Robotic Inference

Noriaki Ibe

Abstract—The first project in term 2 of the Robotics Software Engineer Nanodegree Program requires students to initiate their own inference project inclusive of data acquisition. The project builds on the initial reference project for digit image recognition inside the supplied Nvidia Digits environment.

Index Terms—Robot, IEEEtran, Udacity, LaTeX, deep learning.

1 Introduction

THE Classification is one of the popular tasks in A.I. for now. Many approaches has been done in recent years due to improving of processing power such as GPU etc. Additionally, advanced technology and environments are available for anyone with reasonable prices. Anyone can try Classification and this project is one of the example to show the A.I. technology is close. In this project, NVIDIA's DIGITS platform is used for classification. There are two parts in the project of image classification:

- P1 : The supplied Dataset pictures of candy boxes, bottles, and nothing (empty conveyor belt).
- P2 : The collected Dataset pictures of Dogs, Cats, and nothing (scenery).

2 Background / Formulation

The project requires students to perform:

1) P1: Using the supplied data in the digits image, create a network that achieves at least 75 percent accuracy and an inference time of less than 10 ms.

The supplied data are 256x256 3 channel color (RGB) images. NVIDIAs DIGITS platform provides 3 types of Networks:

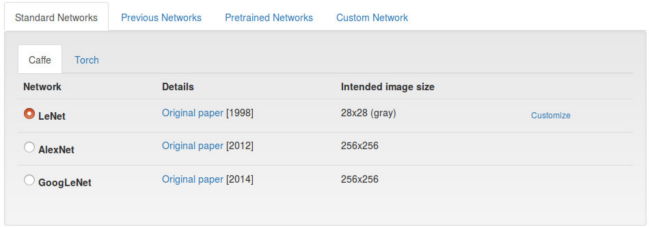


Fig. 1. NVIDIAs DIGITS platform

The LeNet is not applicable due to image scale and the AlexNet would have advantage due to inference time.

So the project shall used the AlexNet mainly, but the GoogleNet shall be tried, too. Regarding accuracy, it's clearly confirmed for P1 such as 75 percent accuracy. It was not necessary to be a 100 percent accurate so that it shall be similar or let's say at least over 70 percent.

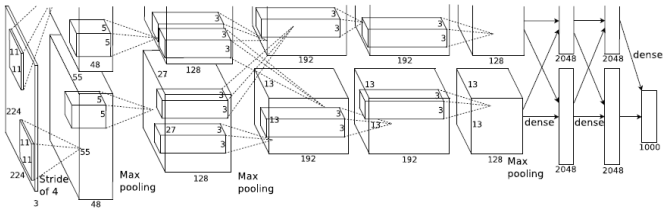


Fig. 2. AlexNet Architecture

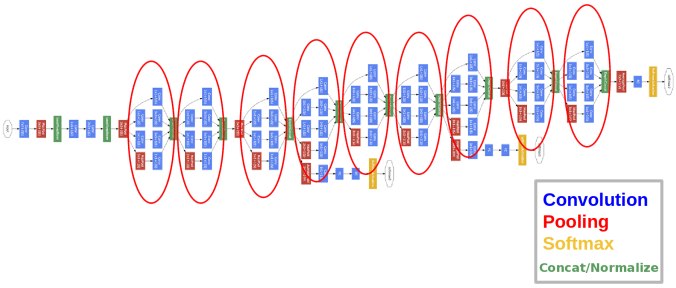


Fig. 3. GoogleNet Architecture

3 Data Acquisition

Regarding P1, the supplied dataset can be found in the data directory ($/data/P1_data$).

The provided data has the following file structure:

```
P1_data/
        Bottle/
            Bottle_1.png
            Bottle_2.png
        Candy_box/
            Candy_box_1.png
            Candy_box_2.png
        Nothing/
            Nothing_1.png
            Nothing_2.png
```

Regarding P2, there was several ways to collect data. The CIFAR-10 dataset etc is open-source. But collecting own data is one of task and the project selected the way to gather images from Web. In case of gathering images from Web, there are three ways as follows:

Additionally, code to collect data per keyword is shared on git by Yutaka N.(https://github.com/nknytk) and the

Fig. 4. Sample images of P1

TABLE 1
Image collecting

	Yahoo(Japan)	Bing	Google
scraping or API	scraping	API	API
Register	no	need	need
Price(/query)	Free	3 USD/1K	3 USD/1K
No of Image(/query)	60	700 900	100

project($https://github.com/nknytk/face-classifier-cnn/blob/master/tools/data_collector/collect_img.py$). Then, the project used several keywords were used such as "dog", "cute dog", "bulldog", "cat" etc. to make Classification task of Dogs, Cats, and nothing (scenery). After collected, those images were resized into 256x256 3 channel color (RGB) images without any consideration of location of target.

The collected dataset can be found in the data directory ($/data/P1_data$).

```
P2_data/
        Dog/
            Dog_0.png
            Dog_1.png
    ...
            Dog_589.png
        Candy_box/
            Cat_0.png
            Cat_1.png
    ...
            Cat_594.png
        Nothing/
            Nothing_0.png
            Nothing_1.png
...
            Nothing_633.png
```

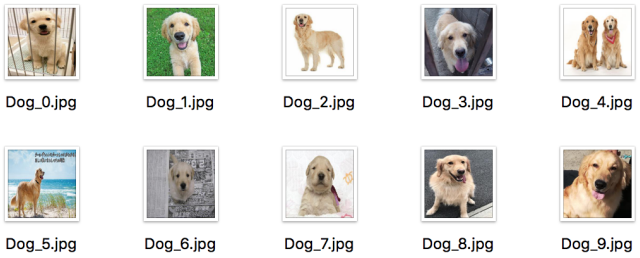


Fig. 5. Sample images of Dogs

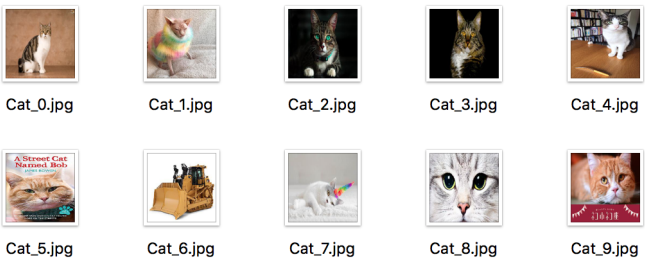


Fig. 6. Sample images of Cats

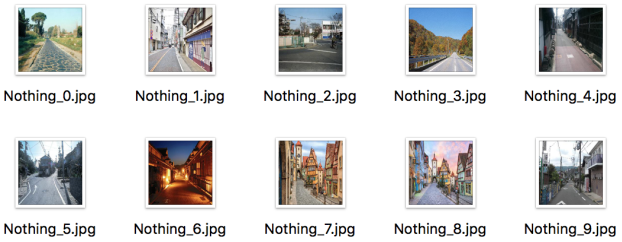
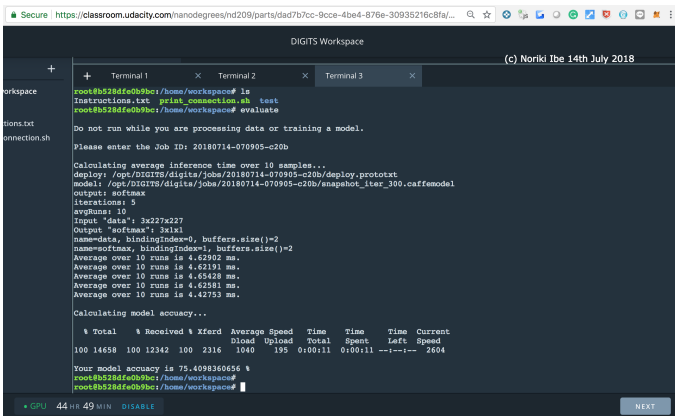


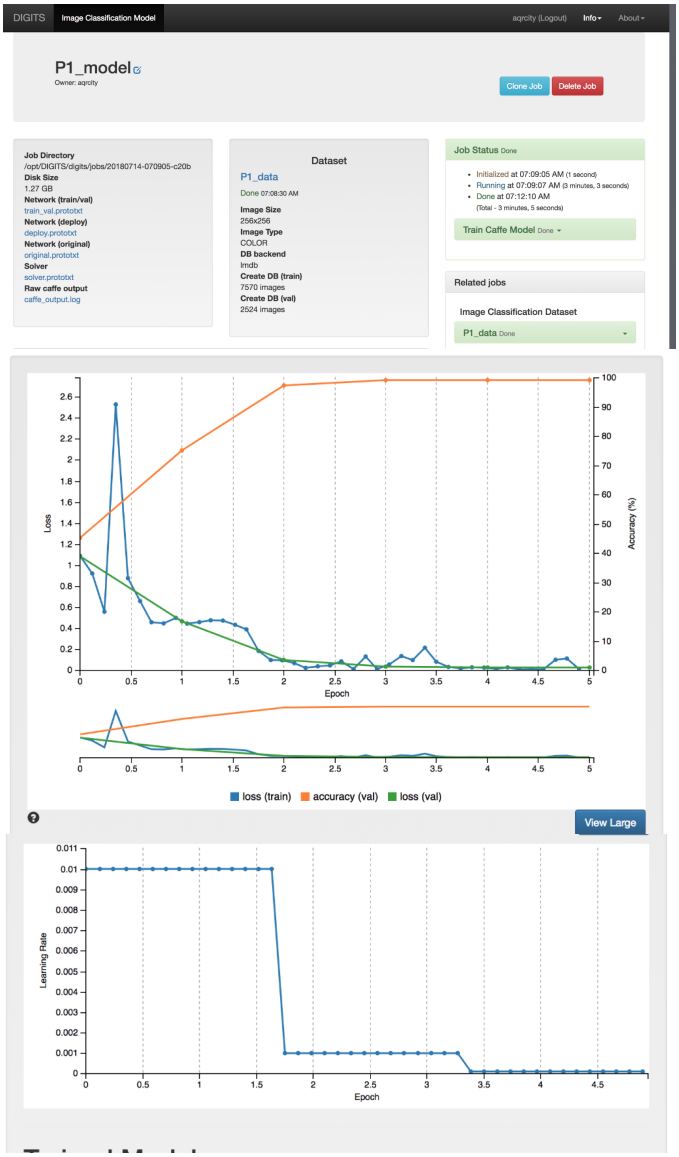
Fig. 7. Sample images of Nothing

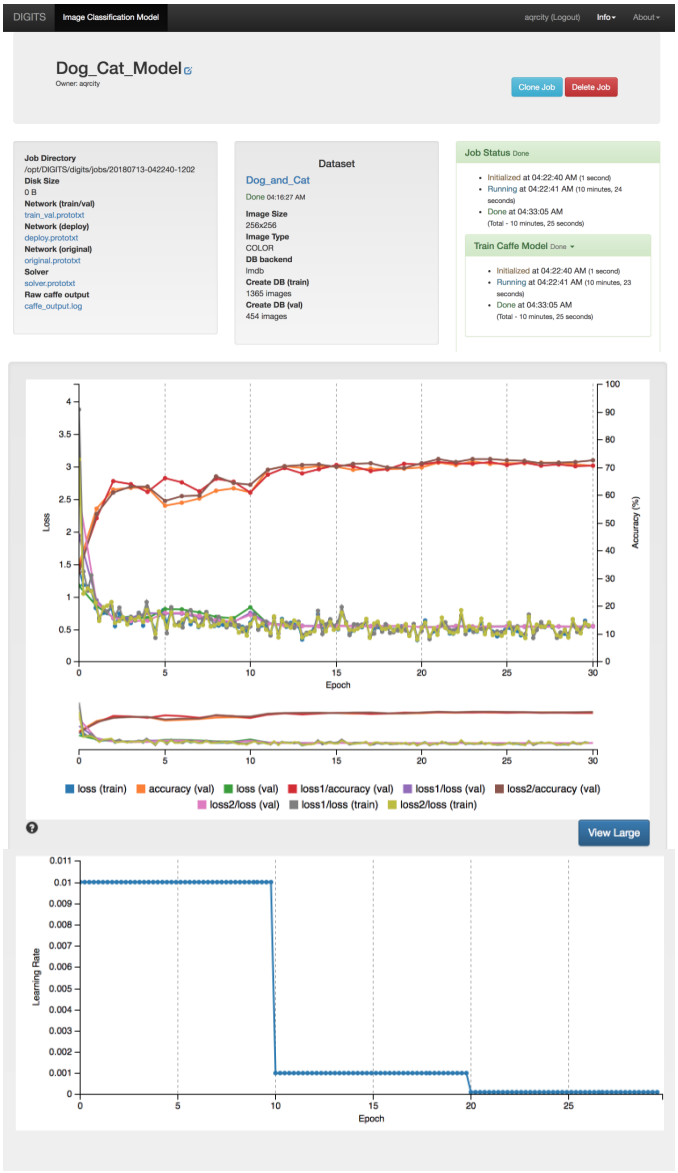
4 RESULTS

Regarding P1, Both AlexNet and GoogLeNet models were applied with P1 dataset provided by the lesson in moving belt image classification. But GoogLeNet was less performed in light of inference time. The results with AlexNet achieved the requirement of Udacity lesson (least 75 percent accuracy and an inference time of less than 10 ms.)



Regarding P2, AlexNet was applied with 30 Epochs.





It's performed over 70 percent anyhow.

And the terminal output after running the evaluate command as follows:

5 DISCUSSION

Even not limited to dog etc., there are several dataset if it's common staff and those shall be available in CIFAR, CoCo dataset etc. But, it's not guaranteed that Image dataset is available for target objects. If Classification's performance was poor without those dataset, Classification would not be one of the popular tasks in A.I. for now. Especially, around 500 images shall be enough to make Classification as shown. In addition, the performance with the images collected by scraping is over 70 percent too, it's not so bad. With Using Nvidia's DIGITS platform, Anyone can apply GUP for model.

6 CONCLUSION / FUTURE WORK

The project could perform at least and it's great practice to use advanced environment such as Nvidia DIGITS. On the other hand, there are several ways to improve the performance itself of this project. Regardless P1 and P2, there should be huge room to improve models in light of network architecture and parameters. Some projects require higher accuracy, it depends on the project. In P2, there are some approaches to improve in collecting and preparing data. The project is just using resizing regardless size of original image. One of the idea to improve dataset is removing too large images and too small images. In scraping, data was collected several images would be outliers. Then, some models should be applied into the real world such as Jetson TX2 etc.