Paper Review for the Computer Science Research Course and the Second Exam

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].
In this report the papers related to the object classification, object detection and object semantic segmentation for the 2D and 3D objects will be discussed. As most of the state of the art algorithms used for those tasks are based on the deep convolutional neural networks, the important papers related to the deep neural networks will also be discussed in this article.

The structure of this article is described bellow: The papers related to the theory part of the deep learning will be discussed in the second session. In the third session, the papers in the 2D object classification will be discussed. In session 4, the 2D object detection papers will be studied. At the same time, an interesting and more challenge related to the 2D image procession or computer vision will be discussed in session 5 which is the 2D object semantic segmentation. In the rest part of this article, the similar tasks in 3D will be discussed as the final goal of this paper review is finding some possible approaches to improve the current 3D computer vision algorithms based on the state of the are 2D computer vision algorithms.

2 Deep Learning Theory

2.1 Approximation with Artificial Neural Networks [2]

In order to build the mathematical theory of the artificial neural networks, several papers are published in the 20 century. In this paper one main contribution is the universal approximation theorem with proof. The universal approximation theorem claims [2] that the standard multilayer feed-forward networks with a single hidden layer that contains finite number of hidden neurons, and with arbitrary activation function are universal approximators in $C(R^m)$. The universal approximation theorem is one of the important theoretical support for the artificial neural networks. However, at that time as the huge size labeled data sets are not available, the updated algorithms haven't been invented and also the limited computation power, these ideas can not be verified.

2.2 Approximation Capabilities of Multilayer Feedforward Networks [3]

The unique value of the Kurt Hornik (1991) [3] paper is it showed that it is not the specific choice of the activation function, but rather the multilayer feedforward architecture itself which gives neural networks the potential of being universal approximators. This is an important contribution which is the foundation for the current state of the art deep learning architecture such as VGG 16 [4] and resnet [5]

2.3 Learning representations by back-propagating errors [6]

The paper of 1986 significantly contributed to the popularisation of BP(Back Propagation) for NNs [6], experimentally demonstrating the emergence of useful internal representations in hidden layers. The Back Propagation algorithm is one of the most critical and fundamental algorithm used in the deep neural network. This paper will be read in the future.

2.4 Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift [7]

The main contribution of this paper as shown in Figure 1 is it greatly reduced the convergence time for the training process and the BN(Batch Normalization) become one of the standard training step for the deep neural network after the publish of this paper.

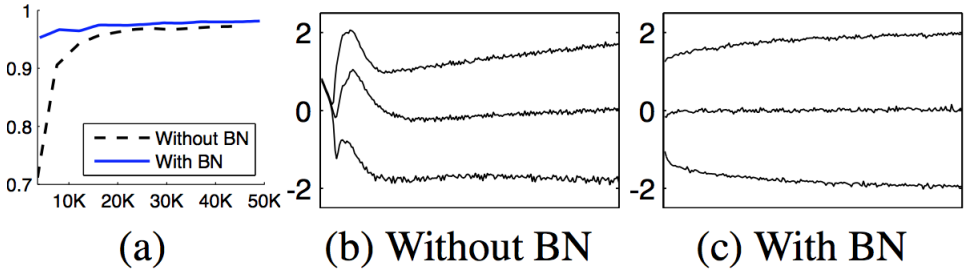

Figure 1: (a) The test accuracy of the MNIST network trained with and without Batch Normalization, vs. the number of training steps. Batch Normalization helps the network train faster and achieve higher accuracy [7]. (b, c) The evolution of input distributions to a typical sigmoid, over the course of training, shown as 15, 50, 85th percentiles. Batch Normalization makes the distribution more stable and reduces the internal covariate shift [7].

3 Object Classification for 2D Images

3.1 Backpropagation applied to handwritten zip code recognition [8]

The first important application of using the BP(Back Propagation) to well resolve the real life problem from the literature. From this paper, one important structure of the neural network as show in Figure 2 including the layers with filters were introduced. The similar structure is used in the modern neural network structures such as Alex Net [9], VGG 16 [4] and resnet [5]. The basic idea of the convolutional neural network was also introduced here.

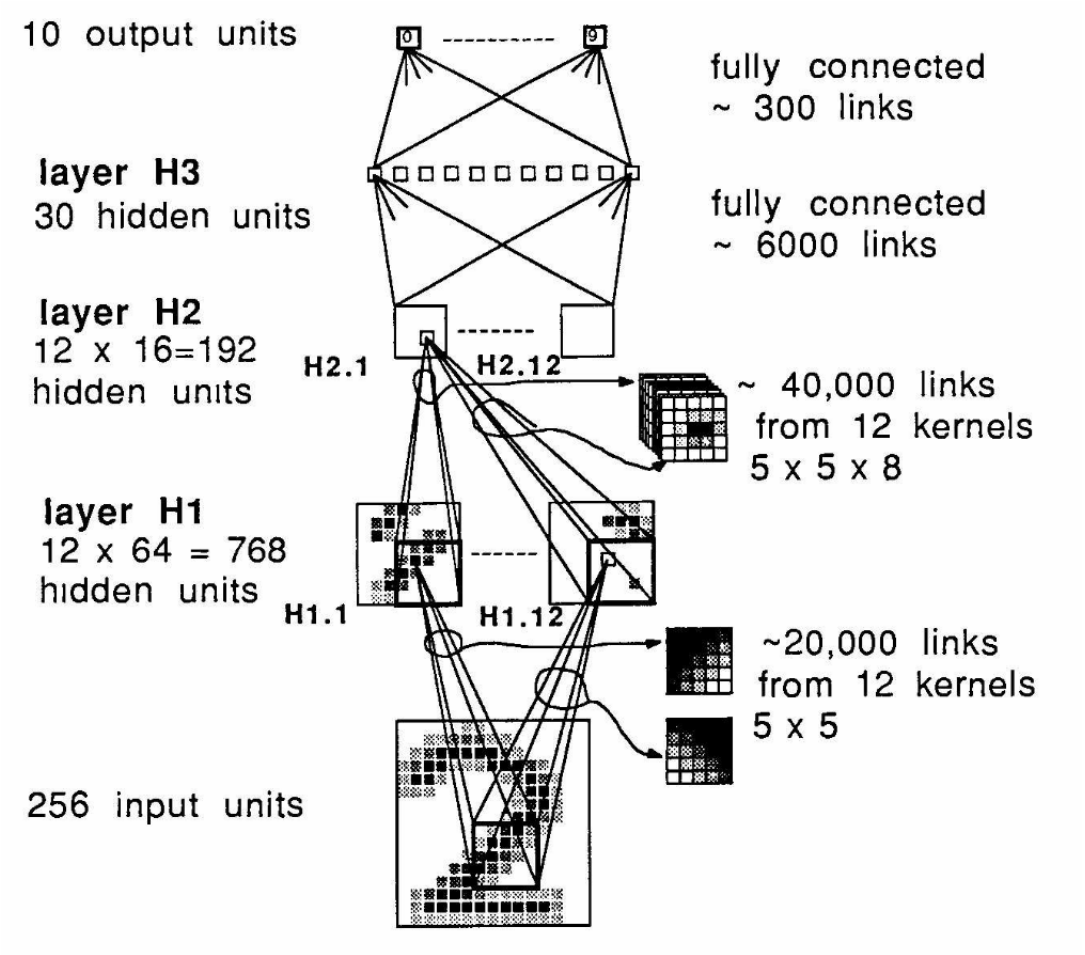

Figure 2: The Neural Network used in [8].

3.2 ImageNet: A Large-Scale Hierarchical Image Database [10]

In the machine learning research area, two kinds of learning approaches can be done: supervised learning and unsupervised learning. For the supervised learning algorithms, the labeled data is required to train the algorithm. So the availability of the labeled data will become very important to the development of the supervised learning based algorithms. The ImageNet dataset [10] provides 1.2 million high-resolution labeled images of 1000 categories. This dataset becomes one of the most important datasets related to the object classification.

3.3 ImageNet Classification with Deep Convolutional Neural Networks [9]

To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" [11] that proved to be very effective. We also entered a variant of this model 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.
The CNN network structure is illustrated in the Figure 3

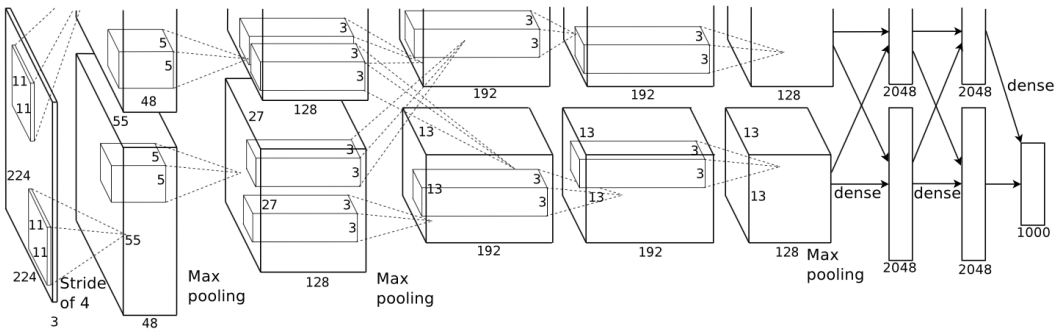

Figure 3: An illustration of the architecture of Alex Net, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network?s 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 [9].

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

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 4: An illustration of the VGG network structure.

3.5 Deep Residual Learning for Image Recognition [5]

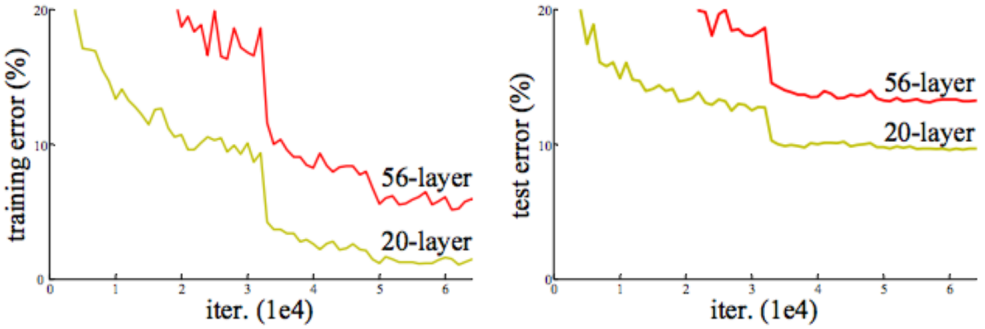

Figure 5: An illustration of deep neural networks.

4 Object Detection for 2D Images

4.1 Microsoft COCO: Common Objects in Context [12]

4.2 Rich feature hierarchies for accurate object detection and semantic segmentation [13]

4.3 Fast R-CNN [14]

4.4 Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [15]

4.5 You Only Look Once: Unified, Real-Time Object Detection [16]

4.6 YOLO9000: Better, Faster, Stronger [17]

References

[1] P. Ji, "Syllabus of the computer science research at cuny fall 2017," 2017.

[2] B. C. Csji, "Approximation with artificial neural networks," pp. 11–12, 2001.

[3] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Networks*, vol. vol. 4, 1991.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[6] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 10 1986.

[7] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015.

[8] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[9] 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.

[10] 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.

[11] 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.

[12] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: common objects in context," *CoRR*, vol. abs/1405.0312, 2014.

[13] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *CoRR*, vol. abs/1311.2524, 2013.

[14] R. B. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pp. 1440–1448, IEEE Computer Society, 2015.

[15] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada* (C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, eds.), pp. 91–99, 2015.

[16] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *CoRR*, vol. abs/1506.02640, 2015.

[17] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," *CoRR*, vol. abs/1612.08242, 2016.