Deep Learning CSCE 636 Project

Chinmay Ajit Sawkar
Department of Computer Science
Texas A & M University
College Station, TX
chinmay997@tamu.edu

Abstract

Image classification task can be achieved by using Convolution Neural networks like Resnet, VGG, GoogleNet. In this project I present a different approach to achieve better accuracy using the modified DenseNet with multiple dense and transition layers along with using bottleneck and reduction. The DenseNet improves on the Resnet approach where instead of using identity mapping for gradient propagation, inputs from all of previous layers are concatenated and given to current layer and it further gives output to all succeeding layers. This approach has shown to help improve the gradient flow, allowing the network to be compact and thinner, improving the computational efficiency. Multiple experiments were done with adding different image augmentations, scheduling, optimizer and using dropouts to achieve a image classification accuracy of 94.66% on the public CIFAR-10 test data.

1 Introduction

Using deep neural networks for Image Classification has been a well research topic in the past decade. Many different approaches have evolved in the last decade like VGG, GoogleNet, Resnet, etc. These approaches stack multiple layers of convolution layers, and try to identify different features in the image. The approach used in Resnet is that we use a identity mapping for propagating the signals front and back between blocks more directly in the network. The advantage this provides is that it makes the training easier and efficient, also improves generalization as gradient flow has increased. Improving on this idea in this project I used DenseNet approach, which uses residual connections from all previous and to all succeeding layers and provides greater training efficiency thereby reducing the width of layers and also improving generalization.

2 Data and Image Augmentation

2.1 Data

The CIFAR-10 data was used to train and test the image classification problem. The model was trained on publicly available training data of 50000 samples and tested on publicly available test data with 10000 samples. Finally the best model was used to predict the private test data. The CIFAR=10 data set consist of 10 classes.

2.2 Image Augmentation

In this project I apply image augmentation to the training data where I selectively crop some of the images from different parts of the image, this will ensure that if some test images are zoomed our model can learn to recognise such images. I also add random degree of rotation in the image so that even if the object in image is tilted the mode tries to generalise that by learning different orientations.

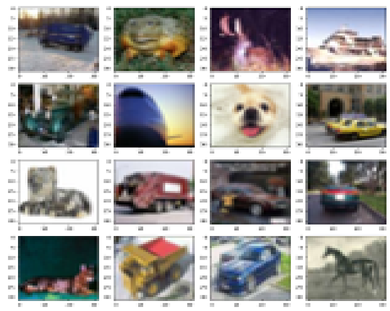

Figure 1: Random CIFAR-10 samples

Finally, I flip some images horizontally by randomly selecting with a probability of 0.25. Given below is a image produced after some image augmentations.

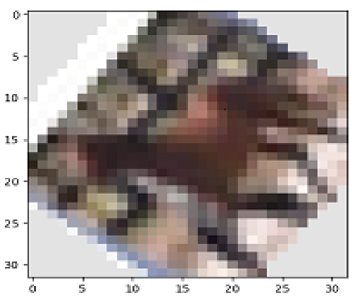

Figure 2: An example of augmented image

3 Model Architecture and Implementation

The DenseNet architecture is such that there are dense blocks which have interconnected residual connections within a block and there are transition layers between dense layers to reduce the feature map size. There are three Dense layers in my model and after each dense layer there is a transition layer which is basically a layer where I add Batch Normalization, ReLU activation, 1x1 convolution for reduction and then finally apply 2x2 average pooling so that the final output is size of the input of original layer. This transition layer downsamples the output of the dense layer thus saving computational cost as the number of parameters are reduced.

In each Dense block the number of layers added are input size time the growth rate. The DenseNet has a depth of 120 and the growth rate is set to 16. This was done after running different experiments and due to computational limitations.

In this I have applied a modified version of DenseNet where reduction and dropout are used for regularization. In this case I have used a compression or reduction of 0.6 in the transition layer. This reduces the feature map size by 60% from the previous Dense layer output feature map size. The dropout rate used is 0.2. This is also used a means of regularization as dropout randomly sets some of the elements of the input tensor to zero based on the random probability generated.

The optimizer used was Stochastic Gradient Decent (SGD) with momentum to push the gradients vectors in the correct directions, which makes the model converge faster. The weight decay used was 0.0001. The learning rate was not kept very low in the start as keeping the learning rate very low in start will not allow us to reach the global minima. Thus the learning rate is kept initially at 0.1 and then using a learning rate scheduler we decrease the learning rate at three time steps by one-tenth (0.1 times) particularly at epochs 70,100 and 170.

Table 1: Hyperparameters

Name	Value
Depth	120
Growth Rate	12
Reduction	0.6
Dropout Rate	0.2
Learning rate	0.1
Momentum	0.9
Epochs	200
Batch Size	64
Weight Decay	1e-4

4 Results

The Model was run on 5 training batches in publicly available training CIFAR-10 data of 50000 samples. The data was divide into training and validation with a train:valid ration of 0.8. The batch size was kept 64 as larger batch size were not able to give good accuracy on the test data. The model is then trained on complete public training i.e. (train + validation). The model was tested on the public CIFAR-10 test data and best classification accuracy was achieved of **94.66%**. The training loss for 200 epochs and the test accuracy of saved models on the public data is given below.

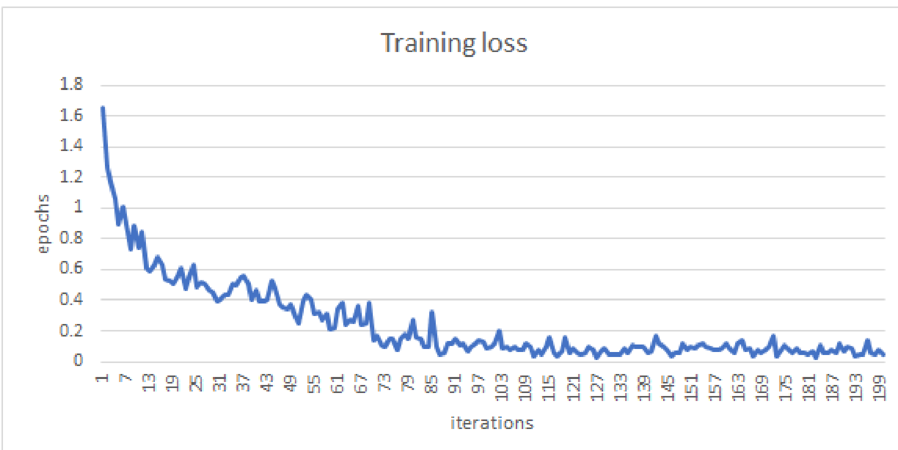

Figure 3: Training Loss

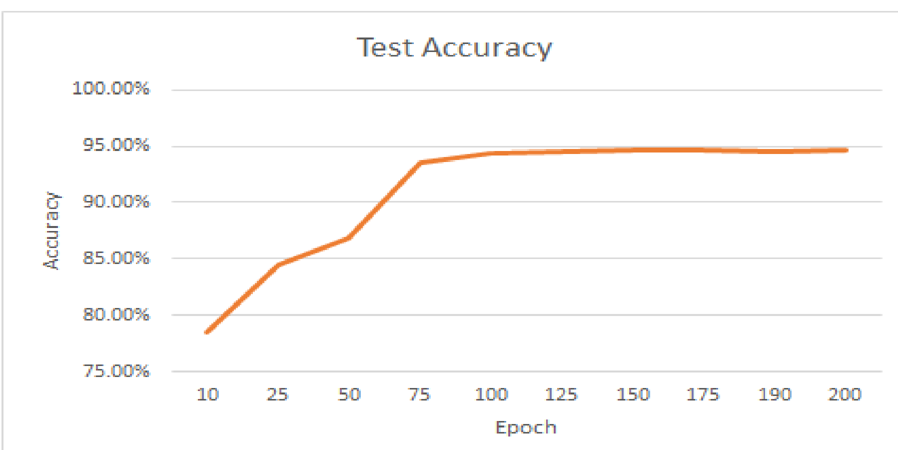

Figure 4: Test Accuracy vs Epoch

References

[1] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Identity mappings in deep residual networks." In European conference on computer vision, pp. 630-645. Springer, Cham, 2016.

[2] Huang, Gao, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. "Densely connected convolutional networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700-4708. 2017.

[3] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016. hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.