
CS 594 Project 2 Proposal: Representation and Regularization Study of Deep and Wide Neural Network for Image Recognition

Yanzi Jin

yjin25z@uic.edu

Yaru Shi

yshi31@uic.edu

We are going to reproduce part of the experiment of Deep Residual Network (resNet) [1] and wide Residual Network (WRN) [6] for image recognition, as well as some follow-up experiments in terms of representation and regularization when training deep neural network. Compared to other framework like GoogLeNet [5], resNet has much simpler architecture, and achieves the state-of-the-art performance. It won the first place of ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation by increasing the network depth to 152 layers on ImageNet. Furthermore, according to the author's recent talk, they are able to achieve 1000 layers on CIFAR-10 dataset. On the other hand, the author of WRN claims that a simple 16-layer-deep wide residual network outperforms in accuracy and efficiency all previous deep residual networks, including thousand-layer-deep networks. Both methods have straightforward architecture, we believe it a good framework to explore the details in deep neural network.

In the first part of our project, we are going to compare “plain” neural network, deep residual network and wide residual network using the parameters provided in above papers. The expected results will be consistent with the papers that deep residual network is better than “plain” network, while wide residual network achieves similar accuracy with much deeper residual network. The results will be displayed as training and testing error plot along with time. In the second part, we are going to play with the parameters of deep and wide residual network. Borrowing the two terms used in [6], there are deepening factor l and widening factor k , where deepening factor l is the number of convolutions in a block and k multiplies the number of features in convolution layers. Since the l is set to 2 in both papers, we are going to increase the number of convolution layers between shortcuts. In the last part, we are going to study the effect of dropout. Inspired by [2] and [3], we are going to integrate the stochastic depth dropout in the resNet and propose a stochastic width dropout scheme. We expect that the dropout will help with generalization, however, may not be trivial on small dataset.

Given the well implemented code on Github, we will study the code and implement the second and third part. Due to the large size of ImageNet dataset, we will mainly work on CIFAR-10 dataset [4], which contains 60000 labeled for 10 classes images 32×32 in size, training set has 50000 and test set 10000. The experiments will run on a NVIDIA TitanX GPU and the maximal network depth would roughly be 20.

References

- [1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.
- [2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. *arXiv preprint arXiv:1603.05027*, 2016.
- [3] Gao Huang, Yu Sun, Zhuang Liu, Daniel Sedra, and Kilian Q. Weinberger. Deep networks with stochastic depth. *CoRR*, abs/1603.09382, 2016.
- [4] Alex Krizhevsky. Learning multiple layers of features from tiny images. 2009.

- [5] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.
- [6] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.