

Robotic Inference

Mustafa Eldeeb

Abstract—Deep Learning is widely used in object classification. Convolutional Neural Networks improved the success of the computers on object detection, recognition, and classification. They are commonly used in Self-Driving Cars, robotics and other fields of machine learning. There exist models such as LeNet, AlexNet, and GoogleNet. Nvidia Digits provides a web interface to utilize these models and create other models. It is possible to upload data and test it with these models. This project includes different classification tasks on image datasets.

Index Terms—Robot, IEEETran, Udacity, L^AT_EX, Deep Learning, Convolutional Neural Networks, Nvidia Digits.

1 INTRODUCTION

THE goal in this project is to utilize Nvidia Digits application to perform classification on image datasets. There are two datasets. The first one is the provided data that contains bottle and candy box images. Besides, there exist samples that do not include any.

The second dataset was The German Traffic Sign dataset [1]. There are 43 different traffic signs. This dataset was selected because it is related with self-driving field and it has many different classes.

Nvidia Digits provide an interface to load the data, create a new model or utilize predefined models. Besides, it enables to customize the models and select hyperparameters. During training, it saves the model on every epoch. This saved models can be used as pretrained models to train more.

2 BACKGROUND / FORMULATION

The problem is to classify objects using Deep Learning. Convolutional neural networks are an excellent choice [2]. Backpropagation provided enhanced learning for Multi-layer Neural Networks.

LeCun et al. developed LeNet to classify digits in 1998 [3]. LeNet was a seven-layer CNN. Krizhevsky et al. built a convolutional neural network which has sixty million parameters, half million neurons and five convolutional layers in 2012 [4]. Szegedy et al. trained another deep neural network in 2014 which has 22 layers [5].

2.1 Task 1: Bottle vs Candy-Box

In the first task, AlexNet and GoogleNet were selected to classify the samples of bottles and candy-boxes. The reason for choosing AlexNet is its speed, and the reason for working with the GoogleNet was its higher accuracy. Five different models trained while completing this task. Since adam and RMSprop help the gradient descent to reach the minima faster, these solver types were allied with AlexNet and GoogleNet. Besides the Stochastic Gradient Descent were selected with AlexNet as the default option. Learning rate 0.01 was a reasonable choice to prevent overfitting and from overshooting. The image sizes were compatible with AlexNet and GoogleNet. Therefore there were no changes in

dimensions. The number of epochs parameter has different values to experiment. The results section discuss the details of the models tried and the outcomes.

2.2 Task 2: Traffic Sign

The selected dataset has 39,209 images in 43 classes for training. Besides, there are 12,530 images to test. The image size is varying between 15x15 and 250x250 pixels. This situation forced the use of LeNet. All samples were resized by squashing to 28x28 during loading. Enlarging the images would result in excessive training. From the experience from the first task, the solver type was selected as stochastic gradient descent. The learning rate was again 0.01 since it provided satisfying results in the previous part of the project.

3 DATA ACQUISITION

The first task does not contain any changes on the data, since the models were compatible with image size of 256x256.

The second task works on the traffic signs data. The main reasons are to select a dataset related to autonomous vehicles and to select a dataset with many classes. A problem with the dataset was, it had varying image sizes, and these sizes were small to use AlexNet or GoogleNet. Therefore, the LeNet model was the remaining choice for this part of the project. Since LeNet takes input grayscale images with size 28x28, all images in the dataset were resized to 28x28 and converted to grayscale. Nvidia Digits automatically performed these operations during loading the data. The number of epochs was selected as ten to prevent overfitting.

4 RESULTS

4.1 Task 1: Bottle vs Candy-Box

The results for the five different model for the Task 1 are given on Table 1.

Figure 1 shows the screenshot of the console when evaluating the AlexNet model for Task1.

Figure 2 shows the training graph of successful AlexNet model. Figure 3, Figure 4, Figure 5, and Figure 6 shows the classification results of testing AlexNet.

```

output: softmax
iterations: 5
avgRuns: 10
Input "data": 3x227x227
Output "softmax": 3x1x1
name=data, bindingIndex=0, buffers.size()=2
name=softmax, bindingIndex=1, buffers.size()=2
Average over 10 runs is 4.63185 ms.
Average over 10 runs is 4.64213 ms.
Average over 10 runs is 4.49077 ms.
Average over 10 runs is 4.11436 ms.
Average over 10 runs is 4.13339 ms.

Calculating model accuracy...

% Total % Received % Xferd Average Speed Time Time Time Current
100 14641 100 12325 100 2316 1026 192 0:00:12 0:00:12 --:--:-- 2503

Your model accuracy is 75.4098360656 %
root@laic7f50b773:/home/workspace#

```

Fig. 1. Screenshot of Task 1 AlexNet

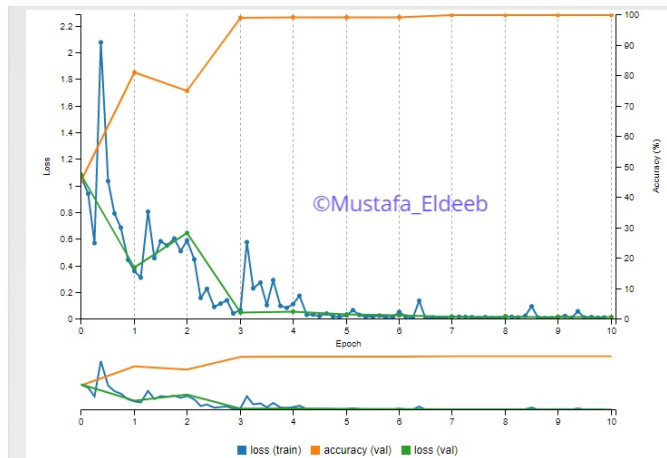


Fig. 2. Training Graph of Task 1 AlexNet

Confusion matrix

	Bottle	Candy Box	Nothing	Per-class accuracy
Bottle	18	0	2	90.0%

All classifications

Path	Ground truth	Top predictions
1 data_P1_data Bottle dressing_940.png	Bottle	Bottle (99.9%), Candy Box (0.1%), Nothing (0.0%)
2 data_P1_data Bottle coconut_1018.png	Bottle	Bottle (100.0%), Candy Box (0.0%), Nothing (0.0%)
3 data_P1_data Bottle dewalla_1050.png	Bottle	Bottle (99.9%), Candy Box (0.2%), Nothing (0.0%)
4 data_P1_data Bottle dressing_1069.png	Bottle	Bottle (100.0%), Candy Box (0.0%), Nothing (0.0%)
5 data_P1_data Bottle dressing_527.png	Bottle	Nothing (99.9%), Bottle (0.1%), Candy Box (0.0%)
6 data_P1_data Bottle dressing_518.png	Bottle	Nothing (99.1%), Bottle (0.9%), Candy Box (0.0%)
7 data_P1_data Bottle coconut_278.png	Bottle	Bottle (99.9%), Candy Box (0.1%), Nothing (0.0%)
8 data_P1_data Bottle dressing_777.png	Bottle	Bottle (100.0%), Candy Box (0.0%), Nothing (0.0%)
9 data_P1_data Bottle dressing_852.png	Bottle	Bottle (99.9%), Candy Box (0.1%), Nothing (0.0%)
10 data_P1_data Bottle dewalla_1352.png	Bottle	Bottle (100.0%), Candy Box (0.0%), Nothing (0.0%)
11 data_P1_data Bottle dressing_79.png	Bottle	Bottle (99.9%), Candy Box (0.0%), Nothing (0.0%)
12 data_P1_data Bottle coconut_380.png	Bottle	Bottle (99.9%), Candy Box (0.0%), Nothing (0.0%)
13 data_P1_data Bottle dressing_265.png	Bottle	Bottle (99.9%), Candy Box (0.0%), Nothing (0.0%)
14 data_P1_data Bottle coconut_265.png	Bottle	Bottle (99.9%), Candy Box (0.0%), Nothing (0.0%)
15 data_P1_data Bottle coconut_1552.png	Bottle	Bottle (100.0%), Candy Box (0.0%), Nothing (0.0%)
16 data_P1_data Bottle coconut_827.png	Bottle	Bottle (99.9%), Candy Box (0.0%), Nothing (0.0%)
17 data_P1_data Bottle coconut_490.png	Bottle	Bottle (99.9%), Candy Box (0.0%), Nothing (0.0%)

©Mustafa_Eldeeb

Fig. 3. Testing 20 Images with Task 1 AlexNet

alex Image Classification Model

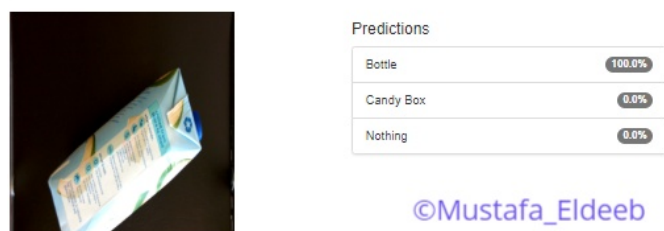


Fig. 4. Testing a Bottle Image Task 1 AlexNet

TABLE 1
Results on Task 1

No.	Model	Epoc	Solver	Rate	Accuracy
1	GoogleNet	6	Adam	0.01	67.21%
2*	GoogleNetE6	4	Adam	0.01	67.21%
3	AlexNet	10	SGD	0.01	75.41%

alex Image Classification Model

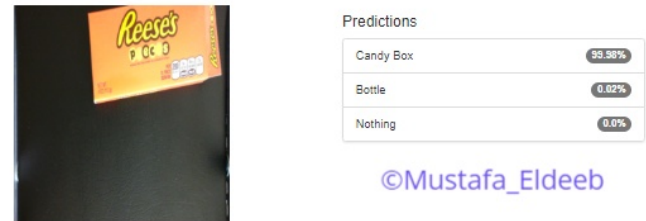


Fig. 5. Testing a Candy-Box Image Task 1 AlexNet

4.2 Task 2: Traffic Sign

Figure 7 show the training graph of the LeNet for the second task. It can be seen from the graph that, the accuracy did not change significantly after ten epochs. The validation accuracy at tenth epoch was 98.7276% and for the thirtieth epoch 99.1191%. The highest accuracy was 99.1191% for the Twentieth epoch.

Figure 11 shows the top 9 predictions for category. Figure 12, Figure 13, and Figure 14 shows sample difficult predictions.

5 DISCUSSION

The Figure ?? and Figure 2 show that SGD worked better on both the validation dataset and the test dataset for the Task 1. The GoogleNet was not successful for this dataset. The reason was the use of Adam optimizer instead of SGD and the small number of epochs. It is required to either increase the number of epochs or select the SGD optimizer.

For the second task, due to time and hardware limitations, the available datasets were limited. Most of the datasets for autonomous driving have larger than 10 GB

alex Image Classification Model

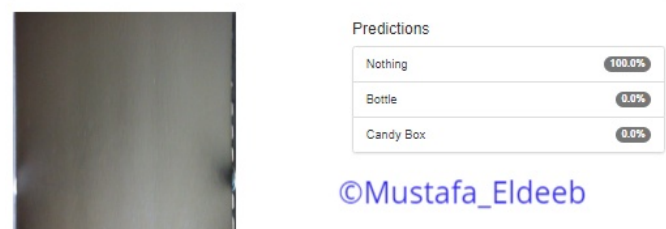


Fig. 6. Testing a Nothing Image Task 1 AlexNet

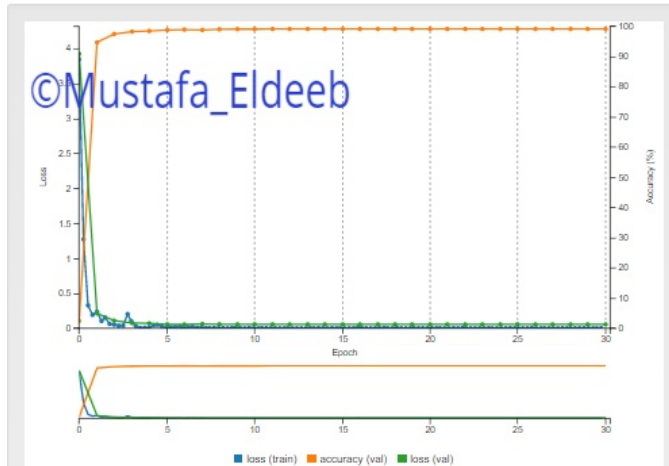


Fig. 7. Training Graph of Task 2 LeNet

data sizes. The Traffic Sign Classification is an old problem, and there are already excellent results. In addition, the size of the images in the dataset made use of AlexNet or GoogleNet irrational. Besides, the LeNet already provide satisfactory results.

6 CONCLUSION / FUTURE WORK

The results of the first task are open for improvement. A deeper network or optimized hyperparameters may provide better accuracy. Due to time limitations, this remains as future work.

For the second part of the project, it is required to access an environment which has more than 10 GB disk space and a GPU with more than 4 GB memory. As a future work, deeper networks will be trained for more complex datasets such as vehicle and pedestrian detection.

Testing the trained networks on a Jetson TX2 is another waiting task.

6.1 Hardware Deployment

The first part of the project was deployed on Udacity Workspace. Udacity Workspace provides a powerful GPU, which offers fast training and enough memory to train large deep networks. The second part of the project was deployed in the local installation of Nvidia Digits. The machine has an I7 CPU, 8GB RAM, and 650M GPU. The configuration was able to train LeNet, and with more powerful CPU and GPU, it will be possible to increase the complexity of the network.

REFERENCES

- [1] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German Traffic Sign Recognition Benchmark: A multi-class classification competition," in *IEEE International Joint Conference on Neural Networks*, pp. 1453–1460, 2011.
- [2] CNN, "Convolutional neural networks — Wikipedia, the free encyclopedia," 2018. [Online; accessed 16-July-2018].
- [3] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," 1998.



Fig. 8. Top 9 Predictions per Category

- [4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012.
- [5] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, "Going deeper with convolutions," 2014.

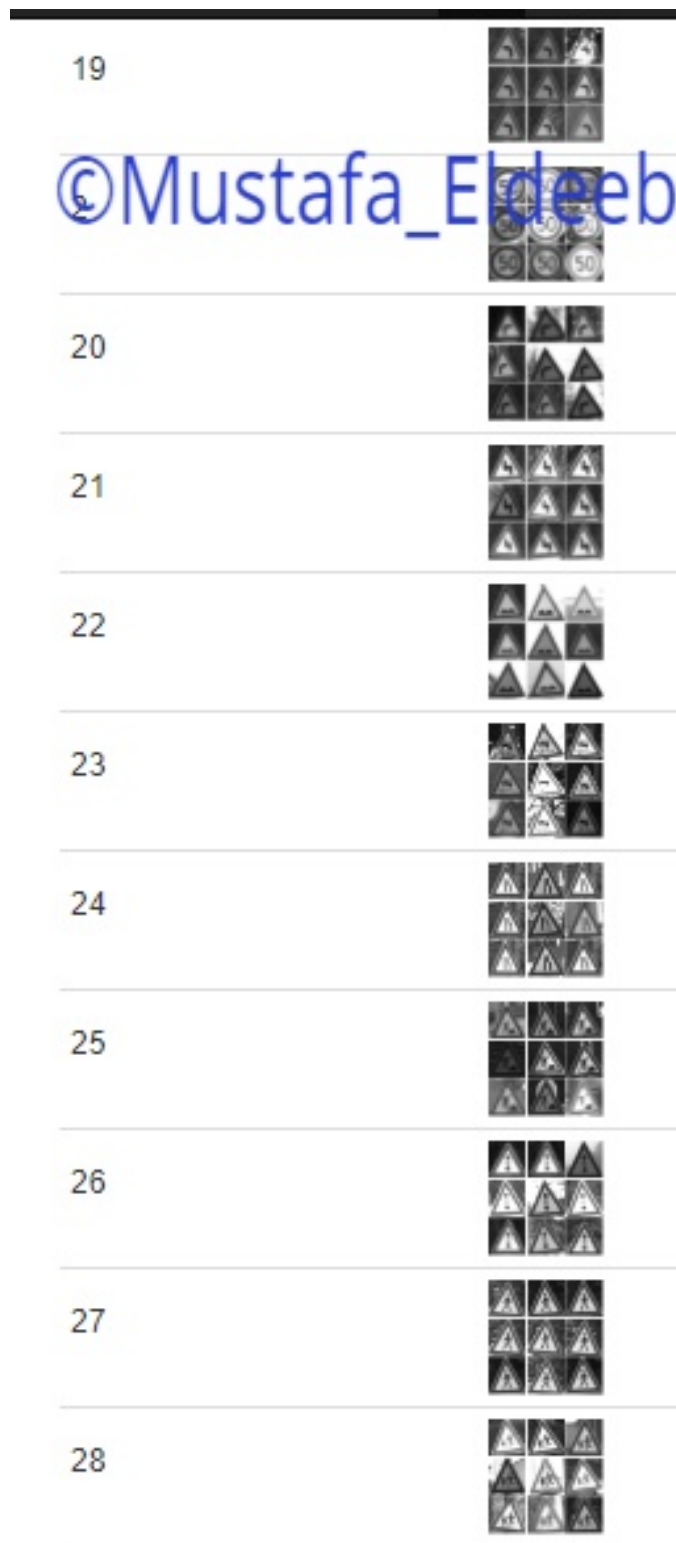


Fig. 9. Top 9 Predictions per Category



Fig. 10. Top 9 Predictions per Category

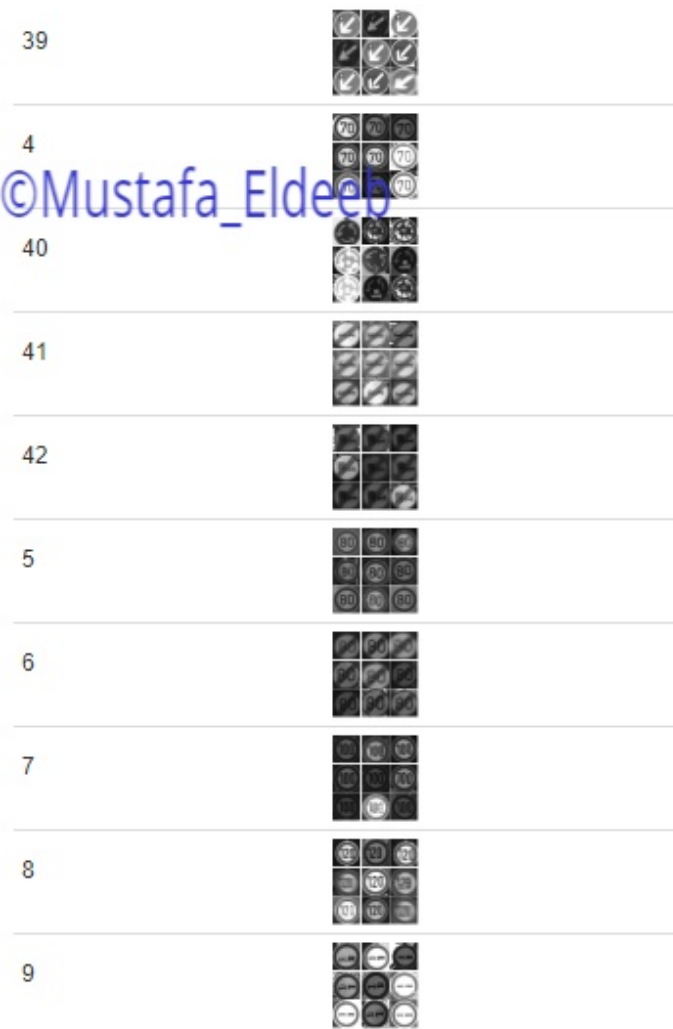


Fig. 11. Top 9 Predictions per Category



Fig. 12. Traffic Sign Test 1



Fig. 13. Traffic Sign Test 2



Fig. 14. Traffic Sign Test 3