Robotic Inference Project

Shotaro Oyama

Abstract—In this project, two neural network based classifiers are built on the supplied data and my own collected data to classify objects in a scene. Training and testing were executed on the NVIDIA DIGITS workspace. First neural network is trained on the supplied data set by Udacity to classify objects in the photos taken from a Jetson mounted over a conveyor belt. Second neural network is trained on the collected data to classify boxes. The accuracy and the inference time of the network for the supplied data meet the requirement.

Index Terms—Robot, Deep Learning, Object Classification

1 Introduction

To classify the objects on the table or board is still one of problem for robotics especially in the logistics industry. There are many type of boxes which packages the items for customers, and these boxes are needed to be identified to be handled correctly. Poor accuracy and inference time may cause trouble for logistic operation or frustrate the customers. The concept of this project is to classify the type of boxes and others (Smart phone and no objects).

2 Background / Formulation

The NVIDIA DIGITS workspace provide three types of deep neural network (LeNet, AlexNet, and GoogleNet) as standard models for image classification. I selected AlexNet, although GoogleNet had almost the same performance. Parameters are determined experimentally, SGD for optimizer, 0.001 for base learning rate, 30 for epoch.

3 Data Acquisition

The dataset is collected by the camera integrated on my laptop. It has 4 type of categories, Normal box, Tissue box, Smart Phone, and None. Each photo is a 8-bit RGB PNG image with the size of 1280x720. Example images are shown in Fig. 1. 75% of the collected dataset is used as a training set and 25% of it is used as a test set.

[1]Normal box [2]Tissue box

[3]Smart Phone [4]None

Fig. 1. The Own Dataset

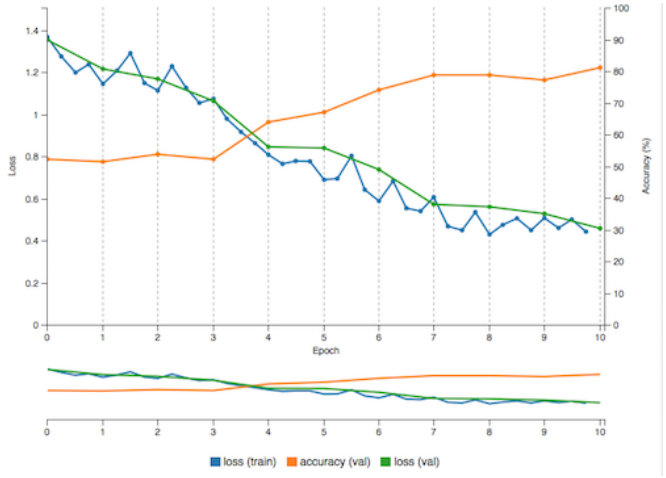

Fig. 2. Result of own dataset

4 Results

The accuracy of the network on the dataset was 82% as Fig. 2 shows. Inference time is not measured on the collected dataset.

5 Discussion

The trained network could classify at almost minimum level of requirement or worse. Also, in real case it is likely to go even worse. Although the photos are taken from different angle, size or positions, the same objects are used for the training and test set. Thus, when the network encounter the objects which is totally new, the accuracy will drop.

6 Conclusion / Future work

There is potential to capture more data or augment the additional data from the Protos taken already, by randomly adding brightness, flipping and so on, as well as taking the photos of new items.