Robotic Inference for Indian currency using Convolutional Neural Networks

Sagarnil Das

Abstract—Cash is used as a medium of exchange in a lots of places in today's world. It might be in a local store, gas station, malls and almost anywhere. Often times, the notes are exchanged for a commodity or a service between two human beings. But what if this whole process can be automated by using a superior technology? In this project, a deep learning approach was implemented using three different neural networks, namely Alexnet, Googlenet and LeNet for classification of different denomination of notes and then they were used for inference. The best results are then published in this article. This method can be implemented in a live robotic environment using NVIDIA Jetson.

Index Terms—Robot, Inference, deep learning.

1 Introduction

The classification method of modern deep learning era follows a diverse range of decision driven methods for the identification of images. The basic assumption is these approaches is that the images contains one or more features which can be used to train a model in order to predict a label, which in this case is the object in the image itself. Often, the features are represented by the pixel values in different regions of the images and the spatial analysis of the image by neural networks often results in a superior classification model for the class of the object depicted in the image. The classes, which acts as labels, can be detected manually by an analyst a-priori for training the model or it can be automatically clustered using an unsupervised algorithm into sets of prototype classes, where the analyst merely mentions the desired number of categories.

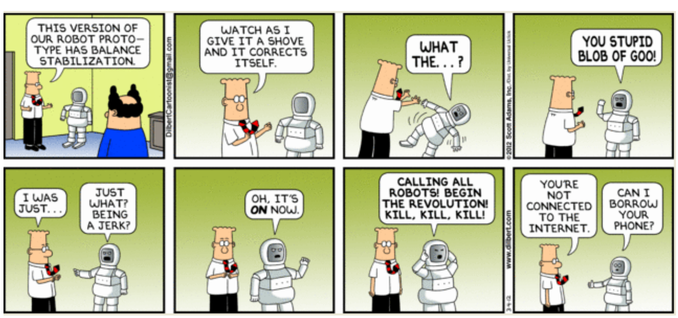

Fig. 1. Robot Revolution.

1.1 Workflow

In this project, NVIDIA's DIGITS workflow was implemented for rapid prototyping of different approaches and methods that can be implemented in NVIDIA Jetson for real time object classification, detection or even semantic segmentation.

1.1.1 *Use Cases*

Two use cases were implemented in this project.

1) A classification network was built for classifying candy wrappers, bottles or nothing (empty conveyor belt) on a moving conveyor belt.
2) A classification network was built for classifying two different denomination of indian currency namely 100 rupee notes, 200 rupee notes or nothing.

2 Background / Formulation

During the first inference task, i.e. classification of objects on a moving conveyor belt, both AlexNet [1] and GoogleNet [2] were chosen as they both had good inference rate per image with reasonable accuracy. The models were trained for 5 epochs and a Stochastic Gradient Descent Optimizer with an initial learning rate of 0.01 was used which was decayed to 0.001 halfway through the training for both of these models and it successfully met the criteria of an inference time of below 10 ms with an accuracy greater than 75 percent.

During the 2nd inference task, i.e. classification of different denomination of currencies, same models, optimizer and initial learning rate was used. The models were trained for 5 epochs using a decaying learning rate as before. The images were kept as RGB images as both the notes look similar, the only major difference being the color of the notes.

Increasing the number of epochs did not improve the accuracy. An alternative approach was taken where a RMSProp optimizer was used. This method also did not improve the accuracy.

3 Data Acquisition

The P1 image dataset consists of images of bottles, candy wrappers and no object on a conveyor belt passing under a camera. A swing arm is used to sort all right objects to correct the bins depending on classifying results. The RGB images were scaled to 256 x 256 pixels for effectively using AlexNet and GoogleNet. This dataset was provided by Udacity. Figure 2 shows the sample conveyor belt dataset.

Fig. 2. Conveyor Belt data

The 2nd dataset, i.e. different denominations of Indian currency namely 100 rupee note, 200 rupee note and nothing were collected with the help of a web camera. The RGB images were scaled to 256 x 256 pixels as these were also trained with the help of AlexNet and GoogleNet. Figure 3 shows the sample currency dataset.

The currency dataset was split into 3 parts: Training, Validation and Test. Figure 4 refers to the distribution of the data. The images were transformed and augmented to generate more training and validation data. After the image augmentation, there are 2679 training images, 622 validation images and 30 test images for the 3 classes.

Fig. 3. Currency data

4 Results

For the initial inference task of conveyor belt classification, both GoogleNet and AlexNet were used as both produced and accuracy of greater than 75 percent. Figure 5 and 6 denotes the evaluation results for AlexNet and GoogleNet respectively.

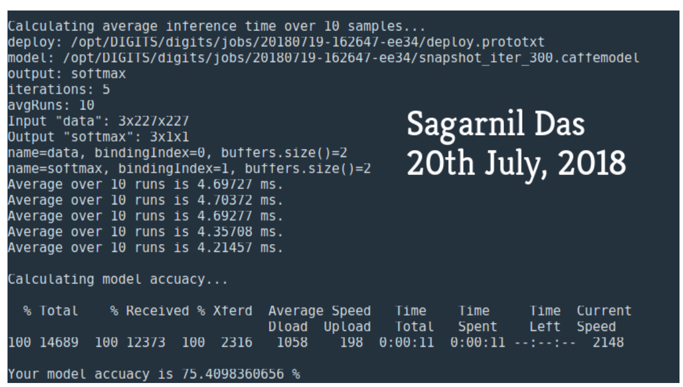

Fig. 4. AlexNet Inference

It is observed that AlexNet produces an average inference time of a little over 4 ms and GoogleNet produces an average inference time of around 5 ms. Both produces an approximate accuracy of 75.41. So in contest, AlexNet performs better than GoogleNet with respect to inference time. So AlexNet was chosen as the final contender.

During the training process, AlexNet reached a validation accuracy of 100 percent within 5 epochs (Figure 7). The validation loss also converged really quickly with the training loss.

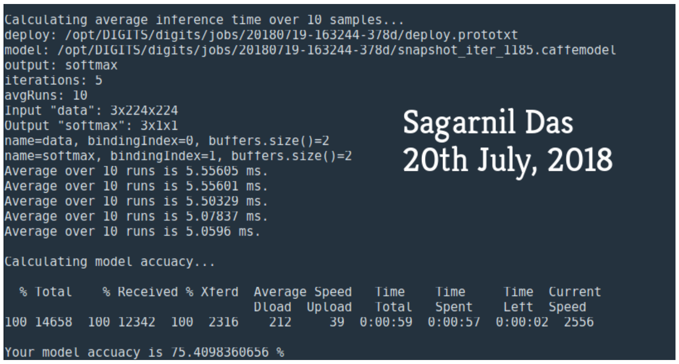

Fig. 5. GoogleNet Inference

Fig. 6. Alexnet Training

However, for the 2nd inference task of currency classification, similar results were not obtained while using AlexNet. The validation accuracy reached about 75 percent at the end of 5 epochs.

During the inference process for the 2nd use case, 30 separate images were kept aside as test images which the model has never seen. Figure 9 shows the test results:

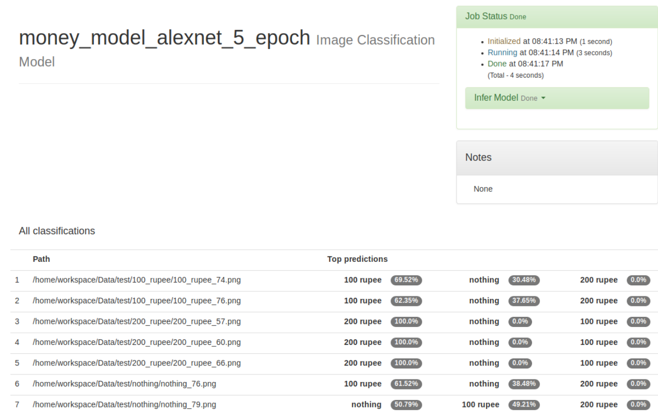

Fig. 7. Alexnet test for currency classification

5 Discussion

In the first inference task, an extremely high validation accuracy of 100 percent was obtained. For the 2nd task, i.e. classification of different denomination of Indian currency, similar level of accuracy was not obtained (75 percent) because the sample size was not big enough. Also, the images were captured using a low resolution web camera. That could have caused some difficulties for the inference task too. Also, during the 2nd inference task, the background was not a solid colored background which added some noise too in the training data. Much better results can be obtained simply by adding more training data and also changing the background to a solid colored one. Also, using a higher resolution camera can also add to an improved result. Still, 6 objects were correctly classified out of 7 in the test set. So the results were significantly acceptable.

6 Conclusion / Future work

Whilst this project did not achieve superlative results, it laid the groundwork for a potential breakthrough in future. The market for this use case is not known yet, but it can be applied to a diverse range of use cases. A supermarket can use this model to calculate the total cash inflow per day, this model can also be trained with a slight variation to detect counterfeit notes etc. The number of use cases are really infinite.

References

[1] A. Krizhevsky, *ImageNet Classification with Deep Convolutional Neural Networks*. 2012.
[2] C. Szegedy, *Going Deeper with Convolutions*. 2014.