Robotic Inference: Classification Network

Hsin-Wen Chang

Abstract—In this project,NVIDIA DIGITS workflow are used for rapidly prototype ideas that can be deployed on the Jetson near real time. With DIGITS prototype classification networks and training a CNN on the supplied data and train another CNN using a combination data set from the GTI vehicle image database, the KITTI vision benchmark suite, and examples extracted from the Udacity Self Driving Car Nanodegree - Vehicle Detection project video itself [7].

Index Terms—Robotic inference, image classification, Udacity, NVIDIA DIGITS, deep learning.

1 Introduction

Machine learning methods base on Deep learning neural network such as Classification Network and Detection Network achieved remarkable result in a wide variety domains such as detect skin cancer,Facial recognition and self-driving car to detect bike, pedestrian and other vehicle to reduce traffic accident. In this paper VGG16, AlexNet and GoogLeNet will be trained with Supplied data set and a combination data set [7][8] to classify cars, pedestrian and not car objects. Selected model will be train and evaluate on Supplied data set from Udacity then train with the mixed public data set.

1.1 Supplied Data set

In this section we will briefly introduce the supplied data set from Udacity. The following are some examples of the data:

Fig. 1. Supplied Data set.

1.1.1 *Supplied Data set*

Udacity supplied data set contain candy boxes, bottles, and nothing. Photos taken from a Jetson mounted over a conveyor belt. By training pictures of candy boxes, bottles, and nothing (empty conveyor belt) for the purpose of real time sorting. This kind of design can be extrapolated to many things that require real time sorting.

2 Background / Formulation

These example images come from a combination of the GTI vehicle image database, the KITTI vision benchmark suite, and examples extracted from the project video itself. Penn-Fudan Database for Pedestrian Detection and Segmentation.

The intuition is from:

- Can you form a model that can tell the difference between cars or not cars?
- Can you form a model that can tell the difference between Pedestrian and cars?

Explain why you chose the network you did for the supplied data set and then why you chose the network used for your robotic inference project. [1]

3 Data Acquisition

Udacity supplied data set contain 3 classes Bottle, candy box and nothing. The combination of public data set is collect from Udacity Self-Driving car Engineer Vehicle detection project[7] and Penn-Fudan Database for Pedestrian Detection and Segmentation [8].

Class	Images	Image Shape
Candy Box	1871	256x256x3
Bottle	3426	256x256x3
Nothing	2273	256x256x3
Total	7570	

TABLE 1
Count of supplied data images

4 Results-

This is typically the hardest part of the report for many. You want to convey your results in an unbiased fashion. If you results are good, you can objectively note this. Similarly, you may do this if they are bad as well. You do not want to justify your results here with discussion; this is a topic for the next session. Present the results of your robotics project model and the model you used for the supplied data with the appropriate accuracy and inference time For demonstrating your results, it is incredibly useful to have some charts, tables, and/or graphs for the reader to review. This makes ingesting the information quicker and easier.

Fig. 2. Combination data set.

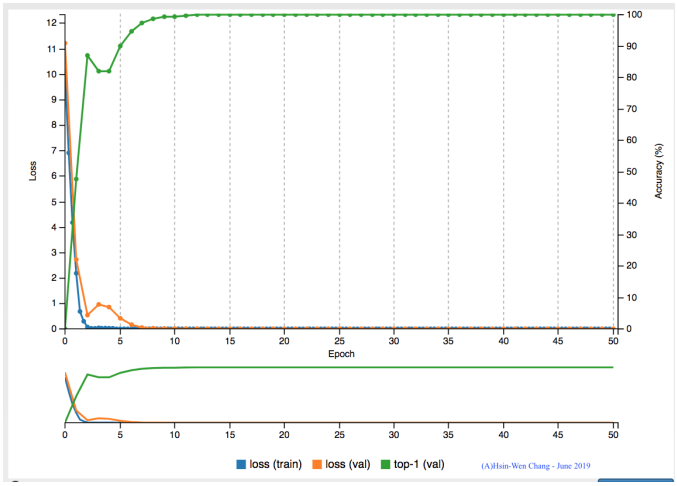

Fig. 3. VGG16 pretrained model

5 Discussion

There is a slightly improvement when switch model from AlexNet to GoogLeNet while GoogLeNet need more time to train and perform inference results. We can use improved hardware like Jetson TX2 and Tensor RT 3.0 to improve this in a very fast time.

6 Conclusion / Future work

GoogLeNet have better accuracy for sensor like camera and video on perform image classification task on detect pedestrian, car and their background environment with higher accuracy than AlexNet but also will cost longer time to inference results and training images. Base on the research and studying in Udacity Self-Driving Car Engineer Nanodegree - Semantic Segmentation Project [9], the inference time to

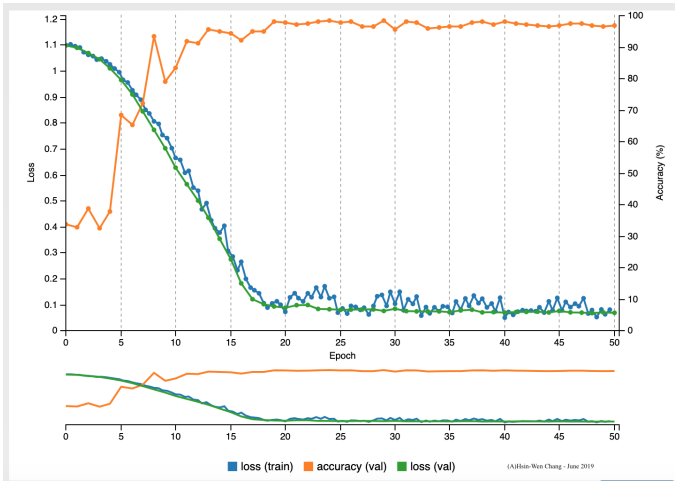

Fig. 4. AlexNet

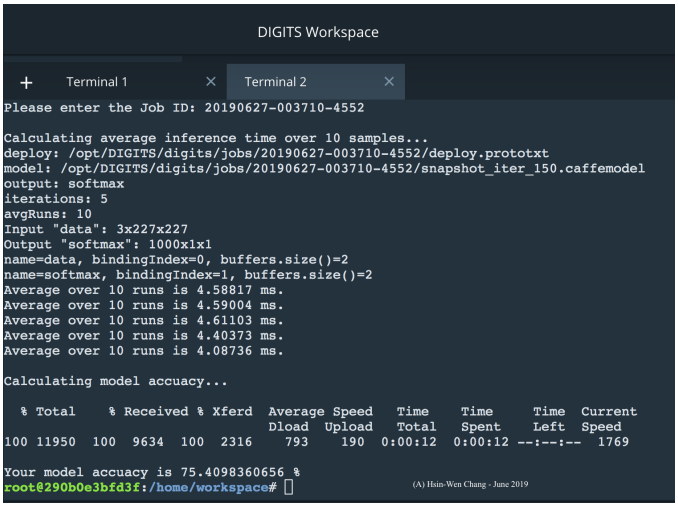

Fig. 5. Evaluate training result in DIGITS Workspace

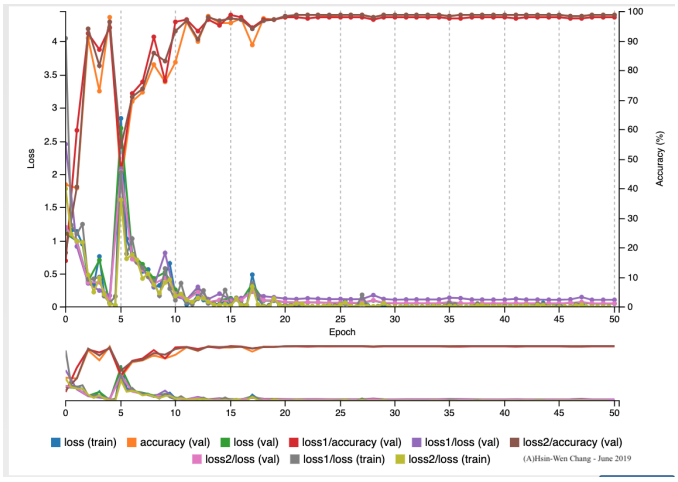

Fig. 6. Combination data train with GoogLeNet

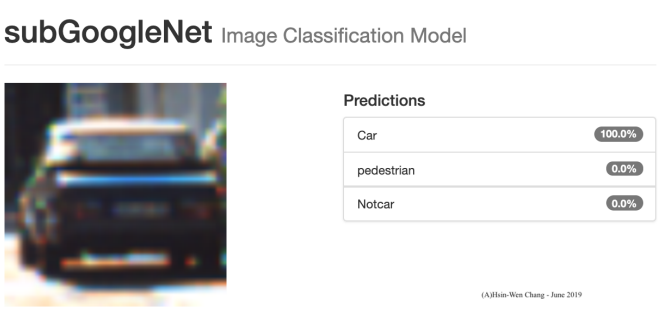

Fig. 7. GoogLeNet Model Inference Result: Car

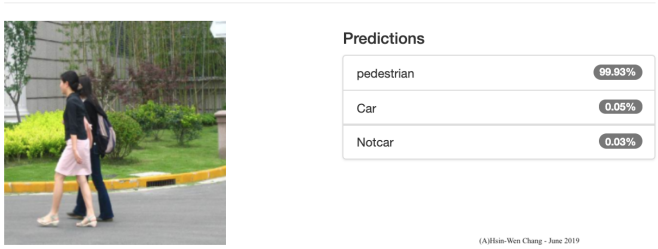

Fig. 8. GoogLeNet Model Inference Result: Pedestrian

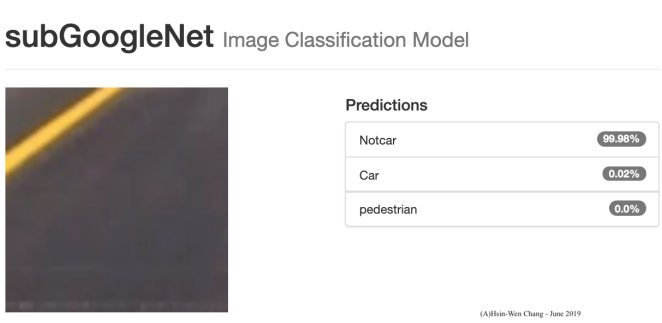

Fig. 9. GoogLeNet Model Inference Result: Not Car

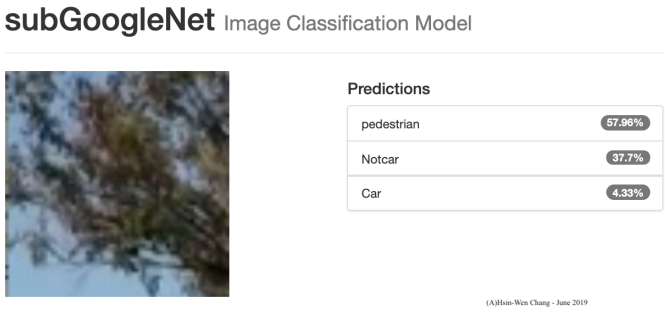

Fig. 10. GoogLeNet Model Inference Result: Not Car

perform real-time computer vision is crucial hence in the future work we can appliy transfer learning skill on VGG16 for less inference time and higher accuracy. Another way to improve accuracy base on our current model we can add more image data but this will prolong training time and cost more storage space.

REFERENCES

[1] L. Lamport, *LATEX: a document preparation system: user's guide and reference manual*. Addison-wesley, 1994.

[2] The Caltech Database (Computational Vision at California Institute of Technology, Pasadena), http://www.vision.caltech.edu/html-files/archive.html. Accessed 14 May 2011.

[3] R Fergus, P Perona, A Zisserman, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, 1622 June 2003.

[4] The TU Graz-02 Database (Graz University of Technology). Accessed 14 May 2011.

[5] A Opelt, A Pinz, in Proceedings of the 14th Scandinavian Conference on Image Analysis, Joensuu, Finland, 1922 June 2005.

[6] Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite, Andreas Geiger, Philip Lenz and Raquel Urtasun, Conference on Computer Vision and Pattern Recognition (CVPR).

[7] Udacity Self-Driving car Engineer Vehicle detection project.

[8] Object Detection Combining Recognition and Segmentation. Liming Wang, Jianbo Shi, Gang Song, I-fan Shen. To Appear in ACCV 2007.

[9] Udacity Self-Driving Car Engineer Nanodegree - Semantic Segmentation Project.