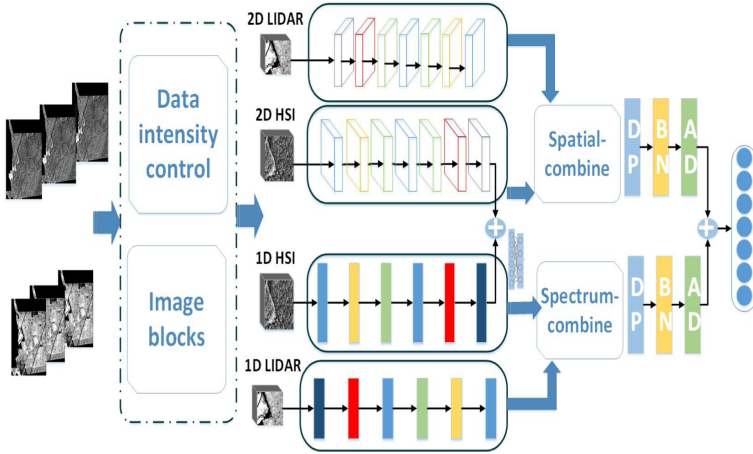


Fig. 1. The proposed collaborative classification framework for forest area.

Based on difficult characteristics of complex tree species, a preprocessing method is proposed for data intensity control which reduces the impact of excessive pixel differences on network training. A similar multi-concentrated network, denoted as SMCN, is further designed for focusing on reducing the mutual interference between spectral and spatial signatures which can effectively combine the respective feature. The similar and a little different structure guarantees the consistency of the features. At the same time, the specific information supplement mode for the spectral features and spatial features makes the network more flexible. A real remote sensing scene has been employed to validate the effectiveness of the proposed SMCN.

2 Proposed SMCN Classification

The proposed SMCN framework is designed to comprehensively learn and reasonably distinguish the difference of multi-sources data in spectral and spatial features. The overall structure is illustrated in Fig. 1.

Firstly, a screening process for original data is designed to ensure the critical control of data intensity. When the network is trained, if pixel range of some

channels is much larger than other channels, it may affect the network only extract features of a large pixel range and lose useful pixel information of small channels. Through comparing the pixel range of the popular remote sensing data sets, a 10-fold difference between the spectral pixel values of hyperspectral image in Belgium data may affect classification. After origin data normalized band-by-band, it has improved visually (as shown in Fig. 2). Because it is a separate normalization operation for each band, difference in the spectrum is also retained while reduces the effect of excessive pixel range at the same time.

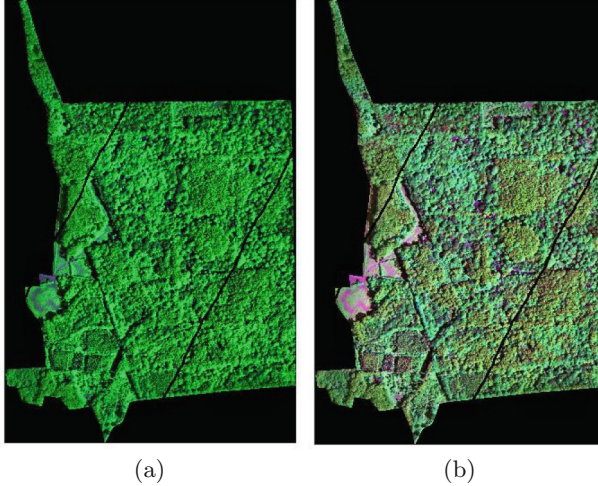


Fig. 2. Example of spectral pixel inspection: (a) the original image, and (b) the image after normalized.

Most previous work only pay attention to the study of spatial information, while proposed network considers the multi-branch to learn different features. Proposed SMCN divided the same location of data into 2D and 1D image block, respectively. The former focuses on spatial features and the latter is concentrating on spectral features. A one-dimensional processing channel for spectral features, including two 1-D convolution layers, batch normalization [3], two activation layers, a max-pooling layer, and the flatten layer. It focuses on the center pixel p_c , through batch normalization to set a high learning rate for accelerating convergence in each training mini-batch. The leaky rectified linear unit (ReLU) [9] is used as activation and the convolutional and max-pooling layer are adopted to solve features simultaneously. To facilitate subsequent processing, the output spectral features $F_{p(ij)}^{spec}$ is solved by flatten layer.

To ensure that the spectral and spatial characteristics of data can be well combined, the structure of spatial branch is as similar as one-dimensional processing channel. It only changed the links and parameter of network. The input data is image block with radius r around the center pixel p_c . After the flatten

layer, the spatial and spectral features were concatenated into the full-connection layer. The output can be further expressed as,

$$\mathbf{L}_{out} = f \left(\mathbf{W} \cdot \left(\mathbf{F}_{p(ij)}^{spc} \parallel \mathbf{F}_{p(ij)}^{spa} \right) + \mathbf{b} \right) \quad (1)$$

where \mathbf{W} and \mathbf{b} are the weights and bias of the full connection, \parallel denotes the simple superposition method of concatenating the spatial and spectral feature vectors.

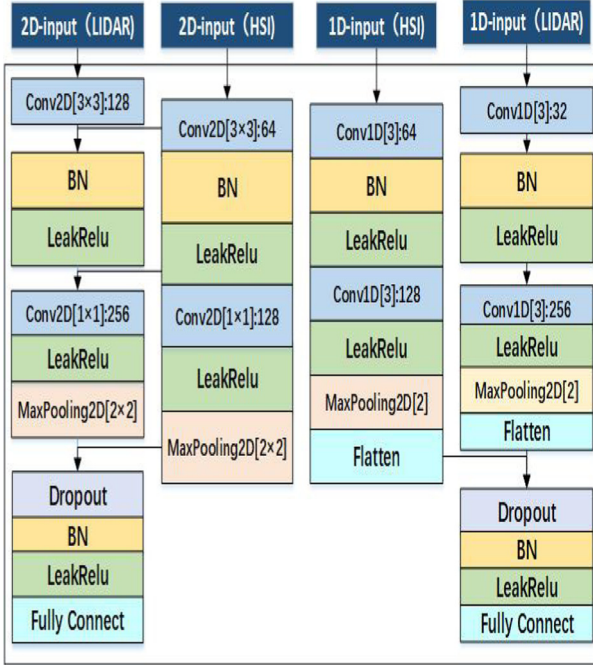


Fig. 3. The parameter of deep-mining module.

During training, use HSI image to train the two branch CNN at beginning. After fixing the weight of trained branch, introduce LiDAR data to fine-tune the network. The extraction and analysis of spectral characteristics is focusing on the central pixel of image block, which is independent of each other and have no corresponding domain information. Therefore, only a simple superposition is used. In the branch of spatial features, it not only focuses on the center pixel, but also considers the spatial features $\mathbf{F}_{p(ij)}^{spa}$ from the surrounding domain of center target. Therefore, LiDAR features are passed to HSI branch in stages continuously to correct the learning of forest information. Finally, perform superposition between different source feature map and passing the fusion features to subsequent layers. The final layer usually has the nodes of classification category, it is denoted as \mathbf{P}_{ij}^n which is a discrete probability distribution values for each category,

Table 1. The classification performance (%) of different window sizes.

Belgium data	3×3	5×5	7×7
OA	86.28	87.68	87.37
AA	84.28	84.82	84.95
Kappa	82.42	84.17	83.79

$$P_{ij}^n = \frac{\exp(\theta_n | \mathbf{L}_{out})}{\sum_{n=1}^N \exp(\theta_n | \mathbf{L}_{out})}. \quad (2)$$

Because the initial random weights are far away from the optimal value, training the specific dual-concentrate network of HSI with a large learning rate in the first stage. When the training of HSI network is completed and the weights of the HSI branches are fixed, LiDAR features are transmitted phase by phase and fine-tuned the network with a small learning rate. The learning rate of the dual-concentrate structure of HSI is set to 0.01, and the network does fine-tune at the learning rate of 0.0001 during the adding of LiDAR data, optimizer is Adam. Figure 3 shows the parameter information of the proposed network in details.

3 Experimental Results and Analysis

TensorFlow is an open source library that can employ Keras as an application interface for machine intelligence. Based on the personal computer equipped with Ubuntu 14.04 and Nvidia GTX 1080, Tensorflow1.3.0 and Keras2.1.2 construct the integral network. Most programs are implemented using Python language, some simple processing use MATLAB language.

Belgium data is used to validate the performance of the proposed network. It represents a forest area reserved at the western part of Belgium. A total of 1450 trees were labeled for the seven species. Tree distribution in the upper canopy was common beech (27.6%), copper beech (5.5%), pedunculate oak (20.6%), common ash (4.6%), larch (8.2%), poplar (28.6%) and sweet chestnut (4.6%). Around 20% samples are used for training, the remaining samples are used for testing. We only use a multi-band image of 11 PH bands (i.e., full-waveform LiDAR data) and 286 band hyperspectral data. It covering the visible and short wave infrared wavelength (372-2498nm). The specific category information can be acquired in Table 2.

Use overall accuracy (OA), average accuracy (AA), and Kappa coefficients as evaluation indicator. Table 1 lists the classification performance of the patch with different sizes. It demonstrates that the size of image block has impact on the classification performance of different data sets, the best size of Belgium data is 5×5.

Table 2. Comparison of the classification accuracy (OA%) among the proposed SMCN.

No.	Class(training/testing)	Classification performance				
		SVM	ELM	Two branch CNN	Contex CNN	Proposed SMCN
1	Beech(88/321)	79.13	75.39	76.64	63.55	84.74
2	Ash(13/54)	14.81	55.56	64.81	38.89	77.78
3	Larch(23/93)	79.57	82.80	91.4	73.12	86.02
4	Poplar(83/333)	97	93.39	97	92.19	96.7
5	Copper beech (16/64)	100	96.88	93.75	93.75	100
6	Chestnut(13/54)	33.33	51.85	48.15	22.22	64.81
7	Oak (60/243)	75.62	75.62	79.75	75.21	85.12
OA(%)		79.59	80.36	83.38	73.56	87.94
AA(%)		68.49	75.93	78.78	65.56	85.02
Kappa		73.45	74.72	78.70	65.84	84.51
Training time (in <i>Seconds</i>)			943.92		140.32	227.68

To demonstrate the performance of the proposed SMCN framework for multi-source remote sensing data classification, some traditional and state-of-the-art methods are compared, such as SVM, ELM [2], Two-Branch CNN [11], Contex CNN [5], paper [6]. Experimental results listed in Table 2 prove that the proposed SMCN performs better than aforementioned methods, all kinds of classification results are excellent.

The distribution of training and testing samples for all the comparison methods is the same as [6], nearly 20% samples are used for training. At the same time, based on the different proportion between train and test samples, Table 3 indicates that the proposed network still has good classification performance on fine classification of tree species. As the number of training samples increases, the classification of the network becomes more accurate.

Table 3. The classification performance (OA%) on the different proportion between training and testing samples.

No.	Compared methods	OA with different proportions		
		1:9	2:8	3:7
1	SVM	75.86	79.59	84.14
2	ELM	78.62	80.36	86.34
3	Contex CNN	72.95	73.56	85.72
4	Two Branch CNN	81.72	83.38	89.31
5	Proposed SMCN	87.86	87.94	91.86

4 Conclusion

A collaborative classification method based on the proposed SMCN using HSI and LiDAR data has been studied for forest area. In the proposed method, each center pixel of the image block was combined with the spatial information of the image for deep analysis after learning the relevant information between the bands. For consensus between the different source information, the structure of each branch was similar and different. Compared with 3D convolution, the proposed SMCN has faster speed and better flexibility without taking up too much memory. Experimental results confirmed that the proposed SMCN was more effective.

References

1. Hartling, S., Sagan, V., Sidike, P., Maimaitijiang, M., Carron, J.: Urban tree species classification using a worldview-2/3 and lidar data fusion approach and deep learning. *Sensors* **19**(6), 1424–8220 (2019)
2. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: theory and applications. *Neurocomputing* **70**(1), 489–501 (2006)
3. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR* (2015)
4. Koetz, B., et al.: Fusion of imaging spectrometer and LIDAR data over combined radiative transfer models for forest canopy characterization. *Remot. Sens. Environ.* **106**(4), 449–459 (2007)
5. Lee, H., Kwon, H.: Going deeper with contextual CNN for hyperspectral image classification. *IEEE Trans. Image Process.* **26**(10), 4843–4855 (2017)
6. Liao, W., Coillie, F.V., Gao, L., Li, L., Chanussot, J.: Deep learning for fusion of APEX hyperspectral and full-waveform LiDAR remote sensing data for tree species mapping. *IEEE Access* **6**, 68716–68729 (2018)
7. Liu, J., Wang, X., Wang, T.: Classification of tree species and stock volume estimation in ground forest images using deep learning. *Sensors* **166**, 0168–1699 (2019)
8. Luo, S., et al.: Fusion of airborne LIDAR data and hyperspectral imagery for aboveground and belowground forest biomass estimation. *Ecol. Indicator.* **73**, 378–387 (2017)
9. Maas, A.L., Hannun, A.Y., Ng, A.Y.: Rectifier nonlinearities improve neural network acoustic models. In: 30th ICML, vol. 30, no. 1 (2013)
10. Tao, R., Zhao, X., Li, W., Li, H.C., Du, Q.: Hyperspectral anomaly detection by fractional Fourier entropy. *IEEE J. Sel. Top. Appl. Earth Observat. Remot. Sens.* **12**(12), 4920–4929 (2019)
11. Xu, X., Li, W., Ran, Q., Du, Q., Gao, L., Zhang, B.: Multisource remote sensing data classification based on convolutional neural network. *IEEE Trans. Geosci. Remot. Sens.* **56**(2), 937–949 (2018)
12. Zhao, X., et al.: Joint classification of hyperspectral and LiDAR data using hierarchical random walk and deep CNN architecture. *IEEE Trans. Geosci. Remot. Sens.* **58**, 7355–7370 (2020)
13. Yokoya, N., Grohnfeldt, C., Chanussot, J.: Hyperspectral and multispectral data fusion: a comparative review of the recent literature. *IEEE Geosci. Remot. Sens. Mag.* **5**(2), 29–56 (2017)