DISCRIMINATIVE FEATURE EXTRACTION AND FUSION FOR CLASSIFICATION OF HYPERSPECTRAL AND LIDAR DATA

Weiwei Song[1]*, Zhi Gao*[2]*, and Yongjun Zhang*[2]

[1] Department of Mathematics and Theories, Peng Cheng Laboratory, Shenzhen, China
[2] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China

IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-2792-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/IGARSS46834.2022.9883136

ABSTRACT

Multisource remote sensing data provide the abundant and complementary information for land cover classification. In this paper, we propose a deep hashing-based feature extraction and fusion framework for joint classification of hyperspectral and LiDAR data. Firstly, HSIs and LiDAR data are fed into a two-stream network to extract deep features after data preprocessing. Then, we adopt hashing technique to constrain single-source and cross-source similarities, i.e., samples with same classes should have small feature distance and samples with different classes should have large feature distance. Furthermore, a feature-level fusion strategy is exploited to fuse the two kind of multisource information. Finally, we design an object function to consider the similarity information between sample pairs and semantic information of each sample, which can deliver the discriminative features for classification. The experiments on Houston data demonstrate the effectiveness of the proposed method over some competitive approaches.

Index Terms— Hyperspectral images (HSIs), Light detection and ranging (LiDAR), classification, deep learning, hashing, feature extraction.

1. INTRODUCTION

The recent advances in sensor technology have made it possible to simultaneously acquire different source remote sensing data of same investigated region. For instance, hyperspectral images (HSIs) are usually composed of several hundreds of spectral bands spanning from the visible-to-infrared spectrum. With the rich spectral information, HSIs are widely used to discriminate the different materials [1–3]. However, due to lacking of the elevation information, it is difficult to distinguish the building roof and the road that made up of concrete using HSIs. On the other hand, light detection and ranging (LiDAR) data record the elevation information about the surveyed area, which can be acquired even under cloud-cover conditions. Though LiDAR data can accurately classify land covers with different heights, it cannot discriminate roofs with different materials (e.g., asphalt and steel). Thus, joint use of HSI and LiDAR data has been widely investigated in the remote sensing field to enhance the interpretation accuracy for the study area.

In order to make full use of the complementary information between HSI and LiDAR data, a large number of works have been investigated in the past several years. Among them, feature-level fusion and decision-level fusion are two widely-used techniques. For the former one, the features of HSIs and LiDAR are separately extracted, and then fused to improve the ability of representation of features. For example, the spatial and spectral features of HSI extracted from a dual-tunnel CNN were combined with the spatial information of LiDAR extracted from a cascade network [4]. Du *et al.* [5] proposed a graph fusion network to extract a fused feature from multisource data, where feature extraction and fusion are integrated into a single network. In addition, the extinction profiles of HSI and LiDAR data were fused with sparse and low-rank component analysis [6]. For decision-level fusion, there are also some successful application. For instance, a decision-level fusion method for HSI and LiDAR data classification was presented in [7], and liner and nonlinear features were fused in a decision phase [8]. In addition, Hang *et al.* [9] adopted a weighted summation strategy to perform the decision-level fusion for the joint classification of HSI and LiDAR data. Although the above feature-level and decision-level fusion approaches have shown satisfactory performance for classification task, the correlations between samples are not well excavated. Due to lacking of the constraints in similarities between samples, the previous methods may deliver an unsatisfactory classification result when dealing with the high intra-class and low inter-class variabilities for some complex classes.

To this end, we propose a novel deep hashing-based feature extraction and fusion method for the joint classification of HSI and LiDAR data. First of all, we build a two-stream CNN to extract the spectral-spatial feature of HSIs and spatial feature of LiDRA data, respectively. Then, the hashing technique is imposed on the above extracted features to constrain the similarities between single/cross-source samples. Furthermore, the features of HSI and LiDAR data are fused in a fully connected layer to improve the ability of representation. Finally, we elaborately design an object to simultaneously consider the similarity information of sample pairs and seman-

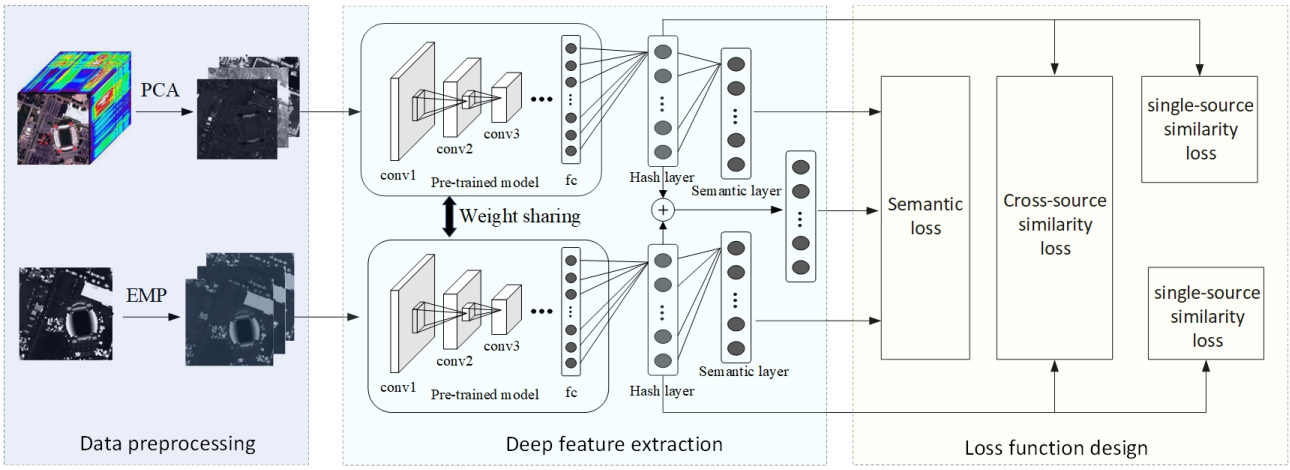

Fig. 1: The architecture of the HDRN. Residual block m, n, and k refers to low, middle, high layer, respectively.

tic information of each sample. The experimental results on Houston data demonstrate that the proposed method can obtain the better classification results.

The rest of this paper is organized as follows. In Section 2, the proposed method is described in detail. The experimental results are presented in Section 3. Section 4 makes some concluding remarks.

2. PROPOSED METHOD

In this paper, we propose a novel feature extraction and fusion framework for the joint classification of HSI and LiDAR data. Fig. 1 shows the whole framework, which can be mainly divided into three steps, i.e., data preprocessing, deep feature extraction, and loss function design. In the following parts, we will describe the procedures in detail.

2.1. Data preprocessing

Let $\mathbf{H} \in \mathbb{R}^{M\times N\times C}$ and $\mathbf{L} \in \mathbb{R}^{M\times N\times 1}$ be a HSI and the corresponding LiDAR data, respectively, where $M \times N$ is the spatial size of image and C is the number of channels of HSI. Considering that the original HSIs usually contain the abundant redundant information within server handers of spectral bands, we perform principal component analysis (PCA) algorithm on the original HSI to extract the most informative components. Here, we preserve three components to meet the need of input of the following pre-trained network. The dimensionality-reduced HSI can be represented by $\mathbf{H}' \in \mathbb{R}^{M\times N\times 3}$. On the other hand, we perform extended morphological profiles (EMP) algorithm [10] on the original LiDRA data to extract the spatial information of LiDAR via morphological operations. Specifically, the processed LiDAR data can be represented by $\mathbf{L}' \in \mathbb{R}^{M\times N\times 3}$.

2.2. Deep feature extraction

We adopt the classic AlexNet [11] as the pre-trained model, where the last classification layer is discarded. Here, we build a two-stream AlexNet model to extract deep spectral-spatial features of HSI and deep spatial features of LiDAR data.

For a pair of sample $(\mathbf{x}_i, \mathbf{x}_j)$, where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the image patches with size of $p \times p$. Here, $\mathbf{x}_i$ and $\mathbf{x}_j$ can be from same source or different sources, i.e., $\mathbf{x}_i, \mathbf{x}_j \in \{\mathbf{H}', \mathbf{L}'\}$. The label of image patch is determined by the label of centered pixel. Then, $(\mathbf{x}_i, \mathbf{x}_j)$ is input into the above two-stream network, which can be represented by the following formulation:

$$\mathbf{f}_t = \Phi(\mathbf{x}_t; \theta), t = i, j \tag{1}$$

where Φ is the network function characterized by the parameter θ existed in the first seven layers of AlexNet.

In addition, we build a hash layer and a semantic layer on the top of AlexNet to extract the hash-like codes and semantic features, respectively [12]. Furthermore, we fuse features of two hash layers to improve the ability of representation. The above procedures can be represented by:

$$\mathbf{u}_t = \sigma(\mathbf{W}_h^t \mathbf{f}_t + \mathbf{v}_h^t), t = i, j \tag{2}$$

$$\mathbf{u}_f = \sigma(\mathbf{W}_h(\mathbf{f}_i + \mathbf{f}_j) + \mathbf{v}_h) \tag{3}$$

$$\mathbf{z}_t = (\mathbf{W}_s^t \mathbf{u}_t + \mathbf{v}_s^t), t = i, j, f \tag{4}$$

where $\mathbf{W}_h$ and $\mathbf{W}_s$ are the weights of hash layer and semantic layer respectively, $\mathbf{v}_h$ and $\mathbf{v}_s$ are the corresponding biases.

2.3. Loss function design

As shown in Fig. 1, there are three kinds of losses are considered in our method. Specifically, for pairs of samples from

Table 1

Class No.	Class Name	Training /Test	CNN-PPF	FDSSCN	CRNN	Context CNN	Two-branch CNN	Coupled CNN	**Our method**
1	Healthy grass	198/1053	83.57	85.53	83.00	84.89	83.10	98.51	85.33
2	Stressed grass	190/1064	98.21	84.61	79.41	87.40	81.20	97.83	86.91
3	Synthetic grass	192/505	98.42	99.43	99.80	99.86	100	70.60	87.23
4	Tree	188/1056	97.73	94.13	90.15	93.49	92.90	99.06	96.99
5	Soil	186/1056	96.50	100	99.71	100	99.81	100	98.46
6	Water	182/143	97.20	98.15	83.21	98.77	100	41.11	91.96
7	Residential	196/172	85.82	92.11	88.06	82.81	92.54	83.14	96.90
8	Commercial	191/1053	56.51	65.03	88.61	78.78	94.87	98.39	95.37
9	Road	193/1059	71.20	66.21	66.01	82.51	83.85	94.81	89.64
10	Highway	191/1036	57.12	71.88	52.22	59.41	69.89	92.98	72.32
11	Railway	181/1054	80.55	84.53	81.97	83.24	86.15	90.88	97.27
12	Parking lot 1	192/1041	62.82	98.54	69.83	92.13	92.60	91.02	81.89
13	Parking lot 2	184/285	63.86	90.41	79.64	94.88	79.30	97.09	92.89
14	Tennis court	181/247	100	100	100	99.77	100	100	100
15	Running track	187/473	98.10	100	100	98.79	100	97.85	85.52
OA	-	-	83.33	86.60	88.55	86.90	88.91	90.43	90.65
AA	-	-	83.21	88.20	90.30	89.11	90.42	90.22	90.48
Kappa	-	-	81.88	85.59	87.56	85.89	87.96	89.68	89.95

single source, the similarity loss is formulated by taking the negative log-likelihood of the observed pairwise labels, i.e.,

$$\begin{aligned}\mathcal{L}_{sin} &= -\sum_{s_{ij}\in S_h} log\ p(s_{ij}|\mathbf{B}) - \sum_{s_{ij}\in S_l} log\ p(s_{ij}|\mathbf{B}) \\ &= -\sum_{s_{ij}\in S_h} (s_{ij}\omega_{ij} - log(1+e^{\omega_{ij}})) \\ &\quad -\sum_{s_{ij}\in S_l} (s_{ij}\omega_{ij} - log(1+e^{\omega_{ij}})).\end{aligned} \tag{5}$$

where s_{ij} is the label of pair of samples ($\mathbf{x}_i$, $\mathbf{x}_j$) such that $s_{ij} = 1$ if $\mathbf{x}_i$ and $\mathbf{x}_j$ share same label and 0 otherwise. $\mathbf{B}$ is the hash code matrix such that the element $\mathbf{b}_t = sign(\mathbf{u}_t)$. $\omega_{ij} = \frac{1}{2}\mathbf{b}_i^T\mathbf{b}_j$

For pairs of samples from different source, the similarity loss is also represented by:

$$\mathcal{L}_{cro} = -log\ p(S|\mathbf{B}) - \sum_{s_{ij}\in S_{hl}} (s_{ij}\omega_{ij} - log(1+e^{\omega_{ij}})). \tag{6}$$

For each sample, the sematic loss is represented by the following formulation

$$\begin{aligned}\mathcal{L}_{sem} = &-\frac{1}{K}\sum_{\mathbf{x}_i\in\mathbf{H}'} < \mathbf{y}_i, log\mathbf{t}_i > -\frac{1}{K}\sum_{\mathbf{x}_j\in\mathbf{L}'} < \mathbf{y}_j, log\mathbf{t}_j) > \\ &-\frac{1}{K}\sum_{\mathbf{x}_i\in\mathbf{H}',\mathbf{x}_j\in\mathbf{L}'} < \mathbf{y}_f, log\mathbf{t}_f >\end{aligned} \tag{7}$$

where $\mathbf{y}$ is the true label of the corresponding sample, $\mathbf{t} = softmax(\mathbf{z})$, K is the number of training samples. In addition, we also consider the quantitative loss, i.e.,

$$\mathcal{L}_{qua} = \sum_{i,j=1}^{K} \|\mathbf{u}_i - \mathbf{b}_i\|_2^2 + \|\mathbf{u}_j - \mathbf{b}_j\|_2^2 \tag{8}$$

Finally, the object function is defined by:

$$\mathcal{J} = \min\mathcal{L} = \min(\alpha\mathcal{L}_{sin} + \beta\mathcal{L}_{cro} + \gamma\mathcal{L}_{sem} + \lambda\mathcal{L}_{qua}) \tag{9}$$

where α, β, γ, and λ are the regularization parameters. The above function can be solved by stochastic gradient descent (SGD) algorithm.

3. EXPERIMENTAL RESULTS

3.1. Data Description

To verify the effectiveness of the proposed method, we conduct classification experiments on Houston data. This scene was acquired by an airborne sensor over the area of University of Houston and neighboring area in 2012. This scene contains a HSI and LiDAR data, both of which consist of 349×1905 pixels with a spatial resolution of 2.5 m. The HSI has 144 spectral bands across the spectral range from 0.38 to 1.05 μm. 15 classes of interest are considered for this dataset.

3.2. Experimental Results

In experiments, we adopt the standard sets of training and test samples which are provided by this dataset. The proposed method is compared with other competitive methods, including CNN-PPF [13], two-branch CNN [4], FDSSCN [14], context CNN [15], CRNN [16], and Coupled CNN [9]. For the proposed method, the size of image patch is set to 15×15, α, β, γ, and λ are set to 10, 0.5, 1, and 5, respectively. Other compared methods are implemented with optimal parameters.

Table 1 shows the classification results on this dataset. Note that the results of all compared methods are reported according to the reference [17]. From this table, we can see that our method is obviously superior to other methods. Specifically, its improvement is about 7.32%, 4.05%, 2.10%, 3.75%,

1.74%, and 0.22% over CNN-PPF, FDSSCN, CRNN, Context CNN, Two-branch CNN, and Coupled CNN, respectively. The above experimental results demonstrate that the proposed method can obtain a better classification accuracy than other comparison baselines due to more effective information utilization including single/cross-source similarity information and semantic information.

4. CONCLUSIONS

In this paper, we propose a novel feature extraction and fusion framework for the joint classification of HSI and LiDAR data. The designed object function takes full consideration of single/cross-source similarity and semantic information, which delivers a discriminative feature representation for subsequent classification. Experimental results demonstrate the proposed method outperforms other competitive methods.

5. REFERENCES

[1] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.

[2] W. Song, S. Li, L. Fang, and T. Lu, "Hyperspectral image classification with deep feature fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3173–3184, Jun. 2018.

[3] L. Fang, Z. Liu, and W. Song, "Deep hashing neural networks for hyperspectral image feature extraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1412–1416, Sep. 2019.

[4] 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. 2017.

[5] X. Du, X. Zheng, X. Lu, and A. A. Doudkin, "Multisource remote sensing data classification with graph fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10062–10072, Dec. 2021.

[6] B. Rasti, P. Ghamisi, J. Plaza, and A. Plaza, "Fusion of hyperspectral and lidar data using sparse and low-rank component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6354–6365, Nov. 2017.

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

[8] 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.

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

[10] J. A. Benediktsson, J. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles [j]," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[12] W. Song, S. Li, F. Zhang, and J. A. Benediktsson, "Deep hashing learning for visual and semantic retrieval of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9661–9672, Nov. 2021.

[13] 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. 2016.

[14] W. Wang, S. Dou, Z. Jiang, and L. Sun, "A fast dense spectral–spatial convolution network framework for hyperspectral images classification," *Remote Sens.*, vol. 10, no. 7, pp. 1068, 2018.

[15] 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.

[16] H. Wu and S. Prasad, "Convolutional recurrent neural networks forhyperspectral data classification," *Remote Sens.*, vol. 9, no. 3, pp. 298, 2017.

[17] M. Zhang, W. Li, R. Tao, H. Li, and Q. Du, "Information fusion for classification of hyperspectral and lidar data using ip-cnn," *IEEE Trans. Geosci. Remote Sens.*, DOI: 10.1109/TGRS.2021.3093334.