Robotic Inference with DIGITS

Ge Yao, alexgecontrol@qq.com

Abstract—An abstract is meant to be a summary of all of the relevant points in your presented work. It is designed to present a high-level overview of the report, providing just enough detail to convey the necessary information The abstract may often mention a one-sentence summary of the results. While the type of voice chosen for the paper (active or passive) may be up for debate, you should avoid the use of I and me in the report. It usually is kept to a length of 150 - 200 words. Example: You should not write, I present two different neural networks for classifying my data. Instead, you should try to say, Two different neural networks are used for classification. In this paper two neural networks for image classification are built using NVIDIA DIGITS work flow. One network is built upon the supplied data to classify objects placed on a conveyor belt. Another is trained on the collected Tibetan character data from Minzu University of China to classify Tibetan handwritten digits. Results show that the average inference time and classification accuracy of the two networks meet the specification.

Index Terms—Image Classification, NVIDIA DIGITS.

1 Introduction

A Basic problem in robotics is the classification of objects from a live camera video stream. Two of possible scenarios are the classification of objects on a conveyor belt and the recognition of Tibetan characters for intelligent scanner. In this cases both inference speed and classification accuracy matter. The advent of embedded GPU computing board like Jetson TX2 makes deploying deep neural network for real time robotics applications possible. Besides, NVIDIA DIGITS work flow further simplifies the task by providing us an easy-to-use web interface that turn data set preparation and model training into a series of button clicks.

In this paper the end-to-end work flow for building real time deep network solutions for the above two scenarios will be illustrated.

2 Background / Formulation

The network architectures for the two applications are determined as follows.

2.1 Supplied Dataset

For object classification on a conveyor belt using supplied data, GoogLe Net is used. It has larger capacity than LeNet while has smaller number of parameters than AlexNet. This combination gives it the great potential to strike a balance between inference speed and accuracy.

2.2 Collected Dataset

For the Tibetan handwritten digits classification, LeNet [1] is used. It has proven success on famous MNIST data set and has high chance to extend its success on similar tasks.

For both networks, Adam optimizer with initial learning rate at 0.001 is used. This is the best practice recommended by Andrew Ng in his famous deep learning courses. Due to the limited complexity of the two tasks, 10 epochs are also used for both cases.

3 Data Acquisition

The specifications of the two data sets are as follows.

3.1 Supplied Dataset

For object classification on a conveyor belt, data set consists of photos taken from a Jetson mounted over a conveyor belt. Each photo is a 8-bit RGB PNG image with the size of 500x500. There are three categories and the number of each is shown in Tab.1.

TABLE 1
Supplied Dataset

Category	Number
Bottle	4568
Candy Box	2495
Nothing	3031

Sample records are shown in Fig. 1.

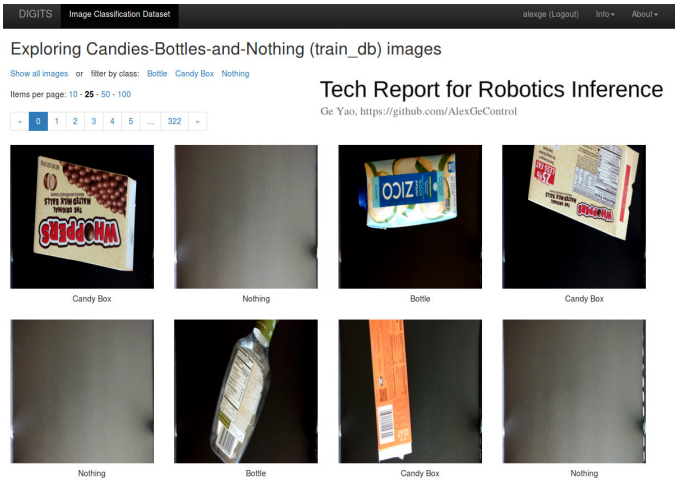

Fig. 1. Samples from Supplied Data

The training set is provided by Udacity and can be accessed from file system. The test set can only be accessed from the provided evaluate script.

3.2 Collected Dataset

For the Tibetan handwritten digits classification, data set consists of scanned Tibetan manuscripts from Minzu University of China. Each photo is a grayscale image with the size of 28x28. There are 10 categories and the number of each is shown in Tab.2.

TABLE 2
Collected Dataset

Category	Number
0	1597
1	1707
2	1498
3	1453
4	1574
5	1080
6	1294
7	1042
8	1314
9	1657

Sample records are shown in Fig. 2.

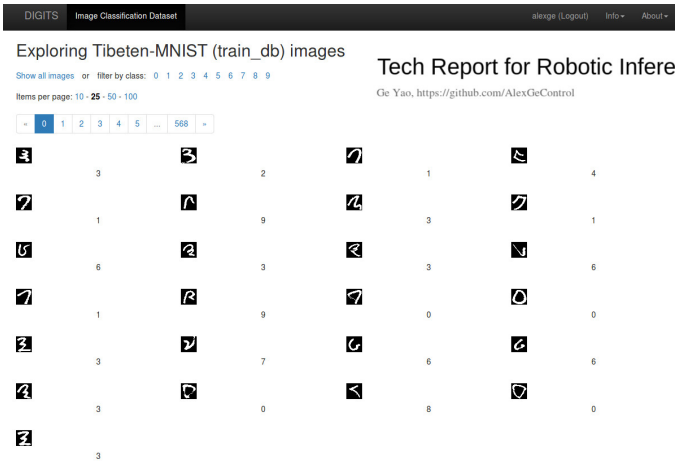

Fig. 2. Samples from Collected Data

The training, validation and testing sets are splitted as 80%:10%:10%.

4 Results

4.1 Supplied Dataset

For the Tibetan handwritten digits classification, the model training curves are shown in Fig. 3.

Its KPIs are evaluated using the script provided by Udacity. The output is shown in Fig. 4.

Its inference speed and average accuracy are listed below for easy reference:

- Mean Inference Speed: 5.13968 ms
- Mean Classification Accuracy: 75.40984%

Which meets the performance requirements of Udacity.

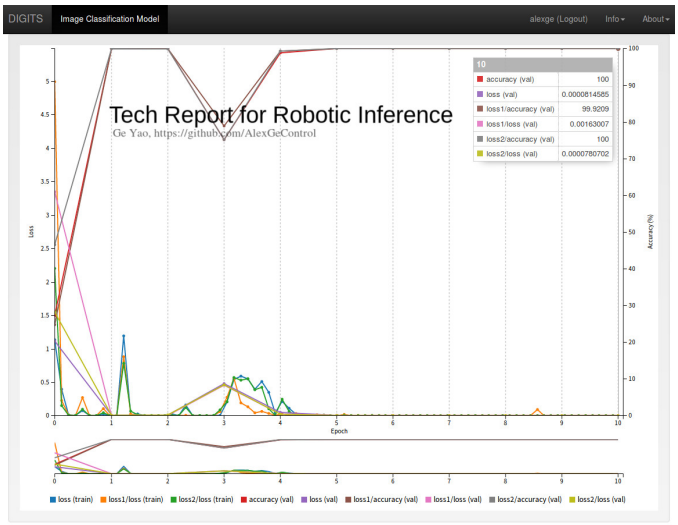

Fig. 3. Model Training Curves for Supplied Data

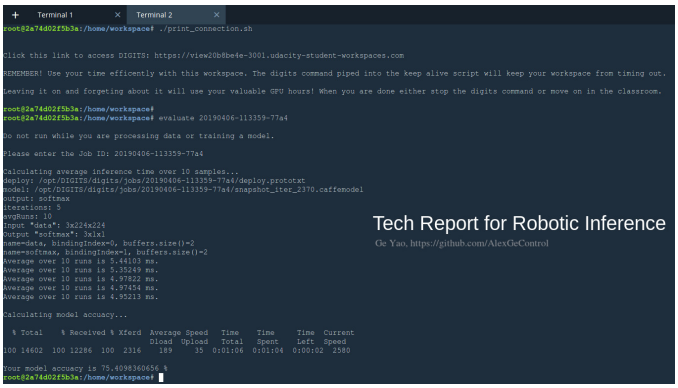

Fig. 4. KPI Evaluation for Supplied Data

4.2 Collected Dataset

For the Tibetan handwritten digits classification, the model training curves are shown in Fig. 5.

Its inference speed is not evaluated. Its average accuracy is listed below for easy reference:

- Mean Classification Accuracy: 98.60424%

5 Discussion

5.1 Supplied Dataset

From the model training curves, the training accuracy and the validation accuracy argees very well with each other. However, a larger difference of around 25% exists between the training accuracy and the test accuracy. This indicates the model is over fitted on training set so it fails to generalize on the test set. One possible reason could be that the test set is collected from a different distribution. Techniques from [2] may solve this situation.

5.2 Collected Dataset

The network performs well on both training and test sets. The reason for its success is the underlying distribution of collected data is similar to that of MNIST and LeNet has a long proven success on MNIST.

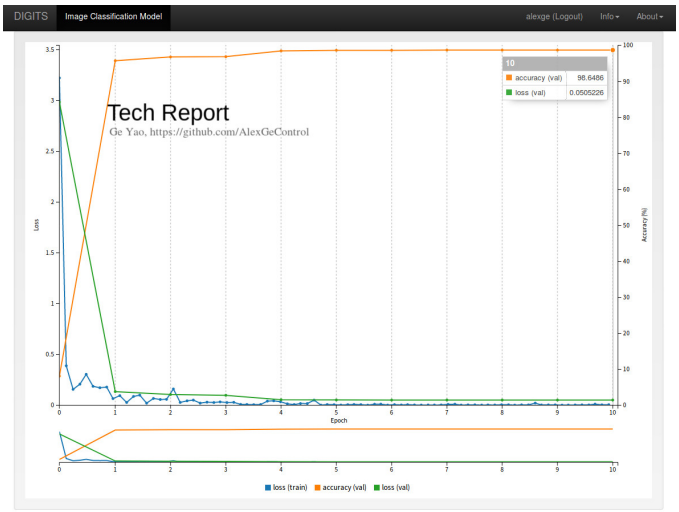

Fig. 5. Model Training Curves for Collected Data

6 Conclusion / Future work

In this paper two neural networks are built using NVIDIA DIGITS work flow. They can classify objects on the conveyor belt and recognize digits from Tibetan characters. The inference speed and the average accuracy of the network for the supplied data meet the requirement. The accuracy of the network for the collected data is also applicable for proposed application.

In future work, the network on supplied data will be deployed on NVIDIA Jetson TX2 board to evaluate its actual performance in real-time scenarios. Besides, more data will be collected in the hope of reducing the performance gap between training and test sets.

References

[1] O. Matan, H. S. Baird, J. Bromley, C. J. C. Burges, J. S. Denker, L. D. Jackel, Y. LeCun, E. P. D. Pednault, W. Satterfield, C. E. Stenard, and T. J. Thompson, "Reading handwritten digits: A ZIP code recognition system," *IEEE Computer*, vol. 25, no. 7, pp. 59–63, 1992.

[2] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," *CoRR*, vol. abs/1605.07678, 2016.