CNN-BASED FEATURE-LEVEL FUSION OF VERY HIGH RESOLUTION AERIAL IMAGERY AND LIDAR DATA

S. Daneshtalab[1], H. Rastiveis[1,*], B. Hosseiny[1]

[1] School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran - (somaye.danesh, hrasti, ben.hosseiny)@ut.ac.ir

Commission III, WG III/6

KEY WORDS: Convolutional Neural Network (CNN), Feature Fusion, Deep Learning, Feature Extraction, Aerial Imagery, LiDAR

ABSTRACT:

Land-cover classification of Remote Sensing (RS) data in urban area has always been a challenging task due to the complicated relations between different objects. Recently, fusion of aerial imagery and light detection and ranging (LiDAR) data has obtained a great attention in RS communities. Meanwhile, convolutional neural network (CNN) has proven its power in extracting high-level (deep) descriptors to improve RS data classification. In this paper, a CNN-based feature-level framework is proposed to integrate LiDAR data and aerial imagery for object classification in urban area. In our method, after generating low-level descriptors and fusing them in a feature-level fusion by layer-stacking, the proposed framework employs a novel CNN to extract the spectral-spatial features for classification process, which is performed using a fully connected multilayer perceptron network (MLP). The experimental results revealed that the proposed deep fusion model provides about 10% improvement in overall accuracy (OA) in comparison with other conventional feature-level fusion techniques.

1. INTRODUCTION

The diversification of geospatial data and the limitations of the RS sensors have attracted the interest of many researchers in developing various data fusion algorithms with greater ability and efficiency (Goshtasby and Nikolov, 2007). Because combining various data sources to integrate various information can help to improve classification results. Among the geospatial data, very high resolution (VHR) images and LiDAR data provide spatial contexture details and elevation information (Schmitt and Zhu, 2016), and fusion of these data is now a successful and active practical. LiDAR can provide height and shape information which is valuable for better describing the scene obtained by optical sensors only (Morsy et al., 2017). Since these source data have specific merits, numerous classification methods have been developed for fusion of VHR and LiDAR data, in the past two decades (Daneshtalab and Rastiveis, 2017; Xu et al., 2018). In this regard, the majority of these approaches are based on relatively simple or highly-customized decision rules that target subject classes or target objects based on specific elevation features, vegetation index, shape, or other information.

Taking advantages of the rich fusion information, a numerous traditional classification methods, such as *k*-nearest-neighbors (*k*-NN) and maximum likelihood as well as advanced classifiers such as Support Vector Machine (SVM) and Neural Networks (NN) have been used for image classification (Li et al., 2007; Makarau et al., 2011). For example, Rastiveis (2015) discussed a decision-level fusion of high-resolution aerial orthophoto and LiDAR data based on the Naïve Bayesian algorithm. In that research, the results of three different strategies for classification of these data using Naïve Bayes algorithm were compared that the results of final decision fusion were provided the best results.

Recently, deep learning (DL) based method using convolutional neural networks (CNNs) have been of great interest of RS researchers. Due to the effectiveness of different CNNs algorithms, they have been used in many RS applications, specifically image classification, and has shown superior performance over traditional methods (Xia et al., 2019; Zhang et al., 2016). CNNs are a type of deep learning that includes a large number of convolutional and sampling layers. The CNN input layer is usually an image matrix with arbitrary dimensions, and its output is a feature vector corresponding to different classes. CNN-based classification methods use these features in a classification algorithm to find the class label.

The statistical characteristics of images with VHR and multispectral images pose significant problems for automated analysis due to their high spatial and spectral redundancy as well as their non-linear nature. Therefore, in this research, we focus on the indirect approach through spatial-spectral feature extraction to study the supplementary information transmitted by a LiDAR and a VHR image. In this case, a CNN-based feature-level framework to integrate these data is proposed for object classification of an urban area. The architecture of the proposed CNN network consists of convolution kernels to extract deep features from surrounding neighbours of a pixel for spatial-spectral feature extraction. All the extracted deep features are then concatenated together in a fully-connected NN to classify the pixels.

Proposing a novel CNN framework for classification of VHR and LiDAR data to make full use of spatial-spectral information of these data is the contribution of this paper. The paper is organized as follows: In Section 2, a review of the related studies is presented. Then, the proposed method is introduced in Section 3. Then, in Section 4, the experimental results of our method in comparison with SVM and NN classifiers are presented. Finally, we will conclude the paper by summarizing our results in Section 5.

* Corresponding author

2. RELATED WORKS

The use of CNN for the fusion of LiDAR data and aerial images have been proposed by a number of researchers (Chen et al., 2017). In terms of multisource RS data, the study of the CNN model is still rare. Längkvist et al. (2016) combined the multispectral bands and digital surface model (DSM) height to conduct per-pixel and per-segmentation land use classification using single and multi-CNN models. Also, Chen et al. (2017) proposed a deep model for RS data fusion and classification. In their research, different CNNs are used to effectively extract abstract and informative features from fusing a multi-spectral image (MSI) or a hyperspectral image (HSI) with LiDAR data, separately. Then, a Deep Neural Network (DNN) is adopted to fuse heterogeneous features obtained by the aforementioned CNNs. Xu et al. (2017) proposed a novel CNN-based approach for the classification of multisource RS data including HSI, LiDAR, and Visible images data. Their CNN is a simple two-tunnel CNN, which consists of the same architecture of 2-D and 1-D CNNs, designed for reinforcing the correspondence spatial-spectral information. In addition, a cascade network was devised to combine features at different levels with a shortcut path. The experimental results with several multisource data demonstrated that in the condition of the same training samples, their CNN outperforms the traditional SVM and extreme learning machine (ELM) (Li et al., 2015).

Using DL-based approaches for building extraction based CNN and auto-encoders using LiDAR and orthophoto integration was studied by Nahhas et al. (2018). Their proposed architecture includes multi-resolution and spectral difference segmentations to create objects by grouping the image pixels according to their shape and spectral properties. Hartling et al. (2019) examined the potentiality of a novel DL-based method, called Dense Convolutional Network (DenseNet), to identify dominant individual tree species in a complex urban environment within a fused image of WorldView-II, Worldview-II and LiDAR data. DenseNet were compared against two popular machine learning classifiers including Random Forest (RF) and SVM, and the results proved the superiority of the DenseNet over the other two methods. Also, Santos et al. (2019) proposed and evaluated a CNN-based approach for detecting tree species from high-resolution images captured by RGB cameras in a UAV platform. In that study, three state-of-the-art CNN-based methods for object detection were tested: Faster R-CNN, YOLOv3, and RetinaNet. In the experiments carried out on a sample dataset comprising 392 images, RetinaNet achieved the most accurate results, having delivered 92.64% average precision.

Feng et al. (2019) proposed a modified two-branch CNN for urban land-use mapping using HSI and LiDAR data. Their network consists of an HSI branch and a LiDAR branch, both of which share the same network structure in order to reduce the burden and time cost of the design. Within the HSI and LiDAR branches, a hierarchical, parallel, and the multi-scale residual block was utilized, which could simultaneously increase the receptive field size and improve gradient flow. Moreover, an adaptive feature-fusion module based on a Squeeze-and-Excitation Net was proposed to fuse the HSI and LiDAR data.

Bigdeli et al. (2019) presented two different feature-learning strategies for the fusion of hyperspectral thermal infrared (HTIR) and visible RS data. First, a Deep Convolutional Neural Network (DCNN)-Support Vector Machine (SVM) was utilized on the features of two datasets to provide the class labels. To validate the results with other learning strategies, a shallow feature model was used, as well. Their experimental results showed that, except for the computational time, the DCNN-SVM model outperformed shallow feature-based strategies in the classification accuracy. Another state-of-the-art method for classification of LiDAR data and WorldView-II image based on DL concept was proposed by Wu et al. (2019). They implement a hierarchical multi-scale super-pixel based classification using the ResNet+SPP network for urban impervious surfaces extraction. The results showed that the hierarchical method based on LiDAR height information significantly improves the extraction of buildings and roads, and sufficiently exerts the superiority of LiDAR height information.

3. METHOD

The literature review has indicated that continued research is still needed to reach a classification framework that can efficiently integrate image and LiDAR data. An overview of the proposed method is shown in Figure 1. As shown in this figure, feature level fusion of the input data (RGB image and LiDAR data) is initially performed to extract low-level features. Then, the resulted features are imported in the designed CNN-based network to extract deep or high-level descriptors. At the end of the framework, the multi-layer perceptron (MLP) is employed to produce the final classification map. The procedure is optimized through the back-propagation with the help of training samples. Details of the proposed framework are elaborated in the following subsections.

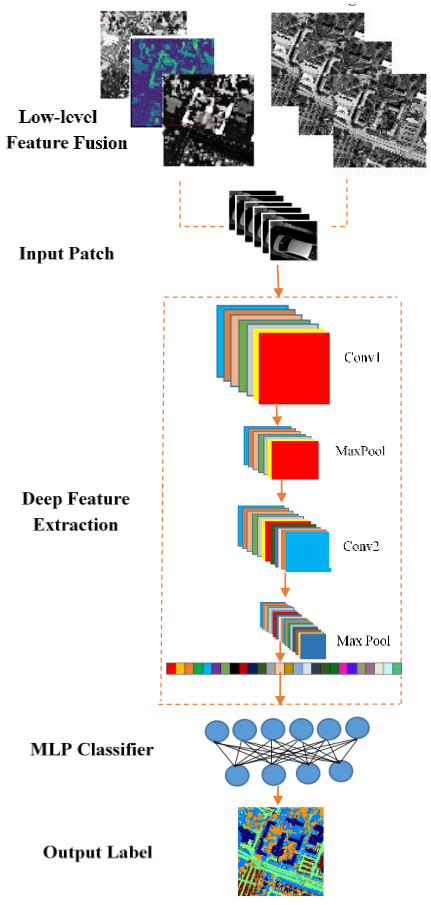


Figure 1. Framework of deep feature fusion for VHR and LiDAR data for accurate classification

3.1 Low-level Feature Fusion

In this step, after geometric correction and co-registration of LiDAR point clouds and orthophoto, the possible features of

these data are used as low-level features. The LiDAR-derived layers can be included normalized digital surface model (nDSM), intensity, first and last pulses differentiation and orthophoto bands may involve red, green, and blue channels. The resulted layers are then normalized and combined before importing in the convolutional layers. Note, the spatial resolutions of all these layers must be equalized using resampling techniques.

3.2 Deep Feature Extraction

The extracted low-level information may not result in superior classification accuracy. Therefore, in this step, a CNN-based framework is used as a deep feature extractor to extract high-level and powerful features by importing the low-level features into a number of convolutional layers. The hidden layers in this network include the convolutional layer, the sampling layer, and the fully connected layer. The neurons in each layer of a convolution are sorted in a three-dimensional (3D) manner, transforming a 3D input to a 3D output. For an image input, the first layer (input layer) holds the images as 3D inputs, with the dimensions being height, width, and the feature channels of the image. The neurons in the first convolutional layer connect to the regions of these images and transform them into a 3D output. The hidden units (neurons) in each layer learn non-linear combinations of the original inputs, which is called feature extraction (Xu et al., 2018). Moreover, the sample data for each class are randomly distributed over the ground truth image, and are divided into two categories of training and experiment data. Patch window around each pixel is defined considering dimensions of 25×25 pixels. Sufficient training data for each class will be randomly selected among the sample data.

The convolutional layer consists of learnable weights and biases that are used in the form of a filter with different dimensions and depths on the patch windows of the input layer. In this research, convolutional filters are 3×3 two-dimensional (2D) filters. Pooling layers follow the convolutional layers for down-sampling, hence, reducing the number of connections to the following layers. They do not perform any learning themselves but reduce the number of parameters to be learned in the following layers. They also help reduce overfitting. A max-pooling layer performs down-sampling by dividing the input data into rectangular pooling regions and computing the maximum of each region. The neuron in the pooling layer combines a small 2×2 patch of the convolution layer.

3.3 Classification

A fully connected layer multiplies the input by a weight matrix and then adds a bias vector. The convolutional (and down-sampling) layers are usually followed by one or more fully connected layers. All neurons in a fully connected layer connect to all the neurons in its previous layer. This layer combines all of the features (local information) learned by the previous layers across the image to identify larger patterns. At the end of the framework, a multi-layer perceptron (MLP) is taken into account to produce the final classification map. The structure of the MLP, here, contains 2 hidden layers with 100 neurons at each layer plus the SoftMax layer.

4. EXPERIMENTS AND RESULTS

4.1 Dataset

The used test data for evaluating the algorithm is the "grss_dfc_2018" dataset which has been provided by the 2018 IEEE (Institute of Electrical and Electronics Engineers) GRSS (Geoscience and Remote Sensing Society) Data Fusion Contest (Xu et al., 2019). It includes a VHR aerial imagery with a spatial resolution of 5 cm, multispectral-LiDAR point cloud data at 1550 nm, 1064 nm, and 532 nm, the intensity raster from first return per channel, and DSMs at a 50 cm ground sample distance (GSD). From the data set, a 325×350 sq. m. area was chosen as sample area to evaluate the algorithm. The available training samples cover five classes of Buildings (B), Roads (R), Trees (T), Grasses (G), and Cars (C). For this data set, 500 samples were randomly selected from each class as training and the rest as test samples.

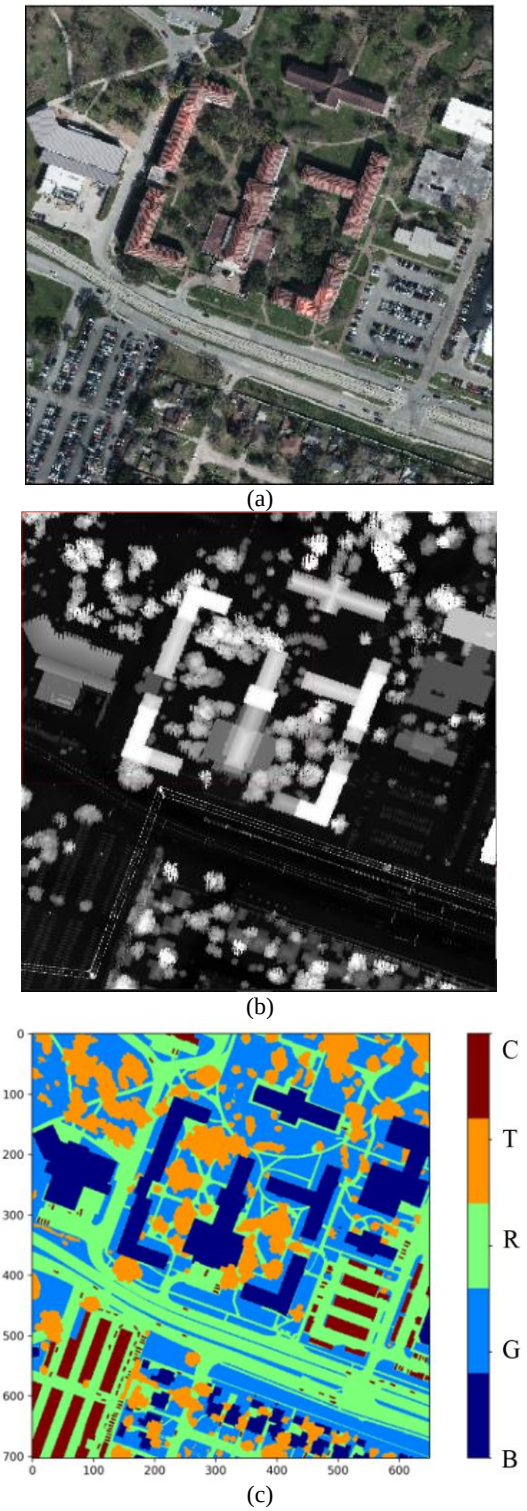


Figure 2. The grss_dfc_2018 data set over Houston. (a) VHR aerial image. (b) DSM (c) Ground truth map

4.2 Results

The proposed CNN was implemented using the Python programming language and TensorFlow deep learning framework[1]. In the first step of the proposed method, feature spaces on VHR and LiDAR data were produced, independently. From the LiDAR data, the intensity image, the nDSM, and the first-pulse and last-pulse differentiation with 5 cm spatial resolution were considered as low-level features. They integrated with the red, green, and blue channels of the image in the feature-level fusion. At the Deep feature extraction step, given the advantages of the 2D CNN for feature extraction and classification, with the help of appropriate CNN architecture, pixel neighbors across all bands were used. CNN teaches spatial features on its own and uses the features learned in classification. For the sample data set, a 25×25×6 patch around each pixel as the obtained low-level feature fusion is used as the input of the CNN. Note, the input images were normalized into [0 1] before importing to the CNN. The size of the mini-batch for training was 100, and the learning rate was 0.0003. In this set of experiments, the number of training epochs CNNs was 500. Also, random weights were used to initialize the network.

There are three factors, (i.e. dropout, ReLU, and patch size) that significantly affect the accuracy of the final classification, and they should be analyzed. The nonlinear layer was added after each convolution operation. It contains ReLU activation function, which brings nonlinear property. Overfitting is the phenomenon when the constructed model recognizes the examples from the training sample, but works relatively poorly on the examples of the test sample. To prevent overfitting, 10% and 50% dropout were considered after each convolutional and fully connected layers, respectively. Also, to train the CNN model, the patch size of the neighborhood window was set to 25×25 pixels, to include more spatial information in the input. Due to the small input size and limited training samples, only two convolution layers and pooling layers were used. Moreover, a layer of BN was inserted after each convolution layer to deal with the vanishing gradient problem as well as accelerate the training procedure. In the training process, mini-batch-based back-propagation was considered. Figures 3 displays the corresponding accuracy vs validation behavior during training process. Figure 4-c displays the obtained classification map using these parameters, which resulted in overall accuracy (OA) of 85%, and kappa coefficient of 82% after 500 epochs.

We also conducted radial basis function (RBF)-based SVM and MLP classifiers with the obtained low-level features for evaluating the proposed CNN-based feature-level fusion framework. In the SVM classifier, the grid search optimizes the SVM parameters (C, gamma, etc.) using a cross validation (CV) technique as a performance metric. Wide ranges of c and gamma values for the SVM were searched with the RBF-SVM method, which they were configured as c = 218 and gamma = 21. Also, the MLP with one hidden layer including 100 neurons was implemented to classify the dataset. The results of the MLP and RBF-SVM classifiers are shown in Figure 4-a and -b.

For any class, errors of commission occur when a classification procedure assigns pixels to a certain class that in fact does not belong to it. Also, errors of omission occur when pixels that in fact belong to one class, are included in other classes. The amount of errors of commission is also described by the Producer's accuracy (PA) indicator (1-PA). User's accuracy (UA) is another index characterizing the number of errors of omission (1-UA). It is the number of the correctly identified pixels of a class, divided by the total number of pixels of the class in the classified image. An Error matrix of each classification strategy, e.g. the MLP, the SVM, and the CNN-based, is illustrated in Tables 1, 2, and 3, respectively.

		Reference					
	Class	**B**	**G**	**R**	**T**	**C**	**UA (%)**
Classified	**B**	**53961**	371	794	10365	1886	80.09
	G	1850	**108375**	5497	19783	5113	77.07
	R	7371	13533	**100682**	9854	21068	66.02
	T	1980	6997	747	**58905**	1827	83.61
	C	335	447	1553	748	**20408**	86.88
	PA(%)	82.39	83.54	92.14	59.11	40.57	

Table 1. Confusion Matrix of the MLP classifier. B=Buildings; G = Grasses; R = Roads; T = Trees; C = Cars; UA = User Accuracy; PA = Producer Accuracy.

		Reference					
	Class	**B**	**G**	**R**	**T**	**C**	**UA (%)**
Classified	**B**	**61592**	819	1218	2088	1660	91.41
	G	2465	**112013**	7686	16187	2267	79.66
	R	5951	14275	**108361**	7981	15940	71.05
	T	2479	9645	1524	**56007**	801	79.49
	C	654	596	1797	313	**20131**	85.7
	PA(%)	84.21	81.55	89.86	67.82	49.34	

Table 2. Confusion Matrix of the SVM classifier. B=Buildings; G = Grasses; R = Roads; T = Trees C = Cars; UA = User Accuracy; PA = Producer Accuracy.

		Reference					
	Class	**B**	**G**	**R**	**T**	**C**	**UA (%)**
Classified	**B**	**64029**	659	1017	1399	273	95.03
	G	2119	**111382**	10857	14846	1414	79.21
	R	4137	11734	**119247**	8717	8673	78.19
	T	1196	6395	1940	**60680**	245	86.12
	C	11	44	643	112	**22681**	96.55
	PA(%)	89.56	85.54	89.19	70.76	68.14	

Table 3. Confusion Matrix of the proposed CNN classifier. B=Buildings; G = Grasses; R = Roads; T = Trees; C = Cars; UA = User Accuracy; PA = Producer Accuracy.

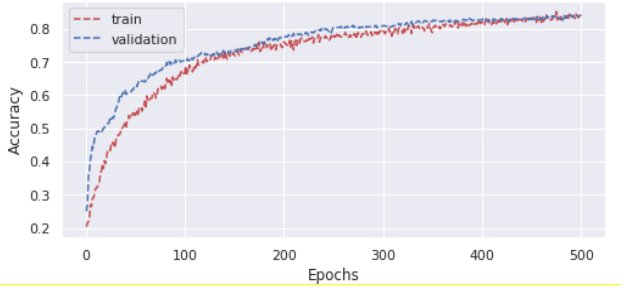
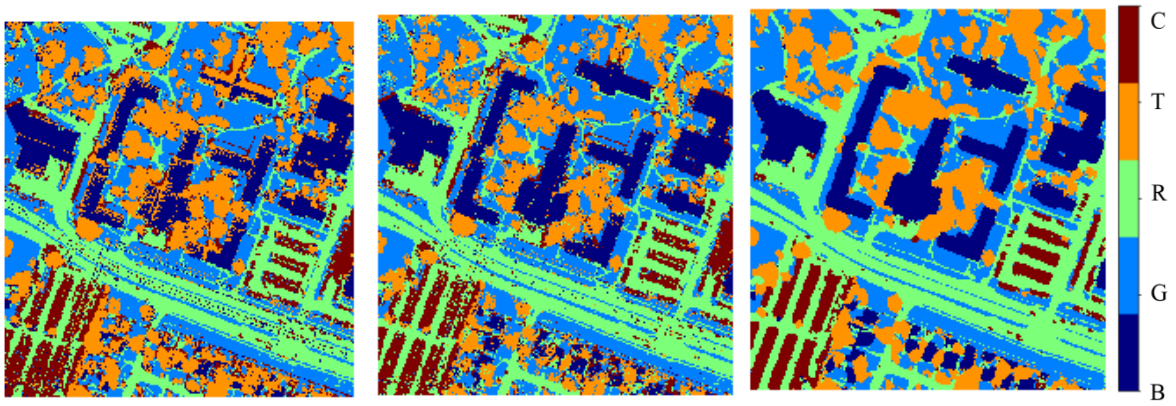


Figure 3. Accuracy of the training and the validation data during the training process of the proposed CNN.

[1] Tensorflow.org



(a) (b) (c)

Figure 4. The resulted classification maps using LiDAR data and VHR imagery (a) The RBF-SVM classifier using low-level features; (b) The MLP classifier using low-level feature; (c) The proposed CNN-based framework.

4.3 Discussion

Figure 5 summarized the resulted OA and Kappa coefficient obtained from these three strategies. As can be seen, extracting high-level and deep features, the CNN has improved the classification accuracy for VHR aerial imagery and LiDAR data fusion. The resulted OA and Kappa coefficient by this strategy about 11%, and 16% is higher than the RBF-SVM, and about 7% and 8% higher than the MLP classifiers which had used low-level features. The average improvement for OA and Kappa coefficient is about 10% and 12%, respectively. Based on the chart shown in Figure 5, the RBF-SVM obtained the worst results in terms of OA and Kappa coefficient.

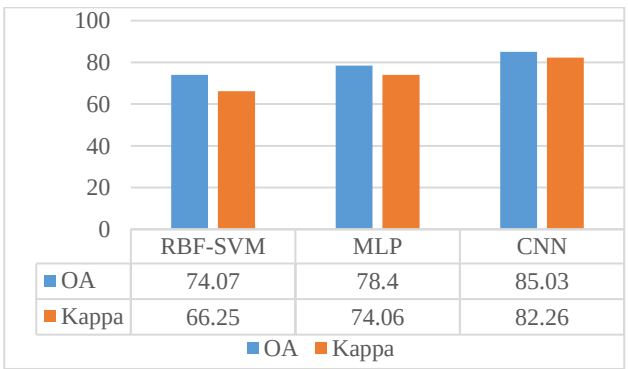


Figure 5. Comparison of different classification results obtained by the RBF-SVM, the MLP, and the proposed CNN-based framework

As it is clear from the error matrices shown in Tables 1-3, the accuracy of the car class is more accurate than the other classes in the CNN classification. It demonstrates the ability and structure of the network to train and identify different types of cars with different colors and spectral characteristics. Also, in classifying buildings and roads, due to User's accuracy and omission error, it is shown that the pixels of other classes are less allocated to these classes.

In most of the previous studies, this value has been chosen as 70% to 30% of the whole study area as training and test data, which is a big number. However, in this study, only 400 training samples for each class were chosen for training the algorithm, and the rest of the data were considered as tests. This can prove the generalization capability of the proposed network in facing with high amount of the test data.

5. CONCLUSIONS

This study developed a DL-based approach using CNN models to classify a fused LiDAR–VHR dataset. In this approach, the extracted low-level features from the image and LiDAR data are normalized and integrated. Then, they are imported in the designed CNN to extract the spectral-spatial descriptors. At the end of the framework, an MLP classifier is employed to produce the final classification map. The algorithm was tested on a sample area from the "grss_dfc_2018" dataset, and compared with two conventional classifiers using the low-level features. The results showed the average about 10% and 12% improvement in OA and Kappa coefficient, respectively. This reveals that the proposed deep fusion method can be a potential tool for the fusion of remote sensing images. Although the results were promising, however, testing the proposed network on other datasets to optimize this framework is suggested for future studies. Also, the algorithm may be tested for multiple data sources.

ACKNOWLEDGEMENTS

The authors would like to thank the National Center for Airborne Laser Mapping and the Hyperspectral Image Analysis Laboratory at the University of Houston for acquiring and providing the data used in this study, and the IEEE GRSS Image Analysis and Data Fusion Technical Committee.

REFERENCES

Bigdeli, B., Amini Amirkolaee, H., Pahlavani, P., 2019. Deep feature learning versus shallow feature learning systems for joint use of airborne thermal hyperspectral and visible remote sensing data. International Journal of Remote Sensing 40, 7048-7070.

Chen, Y., Li, C., Ghamisi, P., Jia, X., Gu, Y., 2017. Deep fusion of remote sensing data for accurate classification. IEEE Geoscience and Remote Sensing Letters 14, 1253-1257.

Daneshtalab, S., Rastiveis, H., 2017. Decision Level Fusion of Orthophoto and Lidar Data Using Confusion Matrix Information for Land Cover Classification. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 42.

Feng, Q., Zhu, D., Yang, J., Li, B., 2019. Multisource Hyperspectral and LiDAR Data Fusion for Urban Land-Use Mapping based on a Modified Two-Branch Convolutional

Neural Network. ISPRS International Journal of Geo-Information 8, 28.

Goshtasby, A., Nikolov, S., 2007. Guest editorial: Image fusion: Advances in the state of the art. Information Fusion 8, 114-118.
Hartling, S., Sagan, V., Sidike, P., Maimaitijiang, M., Carron, J., 2019. Urban Tree Species Classification Using a WorldView-2/3 and LiDAR Data Fusion Approach and Deep Learning. Sensors 19, 1284.

Längkvist, M., Kiselev, A., Alirezaie, M., Loutfi, A., 2016. Classification and segmentation of satellite orthoimagery using convolutional neural networks. Remote Sensing 8, 329.

Li, H., Gu, H., Han, Y., Yang, J., 2007. Fusion of high-resolution aerial imagery and lidar data for object-oriented urban land-cover classification based on svm, ISPRS Workshop on Updating Geo-spatial Databases with Imagery & The 5th ISPRS Workshop on DMGISs.

Li, W., Chen, C., Su, H., Du, Q., 2015. Local binary patterns and extreme learning machine for hyperspectral imagery classification. IEEE Transactions on Geoscience and Remote Sensing 53, 3681-3693.

Makarau, A., Palubinskas, G., Reinartz, P., 2011. Multi-sensor data fusion for urban area classification, 2011 Joint Urban Remote Sensing Event. IEEE, pp. 21-24.

Morsy, S., Shaker, A., El-Rabbany, A., 2017. Multispectral LiDAR data for land cover classification of urban areas. Sensors 17, 958.

Nahhas, F.H., Shafri, H.Z., Sameen, M.I., Pradhan, B., Mansor, S., 2018. Deep learning approach for building detection using lidar–orthophoto fusion. Journal of Sensors 2018.

Rastiveis, H., 2015. Decision level fusion of LIDAR data and aerial color imagery based on Bayesian theory for urban area classification. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 40, 589.

Santos, A.A.d., Marcato Junior, J., Araújo, M.S., Martini, D., Robledo, D., Tetila, E.C., Siqueira, H.L., Aoki, C., Eltner, A., Matsubara, E.T., 2019. Assessment of CNN-Based Methods for Individual Tree Detection on Images Captured by RGB Cameras Attached to UAVs. Sensors 19, 3595.

Schmitt, M., Zhu, X.X., 2016. Data fusion and remote sensing: An ever-growing relationship. IEEE Geoscience and Remote Sensing Magazine 4, 6-23.

Wu, M., Zhao, X., Sun, Z., Guo, H., 2019. A hierarchical multiscale super-pixel-based classification method for extracting urban impervious surface using deep residual network from worldview-2 and LiDAR data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12, 210-222.

Xia, Y., d'Angelo, P., Tian, J., Fraundorfer, F., Reinartz, P., 2019. Self-Supervised Convolutional Neural Networks for Plant Reconstruction Using Stereo Imagery. Photogrammetric Engineering & Remote Sensing 85, 389-399.

Xu, X., Li, W., Ran, Q., Du, Q., Gao, L., Zhang, B., 2017. Multisource remote sensing data classification based on convolutional neural network. IEEE Transactions on Geoscience and Remote Sensing 56, 937-949.

Xu, X., Li, W., Ran, Q., Du, Q., Gao, L., Zhang, B., 2018. Multisource remote sensing data classification based on convolutional neural network. IEEE Transactions on Geoscience and Remote Sensing 56, 937-949.

Xu, Y., Du, B., Zhang, L., Cerra, D., Pato, M., Carmona, E., Prasad, S., Yokoya, N., Hänsch, R., Saux, B.L., 2019. Advanced Multi-Sensor Optical Remote Sensing for Urban Land Use and Land Cover Classification: Outcome of the 2018 IEEE GRSS Data Fusion Contest. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12, 1709-1724.

Zhang, L., Zhang, L., Du, B.J.I.G., Magazine, R.S., 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. 4, 22-40.