

Robotics Inference Project

Tan Kai Xiong

Abstract—In this project, an image classification neural network is trained by the user supplied data. It must capable of classify an object in a scene. Neural networks are trained and tested using the NVIDIA DIGITS work space. Using existing Deep Neural Network, example such as LeNet, AlexNet and GoogleNet for image classification. The neural network must achieves at least 75 percents accuracy and an inference time of less than 10 ms.

Index Terms—Robot, IEEEtran, Udacity, L^AT_EX, deep learning.

1 INTRODUCTION

IN robotics, there is a common problem of real time object classification from a live camera video feed. Speed and accuracy is required for a robot to perform object recognition in a real world environment. In an intelligent conveyor belt system, to identify object on the moving conveyor belt. Server based image classification unable to perform real time inference. The introduction of NVIDIA Jetson is an embedded computer that capable of running deep neural network. The edge device will fulfill the need for real-time inference. A fast and efficient neural network is required to achieve low latency inference time and high accuracy result.

2 BACKGROUND / FORMULATION

The NVIDIA DIGITS is an open source software which provide interactive user interface to build a deep neural network. The NVIDIA DIGITS workspace provide three types of deep neural network (LeNet, AlexNet and GoogleNet) as their standard models for image classification. The training parameter is listed below.

- Solver type : Adam (Adaptive Moment Estimation)
- Base Learning Rate : 0.0001
- Epochs : 30

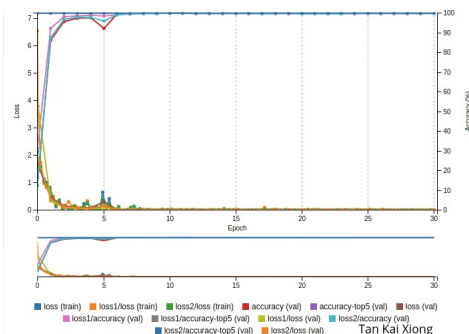


Fig. 1. Training Loss Function Graph.

3 DATA ACQUISITION

Logitech web camera is used for capturing image. They are "game pad", "marker", "can drink", "teddy bear" and

"none". There are total of 55% of the dataset is used for training set, 25% for validation set and 10% for testing set.



Fig. 2. Five Categories of object.

- Number of Categories : 5
- Training Images : 1630
- Validation Images : 627
- Test Images : 250
- Image Dimensions : 256 x 256
- Image Type : RGB

4 RESULTS

Overall, the prediction result for the test set is greater than 90%. The inference speed is around 15 to 20 FPS on Jetson nano embedded computer.

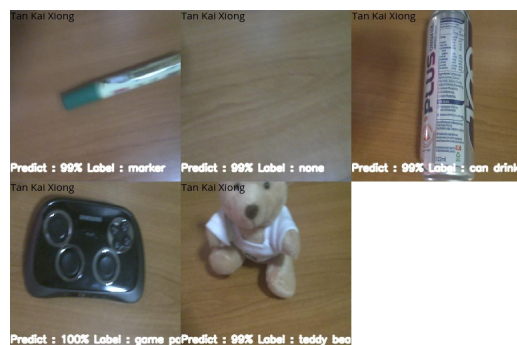


Fig. 3. All Categories Prediction result.

5 DISCUSSION

Both inference speed and accuracy are important for image classification. Fast inference speed improve the robot performance efficient. Accuracy inference result help the robot to make the correct decision,

6 CONCLUSION / FUTURE WORK

Image classification limit to identify 1 object class per frame. The image dataset can be used for further annotation of the object X and Y position. With the position annotation, an object detection network can be built using the following annotated dataset. Object detection network can detect multiple objects and identify their categories.

The use of NVIDIA DIGITS make training of Deep Neural Network to be easy. It come with a lot of data analysis tool.