

# Deep Fusion of Remote Sensing Data for Accurate Classification

Yushi Chen, Chunyang Li, Pedram Ghamisi, Xiuping Jia, Yanfeng Gu

**Abstract**—The multisensory fusion of remote sensing data has obtained a great attention in recent years. In this letter, we propose a new feature fusion framework based on deep neural networks (DNNs). The proposed framework employs deep convolutional neural networks (CNNs) to effectively extract features of multi-/hyper- spectral and light detection and ranging (LiDAR) data. Then, a fully connected DNN is designed to fuse the heterogeneous features obtained by the previous CNNs. Through the aforementioned deep networks, one can extract the discriminant and invariant features of remote sensing data, which are useful for further processing. At last, logistic regression is used to produce the final classification results. Dropout and batch normalization strategies are adopted in the deep fusion framework to further improve classification accuracy. The obtained results reveal that the proposed deep fusion model provides competitive results in terms of classification accuracy. Furthermore, the proposed deep learning idea opens a new window for future remote sensing data fusion.

**Index Terms**—Convolutional neural network (CNN), data fusion, deep neural network (DNN), feature extraction (FE), multispectral image (MSI), hyperspectral image (HSI), light detection and ranging (LiDAR).

## I. INTRODUCTION

OWING to the exponential increase in the number of available remote sensors in recent years, one can simultaneously obtain multi-source data sets of the same region. Multispectral images (MSI) or hyperspectral images (HSI) are usually composed of a number of spectral channels of the same scene. The detailed spectral and spatial information provided by MSI/HSI increases the power of accurately differentiating materials of interest [1]. On the other hand, light detection and ranging (LiDAR) can reveal the elevation information of the surface with respect to the sensor, which is useful to distinguish objects constructed by similar materials. Therefore, the aforementioned sensors can provide detailed spectral-spatial-elevation information to obtain the robust description of a

This work was supported in part by State Key Laboratory of Frozen Soil Engineering under the Grant SKLFSE201614, by the Natural Science Foundation of China under the Grant 61301206, 61371180 and 60972144, and by the National Science Foundation for Excellent Young Scholars under the Grant 61522107. (Corresponding author: *Y. Gu and Y. Chen*)

Y. Chen, C. Li and Y. Gu are all with the School of Electronics and Information Engineering, Harbin Institute of Technology. (e-mails: cheniyushi@hit.edu.cn; lcy\_buzz@mail.dlut.edu.cn; guyf@hit.edu.cn).

P. Ghamisi is with German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF) and with Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM) (e-mail: p.ghamisi@gmail.com).

X. Jia is with the School of Engineering and Information Technology, The University of New South Wales, Canberra, A.C.T. 2600, Australia (e-mail: x.jia@adfa.edu.au).

scene.

To make full use of the MSI/HSI and LiDAR data, various methods and strategies have been proposed in the literature since the last decade [2].

In the early stage, the investigated methods were relatively simple. For instance, one way is to stack the LiDAR as one channel to the HSI [3], and the classification is applied on the stacked data. Because of the discriminant features of LiDAR, the classification performance is increased. However, a simple concatenation of original features may not be powerful enough to separate the classes of interest.

Support vector machines (SVMs) exhibit low sensitivity to high dimensionality, and therefore, SVM-based classifiers for HSI might lead to better classification accuracy than other widely used pattern recognition techniques [4]. Recently, kernel-based methods have attracted the attention of many researchers due to their capability for integrating heterogeneous information provided by multi-source remote sensors [4].

Traditional feature extraction methods usually have limited performance in invariant feature learning. While recently, deep learning has been proposed to extract the features in a hierarchical way [5], which provides a promising direction for deep feature-based fusion. Among deep learning approaches, convolutional neural networks (CNNs) are theoretically sound due to its powerful feature extraction capability inspired by neuroscience [6], and CNNs have successfully been applied to a wide variety of applications [7].

In this letter, we investigate the potential of a deep model for remote sensing data fusion. The proposed deep model uses CNNs to extract the spectral-spatial features of MSI/HSI as well as the spatial-elevation features of LiDAR data. Then, a deep neural network (DNN) is used to fuse the extracted features of the CNNs for classification.

The main contributions of the letter are listed as follows.

1) To the best of our knowledge, CNNs have not been used for LiDAR data feature extraction. Here, we carefully design a deep architecture to extract the contextual features of LiDAR data.

2) A new feature fusion method, which is based on deep neural networks, is proposed to fuse the heterogeneous features extracted by CNNs.

3) Dropout and batch normalization (BN) are used to boost the performance of the proposed method in terms of classification accuracy.

The rest of this letter is organized as follows: Section II elaborates on the proposed deep fusion framework and Section III presents the experimental results on several remote sensing data sets. Finally, Section IV summarizes the observations by

providing the main concluding remarks.

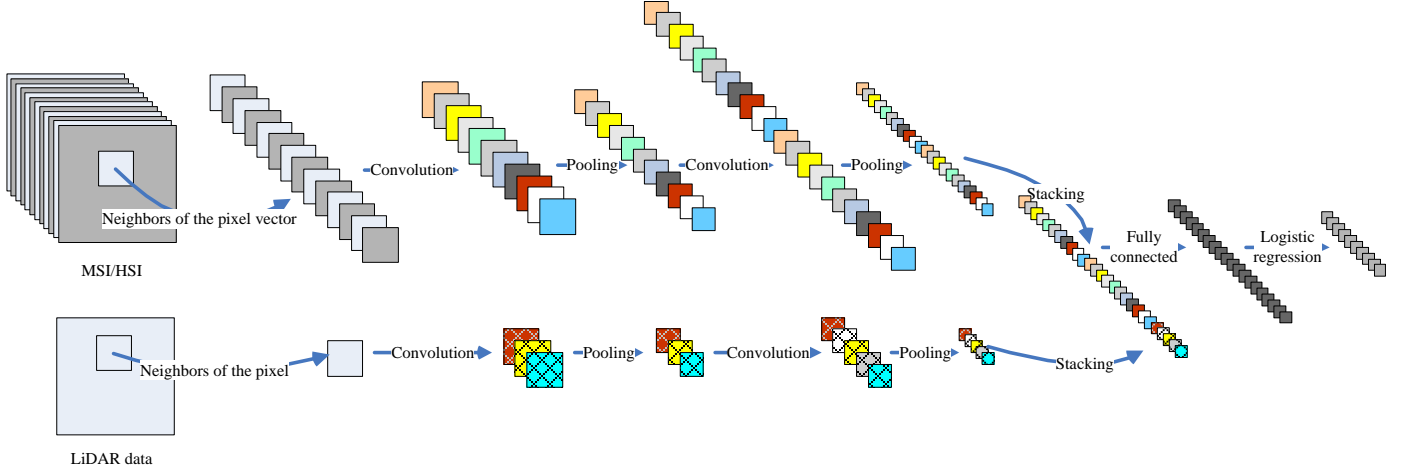


Fig. 1. The framework of deep fusion for MSI/HSI and LiDAR data for accurate classification. There are three deep architectures in the framework: two CNNs and a DNN. The CNNs are designed to extract spectral-spatial-elevation features of MSI/HSI and LiDAR data, and the DNN is designed to fuse the extracted features.

## II. DEEP FUSION OF REMOTE SENSING DATA

### A. The framework of the proposed method

The flowchart of the proposed deep fusion of MSI/HSI and LiDAR data for image classification is shown in Fig. 1. In the CNNs, pixels in a neighborhood region of a given pixel in the original data are taken into consideration. The CNNs are designed to extract the spectral-spatial features of MSI/HSI, and the spatial-elevation features of LiDAR data. Then a fully connected DNN is adopted to fuse the aforementioned features. At the end of the framework, logistic regression is employed to produce the final classification map.

The whole procedure is optimized through back-propagation with the help of training samples. Details of the proposed framework are elaborated in the following subsections.

### B. Convolutional Neural Network Based Feature Extraction

Convolution layers, nonlinear transformation and pooling layers are the three fundamental parts of CNNs [8]. By stacking several convolution layers with nonlinear operation and several pooling layers, a deep CNN can be formulated. Deep CNNs can hierarchically extract features of inputs, which tend to be invariant and robust [9].

Using a specific architecture like local connections and shared weights, CNNs tend to provide good generalization when facing vision problems.

A convolutional layer with nonlinear operation is as follows:

$$\mathbf{x}_j^l = f\left(\sum_{i=1}^M \mathbf{x}_i^{l-1} * \mathbf{k}_{ij}^l + \mathbf{b}_j^l\right) \quad (1)$$

where matrix  $\mathbf{x}_i^{l-1}$  is the  $i$ -th feature map of the previous  $(l-1)$ -th layer,  $\mathbf{x}_j^l$  is the  $j$ -th feature map of the current  $(l)$ -th layer, and  $M$  is the number of input feature maps.  $\mathbf{k}_{ij}^l$  and  $\mathbf{b}_j^l$  are randomly initialized and set to zero, respectively, then they are fine-tuned through back-propagation.  $f(\cdot)$  is a nonlinear function and  $*$  is the convolution operation.

Pooling operation offers invariance by reducing the resolution of the feature maps. The neuron in the pooling layer combines a small  $N \times N$  (e.g.  $N = 2$ ) patch of the convolution layer. The most common pooling operation is max pooling.

The success of CNNs mostly relies on the fact that the networks hierarchically learn the context invariant features, which are particularly useful for image classification.

### C. Deep Neural Network Based Deep Fusion

Through the aforementioned CNNs, the spatial-elevation and spatial-spectral information can be extracted from LiDAR and MSI/HSI, respectively. In this subsection, the extracted features are fused based on fully connected deep neural networks.

Due to the high dimensionality and limited training samples, overfitting is a serious problem one may face. In order to handle the issue, a recently-introduced method named dropout is used [10]. This method is based on setting the output of some hidden neurons to zero, thus the dropped neurons do not contribute in the forward pass and are not used in the back propagation procedure. In different training epochs, the deep net forms a different neural network by dropping neurons randomly. The dropout method prevents complex co-adaptations and the neurons can learn the correct features.

In order to build up a robust network, rectified linear units (ReLUs) are used in the training procedure [11]. In this letter, a ReLU is investigated, which is a simple nonlinear operator. This method accepts the output of a neuron if it is positive, while it returns 0 if the output is negative.

By using ReLU and dropout, the outputs of majority neurons turn to 0. We use several ReLUs and dropouts at several layers to achieve a powerful sparse-based regularization for the deep network and in parallel, address the overfitting problem for MSI/HSI and LiDAR data classification.

To further boost the performance of our networks, an advanced technique named batch normalization is adopted [12]. Batch normalization explicitly forces the activations of each layer to have zero means and unit variances. During the training process, the distribution of activations keeps changing due to the change of network parameters, i.e., this refers to the internal covariate shift, while batch normalization is proposed to tackle this problem.  $\mathcal{B} = \{x_1, x_2, \dots, x_m\}$  contains values over a mini-batch, then the batch normalization mechanism can be

formulated as following:

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad (2)$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (3)$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (4)$$

$$y_i = BN_{\gamma, \beta}(x_i) = \gamma \hat{x}_i + \beta \quad (5)$$

Equation (4) implements the normalization operation. Then the normalized value is scaled and shifted by learnable parameters  $\gamma$  and  $\beta$  to get the final result  $y_i$ . In the implementation, batch normalization can be inserted anywhere into the network just as a normal computational layer since all steps in the batch normalization are based on simple differentiable operations. The batch normalization is a practical tool in training deep neural networks for the following reasons: First, it can alleviate the problem caused by improper network initialization. Second, it can effectively speed up the training procedure by preventing “gradient vanishing”.

### III. EXPERIMENTAL RESULTS

#### A. Data Description and Experimental Settings

Three data sets were used in our study to evaluate the proposed methods. All three data sets used in this letter are benchmark data sets which have been geometrically corrected. Two of them contain multispectral and LiDAR data, while the last is composed of hyperspectral and LiDAR data.

The first two data sets are sub regions of a scene acquired from the downtown area of San Francisco, USA. The former is located at Bayview Park (Fig. 2), and the latter is located at a factory named “Recology” (Fig. 3). The multispectral images with eight bands ranging 400-1040 nm in the wavelength were acquired on October 9, 2011 by WorldView2, and the corresponding LiDAR data were acquired in June 2010. The Bayview Park data set has 300×200 pixels with seven land cover classes while the Recology data set has 200×250 pixels with eleven land cover classes. Both of them have a resolution of 1.8 m. For both sets, 50 samples were randomly selected from each class as training samples and the rest labeled as test samples. Detailed information about the numbers of samples of the above-mentioned data sets can be found in [4].

The third data set (Fig. 4) was captured over the University of the Houston Campus and the neighboring urban area by the NSF-funded Center for Airborne Laser Mapping (NCALM). The hyperspectral image with 144 bands in the spectral range of 380-1050 nm was acquired on June 23, 2012, and the LiDAR data was acquired on June 22, 2012. The Houston data set has 349×1905 pixels with a spatial resolution of 2.5m. The available training samples cover 15 classes. For this data set, 80 samples were randomly selected from each class as training and the rest as test samples [13]. Particularly, LiDAR-derived digital surface models (LiDAR-DSMs) were used to provide elevation information for the convenience of computer processing.

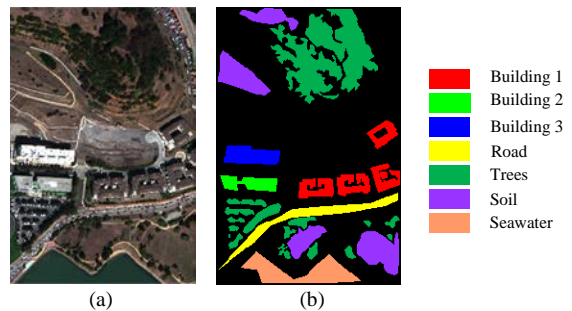


Fig. 2. Bayview Park data set: (a) false color map, and (b) groundtruth map.

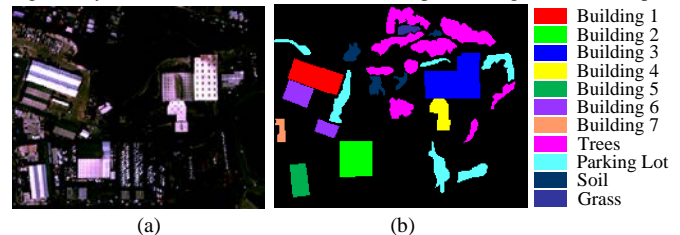


Fig. 3. Recology data set: (a) false color map, and (b) groundtruth map.

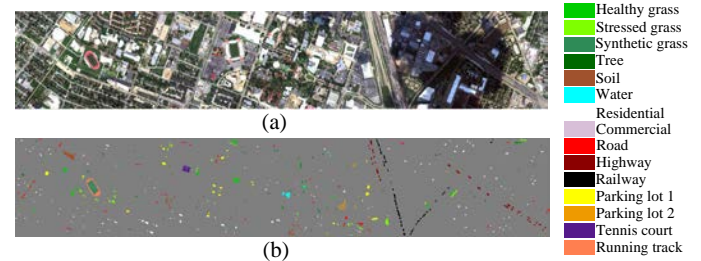


Fig. 4. Houston data set: (a) false color map, and (b) groundtruth map.

Totally, three nets were trained for each data set in the experiments: a designed spectral net, a proposed LiDAR net, and a proposed net for fusion. The spectral net and the LiDAR net are both convolutional nets with similar structures. For the spectral net, the size of the neighborhood window was set to 27×27, while for the LiDAR net, a larger window size (41×41) was adopted to include more spatial information in the input. Both of the spectral and LiDAR-DSM data were linearly mapped into [-0.5, 0.5] for the convenience of training. In consideration of the small input size and limited training samples, only three convolution layers and pooling layers were used. To deal with the vanishing gradient problem and accelerate the training procedure, a layer of batch normalization was inserted after each convolution layer.

Both spectral and LiDAR-DSM data were converted into 128-dimension vectors through CNNs. Then, the stacked vector had 256 dimensions containing spectral, spatial, and elevation information.

In the training process, mini-batch-based back-propagation was taken into account. The size of the mini-batch was set to 25 for all three nets. The initial learning rate of all CNNs was set to 0.005 and the learning rate gradually decreased to a half of its former value with a step size of 30 epochs. The number of training epochs was 240 for convolutional nets and 180 for the fully connected net.

SVM classifiers with radial basis functions using extended multi-attribute profiles [13] and extended multi-extinction profiles [14], which were called EMAP-SVMs and EMEP-SVMs for short, were used for comparison. We used

same parameters as [12] to ensure the effectiveness of EMAPs. For EMEPs, four attributes considered were: 1) area; 2) height; 3) volume; and 4) diagonal of the bounding box. Thresholds values used to generate EPs are automatically given by  $[a^j]$ ,

where  $j = 0, 1, \dots, s - 1$ . Here we set  $a$  and  $s$  to 3 and 7, respectively, as [13].

TABLE I CLASSIFICATION ACCURACIES OF DIFFERENT CLASSES ON BAYVIEW PARK DATA SET

	EMEPs <sub>MSI</sub>	EMEPs <sub>LiDAR</sub>	Spectral Net	LiDAR Net	EMAPs <sub>MSI</sub> +EMAPs <sub>LiDAR</sub>	EMEPs <sub>MSI</sub> +EMEPs <sub>LiDAR</sub>	Proposed
Building 1	93.38±2.63	95.81±1.39	99.86±0.34	98.59±1.36	95.16±1.15	95.98±1.42	<b>99.96±0.08</b>
Building 2	97.34±0.93	95.76±1.02	99.86±0.26	<b>100.00±0.00</b>	95.87±1.90	97.16±1.31	99.98±0.05
Building 3	94.42±2.91	95.88±2.21	99.07±1.73	99.55±1.24	94.45±2.35	95.02±1.66	<b>99.95±0.14</b>
Road	92.70±2.47	8.95±3.39	99.94±0.14	96.07±3.86	94.64±2.15	95.70±2.10	<b>100.00±0.00</b>
Trees	94.85±3.68	82.33±1.96	97.96±0.76	91.10±3.78	93.52±2.92	94.56±3.57	<b>98.95±0.45</b>
Soil	97.13±1.00	0.75±0.61	<b>99.79±0.38</b>	70.48±15.71	97.70±1.12	97.67±1.11	99.41±0.78
Seawater	100.00±0.00	100.00±0.00	99.76±0.55	94.83±7.24	100.00±0.00	100.00±0.00	<b>100.00±0.00</b>
OA	95.29±1.73	64.24±1.08	98.95±0.28	89.55±3.35	95.14±1.24	95.87±1.56	<b>99.41±0.18</b>
AA	95.69±1.13	68.50±0.87	99.46±0.26	92.94±1.98	95.91±0.72	96.58±0.72	<b>99.75±0.10</b>
K	0.9362±0.0228	0.5442±0.0127	0.9857±0.0037	0.8606±0.0438	0.9342±0.0162	0.9441±0.0205	<b>0.9919±0.0025</b>

TABLE II CLASSIFICATION RESULTS ON RECOLOGY DATA SET

	EMEPs <sub>MSI</sub>	EMEPs <sub>LiDAR</sub>	Spectral Net	LiDAR Net	EMAPs <sub>MSI</sub> +EMAPs <sub>LiDAR</sub>	EMEPs <sub>MSI</sub> +EMEPs <sub>LiDAR</sub>	Proposed
OA	94.39±0.82	75.73±4.73	96.41±0.85	94.86±1.05	95.92±0.83	95.11±1.04	<b>98.15±0.02</b>
AA	95.61±0.68	74.00±0.58	98.16±0.33	96.18±0.54	96.58±0.50	95.98±0.61	<b>98.90±0.01</b>
K	0.9337±0.0096	0.7201±0.0519	0.9575±0.0100	0.9392±0.0123	0.9517±0.0098	0.9422±0.0122	<b>0.9781±0.03</b>

TABLE III CLASSIFICATION RESULTS ON HOUSTON DATASET

	EMEPs <sub>HSI</sub>	EMEPs <sub>LiDAR</sub>	Spectral Net	LiDAR Net	EMAPs <sub>HSI</sub> +EMAPs <sub>LiDAR</sub>	EMEPs <sub>HSI</sub> +EMEPs <sub>LiDAR</sub>	Proposed
OA	96.69±0.43	84.62±0.86	97.80±0.54	93.18±0.0085	97.07±0.48	97.78±0.54	<b>98.61±0.27</b>
AA	96.81±0.36	85.36±0.73	98.05±0.44	94.36±0.67	97.22±0.46	97.87±0.57	<b>98.75±0.28</b>
K	0.9642±0.0046	0.8337±0.0092	0.9762±0.0058	0.9262±0.0091	0.9683±0.0052	0.9759±0.0059	<b>0.9850±0.0029</b>

TABLE IV STATISTICAL SIGNIFICANCE OF DIFFERENCES IN CLASSIFICATION ACCURACIES ON THREE DATA SETS

	Deep Fusion vs. EMEPs <sub>Spectral</sub>	Deep Fusion vs. EMEPs <sub>LiDAR</sub>	Deep Fusion vs. Spectral Net	Deep Fusion vs. LiDAR Net	Deep Fusion vs. EMAPs <sub>Spectral</sub> +EMAPs <sub>LiDAR</sub>	Deep Fusion vs. EMEPs <sub>Spectral</sub> +EMEPs <sub>LiDAR</sub>
Bayview Park	25.7439	73.7139	8.2496	45.7069	20.8767	19.7990
Recology	17.0570	49.5132	9.2066	19.2937	7.7185	11.2921
Houston	13.2351	41.7911	7.2524	27.5845	5.1694	1.9192

There are two parameters for SVMs with RBF: the penalty factor  $C$  and the RBF kernel width  $\gamma$ . The best  $C$  was searched within the given set  $\{10^{-1}, 10^0, 10^1, 10^2, 10^3\}$  and best  $\gamma$  was searched within the given set  $\{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1\}$ . Here we adopted the fivefold cross-validation method to search for the optimal SVM parameters.

In order to quantitatively compare and estimate the capabilities of the proposed models, overall accuracy (OA), average accuracy (AA) and Kappa coefficient (K) were conducted independently for 10 times to achieve a statistical evaluation, which aimed to alleviate the bias of the weight initialization of CNNs.

### B. Experiment Results and Analysis

The average classification accuracy of each class and its standard deviation on Bayview Park data set is listed in Table I. Table II and Table III are dedicated to the classification results obtained on Recology and Houston data sets, where we only list average OA, AA and Kappa coefficient along with their corresponding standard deviations.

#### 1) Performances using spectral and LiDAR information separately

All of three classifiers have shown better performances on spectral than on LiDAR data. LiDAR data only contain elevation information which is not sufficient to discriminate different categories for complex classification tasks. From Table I, it can be seen that classifiers using only LiDAR data have demonstrated rather poor performances on classes with similar heights. For classifiers based on the spectral data, more information is available so that there is not any sharp deterioration of accuracies between different classes of interest. However for separating objects made by similar materials, the advantage of using spectral data is not apparent any more. For example, for both EMEP-SVMs and CNNs, LiDAR-based classifiers have revealed comparable abilities with those trained on spectral data in terms of discriminating buildings. The above analysis shows that spectral and LiDAR data provide complementary information for classification tasks. Meanwhile it is worth mentioning that CNN-based classifiers demonstrate better performances than EMEP-SVMs, especially for the

LiDAR data. We attribute this to the better use of spatial information extracted of the CNNs.

### 2) Classification performances after data fusion

For both the attribute profile-based fusion methods and the proposed deep fusion framework, higher accuracies can be achieved compared to the use of each source individually. The improvement is due to the joint use of spectral and elevation information. Meanwhile, our deep fusion and classification networks lead to higher OA, AA, and Kappa coefficients for all three data sets than the ones obtained by the EMEP-SVM and EMAP-SVM classifiers on randomly chosen training sets. In addition, the proposed method achieves the highest class specific accuracies on almost all classes. Besides we also conducted experiments on standard training and test samples of Houston data set. Detailed information about the number of standard training and test samples can be found in [14]. In this case, basic structures of the model remained the same except that a smaller neighborhood window (11×11) was adopted for both spectral and LiDAR networks. TABLE V shows classification accuracies of the deep fusion model and EMEP-CNN methods proposed in [14]. It can be seen that our method outperforms EMEP-CNNs and achieves rather competitive results compared to state-of-art techniques.

TABLE V CLASSIFICATION RESULTS ON HOUSTON DATA SET USING STANDARD TRAINING AND TEST SAMPLES

	Methods in [13]			Proposed
	EMEP <sub>S<sub>HSI</sub></sub> +EMEP <sub>S<sub>LiDAR</sub></sub>	GBFF	Stack	
OA	88.81	91.02	89.71	<b>91.32</b>
AA	90.00	91.82	90.39	<b>91.96</b>
K	0.8788	0.9033	0.8884	<b>0.9057</b>

The competitive performance of our method is mainly obtained due to the following two advantages: First, as mentioned before, CNNs are powerful tools for feature extraction. Through CNNs, spectral and LiDAR data can be converted into highly abstract feature vectors, which are crucial for the following fusion and classification. Second, in our deep fusion method, a fully connected neural network is used for the fusion of spectral and LiDAR data features instead of simply stacking them together. The network can learn nonlinear combinations of the two kinds of features through training which are useful for the final classification.

### 3) Statistical significance about the difference between methods

Besides the above metrics, a McNemar's test [154] was adopted to demonstrate the statistical significance in the classification accuracy obtained by the proposed deep fusion method. The McNemar's test is a non-parametric test based on the standardized normal test statistic:

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \quad (6)$$

where  $f_{ij}$  indicates the number of samples correctly classified by method  $i$  while misclassified by method  $j$ . From Table IV it can be seen that for all of the three data sets, the differences

between the deep fusion method and other methods are statistically significant in most cases. The results further confirm that the proposed fusion framework can effectively improve classification performances.

## IV. CONCLUSIONS

In this letter, a deep model is proposed for remote sensing data fusion and classification. CNNs are used to effectively extract abstract and informative features from MSI/HSI and LiDAR data, separately. Then, a DNN is adopted to fuse the heterogeneous features obtained by the aforementioned CNNs. The fused features are useful for accurate classification. The experimental results with three data sets indicate the usefulness of the proposed deep fusion model and it opens a new window for remote sensing data fusion.

## REFERENCES

- [1] J. A. Benediktsson and P. Ghamisi, *Spectral-Spatial Classification of Hyperspectral Remote Sensing Images*. Boston, MA, USA: Artech House, 2015.
- [2] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of hyperspectral and LiDAR remote sensing data for classification of complex forest areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1416–1427, May 2008.
- [3] E. Puttonen, A. Jaakkola, P. Litkey, and J. Hyypä, "Tree classification with fused mobile laser scanning and hyperspectral data," *Sensors*, vol. 11, no. 5, pp. 5158–5182, May 2011.
- [4] Y. Gu, Q. Wang, X. Jia, and J. A. Benediktsson, "A novel MKL model of integrating LiDAR data and MSI for urban area classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5312–5326, Oct. 2015.
- [5] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [6] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Neural Inf. Process. Syst.*, Lake Tahoe, NV, USA, 2012, pp. 1106–1114.
- [7] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.
- [8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [9] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
- [10] G. Hinton, *et al.*, "Improving neural networks by preventing co-adaptation of feature detectors," *Comput. Sci.*, vol. 3, no. 4, pp. 212–223, Apr. 2012.
- [11] V. Nair, and G. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, Haifa, Israel, 2010, pp. 807–814.
- [12] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, Lille, France, 2015, pp. 448–456.
- [13] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5975–5991, Jul. 2010.
- [14] 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 Observ. Remote Sens.*, vol. 8, no. 6, pp. 2971–2983, Jun. 2015.
- [15] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, May 2004.