

Review of Week 3

Xiaoke(Jimmy) Shen

September 22, 2017

1 Introduction

This paper review report is provided based on the requirement of the computer science research course provided by the Graduate Center of the City University of New York during the 2017 Fall Semester. The main objective of this course is helping the PHD candidates to prepare their second exam and identify their thesis topic as early as possible [1].

2 Object Classification for 2D Images

2.1 ImageNet Classification with Deep Convolutional Neural Networks [2]

In the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry [2]. The outstanding performance of the deep neural network used in this paper brought the focus of both the academic and industry back to the neural network again. In this paper, the CNN network is used and the CNN network structure is illustrated in the Figure 1. Two GPUs were used to speed up the calculation. Dropout [3] was used here and was proved to be effective to reduce the overfitting problem. The cons of this network is it is a bit complicated and the structure is not so elegant and it will be addressed in the future works by the deep learning researchers.

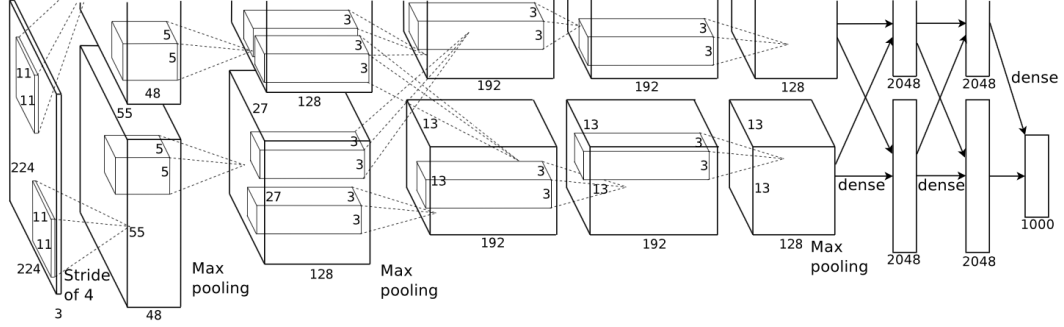


Figure 1: An illustration of the architecture of Alex Net, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253, 440-186, 624-64, 896-64, 896-43, 264-4096-4096-1000 [2].

2.2 Very Deep Convolutional Networks for Large-Scale Image Recognition [4]

One important contribution of the VGG network is the structure of this network is very tidy and elegant. Different to the Alex Net [2] with different size of the convolutional kernels, in the VGG CNN network [4], only size of 3 by 3 kernels were used for the whole network. At the same time, the trained network weights based on the ImageNet [5] dataset were shared to public. The cons of this model is there are too many parameters and the computation took a long time. Weeks for doing the ImageNet [5] dataset by using a powerful GPU.

This elegant design can also be used in the 3D CNN network to address the 3D object classification problem which is my current working topic.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 2: An illustration of the VGG network structure.

2.3 Deep Residual Learning for Image Recognition [6]

In motivation of this work is trying to find a deeper neural network structure to achieve better performance as the deepest network prior this work is around 20 to 30 layers. The authors had a basic assumption that if more layers are added, the performance should be at least the same as the less layer networks as the new added layers can be designed as the identify layer then the performance can be at least the same. However, the traditional network's performance decreases when the layers increases to a higher level as shown

in Figure 3. In order to address this anti-intuitive problem, the authors of this paper designed a deep residual network by adding a short cut between every other layers and finally it shows that this network can achieve a better performance. The problem is addressed. In this paper, the layer of the neural network can go to 110 layer and have a better performance than the previous state-of-the-art neural network such as VGG16 [4], which only has 16 layers neural network. However, as mentioned in [6], the parameters used for the 110 layers Resnet is even less than the 16 layers VGG16 network. This seems amazing. However, by reading this article carefully, the main contribution of the reduction of the parameters was using the convolutional network layer instead of the fully connected layer. There is no contribution on the parameters reduction by the Resnet itself.

Of course, the Resnet structure used for the 2D images can be further studied to see whether can also be used for the 3D cloud point data.

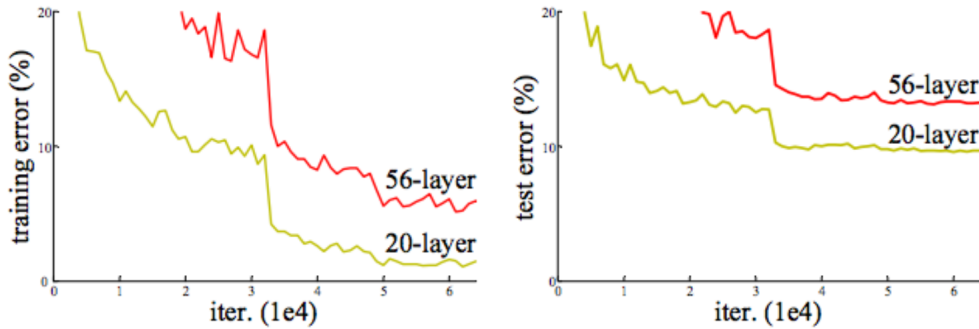


Figure 3: An illustration of deep neural networks fails by using the traditional network structure

References

- [1] P. Ji, “Syllabus of the computer science research at cuny fall 2017,” 2017.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012.

- [3] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” *CoRR*, vol. abs/1207.0580, 2012.
- [4] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *CoRR*, vol. abs/1409.1556, 2014.
- [5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” in *CVPR09*, 2009.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *CoRR*, vol. abs/1512.03385, 2015.