The Comparison of Different Machine Learning Platform and Models for CIFAR 100

Transfer learning by Tensorflow (Keras API)

Group 2
Li Du

Professor Amir Jafari

DATS6203 Machine Learning II
December 2018

Introduction

One of the major challenges for image classification, especially involving complex and large datasets, is the high computational cost. Such demands for computational resources are usually beyond the capability any standard computers. In addition, new ideas and new architectures are driving the progress for better network performance and higher accuracy. Fortunately, there are research groups and/or companies that have taken great efforts to develop and improve neural networks with the advantage of having access to high-performance facilities and talented team of researchers and scientist. As a result, these trained models are now accessible to the public for many individual projects.

In this project, we used the pre-trained deep neural network "Inception 5h" (Inception V1), or popularly known as "GoogLeNet". It originated from the work done by a group of scientists with Google as a submission to the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC 2014)[1]. Inception module had more than one filters at the same level and then concatenated them before stacking all the layers up. This is particularly helpful in dealing with images having the key features occupying varying portion of the images. The entire net has 22 layers (27, including the pooling layers).

Figure 1 is a simple demonstration of the structure of the Inception V1 network.

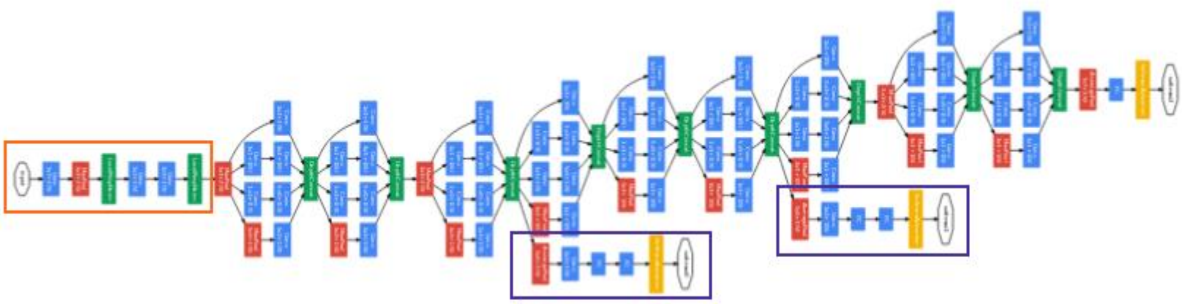


Figure 1. Structure of the InceptionV1 network[2]

Implementation of transfer learning neural network model

[1] Going Deeper with Convolutions. Accessed at: https://www.cs.unc.edu/~wliu/papers/GoogLeNet.pdf on Dec 3, 2018.

[2] A Simple Guide to the Versions of Inception Network. Accessed at: https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202 on Dec 3, 2018.

The pre-trained InceptionV1 nets can be downloaded from Google' developers' tool deposit and easily imported into tensorflow's framework (please refer to the python scripts for the details).

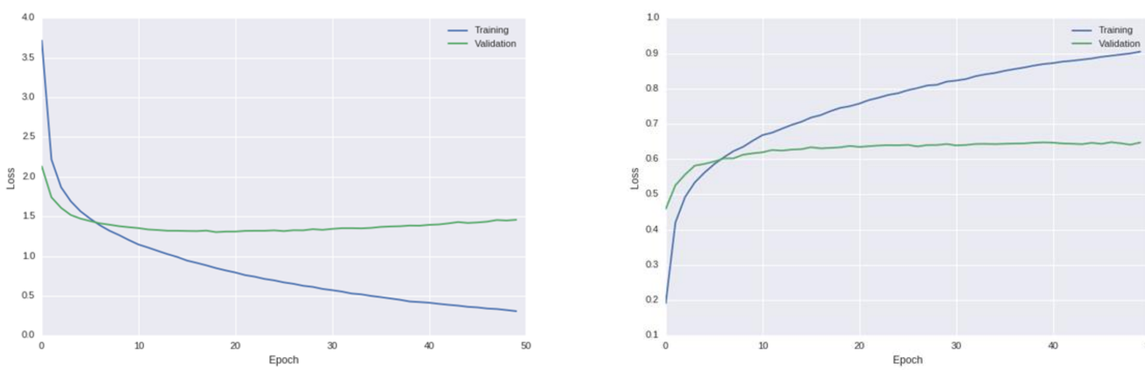
The idea is to first run the CIFAR100 images through this pre-trained network and export the tensors from a pooling layer before the fully connected layers. This requires reshape and zooming of the input images because Inception V1 takes inputs as 224x224. Then, these tensors are subsequently fed to an MLP with a dropout layer that exclude 50% of the neurons followed by an activation layer using softmax. The first step took about 1.5 hours to finish, but it only needs to be executed once and serialized and saved locally.

Results and conclusion

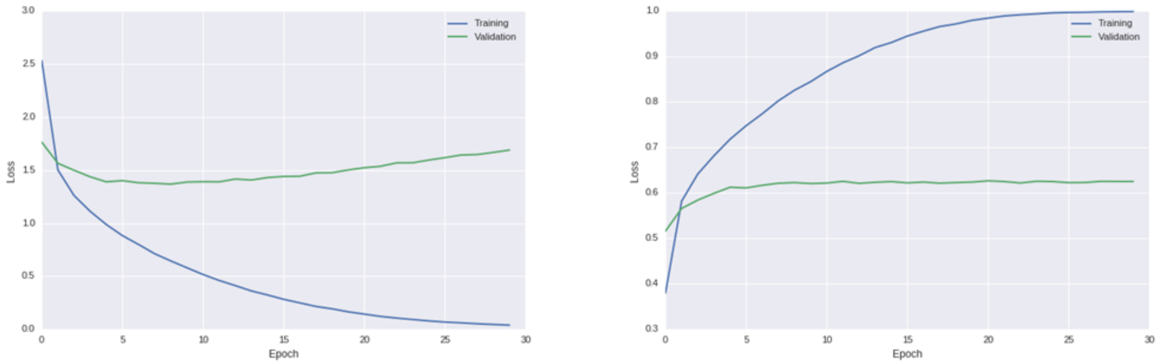
As discussed previously, the output from the convolution and pooling layers were first exported and imported to feed an MLP. The base case has 2 layers with 1024 neurons each, 1 dropout layer with 50% neurons left out and 1 more layer having softmax as the activation function. With this design, the model will predict the probability of each image being classified as each of the 100 classes. For the purpose of evaluating the accuracy, simple post-processing was done to identify the class with highest probability.

After tuning the hyperparameters of the MLP, it was found that the performance of the entire transfer learning model was dominated by the pre-trained deep nets. The below graphs demonstrate the effects of the multiple hyperparameters.

Base case (loss, accuracy vs. epoch)



Without dropout layers (loss, accuracy vs epoch)



With 128 neurons

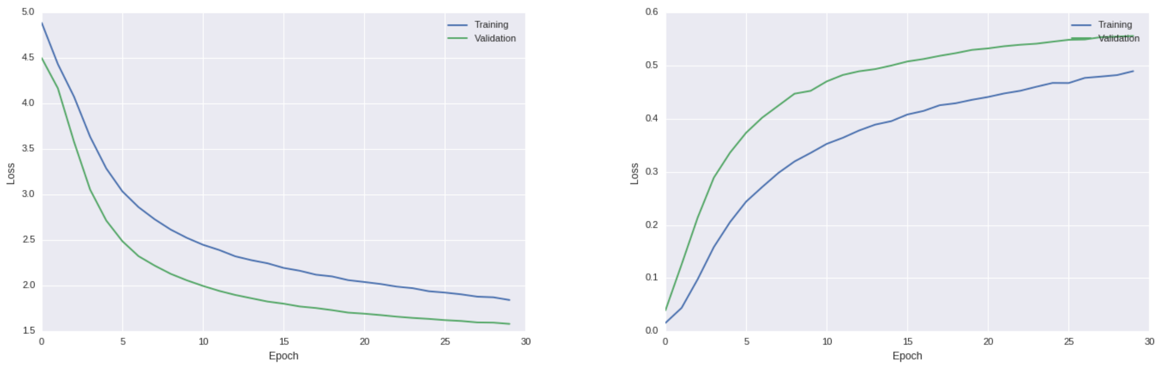


Figure 2. Effect of hyperparameter tuning on the model performances

The best performance of this transfer learning model achieves a total accuracy of approximately 66%. The relatively higher mis-classification rates represented as darker spots that are not on the diagonal line are loosely distributed and sparse, which, again, indicated no systematic misclassification. The class “motorcycle” has the highest accuracy (> 90%) whereas the class “seal” has the lowest accuracy (< 30%). By investigating several, classes that have visually similar textures and shapes are prone to misclassification (such as snake vs. worm, baby vs. girl).

Even if the Inception model was not trained using CIFAR100 dataset, it still shows superior performance compared to MLP and shallow CNN constructed in this project. 66% accuracy is a very good starting point for further deep neural nets development and tuning, which demonstrated the potential of transfer learning in many applications. In terms of future work, the performance can be evaluated by attaching a more complex net (such as complex CNN instead of just MLP) after the output from the Inception model.

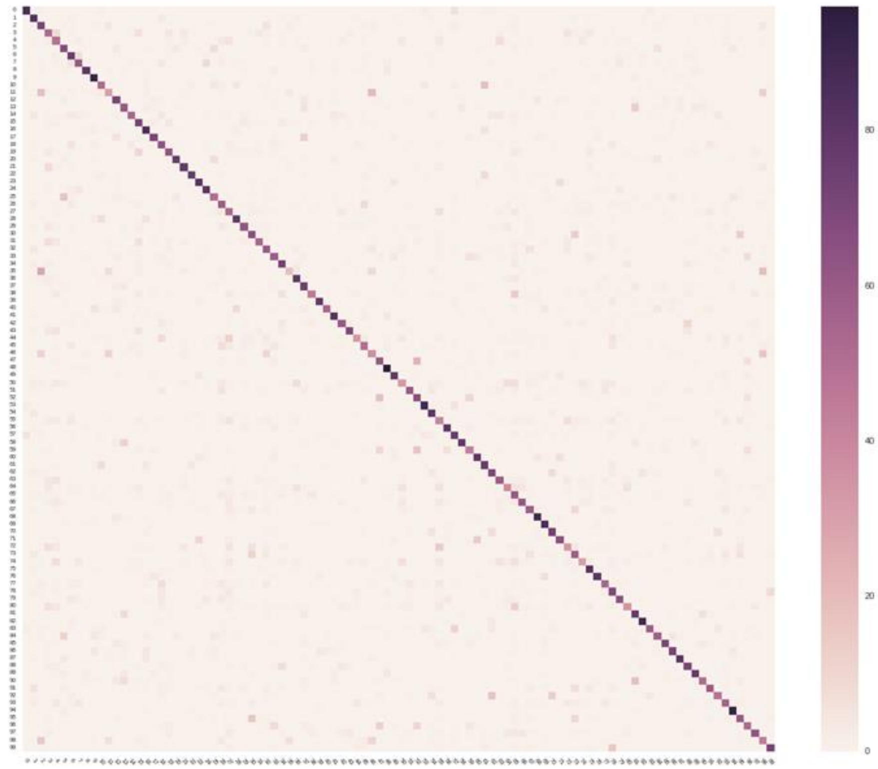


Figure 3. Confusion matrix