FEBRUARY 1, 2019

RoboND-Robotics-Inference

ABSTRACT

In this project, fully convolutional neural network made by AlexNet was trained with 30 epochs to classify sign-language numbers and translate them to the correct written form. DIGITS which is a platform made by nvidia for Neural Networks application helps in simplifying implementation of AlexNet, dealing with datasets, training and evaluating models. Also, to provide needed process power GPU workspace made by Udacity is used.

INTRODUCTION

Although, people with difficulties in hearing can understand sign language easily. It is still hard for other people to understand them, this is because only very small percentage of us learn this language.

Now with this advance in neural network. Neural networks can be used to translate sign language to its written format easily; it only needs time to be trained at the beginning, and after that with a good network it will perform in a Real time with no noticeable delays.

Many Researches were made around this idea in the latest 10 years. Indeed there is already some applications that came to light.

This project represent a prototype of neural network that could be implemented in Jetson TX2.This Network with jetson enables us to build a robot that can aid people with difficulties in their life.

BACKGROUND / FORMULATION

There were three options for this project neural network: AlexNet, GoogleNet and LeNet.

AlexNet is chosen here based on some reasons. First, the size of the image of the dataset which is 256x256, this size allows us to obtain much data from training than LeNet. Also, it is the fastest Neural network available which makes it preferable than GoogleNet. Moreover, with small tuning in parameters the needed efficiency and speed is obtained.

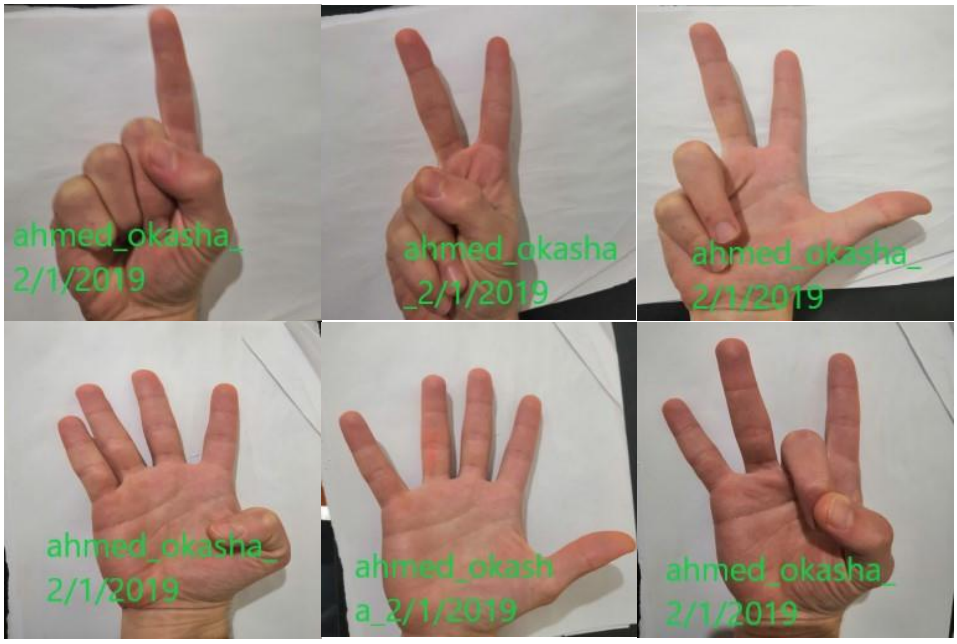
DATA ACQUISITON

Dataset used in this project is free dataset obtained from https://www.kaggle.com/ .Each class in this data sets has range of 210 samples. After downloading this dataset it was uploaded to Udaity workspace at path"/home/workspace". Then unzipped after enabling GPU option to data directory "~/data".Data is arranged in this form.

```
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
```

After arranging data it is trained in DIGITS workspace and resized during this step to match AlexNet required size. Finally grayscale is chosen because color isn't important in our case and can lead to over fitting.

Also, some samples is collected for testing from mobile camera.

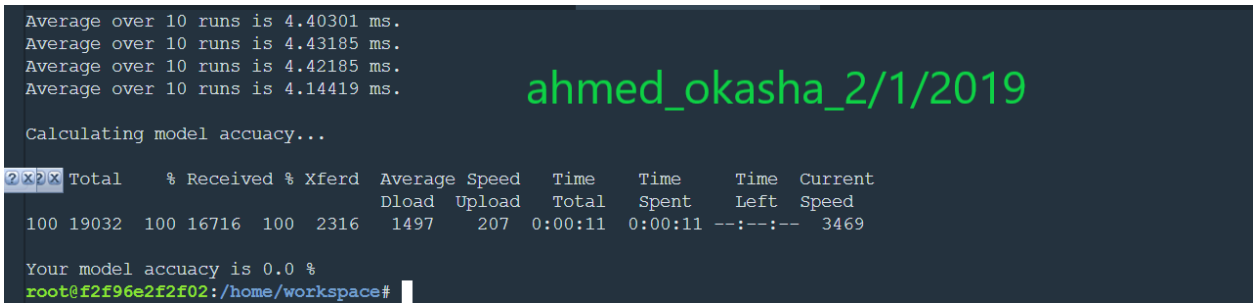


RESULTS

After model is trained, the output obtained results was good and above the required results.

AVERAGE TIME

Obtained Average time from model over 10 times =4.1 ~ 4.4 ms.



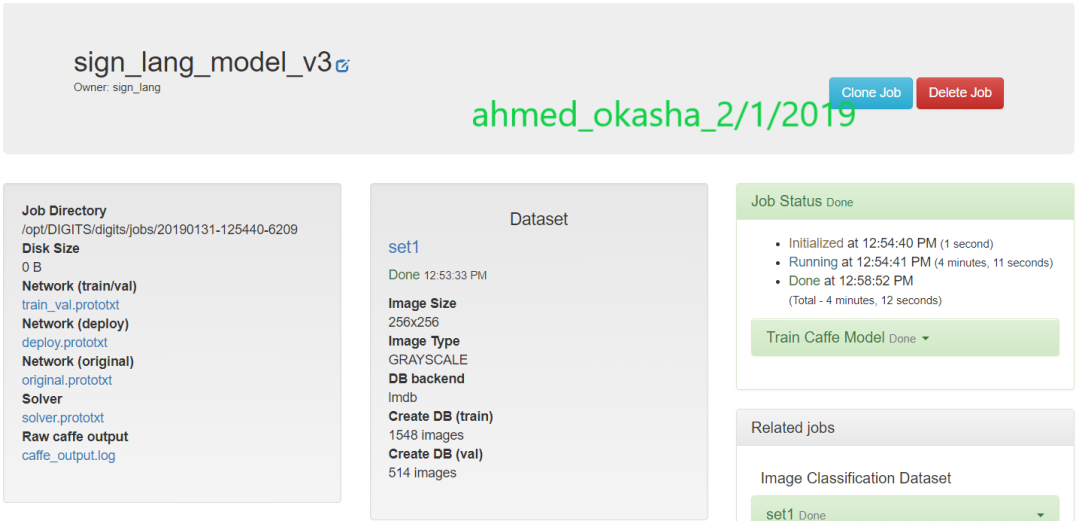
MODEL ACCURACY

Final accuracy at epoch 30 was above 95 % from charts.

From experiment with testing set of sample took from mobile camera we got 2 wrong predictions from 10 samples so accuracy can be considered 90% and roughly 80%.

TRAINING MODEL TIME & STATISTICS

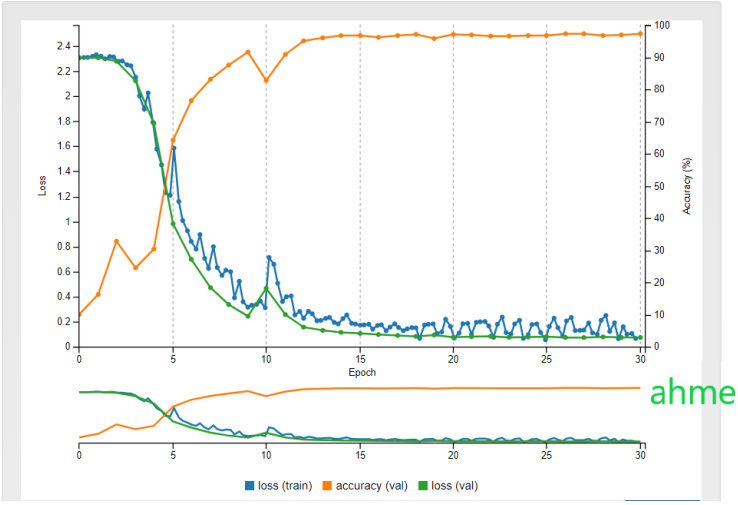
This model took 4 minutes, 12 seconds to be trained.



LOSS STATISTICS & LEARNING RATE

Loss begins from 2.4 and decreases over epochs until it become in 0.1 range at epoch 30.

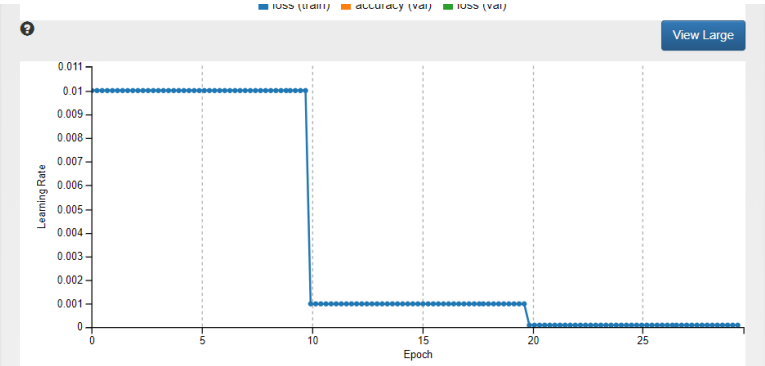
To avoid overfitting of model, and to get high accuracy we start with high learning rate at the beginning but when accuracy starts to decrease learning rate is decreased.



Notes

None

ahmed_okasha_2/1/2019



Trained Models

ahmed_okasha_2/1/2019

TESTED SAMPLES

sign_lang_model_v3 Image Classification Model

Predictions

1	83.75%
9	14.07%
3	1.11%
8	0.5%
2	0.48%

Job Status Done

- Initialized at 01:09:04 PM (1 second)
- Running at 01:09:05 PM (3 seconds)
- Done at 01:09:08 PM
(Total - 4 seconds)

Infer Model Done

- Initialized at 01:09:04 PM (1 second)
- Running at 01:09:05 PM (3 seconds)
- Done at 01:09:08 PM
(Total - 4 seconds)

Notes

ahmed_okasha
2/1/2019

sign_lang_model_v3 Image Classification Model

Predictions

2	33.15%
7	31.25%
3	19.06%
6	14.39%
8	1.52%

Job Status Done

- Initialized at 01:09:23 PM (1 second)
- Running at 01:09:24 PM (3 seconds)
- Done at 01:09:27 PM
(Total - 4 seconds)

Infer Model Done

- Initialized at 01:09:23 PM (1 second)
- Running at 01:09:24 PM (2 seconds)
- Done at 01:09:27 PM
(Total - 4 seconds)

Notes

ahmed_okasha_2/1/2019

sign_lang_model_v3 Image Classification Model

Predictions

3	95.47%
2	4.05%
1	0.4%
7	0.06%
9	0.01%

Job Status Done

- Initialized at 01:10:21 PM (1 second)
- Running at 01:10:22 PM (3 seconds)
- Done at 01:10:26 PM
(Total - 5 seconds)

Infer Model Done

Notes

None

ahmed_okasha
2/1/2019

sign_lang_model_v3 Image Classification Model

Predictions

4	72.96%
8	14.23%
9	5.51%
5	4.34%
6	1.3%

Job Status Done

- Initialized at 01:09:26 PM (1 second)
- Running at 01:09:28 PM (3 seconds)
- Done at 01:09:31 PM
  (Total - 4 seconds)

Infer Model Done

- Initialized at 01:09:26 PM (1 second)
- Running at 01:09:28 PM (2 seconds)
- Done at 01:09:31 PM
  (Total - 4 seconds)

Notes

ahmed_okasha_2/1/2019

sign_lang_model_v3 Image Classification Model

Predictions

5	99.93%
4	0.05%
9	0.02%
8	0.0%
3	0.0%

Job Status Done

- Initialized at 01:09:34 PM (1 second)
- Running at 01:09:35 PM (3 seconds)
- Done at 01:09:39 PM
  (Total - 4 seconds)

Infer Model Done

Notes

None

ahmed_okasha_2/1/2019

sign_lang_model_v3 Image Classification Model

Predictions

4	75.87%
6	12.84%
8	4.36%
5	3.02%
7	1.56%

Job Status Done

- Initialized at 01:09:38 PM (1 second)
- Running at 01:09:39 PM (3 seconds)
- Done at 01:09:42 PM
  (Total - 4 seconds)

Infer Model Done

Notes

None

ahmed_okasha_2/1/2019

sign_lang_model_v3 Image Classification Model

Predictions

7	64.82%
4	21.04%
8	7.77%
3	4.7%
5	1.08%

Job Status Done

- Initialized at 01:09:43 PM (1 second)
- Running at 01:09:44 PM (3 seconds)
- Done at 01:09:48 PM
  (Total - 4 seconds)

Infer Model Done

Notes

None

ahmed_okasha_2/1/2019

sign_lang_model_v3 Image Classification Model

Predictions

8	99.92%
9	0.04%
7	0.02%
4	0.01%
3	0.0%

Job Status Done

- Initialized at 01:09:55 PM (1 second)
- Running at 01:09:56 PM (3 seconds)
- Done at 01:10:00 PM
(Total - 4 seconds)

Infer Model Done

- Initialized at 01:09:55 PM (1 second)
- Running at 01:09:56 PM (3 seconds)
- Done at 01:10:00 PM
(Total - 4 seconds)

Notes

sign_lang_model_v3 Image Classification Model

Predictions

9	99.06%
8	0.86%
4	0.07%
5	0.0%
3	0.0%

Job Status Done

- Initialized at 01:10:01 PM (1 second)
- Running at 01:10:02 PM (3 seconds)
- Done at 01:10:05 PM
(Total - 4 seconds)

Infer Model Done

- Initialized at 01:10:01 PM (1 second)
- Running at 01:10:02 PM (3 seconds)
- Done at 01:10:05 PM
(Total - 4 seconds)

Notes

sign_lang_model_v3 Image Classification Model

Predictions

0	99.52%
6	0.18%
4	0.13%
1	0.07%
9	0.04%

Job Status Done

- Initialized at 01:10:05 PM (1 second)
- Running at 01:10:06 PM (3 seconds)
- Done at 01:10:09 PM
(Total - 4 seconds)

Infer Model Done

- Initialized at 01:10:05 PM (1 second)
- Running at 01:10:06 PM (2 seconds)
- Done at 01:10:09 PM
(Total - 4 seconds)

Notes

CONCLUSION / FUTURE WORK

To summarize everything up, this report can be considered as small prototype of sign language translator that takes images as input with a sign-language number and translates it to its written form through Neural Network trained models made by DIGITS & Udacity GPU workspaces. This Idea can be used in multi-tasking robots to help people with difficulties in hearing, Also, This idea can be implemented to make sign-language translator application. Moreover, this idea can go further and further to make a multi-language translator that can make communication between us easier and our life happier.