Robotics Inference: Object Classification

Shilpaj Bhalerao

Abstract – *Deep Learning has made great progress in object classification and detection tasks. The breakthrough happened in 2012, when Geoffrey Hinton along with Alex and Ilya published a network architecture, AlexNet, which performed better than humans on ImageNet dataset. In this paper, the same AlexNet architecture is used for object classification task. The model was trained using a Deep Learning Training system - Nvidia Digits. This paper shares the training process and results of the model trained using Nvidia Digits.*

Keywords: Nvidia Digits, Deep Learning, Convolutional Neural Network (CNN), Machine Learning, ImageNet dataset, etc.

1. Introduction

Image classification is not a new problem. Computer vision researchers had tried lot different algorithms like Histogram of Gradient (HOG), Color Histogram, etc. feature extracting techniques for object classification. Before the age of Deep Learning, researchers used to design the kernels which would extract features from the images. These features where trained on a Machine Learning algorithm like Support Vector Machine (SVM) to classify the objects. The problem with this approach was they were not robust. These models performed very poorly in different lighting conditions and images with noise. The speed of execution was also very high because of which real-time inference was not possible. Since real world applications require higher accuracy and real-time inference speed, it is important to solve this problem. Using Deep Learning models, models can achieve higher accuracy as well as real-time inference speed. AlexNet is one of the first network which used Convolution Neural Networks, a type of Deep Learning networks, and achieved high accuracy on ImageNet dataset. This paper discusses the use of AlexNet to solve object classification problem

2. Formulation

The figure below shows the model architecture of the AlexNet [1].

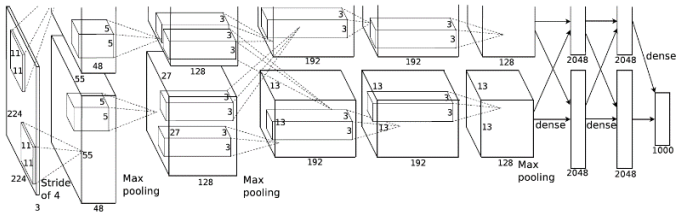

Fig 1: AlexNet model architecture

Nvidia Digits [2] provides some parameters which can be assigned to the model before training. The model was trained for 100 epochs with Stochastic Gradient Descent (SGD) with a batch size of 128. Nvidia Digits provides facility to use learning rate scheduler. This will reduce the learning rate after certain number of epochs. The initial learning rate for training of the model was set of 0.01. Digits provides two award winning models on ImageNet dataset viz. AlexNet and GoogLeNet. GoogLeNet is a much deeper network compared to AlexNet. This means total number of parameters of GoogLeNet are more than that of AlexNet. Thus, inference speed of AlexNet is better than that of GoogLeNet. Hence

AlexNet is chosen for this project. From empirical data, learning rate of 0.01 is a good starting point for the CNNs hence initial learning rate was set to 0.01. As more data is fed to the network in a batch, the generalization capability of the network increases. Hence a batch size of 128 was selected while training the network.

3. Data Acquisition

Data Acquisition is as important as training a model. The data is collected using webcam of a laptop. Data can also be collected using camera on Nvidia Jetson TX boards. The size of the images is 640x480. Prior to the model training, Digits provide an option of resizing the images. Since, AlexNet requires the images of size 256x256, the images are resized before sending it to the network. Color of an object is important feature to classify the object hence the images collected are RGB images.

Fig. 2: Samples of collected data for training

Figure shows the resized collected data for training. Book, bottle and helmet are the three objects which the model has to classify.

4. Results

The trained model got a training accuracy of around 99% on the training dataset while the validation accuracy of 60%. The model can process each image within 5 milliseconds. Thus, the model can be used for real-time inference applications.

Fig 3: Model Results

After testing it on a single image sample,

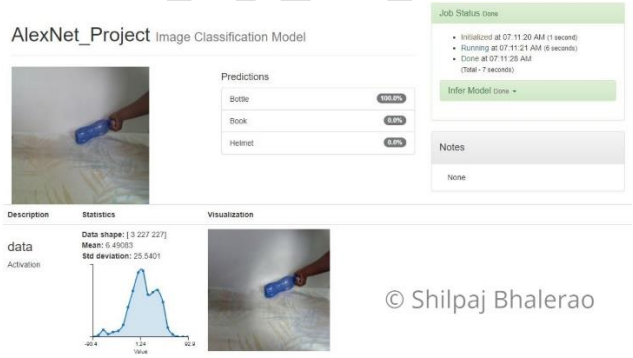

Fig 4: Model predictions on single sample

5. Discussion

One may think that the results are poor on validation dataset but given the data size without using image augmentation techniques, the results are good. Using image augmentation technique like image normalization, pixel normalization and image equalization, the perform will increase. Also, using cutout will surely increase the model accuracy. Batch Normalization, the concept introduced three years after the AlexNet paper was published can be added to the network to improve model accuracy. A custom model can also be trained with network architecture to improve the performance of the model.

6. Future Work

The model is able to classify the objects correctly. It takes under 5 milliseconds for the model to classify

object. Thus, it can be used for real-time applications.

This proved the concept of using AlexNet for real-time object classification application. Future work involves

1. Adding a greater number of objects to the dataset
2. Using custom designed CNN model for higher accuracy
3. Using the model for keeping track of objects in a store

7. References:

1. Geoffrey E. Hinton, Alex Krizhevsky, Ilya Sutskever. ImageNet Classification with Deep Convolutional Neural Networks. Paper Link
2. Luke Yeager, Julie Bernauer, Allison Gray, Michael Houston. DIGITS: the Deep learning GPU Training System. ICML 2015 AutoML Workshop. Paper Link