

Robotics Software Engineer Nanodegree Program: Term Two: Project One: Robotic Inference

Li Feng,
acumenta@yahoo.com

ABSTRACT

This is the first project in term 2 of the Udacity Robotics Software Engineer Nanodegree program. This report presents robotics inference project using NVIDIA's DIGITS workflow and TX2 Jetson environment. It first used the supplied data in the digits image, created a model using GoogLeNet network that achieved 75.4098 % accuracy and average inference time over 10 run were between 4.9649 ms - 5.52061 ms on p2.xlarge instances. In the second part, the robotic inference ideas were author's own and data collected. The dataset contained 5 classification categories and used two different neural networks models (i.e. AlexNet and GoogLeNet) for classification, inference, and comparison on the DIGITS and TX2 Jetson. The model used GoogleNet had higher probability estimations.

1 Introduction

Due to personal experience, the author's goal is to design and create robots that can provide partial or full care to limited mobility individuals. This kind of robots can be used in residential home, hospitals, or senior care facilities that provide routine lifting and transporting between various locations (ex: bed, bathroom, car, ...etc), and holding and balancing fragile bodies for daily hygiene activities. A soft robot is therefore preferred over hard robot because of the interfaces with human. Although there are many equipments or machines provide (partial) services currently available in the market, but there is no soft robot yet, no one single smart soft robot providing integral functionalities as a care provider. However, the desired goal is very big and can only be achieved through series of small incremental milestones or goals. A robotic inference project is ideally the first step, so that the robot can recognize various household and personal items (like: gloves, shoes, coat, nail clippers, vitamin bottle,...etc) before anything else.

2 Background/Formulation

For the first inference task on supplied data, the dataset was created with: image size of 500x500, image type: color, and image encoding: png, to match the images in the supplied (P1_data) data. The standard networks of GoogLeNet, learning rate of 0.01, and 9 epochs were chosen in the classification model. The rest of the parameters used the default settings. This configuration achieved the numerical requirements of inference time below 10 ms with accuracy > 75%.

GoogLeNet, in general, has good inference rate per image with reasonable accuracy. It is a newer model, consisted of 22 layer deep CNN (Convolutional Neural Network) and 4 million of parameters. It introduced the idea that CNN layers didn't have to be stacked up sequentially. The parallel arrangement of several very small convolutions in the 9 inception modules had improved performance and reduced the number of parameters to 4 million (from 60 million in AlexNet) [4, 5, 6]. Therefore, the author selected GoogLeNet in both the supplied data and collected data classification models.

The collected data contained images of five different items, each item with > 400 images. The images were resized to 256x256 RGB to match the sizes in GoogLeNet and AlexNet. Images encoding is png also. A dataset and two models using different neural networks (i.e. AlexNet and GoogLeNet) were created for classification, inference, and comparison.

Experimentations with various parameters values (like: epochs, learning rate, network, image sizes) were done on the above 3 models. After some analysis and comparison, the configurations presented here are the optimal settings (from author's point of view) which resulted highest accuracies and lowest inference times.

3 Data Acquisition - Supplied Data

The supplied data images are photos taken from a Jetson mounted over a conveyor belt. It's training pictures of candy boxes, bottles, and nothing (empty conveyor belt) for real time sorting. They are of size 500x500 color images in PNG file format. Here is a table of relevant information:

ItemofImages	NumberOfImages	SizeofTheImages	TypeofImage
Bottle	4568 images	500x500	RGB
Candy_Box	2495 images	500x500	RGB
Nothing	3031 images	500x500	RGB

And below are some examples or images of the supplied data. The supplied data are located at: /Data/P1_data folder.



3 Data Acquisition - Collected Data

Three different pairs of gloves, a vitamin bottle, and a nail clipper (i.e. total of 5 classes) were chosen to be photographed to create the collected data. They are the kind of items that a care provider often needs to recognize. The "Example Python Data Capture Script" from the lecture was used. The images were resized to 256x256 RGB to match the sizes in GoogLeNet and AlexNet. A laptop equipped with a webcam was used to collect data. Light or neutral colors were arranged to be the background colors. Here is a table of relevant information:

Item Of Images	Number Of Images	Size Of The Images	Type Of Image
Green PopTop Glove	495 images	256x256	RGB
Centrum Vitamin Bottle	501 images	256x256	RGB
Pink Nail Clipper	496 images	256x256	RGB
Red GreyWhite Glove	448 images	256x256	RGB
Tiger Black Glove	427 images	256x256	RGB

And below are photos of items images:

Image of Red Grey White Gloves



Image of Green PopTop Gloves



Image of Pink Nail Clipper

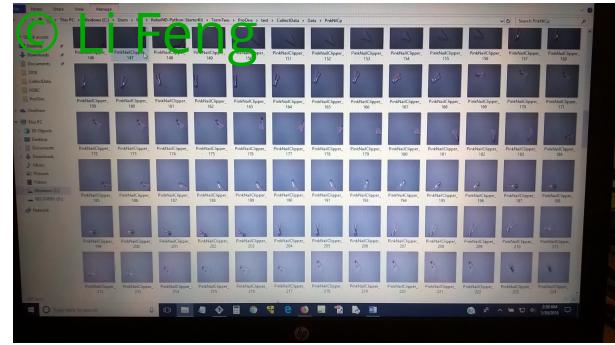


Image of Tiger Black Gloves

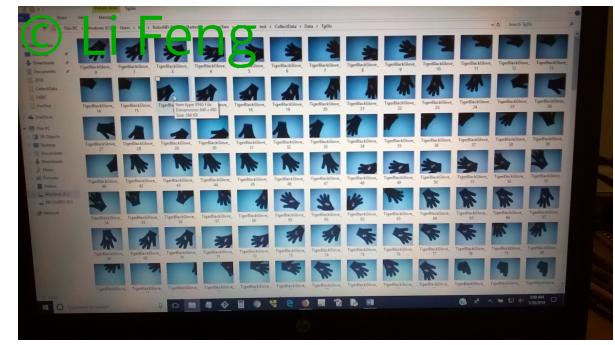
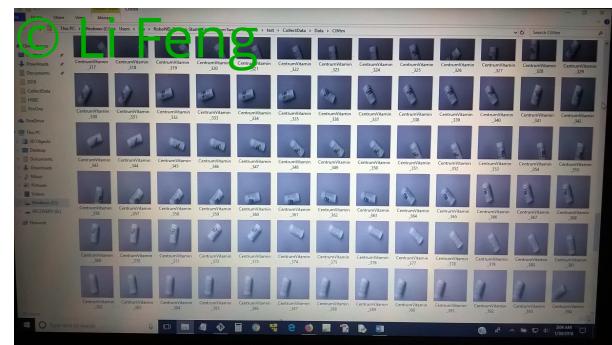
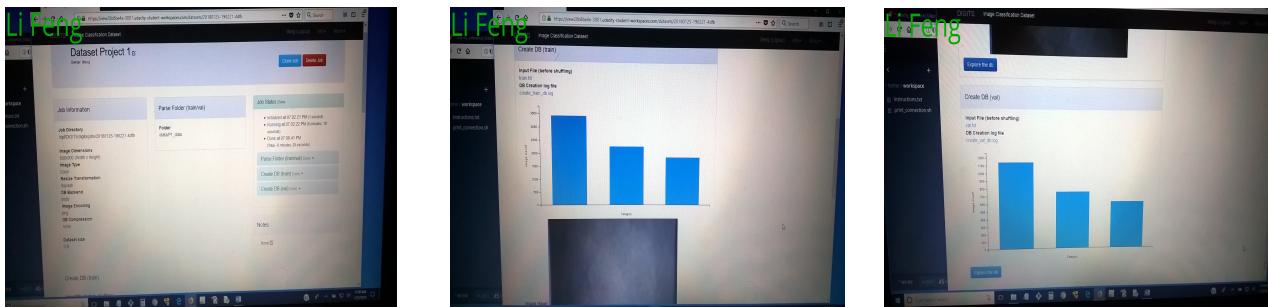


Image of Centrum Vitamin Bottle

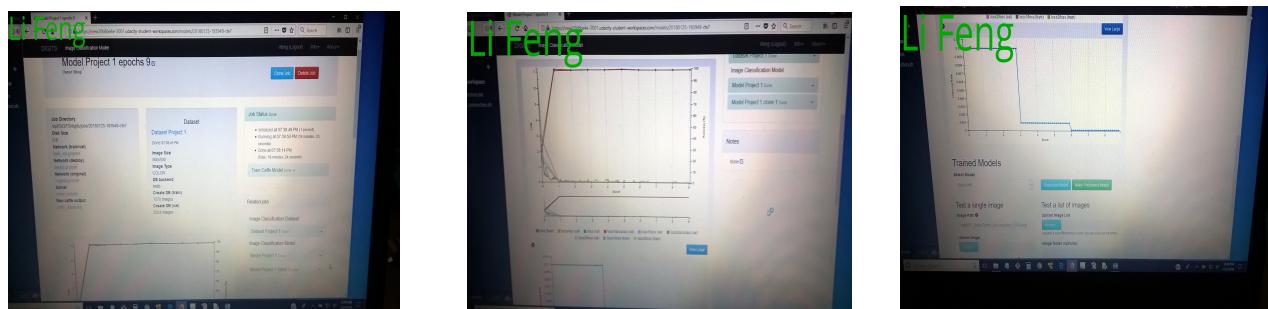


4 Results on Supplied Data

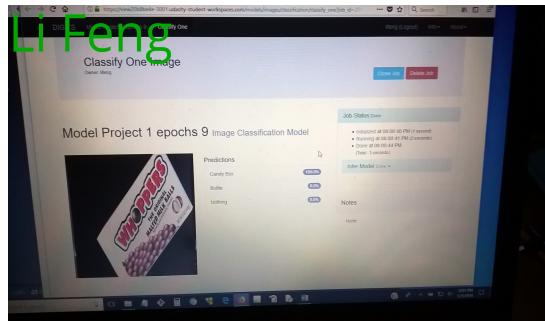
Supplied Data are located at: /Data/P1_data folder. There are three classes: Bottle (4568 items), Candy_box (2495 items), and Nothing(3031 items). A dataset (Dataset Project 1) was created with the supplied data. Below are the images of the created DataSet.



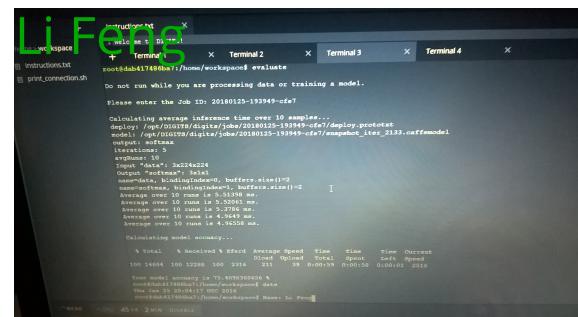
A model (Model Project 1 epochs 9) was created using (Dataset Project 1) with 500x500 image size, Googlenet, and 9 epochs. Below are the images of the created model (Model Project 1 epochs 9).



The model (Model Project 1 epochs 9) was tested with the (Classify One Image) on a whoppers image and the predictions for Candy Box was 100% as seen in below picture.



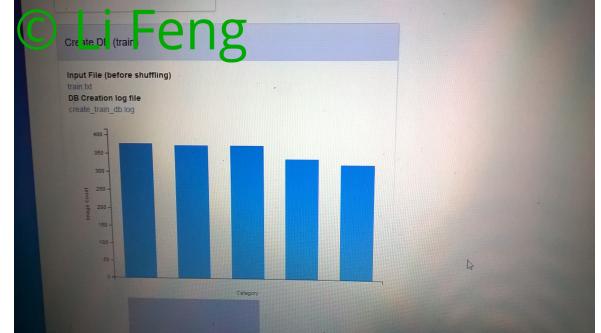
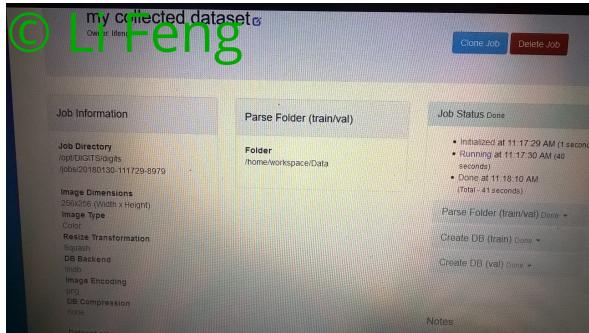
The model (Model Project 1 epochs 9) was also tested with the (evaluate) command in a terminal window and model accuracy was 75.4098360656% and average inference time over 10 run were between 4.9649ms - 5.52061 ms on p2.xlarge instance.



This trained model (Model Project 1 epochs 9) files are located at: ./supplied_data/model/ folder which contains: deploy.prototxt, info.json, labels.txt, mean.binaryproto, original.prototxt, snapshot_iter_2133.caffemodel, solver.prototxt, and train_val.prototxt.

4 Results on Collected Data - Dataset

A dataset (i.e. my collected dataset) was created from the collected data as can be seen below. The collected data contained images of five different items, each item with > 400 images. The image sizes were 256x256 RGB. Images encoding is png.



Images of (my collected dataset) were explored and two sample images are displayed below.

