Robotic Inference using Deep Neural Networks

Navin Rahim, navin.rahim@gmail.com

Abstract—This paper presents two deep neural network solutions to two different problems - one to identify objects on a conveyor belt and to identify vegetables placed on a table. For both the problems, AlexNet model was used. The data collection, training and results are discussed in this paper.

Index Terms—Robot, Udacity, deep learning.

1 Introduction

ROBOTS require to understand the environment around them to perform actions. Inorder to understand the surroundings, robots need to identify the objects around them. Image classification is a task that majority of the robots perform to identify objects in their surroundings. Images of objects are taken using a camera and they have to be classified to understand what it is.

Deep Learning is a machine learning task which takes in a lot of data and learns from them. Later, new data can be provided to get meaningful output. This technique can be used for image classification tasks and is the widely used method for this task.

This paper discusses the implementation of two such image classification tasks. First one involves identifying bottles, candy boxes or nothing from images taken over a conveyor belt and the second one involves identifying onion, garlic or nothing from images taken over a table.

2 Background / Formulation

NVIDIA Digits platform was used to train the models for the two tasks. It provides three models by default - LeNet, AlexNet and GoogLeNet.

AlexNet was chosen to train for both the tasks and provided good results. SGD (Stochastic Gradient Descent) optimizer with learning rate of 0.01 was used as the starting points. Since, the model provided good results with these values, these were finalized.

The conveyor belt problem gave the required results in 5 epochs, but the vegetable classification problem was trained for 50 epochs, and the best model was obtained at epoch 32.

3 Data Acquisition

3.1 Conveyor Problem

For the conveyor problem, a dataset was provided by Udacity. The dataset had RGB images with three classes - bottles, candy bar and nothing. The details of the dataset can be seen in TABLE 1. The dataset contained a total of 10094 images out of which 25% were used as validation set.

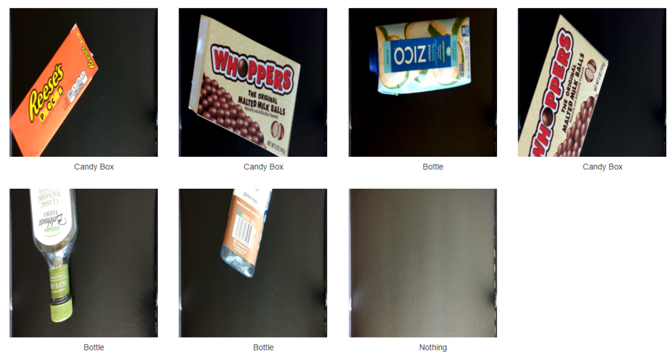

Fig. 1: Conveyor Data

Labels	Training Count	Validation Count	Total
Bottles	3426	1142	4568
Candy Bar	1871	624	2495
Nothing	2273	758	3031
Total	7570	2524	10094

TABLE 1: Conveyor problem data details

3.2 Vegetable Classification

For the vegetable classification problem, dataset was collected using an Android phone. The vegetable to be classified were Onion, Garlic and nothing on a table. The dataset has RGB images. There were 615 images collected. More details on the dataset is given in TABLE 2. 25% of these images were used as validation images and 10% were used as test set.

Fig. 2: Vegetable Data

Labels	Training Count	Validation Count	Test Count	Total
Onion	152	59	23	234
Garlic	125	48	19	192
Nothing	123	47	19	189
Total	400	154	61	615

TABLE 2: Vegetable data details

4 Results

4.1 Conveyor Problem

The requirements of this problem was to train a model that has an inference time of 10 ms or less and have an accuracy greater than 75%. Udacity provided an evaluate command to find these metrics. 5 runs, each with 10 iterations, provided an average inference time of less than 5 ms on the Udacity workspace. The model also gave a test accuracy of 75.41%.

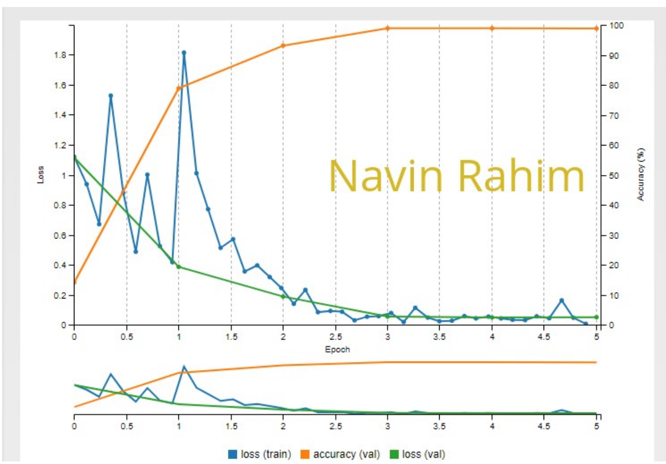

Fig. 3: Conveyor Problem loss graph

4.2 Vegetable Classification

The vegetable classification over onion, garlic and nothing obtained a validation accuracy of 93.125% and test accuracy of 91.8%. The model obtained this accuracy in the 32th epoch. The validation loss was 0.22. The training graph is shown in Fig. 4.

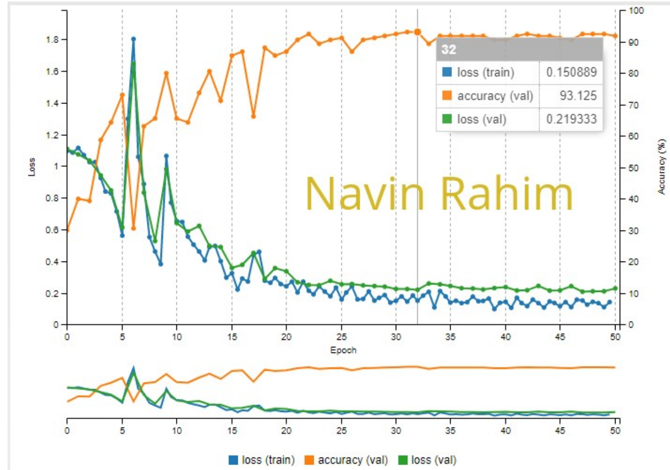

Fig. 4: Vegetable classification loss graph

5 Discussion

5.1 Conveyor Problem

The model obtained the desired results with the AlexNet model in the 5th epoch itself. Although the model obtained close to 100% validation accuracy, the test accuracy was only around 75%. This can be improved by adding more data by looking into the test image results. It may also be improved by changing the model or tuning the hyperparameters.

5.2 Vegetable Classification

The model gave good results on the validation and test set. The accuracy on both of these sets were over 90%. The data were collected on different positions and angles of the images. More data can be added to obtain for more accurate results. The model gave good results on images from internet as well as shown in the Fig. 5. Still more testing has to be done in different real time conditions.

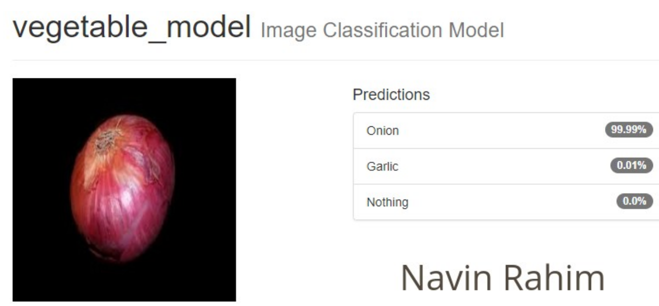

Fig. 5: Inference on image from internet

6 Conclusion / Future work

Two neural networks were trained on the supplied dataset and the collected datasets and satisfactory results were obtained on the same. The networks work well on seen dataset but may perform poorly on real time. Thus, more data has to be added and more tests are to be performed to obtain reliable networks. After obtaining confident results in real world test, these models can be deployed in real world robotic applications.