A Deep Residual Neural Network for Low Altitude Remote Sensing Image Classification

Amin Fadaeddini
Department of Computer Engineering
Khatam University
Tehran, Iran
m.fadaodini@khatam.ac.ir

Mohammad Eshghi
Computer Engineering Department
Shahid Beheshti University
Tehran, Iran
m-eshghi@sbu.ac.ir

Babak Majidi
Department of Computer Engineering
Khatam University
Tehran, Iran
b.majidi@khatam.ac.ir

***Abstract*—In the recent years the applications of deep neural networks are increasing rapidly. There are two important factors determining the efficiency of training a computer vision system using deep neural networks. The first factor is the difficulty of training a very deep neural network with large number of parameters. The second factor is the efficiency of the trained network for decreasing the computational cost. In this paper an efficient deep neural network which uses the grid size reduction, factorization and hyper parameter tuning is proposed. In order to deal with large number of layers the residual units are used. A series of experimental simulations are performed on the application of the proposed deep neural network for classification of aerial images. The experimental results show that the proposed architecture has acceptable accuracy for aerial scene classification.**

***Index Terms*—Convolutional neural networks, Remote sensing, Image classification, Residual network**

本实验主要是采用残差网络构建新的结构，减少训练时间以此达到可接收的误差（并非改善现有算法的准确度）

I. Introduction

In the last decade many advances in both deep learning and remote sensing fields has been achieved. These achievements created a variety of research opportunities, namely scene classification, object detection, surveillance and structural inspection. A significant body of research in remote sensing is focused on image descriptors. These descriptors use low-level feature extractors such as Scale Invariant Feature Transform (SIFT) or local binary patterns (LBP). In recent years many studies has been devoted to the same problem using high-level feature extractors like Convolutional Neural Networks (CNN) [1] [2]. There are number of problems in training a neural network and one of these problems is that as the depth of the neural network increases the time and computational cost for the training process is also increases. This is one of the limiting factors for using the deep neural networks in various applications. In order to solve this problem an identity mapping named residual unit is proposed [3].

指出了深度学习的瓶颈所在

In this paper by using the residual units the classification of the remote sensing images achieved faster convergence rates.

978-1-5386-2836-2/18/$31.00 ©2018 IEEE

Compared with low-level methods, deep learning methods can learn more abstract and discriminative semantic features and achieve far better classification performances. The main contribution of this paper is proposing a novel and compact architecture for convolutional neural networks using a combination of residual units, factorization and grid size reduction. An important application of scene classification is low altitude aerial scene interpretation [4] [5] [6] [7]. The proposed architecture is able to classify the aerial images using reduced number of network layers.

The rest of this paper is organized as follows: Section II discusses some novel and related works in the context of remote sensing classification, Section III describes the architecture of the proposed deep neural network. Section IV presents the experimental results and Section IV concludes the paper.

II. Literature Review

In the last few years many research studies has been devoted to remote sensing and aerial image understanding. A number of these studies are related to classification [1] [2] and other studies are focusing on object detection and segmentation [8] [9]. Xie et al. [10] by extracting illumination in aerial images, maps the poverty of countries in order to provide better development in the less developed areas. Using the interpretation of the aerial images this study provides information such as food security data in the target areas.

In general, there are two way for training a remote sensing classifier. The first method is to train the network from scratch which results in better accuracy. The computational cost of this method is significantly high. The second method is by leveraging transfer learning. In transfer learning a pretrained network that was previously trained on a relatively similar input data is used. The feature maps from penultimate layers are then extracted and are used as feature vectors for the new network.

As an example of transfer learning Castelluccio et al. [1] used pretrained ConvNet with fine-tuning for high accuracy remote sensing image classification. Xia et al. [2] introduced a novel dataset and a benchmark for remote sensing classification evaluation with 10,000 samples and 30 class labels with high intra-class similarities and high inter-class dissimilarities. It is shown that training aerial images with this dataset produces higher accuracy and far better generalization on unseen samples. In terms of object detection many researches have been performed on Structural Health Monitoring (SHM). Cha et al. [8] investigated a framework for detecting civil infrastructure defects in order to partially replace human-conducted on-site inspections. In that paper the aerial images are used for detection of the cracked concrete surfaces of dams.

III. The Deep Residual Neural Network for Aerial Image Classification

The architecture of the proposed residual neural network is presented in Fig. 1. The proposed architecture consists of sixteen layers. From these sixteen layers thirteen layers are convolution layers which are followed by a max pooling layer to reduce the spatial dimension of the feature maps. The penultimate layer in the proposed architecture is a fully connected neural network with 1024 neurons and finally a Softmax classifier concludes the proposed architecture. In order to test the efficiency of the proposed network and to investigate whether it can produce acceptable results in few epochs, the network is trained in only 20 epochs. In the proposed architecture the convolution layers use Rectified Linear Unit (RLU) as activation function to transform the linearity of weighted sum into non-linearity. The proposed deep neural network architecture uses some techniques in order to increase the performance of the network. The first technique is the factorization of the kernels into smaller convolutions. Convolution that uses large filter size for example 5×5, 7×7 or higher, have significant computational cost. Based on Szegedy et al. [11], it is more efficient to transform a larger kernel into a series of smaller kernels. In the proposed architecture the 5×5 kernel is transformed into two 3×3 kernels as follows:

$$Depth_{n-1} \times (5 \times 5 \times Depth_n) = 25 Depth_{n-1} Depth_n \quad (1)$$

$$2 \times Depth_{n-1} \times (3 \times 3 \times Depth_n) = 18 Depth_{n-1} Depth_n \quad (2)$$

Also as suggested by [11] the factorization before 7th convolution is not performed due to the fact that the model will generate better result by using this factorization in medium grid size ranging between 12 and 20.

The second technique used in the proposed architecture is efficient grid size reduction. In order to make the proposed architecture more efficient a technique called asymmetric convolutions is used. In this method the 3×3 kernels are transformed to a 3×1 kernel followed by a 1×3 kernel. In addition to this transform and factorization of the larger filter size 1×1 kernels with half of the previous kernel depth in some layers are used. Using this method, the computational cost decreases rapidly.

Krizhevsky et al. [12] proposed a deep network which is trained on Imagenet dataset containing 15 million images of various categories. They proposed training the network on multiple GPUs, using local response normalization and overlapping pooling. In this paper overlapping pooling with grid size of 3 pixels and striding of 2 pixels is used to reduce overfitting. 这不就在调参吗？

The final technique used in the proposed architecture is the residual units. As the number of layers in the deep neural networks increases the accuracy of the predicted model is also increases. The residual units give the deep neural network the ability to converge with acceptable computational costs while having a large number of layers. Residual units are proposed by He et al. [3] and in the proposed model, as shown in (3), instead of expecting that each layer of the deep neural network directly fit to a desired underlying mapping, we use an identity mapping. This process makes the training time faster with better results. In the proposed architecture the dropout with keep probability of 80 percent and He initialization [13] is used in order to avoid overfitting and vanishing and exploding gradient problems.

$$F(x) = H(x) - x \quad (3)$$

We also used dropout with keep probabilities of 80 percent to reduce overfitting.

IV. Experimental Results

One of the the most commonly used datasets for remote sensing imagery is UC Merced dataset [1]. This dataset is provided by the US Geological Survey, taken over various regions of the United States. This dataset consists of RGB images with the size of 256×256. This dataset consists of 2,100 samples that belong to 21 label classes comprise of agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis courts. Examples of the images provided by this dataset is presented in Fig. 2.

We have used 10 percent of the dataset samples for test set and 90 percent of remaining samples for training. The results of the preprocessing and hyper parameter tuning on the proposed architecture is presented in Fig. 3. Using subtracting method in our input data the network converges faster compare to not having any kind of preprocessing methods. The different initialization of weights by comparing He initialization and Xavier initialization are investigated and the results are shown in Fig. 3. As shown in Fig. 4 in every 2 epochs the trained network is validated on the test set. In Fig. 4 the blue line shows the test set results and the red line shows the training accuracy. The overall accuracy for the proposed model is 78 percent.

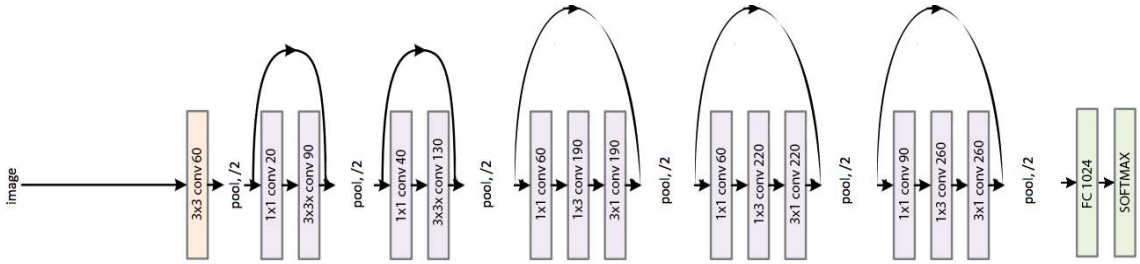

Fig. 1. The proposed deep residual neural network for aerial image classification

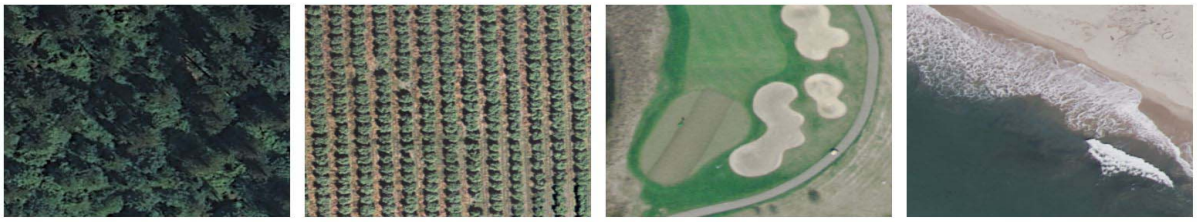

Fig. 2. The samples from UC Merced dataset

Although the accuracy of the proposed model is less than the state of the art results, the number of iterations and the training cost is significantly lower than the networks with higher accuracy. Table 1 compares the proposed model with similar models. The Nvidia GeForce GTX 750 Ti graphic card is used for implementation of the proposed model.

CONCLUSION

In this paper an efficient deep network based on grid size reduction, factorization and hyper parameter tuning is proposed. The proposed deep neural network uses residual unit to achieve better accuracy. A series of experimental simulations are performed on the application of the proposed deep neural network for classification of aerial images. The experimental results show that the proposed architecture has the accuracy of 78 percent for aerial scene classification in the UC Merced dataset. The next stage of this project is to train the proposed network on the newly released Aerial Benchmark [2] which contains 10,000 samples in parallelized GPUs implementation to increase the models accuracy.

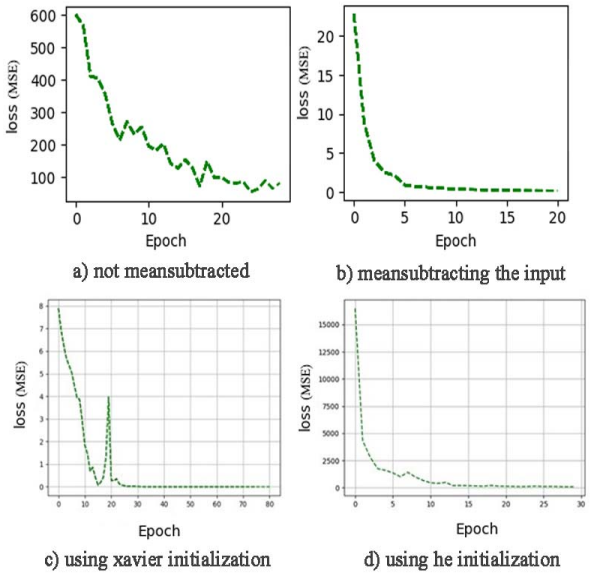

Fig. 3. Preprocessing and hyper parameter tuning on the proposed architecture

TABLE I
COMPARISON WITH OTHER CLASSIFIER

Model	Accuracy (percent)	Epochs
CaffeNet from scratch	85.71	100,000
CaffeNet transfer learning	95.48	20,000
GoogLeNet from scratch	92.86	100,000
GoogLeNet transfer learning	97.10	20,000
Proposed Architecture	78	20

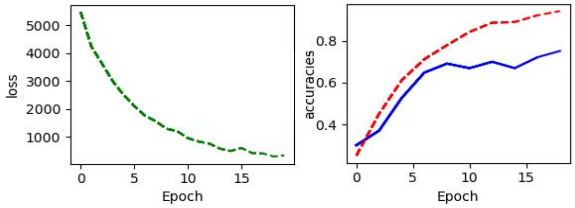

Fig. 4. Training result

REFERENCES

[1] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land use classification in remote sensing images by convolutional neural networks," arXiv preprint arXiv:1508.00092, 2015.

[2] G.S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, "AID: A benchmark data set for performance evaluation of aerial scene classification," in IEEE Transactions on Geoscience and Remote Sensing, 2017, pp. 3965-3981.

[3] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778.

[4] B. Majidi, J. C. Patra, and J. Zheng, "Modular interpretation of low altitude aerial images of non-urban environment," Digital Signal Processing, vol. 26, pp. 127-141, 2014/03/01/ 2014

[5] B. Majidi and A. Bab-Hadiashar, "Real time aerial natural image interpretation for autonomous ranger drone navigation," in Digital Image Computing: Techniques and Applications (DICTA'05), 2005, pp. 65-65.

[6] B. Majidi and A. Bab-Hadiashar, "Land cover boundary extraction in rural aerial videos," 2007, pp. 311-314.

[7] I. evo and A. Avramovi, "Convolutional neural network based automatic object detection on aerial images," IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 5, pp. 740-744, 2016.

[8] Y. J. Cha, W. Choi, and O. Bykztrk, "Deep learning-based crack damage detection using convolutional neural networks," ComputerAided Civil and Infrastructure Engineering, vol. 32, no. 5, pp. 361-378, 2017.

[9] M. Shi, F. Xie, Y. Zi, and J. Yin, "Cloud detection of remote sensing images by deep learning," in 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2016, pp. 701-704.

[10] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon, "Transfer learning from deep features for remote sensing and poverty mapping," arXiv preprint arXiv:1510.00098, 2015.

[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818-2826.

[12] 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.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1026-1034.