# 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 [6][7] 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?

NeuralNetworkArchitectures.png

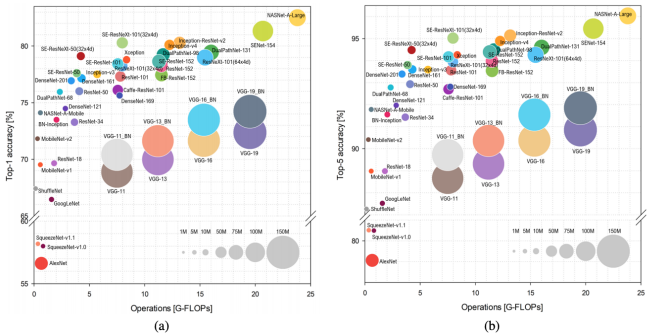


Fig. 2. Benchmark Analysis of Representative Deep Neural Network Architectures [9].

Compare to different Neural Network architecture(see Fig. 2)and accuracy. VGG16 pretrained model, AlexNet and GoogLeNet seems to be a good choice.

## 3 Data Acquisition

Udacity supplied data set contain 3 classes Bottle, candy box and nothing. The combination of public data set is collect

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

TABLE 1
Count of supplied data images

from Udacity Self-Driving car Engineer Vehicle detection project[6] and Penn-Fudan Database for Pedestrian Detection and Segmentation [7].


Fig. 3. Combination data set.

## 4 Results-

In this section, we will discuss the result as two part. The training result on Udacity supplide data set and on the mixed public data set from Penn-Fudan Database for Pedestrian Detection and Segmentation and the GTI vehicle image database, the KITTI vision benchmark suite, and examples extracted from the Udacity Self-Driving car Engineer Vehicle detection project [6] video itself. Supplied data set was used on test VGG16 pretrained model and AlexNet. Acurracy of VGG16 pretrained model is really close to near 100 percent slightly higher than AlexNet which accuracy is close to 96-97 percent when applied 50 epochs. After test on supplied data set as reference, then we choose GoogLeNet to train on our mixed public data set. The accuracy is higher than AlexNet when using 50 epochs same as reference paper [10]. Result are show in figure 4-11.

## 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 performance 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

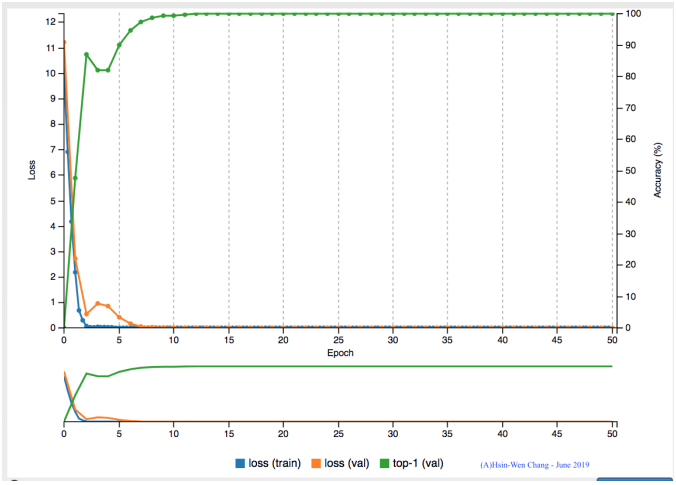


Fig. 4. VGG16 pretrained model

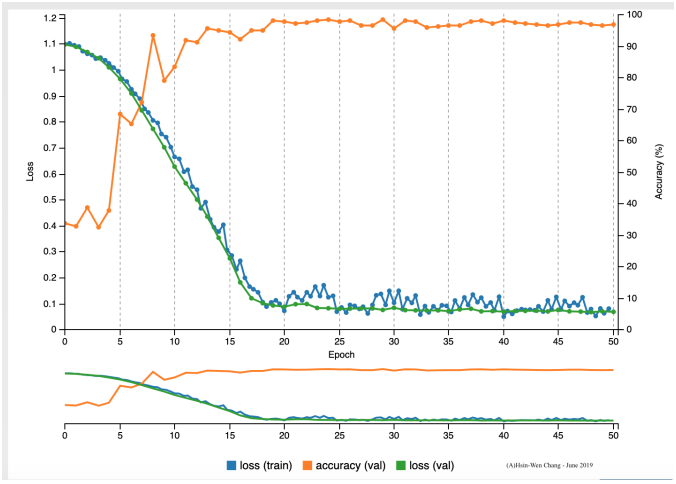


Fig. 5. AlexNet

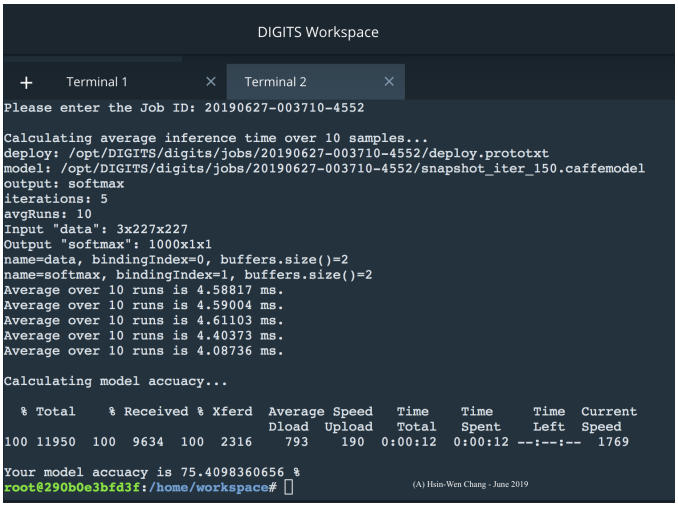


Fig. 6. Evaluate training result in DIGITS Workspace

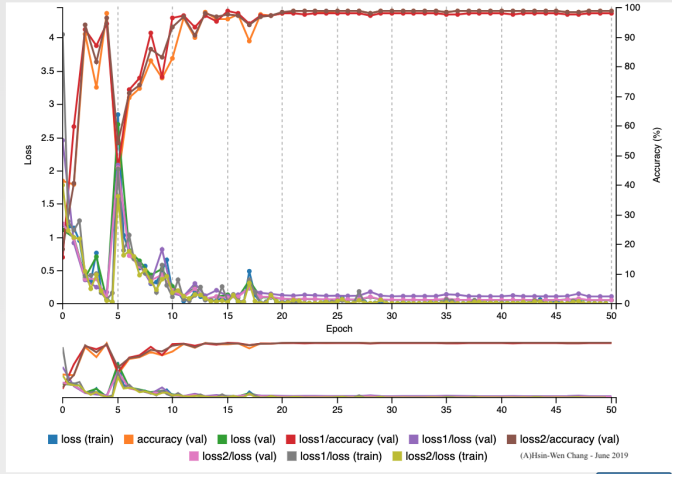


Fig. 7. Combination data train with GoogLeNet

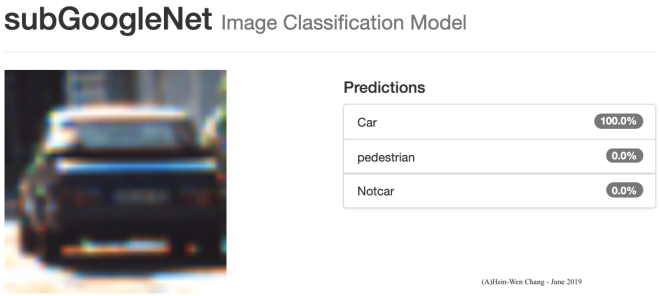


Fig. 8. GoogLeNet Model Inference Result: Car

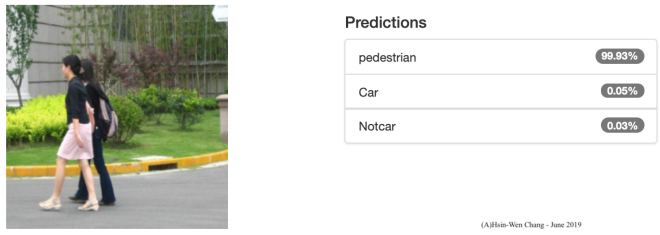


Fig. 9. GoogLeNet Model Inference Result: Pedestrian

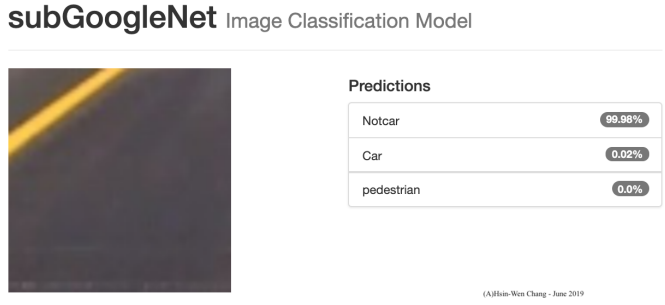


Fig. 10. GoogLeNet Model Inference Result: Not Car


Fig. 11. GoogLeNet Model Inference Result: Not Car

higher accuracy than AlexNet but 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 perform real-time computer vision is crucial hence in the future work we can apply 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] The Caltech Database (Computational Vision at California Institute of Technology, Pasadena), http://www.vision.caltech.edu/html-files/archive.html. Accessed 14 May 2011.

[2] 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.

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

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

[5] 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).

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

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

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

[9] Benchmark Analysis of Representative Deep Neural Network Architectures. SIMONE BIANCO1,REMI CADENE2,LUIGI CELONA1 AND PAOLO NAPOLETANO,https://arxiv.org/pdf/1810.00736.pdf,ACCESS.2017.