DEEP RESIDUAL NETWORKS FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Zilong Zhong[1*], Jonathan Li[1], Lingfei Ma[1], Han Jiang[1], He Zhao[1]
[1] Mobile Mapping Lab, Department of Geography & Environmental Management, University of Waterloo, Ontario N2L 3G1, Canada - (z26zhong, junli, l53ma, h64jiang, h224zhao)@uwaterloo.ca

ABSTRACT

Deep neural networks can learn deep feature representation for hyperspectral image (HSI) interpretation and achieve high classification accuracy in different datasets. However, counterintuitively, the classification performance of deep learning models degrades as their depth increases. Therefore, we add identity mappings to convolutional neural networks for every two convolutional layers to build deep residual networks (ResNets). To study the influence of deep learning model size on HSI classification accuracy, this paper applied ResNets and CNNs with different depth and width using two challenging datasets. Moreover, we tested the effectiveness of batch normalization as a regularization method with different model settings. The experimental results demonstrate that ResNets mitigate the declining-accuracy effect and achieved promising classification performance with 10% and 5% training sample percentages for the University of Pavia and Indian Pines datasets, respectively. In addition, t-Distributed Stochastic Neighbor Embedding (t-SNE) provides a direct view of the extracted features through dimensionality reduction.

Index Terms — Deep residual networks, deep learning, hyperspectral image classification

1. INTRODUCTION

The pixel-wise image classification lays a solid foundation of geoscience application and analysis pertaining to multiple kinds of remotely sensed data, including hyperspectral images (HSIs) [1]. The increase of spectral and spatial resolution of HSIs poses two major challenges exist for accurate HSI interpretation. First, the Hughes Phenomenon, which means the recognition accuracy decreases drastically with the increase of the dimensionality of training data, derives from hundreds of spectral bands [2]. Second, their high spatial resolution makes the recognition of small objects possible but increases the high correlation between neighbouring pixels.

In the face of these challenges, recent studies have tried to apply supervised deep learning (DL) models to extract robust and discriminant features in the context of remotely sensed image classification [3, 4]. In 2016, convolutional neural network (CNN) was used to extract spatial features that integrated with spectral features learned from a embedding method [5]. The CNN model inherently takes spatial correlations of neighboring pixels into account, and the spatial features are complementary to the spectral features of hyperspectral imagery. However, the input of the CNNs are the three principle components of the original hyperspectral image, which means the input data still loses some spatial information.

Furthermore, 3D CNNs were adopted to extract deep spectral-spatial features directly from raw HSI and delivered promising classification outcomes [6]. These models generate semantic maps from an end-to-end structure that can directly process raw HSIs without any hand-crafted feature extraction step, whereas, the classification accuracy of the CNN models decreases with the increase of layers. The recent application of DL models indicates a new trend in utilizing features learned by models, rather than hand-crafted features, for HSI classification.

Since 2015, [7] proposed deep residual networks through connecting between every other convolutional layers for identity mapping and achieved state-of-the-art results for multiple computer vision tasks. Residual Networks can be regarded as an extension of Convolutional Neural Networks with skip connections that facilitate the propagation of gradients and performed robustly with very deep architecture.

In this paper, therefore, we applied and investigated deep learning models for HSI feature extraction and pixel-wise classification using two widely studied datasets. Three major contributions of this paper are: 1) assessing the influence of depth and width of ResNets for HSI classification accuracy; 2) validating the effectiveness of residual architectures and batch normalization strategy for mitigating the decreasing-accuracy phenomenon; and 3) visualizing the distribution of learned representations in 2D projected spaces through a embedding method.

2. RELATED WORK

Deep learning models are composed of multiple layers of nonlinear neurons that can learn hierarchical representation out of large amounts of labelled images [8]. CNN is the most popular supervised deep learning network at present and has shown its deep feature extraction power in computer vision contests [9]. Typically, CNN models include convolutional layers, pooling layers, fully connected layers, and multiclass logistic regression layers. The most prominent characteristic of CNNs in contrast to other DL models is their special convolutional structure, which imposes sparsity inherently and reduces the number of parameters significantly. The

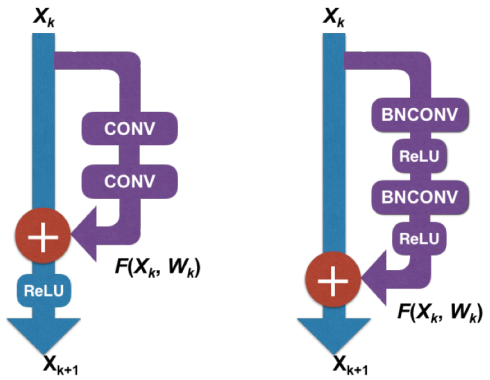

Fig. 1. Basic residual blocks. Residual architecture without batch normalization (left) and with batch normalization (right).

convolutional layers can be formulated as follows.

$$\mathbf{H}^{\mathrm{k}} = \mathrm{G}\left(\mathbf{F}^{\mathrm{k}-1} * \mathbf{H}^{\mathrm{k}} + \mathbf{b}^{\mathrm{k}}\right) \tag{1}$$

In Eqn. (1), $\mathbf{H}^{\mathrm{k}}$ represents the output of the kth layer in the model, $\mathbf{F}^{\mathrm{k}}$ is the kth convolutional filter banks, b^{k} denotes the bias of the kth layer, and $G(\bullet)$ is a rectified linear unit (ReLU). Given sufficient labelled data, CNNs can generate more accurate classification results than traditional machine learning methods.

To regularize and speed up the training process, batch normalization layers can be inserted into the deep learning models to impose Gaussian distribution on intermediate batch features [10]. This technique allows deep learning networks to converge smoothly to an acceptable local minimum and does not require a delicate initialization setting of parameters. The batch normalization is defined as follows.

$$\widehat{\mathbf{X}}^{(\mathrm{i})} = \frac{\mathbf{X}^{(\mathrm{i})} - \mathrm{E}(\mathbf{X}^{(\mathrm{i})})}{\mathrm{VAR}(\mathbf{X}^{(\mathrm{i})})} \tag{2}$$

In Eqn. (2), $\mathbf{X}^{(\mathrm{i})}$ denotes the ith dimension of feature batch $\mathbf{X}$, $\mathrm{E}(\bullet)$ represents the expected value and $\mathrm{VAR}(\bullet)$ is the variance of the features.

Dimensionality Reduction (DR) plays a significant role in HSI visualization and classification [11]. The t-SNE, which constructs probability distribution over similar samples and preserves local structure of high dimensional data, is a dimensionality reduction method that can embed the HSI data into a lower dimensional space. The t-SNE results of each iteration are carried out by minimization of Kullback-Leibler (KL) divergence, which can be written as follows.

$$KL(P||Q) = \sum_{i \neq j} p_{ij} \log\left(\frac{p_{ij}}{q_{ij}}\right) \tag{3}$$

In Eqn. (3), p_{ij} represents the similarity of the ith and the jth samples, and q_{ij} denotes the corresponding similarity in the projected feature space.

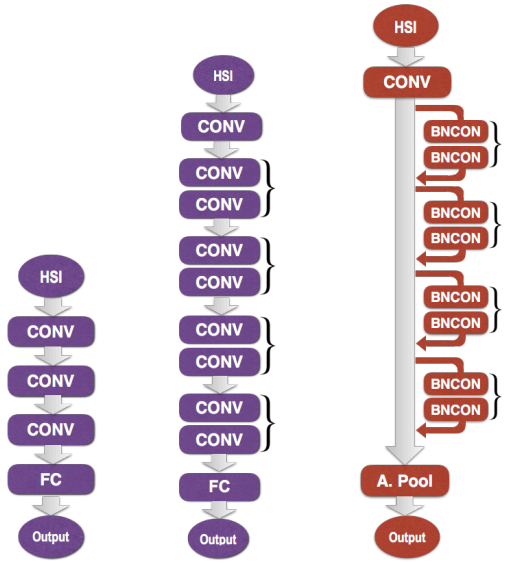

Fig. 2. Examples of CNN and ResNet models. CNN-4 (left), CNN-10 (middle), and ResNet-10 (right).

Table I. ARCHITECTURES OF DEEP LEARNING MODELS

	Convolutional Layers	Residual Blocks
Conv-4	3	
ResNet-4	1	1
Conv-6	5	
ResNet-6	1	2
Conv-8	7	
ResNet-8	1	3
Conv-10	9	
ResNet-10	1	4

3. DEEP RESIDUAL NETWORKS

Although CNN models have been used for HSI classification and achieved state-of-the-art results, it is counterintuitive that the classification accuracy decreases with the increase of convolutional layers after four or five stacked layers [6]. Inspired by the latest deep residual learning framework proposed in [7], this deteriorating issue can be solved by adding shortcut connections between every other layer and propagating the value of features. As shown in Fig. 1, the residual network can be formulated as follows.

$$\mathbf{X}_{\mathrm{k}+1} = \max\{0,\ \mathbf{X}_{\mathrm{k}} + \mathrm{F}(\mathbf{X}_{\mathrm{k}}, \mathbf{W}_{\mathrm{k}})\} \tag{4}$$

In Eqn. (4), X_k is the output of kth unit, W_k is denotes the parameters of the residual structure. The stacked nonlinear layers aim to construct the function $F(X_k, W_k)$ instead of mapping the desired X_{k+1} directly. Comparing to CNNs, the deep Residual Networks 1) are easier for optimization; 2) have more representative capacities; and 3) deliver higher recognition accuracy with deeper layers.

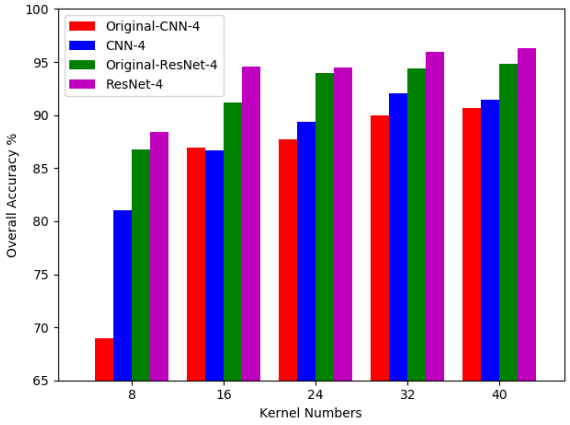

Fig. 3. Overall accuracy of CNN-4 and ResNet-4 with 8 kernels using IN dataset with or without regularization.

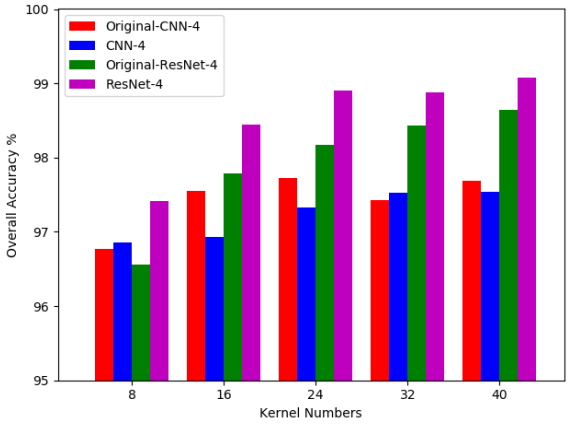

Fig. 4. Overall accuracy of CNN-4 and ResNet-4 with 8 kernels using UP dataset with or without regularization.

4. EXPERIMENTAL RESULTS

In this paper, as illustrated in Fig.2, we explored and evaluated ResNet models and their CNN couterparts with different numbers of convolutional kernels and different layers of depth for HSI classification by overall accuracy (OA) in two commonly used HSI datasets. Inspired by [12], we adopted the improved residual networks for HSI classification. Tabel I show ResNets and CNNs with kernel number of convolutional kernels from {8, 16, 24, 32, 40} and with layer range from {4, 6, 8, 10} that are employed.

We used the Indian Pines (IN) and University of Pavia dataset (UP) datasets for evaluating the ResNets and CNNs. The IN dataset includes 16 land cover classes of a vegetarian area, and the UP dataset contains 9 land cover categories of an urban area. In the IN dataset, we adopted 10%, 10%, and 80% of available annotated data for the learning, validation, and predicting of deep learning models, respectively. Similarly, in the UP dataset, we used 5%, 5%, and 90% for the same purposes. The input data of all deep learning models are 7×7×b HSI volumes, wherein b is the band number.

Table II. OVERALL ACCURACY OF IN DATASET WITH 10% DATA FOR TRAINING

	8	16	24	32	40
CNN-4	.81	.87	.89	.92	.91
CNN-6	.83	.86	.88	.88	.89
CNN-8	.77	.85	.88	.87	.91
CNN-10	.80	.84	.86	.86	.86
ResNet-4	.88	**.95**	.94	**.96**	**.96**
ResNet-6	.88	.94	**.94**	.95	.95
ResNet-8	**.90**	.92	.94	.93	.94
ResNet-10	.87	.90	.93	.93	.93

Table III. OVERALL ACCURACY OF UP DATASET WITH 5% DATA FOR TRAINING

	8	16	24	32	40
CNN-4	.97	.97	.97	.97	.98
CNN-6	.96	.96	.97	.97	.98
CNN-8	.96	.96	.97	.97	.97
CNN-10	.95	.97	.97	.98	.97
ResNet-4	.97	**.98**	**.99**	**.99**	**.99**
ResNet-6	.97	.98	.98	.99	.98
ResNet-8	**.97**	.98	.98	.98	.98
ResNet-10	.96	.98	.98	.98	.98

4.1. The Influence of Batch Normalization

As shown in Fig. 2, we applied a 10-layer ResNet and compared the model with the corresponding 10-layer CNN, the 4-layer CNN proposed in [6]. Given a small number of training data, we need regularization methods to prevent the learning process from overfitting. Thus, we trained CNNs and ResNets with or without batch normalization for each convolutional layers. As illustrated in Fig. 3-4, the deep learning models with regularization methods consistently generated higher classification accuracy than the original models.

4.2. The Influence of Width and Depth

To study the influence of model sizes, we tested ResNets and their corresponding CNNs with different width and depth. According to quantitative classification results in Table II-III, CNNs with deeper layer tend to have worse classification results. Moreover, the ResNets obtained superior classification performance to the CNNs. This means residual architectures alleviates the decreasing-accuracy phenomenon. Although this phenomenon still exists in ResNet models, the classification accuracies of which are more robust than their CNN counterparts.

4.3. Classification Maps of CNNs and ResNets

Fig. 5 and Fig. 6 illustrates the classification results of CNN-4 and ResNet-4 for IN and UP datasets, respectively. The classification maps of ResNets in both datasets are much smoother than those of CNNs.

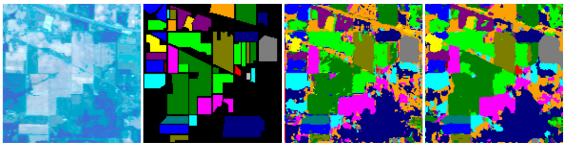

Fig. 5. IN dataset, ground truth image, and classification maps of CNN-4 and ResNet-4 with 32 kernels.

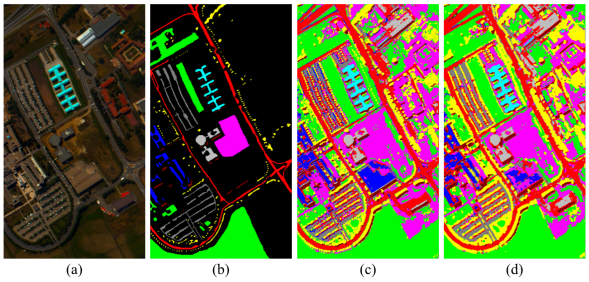

Fig. 6. UP dataset, ground truth image, and classification maps of CNN-4 and ResNet-4 with 32 kernels.

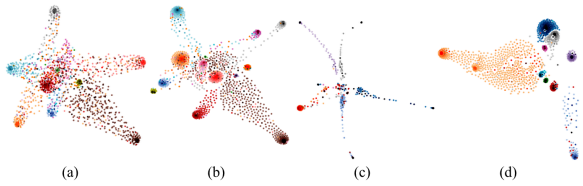

Fig. 7. Visualization of extracted features learned from deep learning models for IN and UP datasets. (a)-(d) t-SNE results of CNN-4 and ResNet-4 for IN and UP datasets.

4.4. HSI Visualization based on t-SNE

We employed t-SNE method to visualize the learned representative features of CNN-4 and ResNet-4 with 8 kernels for each convolutional layer. As shown in Fig. 7, the embedding of ResNets are more separable than CNNs. in both datasets. The perplexity for all four cases is 12. Within 1000 iteration, the t-SNE in all four cases converge to stable states.

5. CONCLUSION

In this paper, we have used deep ResNets with different depth and width for spectral-spatial classification using two commonly used HSI datasets with 10% and 5% of the labeled samples as training data for IN and UP datasets, respectively. According to the experimental reports, we can draw three major conclusions. First, batch normalization enhances the HSI interpretation performance of both CNNs and ResNets. Second, the ResNets achieved very competitive classification results and increase the accuracy of their corresponding CNNs. And third, the ResNets have alleviated but not fully overcome the decreasing-accuracy effect.

We have adopted t-SNE that projects the raw HSI data and extracted spectral-spatial features into a 2D plane to get a direct impression about feature representation. Figs. 7 shows that the proposed ResNets have learned a more discriminative representation of HSIs than those of CNNs. Moreover, the deep learning models with wider architectures tend to deliver higher classification accuracy under the same regularization methods, but the increases are not obvious when kernel number is larger than 24. It is worth noting that the ResNets with 4 layers perform the best in both HSI datasets, owing to small numbers of training samples and land cover categories.

6. REFERENCES

[1] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral Image Classification Using Dictionary-Based Sparse Representation," *IEEE Trans. Geosci. Remote Sens.,* vol. 49, no. 10, pp. 3973-3985, 2011.

[2] L. Zhang, Y. Zhong, B. Huang, J. Gong, and P. Li, "Dimensionality reduction based on clonal selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.,* vol. 45, no. 12, pp. 4172-4186, 2007.

[3] Y. Chen, X. Zhao, and X. Jia, "Spectral-spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 8, no. 6, pp. 2381-2392, 2015.

[4] Z. Zhong, J. Li, W. Cui, and H. Jiang, "Fully convolutional networks for building and road extraction: Preliminary results," in *proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 1591-1594.

[5] W. Zhao and S. Du, "Spectral–Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach," *IEEE Trans. Geosci. Remote Sens.,* vol. 54, no. 8, pp. 4544-4554, 2016.

[6] 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, 2016.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385,* 2015.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, no. 7553, pp. 436-444, 2015.

[10] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167,* 2015.

[11] D. Lunga and O. Ersoy, "Spherical stochastic neighbor embedding of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.,* vol. 51, no. 2, pp. 857-871, 2013.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," *arXiv preprint arXiv:1603.05027,* 2016.