Training Variants of ResNet for the Classification of Tiny-ImageNet and SVHN Dataset

Abraham Jose
ID :5068109, CAP6614
abraham@knights.ucf.edu

Abstract

To learn the effect of the number of image samples and the training model(number of parameters) in the training process, I demonsrates the effects in Resnet 50 model and it's variations(Resnet 18, Resnet 34, Renset 50) with varying image sample sizes(400,280 and 120) under prolonged training sessions. Each model is trained for 2000 iterations and the data collected to infer on the effects of the training model and the data. From the experimentation, it is clear that the ability to learn from a dataset is not limited to the size of the model, but also to the complexity of the dataset and the no of samples available for training the model. Prior information in the model(the data distributed in the model) is really helpful and hence the pretrained model requires a lot less time to learn and adapt from the pretrained domain to current domain. From the experiments it is clear that the Resnet 18 model was able to learn as much as the other model. It is attributed to the fact other models were not able to learn the features(for example pretrained model on a larger dataset will give better performance) with data alone.

1 Introduction

Deep learning in computer vision has helped us make great strides that we couldn't have done without it. ImageNet is a large visual database created for image recognition task hosted by Stanford professor Fei-Fei Li and her team in 2010. This huge dataset for image classification sees the big breakthrough with Alexanet, a CNN architecture which reduced the state-of-the-art error rate of 26% to 16%. The challenge had significant impact and enduring relevancy on computer vision research over a 10-year period. The Street View House Numbers (SVHN) Dataset is another dataset from Stanford dataset which had images obtained from house numbers in Google Street View images for recognition of house number.

2 Pre-processing of Data-set

2.1 Tiny-ImageNet Dataset

Tiny-ImageNet dataset has 200 classes of images with training set having 500 images per class and validation set having 50 images per class. The testing set given in the website is not labelled. So I will be considering the actual validation image set as the test set and the training set will be splited to validation and training at 0.8 ratio by default.

We will be training the following models with the given number of training images(including training and validation images used for the model). The corresponding models trained and the number of images used for training are tabulated in the figure below. For example, ResNet 34 with 500 images uses 400 images for training and 100 images for testing the model. The initial images are 64x64x3 RGB images. The Tiny-ImageNet dataset has tight bounding box annotation available for each image.

Table 1: Table with no of images per class for each CNN model.

Model Name	Images/class	Training /class	Validation /class	Testing /class
ResNet 18	500	400	100	50
ResNet 34	500	400	100	50
ResNet 50	500	400	100	50
ResNet 50	350	180	70	50
ResNet 50	150	120	30	50

Inorder to ensure that the images are tightly annotated, we will use them to crop the images using the tight annotations available in the dataset. The images are cropped using the class_name_boxes.txt file in each class files and the dataloader resizes the images to 42x44x3 images(average of the shapes of training images). Further the following set of augmentation is applied to the training dataset with a probability of 0.25.

Table 2: Augmentations applied to training images.

Augmentation	Variable range
Brightness	+/- 0.1
Contrast	+/- 0.25
Saturation	+/- 0.25
Rotation	+/- 10
RandomHorizontalFlip	True
RandomVerticalFlip	True

2.2 SVHN Dataset

The SVHN dataset has images of size 32x32x3 with train and test split. The training image is splitted to training and validation images with a ratio 8:2(0.8 split). The number of training image is 58605, validation is 14652 images and testing is 26032 images. The images has the same augmentation as Table 2 with a probability of 25 percent for occurring.

3 Model

The resnet model was picked because the model offers smooth loss function where the model can back propagate the error better using the residual networks in every CNN block in the network. However, the performance of the model was not as good as expected. Also the model has better performance compared to other models. The block diagram is as follows. We have tried 3 variants for Resnet architecture including Resnet 18, Resnet 34 and ResNet 50 which has 18, 34 and 50 blocks correspondingly.

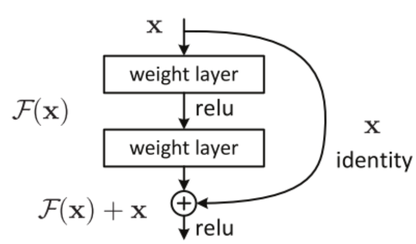

Figure 1: Resnet Block used in Resnet18, Resnet34 and Resnet51

4 Experiments

There improvements in the performance with respect to the variants of Resnet is non existent and the variations of the model has less effects on the overall accuracy. The performance deterioration is associated with the less number of classes per class and many number of classes in the imagenet datset.

Accuracy and Training time V/S Training Samples(ImageNet, Resnet 50) : The number of training samples for training is 400,280 and 120 respectively. I have not recorded the training accuracy and all accuracy will be validation accuracy as in Fig 2.

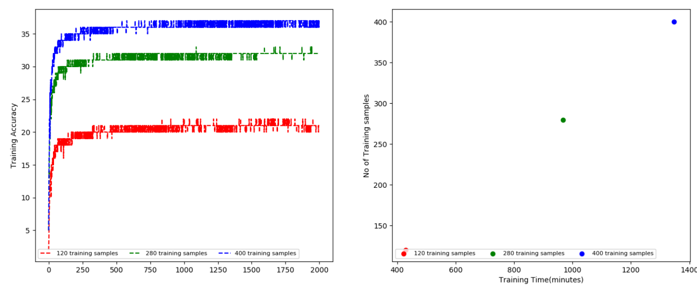

Figure 2: Accuracy and Training time v/s Training samples

Accuracy and Training time V/S Training model(ImageNet, Resnet 50/34/14) : As depicted in the Fig 3., the Resnet 18 model performs really well as opposed to resnet 34 and resnet 50. Resnet 50 has really poor performance. The number of parameters is in 12K to 25K million range.

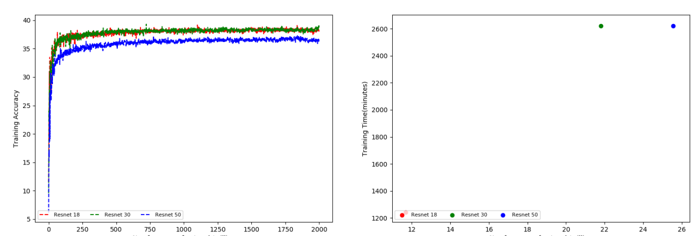

Figure 3: Accuracy and Training time v/s Training models

Accuracy and Loss (ImageNet, Resnet 50) : Fig 4. shows the accuracy and loss for Renset 50 model trained on 400 image samples. From the loss graph it is clear that the model is learning the data distribution in the training set. However, not so in the validation dataset. The validation loss increases as the data distribution that the model tries to learn from the training data diverges from the validation dataset. Th accuracy saturates to 37% in case of Resnet 50.

Accuracy and Loss vs number of iterations (SVHN, Resnet 101) : In Fig 5. we can find that the model starts from 84% accuracy and attains 92% in less that 100 iterations.We can see that the model has learned the features and prolonged training does not improve the performance of the model.

From the given samples it is clear that the Resnet 18 model was able to learn as much as Resnet 50 and Resnet 34 was able to learn and as the size of dataset increases the accuracy tends in increase(however with a logarithmic relationship). The pretrained model achieved the desired performance in least number of iterations.

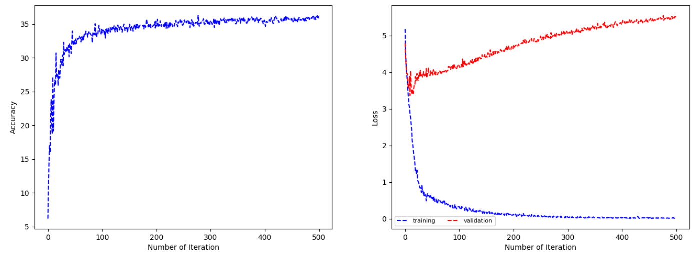

Figure 4: Accuracy and Loss for Resnet 50(400 samples)

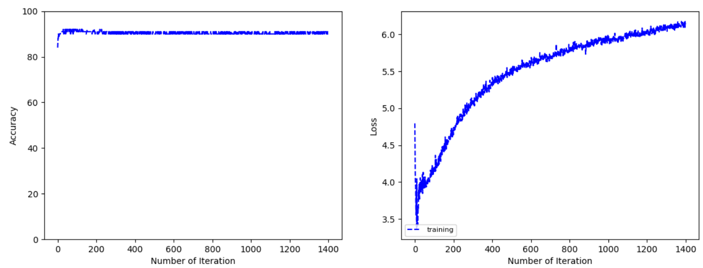

Figure 5: Accuracy and loss for SVHN Resnet 101 model(Validation)

5 Conclusion

From the experiments it is clear that the tiny imagenet datset is a really challenging dataset as there are many number of classes and limited number of images per class. Training for longer time using Resnet could not help us achieve better performance as the model is saturated with the given data/information. It is notable that the model is capable of learning more however more data is required to do so. The accuracy achieved from different training samples are 25, 32 and 38 respectively. From the accuracy results it is clear that all the variations of Resnet model learned almost same amount of information and interestingly larger network find it difficult in reaching the same accuracy level as Resnet 18. It can be attributed to the difficulty and the complexity of the dataset. This is evident as the pretrained models were able to achieve better performance(88+% accuracy using Resnet 50). Pertained models in Resnet did really well in increasing the performance in less number of iterations as the model already has learned features in the CNN as we can see in the SVHN resnet 101 pertained model.

References

[1] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, Hartwig Adam.: "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation" In *CVPR* (2018)

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun.: "Deep Residual Learning for Image Recognition" (2015)

[3] Fig 1 : https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035