# DIMENSIONALITY-VARIED CONVOLUTIONAL NEURAL NETWORK FOR HYPERSPECTRAL IMAGE CLASSIFICATION WITH SMALL-SIZED LABELED SAMPLES

*Xuejian Liang*[1], *Wanjun Liu*[2], *Ye Zhang*[1], *Jie Yu*[2], *Haicheng Qu*[2]

1. School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, China
2. School of Software, Liaoning Technical University, Huludao, China

## ABSTRACT

Recently, convolutional neural network (CNN) achieves outstanding performance for hyperspectral image (HSI) classification. However, CNN requires large-sized labeled samples for training and tremendous computations for 3-D spectral-spatial feature extraction. In practice, most HSI data are unlabeled and the availability of samples for training is limited. In order to improve the classification performance with a simplified structure, a dimensionality-varied convolutional neural network (DV-CNN) is proposed. DV-CNN reduces the complexity and calculations by changing the dimensionalities of feature maps and improves the classification accuracy of small-sized labeled HSI by deep network. The experiments are compared with some state-of-the-art methods on Indian pines and Pavia University data sets. The experimental results demonstrate that the proposed DV-CNN outperforms other methods and achieves better classification performance.

***Index Terms***— Convolutional neural network, hyperspectral image classification, small-sized samples, dimensionality variation, feature extraction

## 1. INTRODUCTION

With the development of remote sensing techniques, hyperspectral image (HSI) has received considerable interest in the field of remote sensing [1]. HSI data contain spectral and spatial information simultaneously [2], which are widely used in many fields such as agriculture, surveillance, geography, industry, environment protection [3], etc. Therefore, the analysis of remote-sensing images is paramount in practical application of the aforementioned fields, and the classification of HSI is one of the most popular topics in the remote-sensing community.

Many methods of classification based on spectral-spatial information were proposed and achieved significant improvements on performance. There were two different categories of spectral-spatial classification methods generally. The first one extracted spectral and spatial features separately, and the classifier of model was responsible for feature fusion, which contains remote sensing image fusion [4] and two-channel convolutional neural network [5]. The second one, which extracted spectral-spatial features synchronously and processed by classifier [6], achieved better performance than methods in the first category [7]. Traditional methods of HSI classification were developing continuously [8] and many innovative algorithms based on deep learning were proposed recently. Specially, convolutional neural network (CNN) were applied to tasks of image classification and achieved more outstanding performance on HSI classification than others. The superiority of CNN was based on big data and deep network structure, and large labeled samples of HSI could improve the classification performance evidently. However, the collection of labeled data involved difficult, expensive and time-consuming ground campaigns for researchers [9]. Thus, a small number of labeled samples were available for image classification in the literature for hyperspectral scenes [10]. The well-known Hughes effect also denoted that the accuracy of classification decreased when the dimensionality increased in the case of limited training samples. Besides, CNN based on spectral-spatial features achieved better classification performance which required more complex networks and tremendous computations. Thus the efficiency of HSI classification model based on CNN could be improved.

In this letter, a dimensionality-varied CNN (DV-CNN) for small-sized labeled samples is proposed. The dimensionalities of feature maps go through four stages variation in DV-CNN. In terms of small-sized samples, DV-CNN is a simpler classifier than other methods based on CNN and achieves better classification performance. We compare the classification accuracy of DV-CNN with some state-of-the-art classifiers for small-sized labeled HSI in this letter to demonstrate that DV-CNN can achieve higher accuracy with a simpler CNN structure.

## 2. PROPOSED METHOD

### 2.1. DV-CNN model

This section elaborates the structure of DV-CNN as illustrated in Figure 1. DV-CNN composes of four stages based on the dimensionality variation of feature map, and all

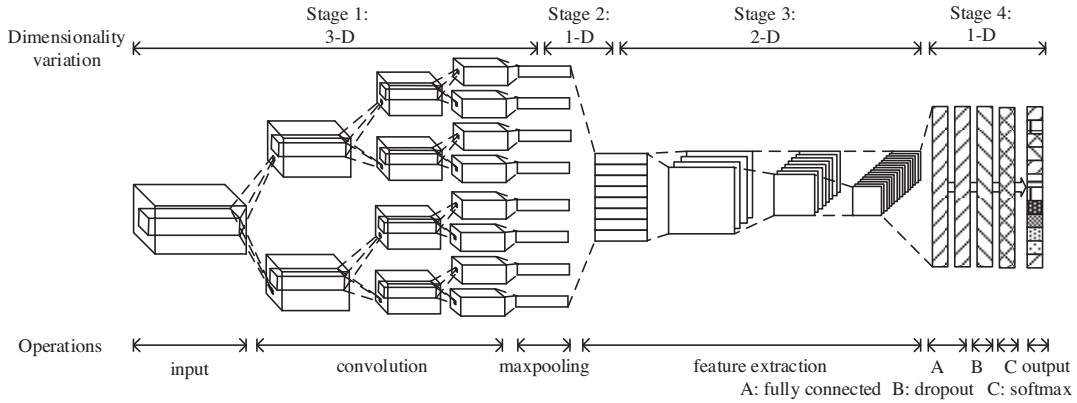



Figure 1 DV-CNN structure

stages are related to feature extraction. The stage 1 is responsible for extracting 3-D features from raw HSI data. The input of network is a 3-D cube of HSI which remains original spectral-spatial features. The input is convolved by 3-D kernels to extract spectral-spatial feature maps simultaneously. There are three convolution layers and one maxpooling layer in stage 1 for feature extraction, and the number of kernels increases as the layers increase. The results of stage 1 are a number of one-dimensionality (1-D) vectors that generated by 3-D feature extraction. Stage 2 assembles all 1-D vectors into a 2-D matrix which is input into stage 3. The 2-D matrix contains spectral-spatial features simultaneously with lower dimensionality than HSI data. The last two stages constitute a general CNN for 2-D image classification. There are three convolution layers and three maxpooling layers in the feature extraction process of stage 3, and the results of stage 3 are a series of 2-D feature maps after dimensionality reduction and feature extraction. Two fully connection layer, a dropout layer and a softmax layer for classification are in Stage 4. All results of stage 3 are reorganized into one 1-D vector in fully connection layer, so that all feature maps in stage 4 are 1-D vectors. Besides, stage 4 also contains an output layer for classification results of DV-CNN.

The 3-D convolution kernels can extract more accurate features for HSI classification methods according to literature of CNN, and classification model achieves higher accuracy when it has more numbers of layers and convolution kernels. Consequently, the complexity of model is increased because lots of calculations. 2-D convolution kernels obtain lower classification accuracy than 3D kernels because of inadequate use of spectral-spatial information, but largely reduce the calculations and complexity of network. Thus, in DV-CNN, 3-D kernels extract spectral-spatial features and 2-D kernels simplify the further processes of feature extraction. And the conversions of feature maps with the variations of dimensionalities ensure the accuracy of feature extractions and reduce the complexity and lots of calculations of model.

### 2.2. Feature extraction

Two different feature extractions, namely 3-D feature extraction and 2-D feature extraction, are used in DV-CNN, and each feature extraction contains convolution and maxpooling operations.

The convolution process in 3-D feature extraction can be formulated as

$$\text{output}\left(m_{l,i}^{h}\right)=\sum_{x=1}^{L}\sum_{y=1}^{W}\sum_{z=1}^{H}\text{input}\left(n_{(l-1),i}^{g(x),g(y),g(z)}\right)k_{l,i}^{x,y,z} \quad (1)$$

$$g\left(x\right)=x+s_{x}\left(h-1\right) \quad (2)$$

In Eq. (1) and Eq. (2), $k_{l,i}^{x,y,z}$ is one of the elements of convolution kernel $k_{l,i}$ with $i$ being the index of kernel in $l$ th layer, and $n_{(l-1),i}^{g(x),g(y),g(z)}$ is one of the elements of $n_{(l-1),i}$. $(x,y,z)$ and $\left(g(x),g(y),g(z)\right)$ indicate the position of elements in them respectively, and $i$ is the index of input 3D feature map $n_{(l-1),i}$ in $\left(l-1\right)$th layer with $L$, $W$ and $H$ being the length, width and height separately. The $m_{1,i}^{h}$ is one of the elements of feature map generated by convolution in the 3D process with $h$ being the index. Besides, $s_x$ stands for the stride of convolution in the direction of $x$. The next operations can be formulated as

$$\text{output}\left(\boldsymbol{m}_{(l+1),i}\right)=f\left(\boldsymbol{m}_{l,i}+\boldsymbol{b}_{l,i}\right) \quad (3)$$

where $l$ indicates the layer that is considered, and $\boldsymbol{m}_{l,i}$ is the input. $\boldsymbol{b}_{l,i}$ is the bias corresponded to $\boldsymbol{m}_{l,i}$. The function $f(\cdot)$ represents Rectified Linear Unites (ReLU), which is a

non-saturation function and faster than other saturating functions. $\boldsymbol{m}_{(l+1),i}$ is the result of convolution.

The convolution process in 2-D can be formulated as

$$m_{(l+1),i,j}^{\alpha,\beta} = \sum_{x=1}^{L}\sum_{y=1}^{W} k_{(l+1),i,j}^{x,y} n_{l,i}^{g(x),g(y)} \quad (4)$$

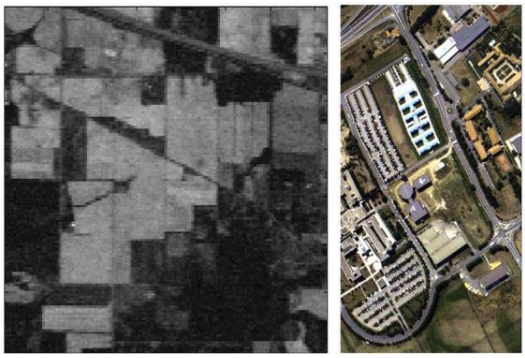
where $(\alpha,\beta)$ stands for the coordinate in 2D feature matrix. The meanings of other symbols, which should be mapped to 2D convolution, are basically the same as Eq. (1)-(3).

## 3. EXPERIMENTS

### 3.1. Experimental data

In this letter, two different HSI data sets are used for experiments to test the classification performance as shown in Figure 2. The two labeled data sets are randomly divided into training sets and test sets with a ratio of 1:9.

The first data set was the Indian Pines which was gathered by AVIRIS sensor over the Indian test site in northwestern Indiana. There were 16 ground truth classes and 220 spectral bands in the wavelength range 0.4 to 2.5$\mu$m in Indian Pines data set. 200 bands left after removing the 20 bands because of water absorption. Thus the input size of Indian Pines data set was 200×145×145. The second data set was Pavia University which was gathered by the ROSIS sensor over the city of Pavia, Italy. There were 9 classes and 103 spectral bands in the wavelength range 0.43 to 0.86$\mu$m in Pavia University. So the input size of Pavia University data set is 103×610×340.



(a) (b)

Figure 2 Data set: (a) Indian Pines and (b) Pavia University

### 3.2. Parameters setting

In 3-D feature extraction process of DV-CNN, there are three convolution layers and one maxpooling layer, and in 2-D feature extraction process, there are three convolution layers and three maxpooling layers. The numbers and sizes of kernels in feature extractions are shown in Table 1.

Besides, the adaptive moment estimation (Adam) algorithm was applied to DV-CNN for stochastic optimization. The dropout of DV-CNN and learning rate were set as 0.8 and $1\times10^{-3}$ respectively. DV-CNN selected 10×10 spatial area of HSI cube for Indian Pines and the batch size was set as 40. The spatial area size of Pavia University was set as 8×8 and the batch size was set as 150. When the value of cross entropy was smaller than $5\times10^{-6}$, the training stopped.

Table 1 Parameters of feature extraction

| Data set | 3-D feature extraction | | | | 2-D feature extraction | | | |
|---|---|---|---|---|---|---|---|---|
| | Convolution layer-1 | Convolution layer-2 | Convolution layer-3 | Maxpooling layer | Convolution layer-1 | Convolution layer-2 | Convolution layer-3 | Maxpoolin layers |
| Indian Pines | 5, 5×5×5 | 10, 5×3×3 | 22, 5×3×3 | 2×2×2 | 5, 3×3 | 10, 3×3 | 20, 3×3 | 2×2 |
| Pavia University | 6, 4×3×3 | 12, 4×3×3 | 22, 4×3×3 | 1×2×2 | 6, 3×3 | 12, 3×3 | 36, 3×3 | |

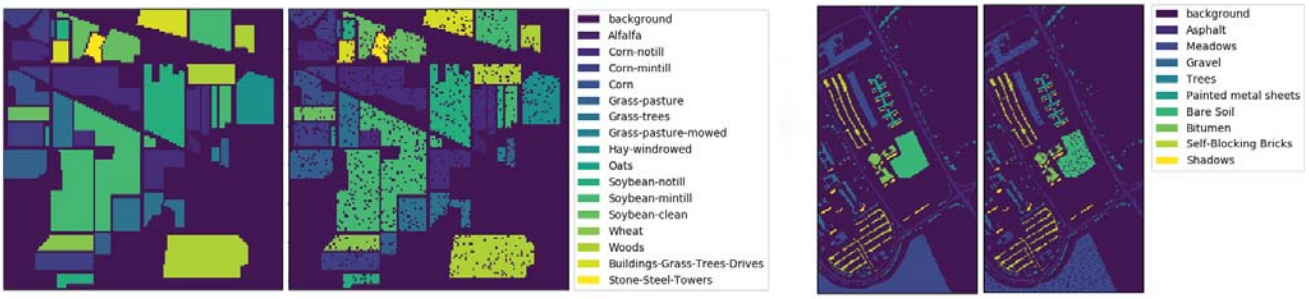
DV-CNN was implemented based on the TensorFlow framework and experiments were performed on Intel Core i7-6700@3.4GHz of CPU, NVIDIA GTX 1080Ti of GPU and 16GB of RAM. Each experiment was performed 10 times to achieve an average accuracy. Several state-of-the-art classification methods were considered for comparison.

### 3.3. Experimental results

In the experiments, we adopted overall accuracy (OA), average accuracy (AA) and kappa coefficient ($\kappa$) to assess the classification performance of methods. The classification results of DV-CNN on Indian Pines and Pavia University data sets are shown in Table 2 and Figure 3. The OA of DV-CNN on Indian Pines is 87.87%, with a 82.88% of AA and 0.862 of $\kappa$. The OA on Pavia University is 98.18%， with a 97.20% of AA and 0.976 of $\kappa$.

Table 2 Classification results of DV-CNN

| | Inian Pines | Pavia University |
|---|---|---|
| OA (%) | 87.87±0.41 | 98.18±0.29 |
| AA (%) | 82.88±1.44 | 97.20±0.31 |
| $\kappa$ | 0.862±0.005 | 0.976±0.004 |




(a) (b)

Figure 3 Classification results: (a) Indian Pines data set and (b) Pavia University data set

Table 3 Classification results of Indian Pines with different methods (%)

| Classes | Total samples | Methods | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | SVM | OMP | NRS | CRT | KCRT | DKCRT | **DV-CNN** |
| Alfalfa | 46 | 42.32 | 47.20 | 52.68 | 16.71 | **72.32** | 71.34 | 64.29 |
| Corn-notill | 1428 | 78.77 | 65.36 | 48.42 | **83.96** | 80.63 | 83.34 | 83.28 |
| Corn-mintill | 830 | 63.72 | 47.26 | 25.40 | 49.18 | 68.95 | 73.35 | **77.64** |
| Corn | 237 | 72.37 | 40.56 | 35.54 | 25.40 | 63.33 | **75.21** | 68.22 |
| Grass-pasture | 483 | 89.55 | 85.10 | 88.18 | 90.10 | 92.89 | 93.43 | **93.56** |
| Grass-trees | 730 | 96.64 | 91.41 | 95.25 | 98.65 | 98.07 | **98.71** | 96.80 |
| Grass-pasture-mowed | 28 | 67.00 | 26.20 | 54.00 | 5.60 | 64.80 | 63.80 | **80.77** |
| Hay-windrowed | 478 | 98.92 | 98.70 | 99.67 | 100.00 | 99.73 | **99.78** | 99.77 |
| Oats | 20 | 35.28 | 29.44 | 33.61 | 5.83 | 50.83 | 50.83 | 33.33 |
| Soybean-notill | 972 | 70.53 | 48.73 | 28.10 | 64.45 | 76.11 | 84.87 | **85.94** |
| Soybean-mintill | 2455 | 86.17 | 70.52 | **98.14** | 84.67 | 83.24 | 82.95 | 90.90 |
| Soybean-clean | 593 | 80.48 | 51.74 | 35.28 | 67.58 | 82.07 | **89.50** | 75.47 |
| Wheat | 205 | 97.66 | 94.35 | 95.27 | 99.27 | 99.24 | **99.29** | 92.97 |
| Woods | 1265 | 96.59 | 93.33 | **97.59** | 96.97 | 95.62 | 96.54 | 94.91 |
| Buildings-Grass-Trees-Drives | 386 | 52.44 | 48.40 | 59.03 | 55.26 | 63.63 | 61.04 | **79.60** |
| Stone-Steel-Towers | 93 | 87.80 | **88.04** | 81.67 | 86.67 | 87.38 | 86.55 | 80.95 |
| OA (%) | | 82.60 | 69.87 | 70.81 | 79.44 | 83.78 | 86.06 | **87.60** |
| AA (%) | | 76.01 | 64.15 | 64.25 | 64.39 | 79.93 | **81.91** | 81.15 |
| $\kappa$ | | 0.800 | 0.654 | 0.653 | 0.763 | 0.815 | 0.841 | **0.858** |

Table 4 Classification results of Pavia University with different methods (%)

| Classes | Total samples | Methods | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | SVM | OMP | NRS | CRT | KCRT | DKCRT | **DV-CNN** |
| Asphalt | 6631 | 91.18 | 70.29 | 93.01 | 85.32 | 83.29 | 86.49 | **98.73** |
| Meadows | 18649 | 98.51 | 92.60 | 98.75 | 98.59 | 95.92 | 97.98 | **99.67** |
| Gravel | 2099 | 33.26 | 57.49 | 58.59 | 56.00 | 67.19 | 67.86 | **91.80** |
| Trees | 3064 | 85.80 | 82.54 | 78.53 | 83.98 | 84.09 | 83.78 | **98.19** |
| Metal sheets | 1345 | 98.87 | 99.77 | 98.85 | 99.70 | 99.47 | 99.22 | **100.00** |
| Bare soil | 5029 | 49.98 | 58.37 | 53.25 | 55.86 | 75.88 | 74.02 | **97.92** |
| Bitumen | 1330 | 14.87 | 54.33 | 41.92 | 63.18 | 74.53 | 75.86 | **94.65** |
| Bricks | 3682 | 89.99 | 63.09 | 84.78 | 81.87 | 79.28 | 86.50 | **95.96** |
| Shadows | 947 | 99.83 | 83.62 | 91.08 | 90.14 | **99.95** | 99.74 | 95.78 |
| OA (%) | | 84.26 | 78.97 | 85.96 | 85.68 | 87.45 | 89.29 | **98.28** |
| AA (%) | | 73.59 | 73.57 | 77.64 | 79.40 | 84.40 | 85.72 | **96.97** |
| $\kappa$ | | 0.785 | 0.718 | 0.808 | 0.805 | 0.833 | 0.856 | **0.977** |

Besides, we also choose some state-of-the-art methods for HSI classification with small-sized labeled samples to compare with DV-CNN. In Table 3 and Table 4, the classification results of support vector machine (SVM) [11], OMP [12], NRS [13], CRT [14], KCRT [14] and DKCRT [15] are listed and compared with DV-CNN on Indian Pines

and Pavia University data sets, respectively. All of these methods are effective for hyperspectral classification with small-sized labeled samples and well received by researchers.

It can be obviously observed from Table 3 and Table 4 that DV-CNN can achieve higher classification performance than other methods. There are five of sixteen classes in Indian Pines and eight of nine classes in Pavia University that DV-CNN obtains better classification accuracy than other methods. Moreover, it is obvious for DV-CNN that the more samples of each class, the higher classification accuracy. The classification performance of DV-CNN is closely related to the number of samples.

## 4. CONCLUSION

In this letter, we proposed a DV-CNN for HSI classification with small-sized labeled samples, which can reduce the calculations by various dimensionalities of feature maps and improve the classification performance of small-sized labeled HSI by deep network structure. Besides, DV-CNN can make use of spectral-spatial information sufficiently to extract fusion features. The experiments compared with other classification methods demonstrate that DV-CNN is suitable for HSI classification with small-sized samples and able to achieve better classification performance.

Moreover, it should be noticed that the classification accuracy of DV-CNN are closely related to the amount of each class. In the future research works, we plan to aim at the data enhancement by Generative Adversarial Network to enrich the samples of classes and improve the classification accuracy further.

## 5. REFERENCES


[1] Y. Zhan, D. Hu, Y. Wang et al. "Semisupervised Hyperspectral Image Classification Based on Generative Adversarial Networks." IEEE Geoscience & Remote Sensing Letters, PP.99(2017):1-5.

[2] W. Z. Zhao and S. H. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach." IEEE Transactions on Geoscience and Remote Sensing, 54(8), 4544-4554 (2016)

[3] H.C. Qu, X.J. Liang, S.C. Liang, W.J. Liu, "Dimensionality-varied deep convolutional neural network for spectral–spatial classification of hyperspectral data," Journal of Applied Remote Sensing, 12(1), 016007 (2018),

[4] J. Zhong, B. Yang, G. Huang et al. "Remote Sensing Image Fusion with Convolutional Neural Network." Sensing & Imaging, 17.1(2016):10.

[5] H. Zhang, Y. Li, Y. Zhang et al. "Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network." Remote Sensing Letters, 8.5(2017):438-447.

[6] Palsson, Frosti, J. R. Sveinsson, and M. O. Ulfarsson. "Multispectral and Hyperspectral Image Fusion Using a 3-D-Convolutional Neural Network." IEEE Geoscience & Remote Sensing Letters, 14.5(2017):639-643.

[7] Li, Ying, H. Zhang, and Q. Shen. "Spectral-Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network." Remote Sensing, 9.1(2017):67.

[8] X. Cao, L. Xu, D. Meng, et al. "Integration of 3-dimensional discrete wavelet transform and Markov random field for hyperspectral image classification." Neurocomputing, 226. C(2017):90-100.

[9] Aydemir, Muhammet Said, and G. Bilgin. "Semisupervised Hyperspectral Image Classification Using Small Sample Sizes." IEEE Geoscience & Remote Sensing Letters, PP.99(2017):1-5.

[10] Li, Jun, J. M. Bioucas-Dias, and A. Plaza. "Semisupervised Hyperspectral Image Classification Using Soft Sparse Multinomial Logistic Regression." IEEE Geoscience & Remote Sensing Letters, 10.2(2012):318-322.

[11] R. Archibald and G. Fann, "Feature selection and classification of hyperspectral images with support vector machines," IEEE Geoscience & Remote Sensing Letters, vol. 4, no. 4, pp. 674–677, 2007.

[12] J. Ma, C. Chen, J. Huang et al. "Infrared and visible image fusion via gradient transfer and total variation minimization." Information Fusion, 31.C(2016):100-109.

[13] W. Li, E.W. Tramel, S. Prasad et al. "Nearest Regularized Subspace for Hyperspectral Classification." IEEE Transactions on Geoscience & Remote Sensing, 52.1(2013):477-489.

[14] W. Li, Q. Wei, and M. Xiong. "Kernel Collaborative Representation With Tikhonov Regularization for Hyperspectral Image Classification." IEEE Geoscience & Remote Sensing Letters, 12.1(2015):48-52.

[15] Y. Ma, C. Li, X. Mei et al. "Hyperspectral Image Classification with Discriminative Kernel Collaborative Representation and Tikhonov Regularization." IEEE Geoscience & Remote Sensing Letters, PP.99(2018).