ECE 570 Paper

Fall 2019

1 Introduction

This paper will be implementing a deep residual network on the MNIST and CIFAR10 datasets. This paper will also discuss 3 research papers related to deep learning.

2 Research Papers

2.1 Deep Residual Learning for Image Recognition

The first paper that was chosen to be read is Deep Residual Learning for Image Recognition. The project associated with this paper will be implementing deep residual learning on MNIST and CIFAR10 datasets. This is an important in the field of deep learning as it gave way to a new way to do neural networks. The paper is from 2016 and won the 1st place on the ILSVRC 2015 classification task.

In the paper the authors primarily used residual nets with a depth of 152 layers on the ImageNet dataset. This paper mainly studies the vanishing gradient problem faced when a neural network has many layers and crafts a solution to allow for having ultra-deep networks. It has become known that the network depth is important for good results. When the authors of this paper were testing this, they found out that the 20-layer network produced better results than the 56-layer deeper network. This experiment carried out on the CIFAR10 dataset shows that blindly adding layers to the network to make it deeper does not necessarily guarantee better results. The authors of the paper show that this problem with deep network arises because of vanishing gradients.

The solution proposed in this paper is the use of residual connections along with regular connection between layers to guarantee better results with deep networks when compared to their shallower counterparts. There are two types of residual connections, firstly the identity shortcuts are used when the input and output layers have the same dimension. Secondly, when the dimensions of layers increase, a shortcut with extra padded zeroes is used for increasing dimensions or a projection shortcut is used to match dimensions. It is claimed that the usage of residual blocks help because they are able to bring in information in not only from the previous layer but also the layers before that which allows the network to have better context to counter information loss that comes with using many layers. *Figure 1* shows the building block of residual learning[1].

2.2 Understanding deep learning requires rethinking generalization

This ICLR 2017 submission tries to challenge the way we think about and understand deep learning.

The authors of the paper list their central findings as follows [3]:

- Deep neural networks easily fit random labels.
- Explicit regularization may improve generalization performance, but is neither necessary nor by itself sufficient for controlling generalization error.

The authors of this paper discovered that deep neural networks can effortlessly fit random labels, in other words, neural networks achieve zero training error when they are trained on random labeling of true data. By just randomizing the labels the authors of this paper found out that the generalization error (difference between training and test error)

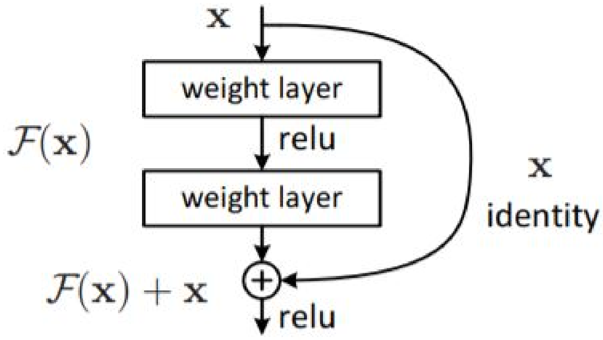

Figure 1: The Building Block of Residual Learning [1]

climbs up substantially. Secondly , the paper shows that the explicit forms of regularization do not amply explain the generalization error of networks[3].

There are a considerable number of people who believe that this paper is a good read for getting started with generalization and regularization, and they feel that this paper is interesting. Some people in the deep learning community on the other hand argue that the results in this paper are un-surprising/obvious and there are also some people who feel that the claims in the paper are not adequately backed up. Me being new to machine learning have very little to critique about this paper and I personally thought it was a good read.

2.3 Going Deeper With Convolutions

This paper proposes a new architecture for deep convolutional neural networks named Inception. It was state of the art for classification and detection algorithm in ILSVRC 2014 (ImageNet Large-Scale Visual Recognition Challenge). This paper focuses on the improved usage of computing resources in the network. This is achieved by allowing for increasing depth and width of the network. The author of this paper used Hebbian principle and intuition of multi scale processing to optimize quality. The GoogLeNet is also a type of network used in the ILSVRC submission[2].

The paper discusses the drawbacks of uniformly increased network size on drastically increased computational resource requirement. It discusses how the Deep Vision's Network increase in filter size increases computation requirement in a quadratic manner. The papers propose the move from fully connected to sparsely connected networks to solve the drawbacks of deep neural networks. The architecture optimizes for computing efficiency by approximating sparse structures by using the readily available dense blocks. The advantage of this is a notable quality gain at a reasonable increase in computational requirement. The inception network used 1x1 convolutions to compute reductions before the expensive 3x3 and 5x5 convolutions. The inception network consists of modules of this type stacked upon each other. The depiction of the inception modules is shown in *Figure 2*[2].

Even though this paper influenced deep learning in a very positive way, some believe that the design decision taken by the authors are somewhat arbitrary. I personally think that this paper is very important because it showed that good deep neural networks don't have to be extremely demanding of computational resources and, I thought the paper was very readable even though it was somewhat hard to follow at some places.

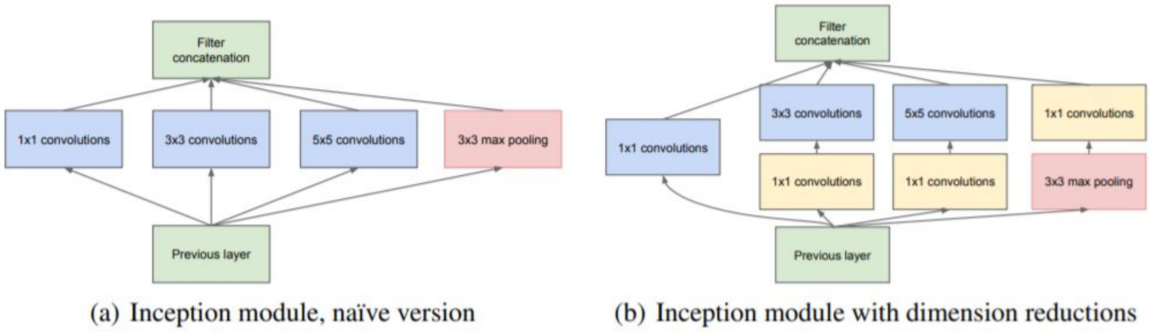

Figure 2: Inception Module.[2]

3 Implementation(Deep Residual Learning)

The code of this project is a re-implementation of the idea in section 4.2 of the research paper which are the ideas discussed in section 2.1 of this paper. The authors of the paper say that their focus in this section was study the behavior of extremely deep networks and to not push the state of the art results.

This implementation was converted to also work with the MNIST dataset, this was added to the implementation by making changes to number of channels being considered (from 3 to 1) and resizing the images in the data set by padding 2 pixels on each side to convert the 28x28 MNIST image to the 32x32 CIFAR10 format. This yielded good results as discussed in the testing section below.

4 Testing

The code had both the training and testing features. 5 epochs were run on the MNIST data set and the accuracy was already 98% after the first epoch. 80 epochs were run on the slightly heavier 3 channeled CIFAR10 data set and an accuracy of 86.63% was achieved. The loss and accuracy functions were plotted for each data set and observations were made. Please note that all the plots are added just after the conclusion section in this paper.

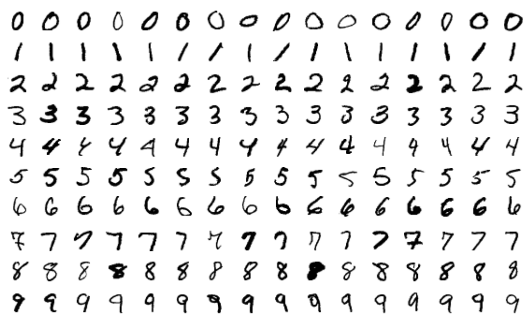

Figure 3: MNSIT Example Images.

4.1 MNIST Dataset

The MNIST data set is a large well known data set containing 60,000 training and 10,000 testing grayscale images of hand written digits(0-9). The loss vs epoch graph (*Figure 7*) for the MNIST data set was almost useless because the accuracy was high since the first epoch, there are very small changes to loss in case of the MNIST data but we do observe a slight increase in accuracy from 98.2% to 99.3% and this can be observed from the shape of the accuracy vs epoch graph (*Figure 5*). The loss and accuracy graphs are given below.

4.2 CIFAR10 Dataset

The CIFAR10 data set is another well known data set containing ten classes of objects and is commonly used in image recognizing tasks. This data set contains 60,000 images to train and test with. The loss vs epoch graph(*Figure*

8) is very meaningful here, we see a decline in loss as the epochs increase and we see a clear increase in accuracy ranging from less than 60% to 86%(*Figure 6*). I believe the accuracy would have increased even further but I was limited by computational resources in the testing phase of this project.

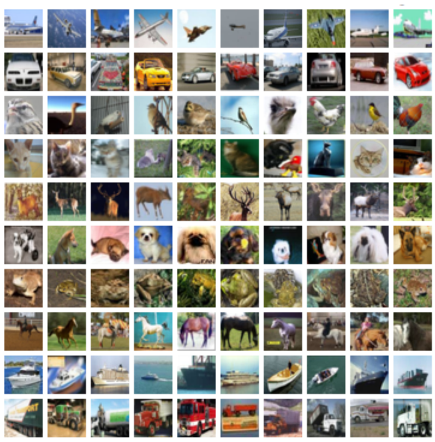

Figure 4: CIFAR10 Example Images.

5 Conclusion

Residual networks do require a substantial amount of memory to operate but provide a good solution to the vanishing/exploding gradient problem. The usage of residual networks guarantees deeper networks to perform better than their shallower counterparts making this network an important player in the deep learning.

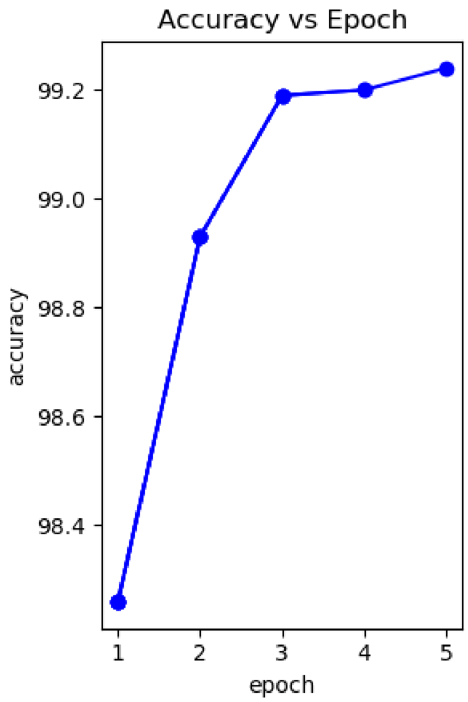

Figure 5: Accuracy each epoch on MNIST Dataset

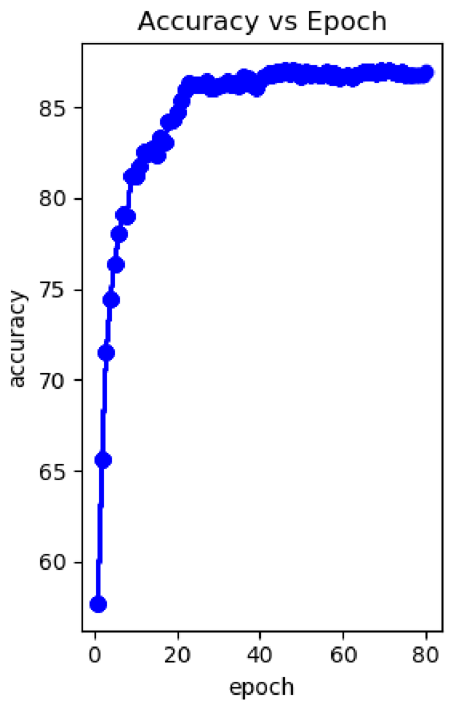

Figure 6: Accuracy each epoch on CIFAR10 Dataset

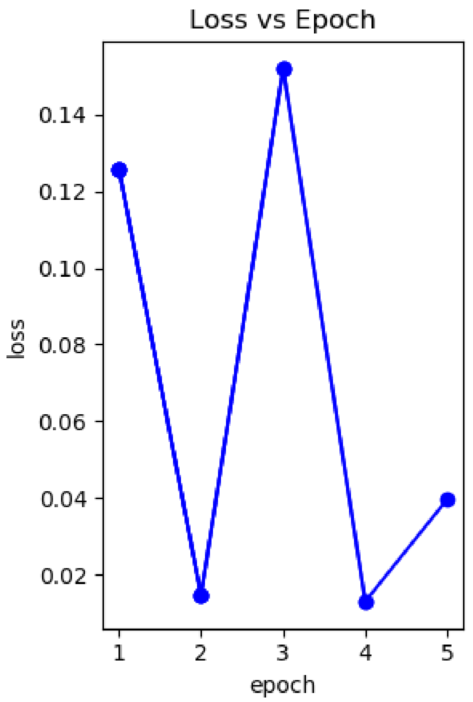

Figure 7: Loss each epoch on MNIST Dataset

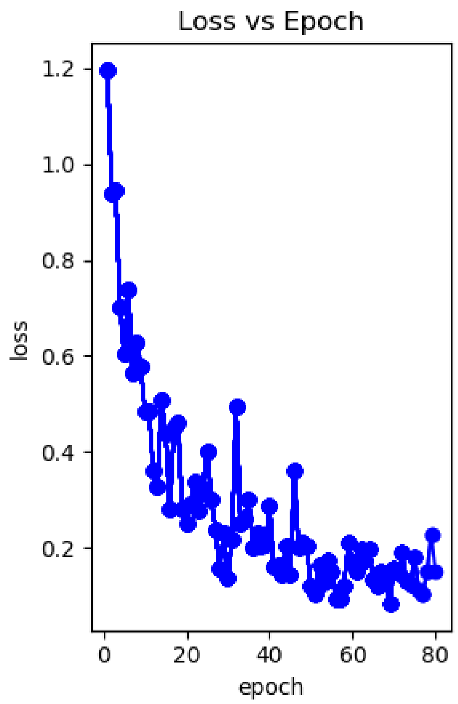

Figure 8: Loss each epoch on CIFAR10 Dataset

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, June 2016. 27785 citations.

[2] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, June 2015. 15281 citations.

[3] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization. *CoRR*, abs/1611.03530, 2016.