Common tools recognition algorithm based on CNN with realsense D435

Ming Lin

Abstract—Object detection is a priority task that helps manipulator and ground vehicle to complete their missions. Object detection should provide object position and object class for implementing specify operation. Especially, when a robot want to do a operation with a tool, robot needs to know what the kind of the tool is. So an automated tool recognition solution is needed. In this paper, a method based on convolutional neural network is presented. Using supervised learning on the NVIDIA deep learning platform which name is DIGITS. The hardware to gathering the data is a depth camera d435 from INTEL. The common deep learning layer structure used in this paper is 'GoogLeNet' and 'Alexnet'. With the GPU acceleration parallel computing service which is provided from Udacity reduces a lot of time. With this method, algorithm can gives a reliable recognition result with only RGB input. The result is only verified in single picture level.

Index Terms—Robot, Convolutional Neural Network, Deep learning, Udacity, LaTeX, INTEL realsense d435

1 Introduction

OBJECT detection becomes hot topic again because of the deep learning algorithm. Since Alexnet gave a extremely good result compare to others, deep neural network based CNN(Convolutional neural network) became a main method to detection problem. Many researchers forcus on the topic of the deep neural network. One of the most exciting method is deep convolutional neural network. It makes machine can learn a complex features of images. Alexnet, GoogLeNet, VGG are very famous in nowdays. Object recognition problem has many sub-fields, such as segmentation, classification and detection, etc. In this paper, a deep neural network based image classification algorithm is proposed to classify the common tools. The data was collected by INTEL realsense D435 depth camera and the network was trained in NVIDIA DIGITS platform.

Fig. 1. Sensor and Tools to be classified.

2 Neural network layer

In order to make a supervised learning algorithm, a computing platform is needed. In this paper, NVIDIA DIGIS computing platform is used. NVIDIA provides a great GUI to realize AI computing. All research should do is concentrated on the parameters effection and layers structure. In this paper, GoogLeNet and Alexnet are used to trained the data. [1]. [2]

The GoogLeNet has more parameters to tuning and more layers. But with the deeper layer, the performance of GoogLeNet is better than Alexnet.

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Fig. 2. GoogLeNet Layer Structure.

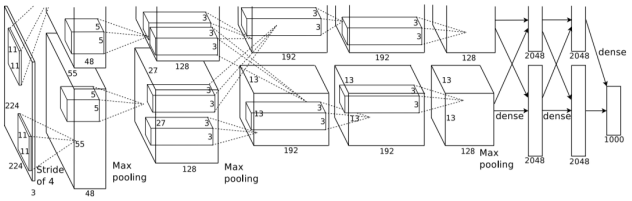

Fig. 3. Alexnet Layer Structure.

3 Data Acquisition

Data was collected by using the INTEL realsense D435 depth camera. In order to maintain the good detect performance,

the environment was set to statistic white background. Since the camera is depth camera, a simple script was programmed to extract RGB matrix from depth camera stream. The statistic data of gathered data is shown as below.

Fig. 4. Data acquisition environment.

Fig. 5. Tools statistic description.

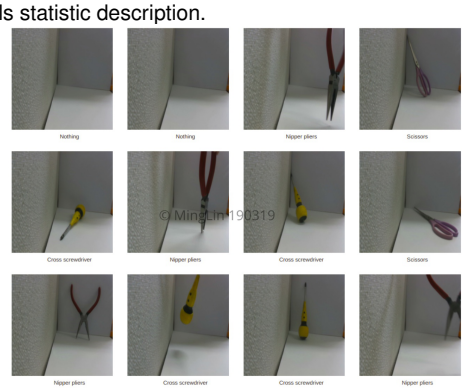

Fig. 6. Data sample.

There are four class of acquired data.

1) Cross screwdriver
2) Nipper pliers
3) Scissors
4) Nothing

The data was collected in color format and resized into size of [256,256] for feeding to the Alexnet and GoogLeNet. Tools were rotated when gathering the training data which is in order to increasing the recognition stability.

4 Results

The data source are P1 data from Udacity and Tools data from INTEL realsense d435.

4.1 P1 data

4.1.1 Loss and accuracy performance

The p1 data provided from Udacity and trained based on GoogLeNet. [3]. The loss and accuracy performance graph are shown as below. It can be found that object detection accuracy is higher than 70 percentage and the latency is 5ms in average. From the results, one can be confirmed is that GoogLeNet is good enough to finish the classification problem.

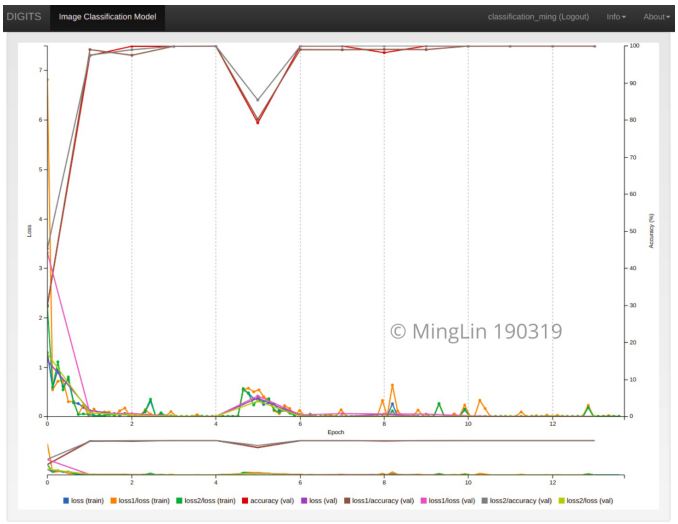

Fig. 7. Training loss graph.

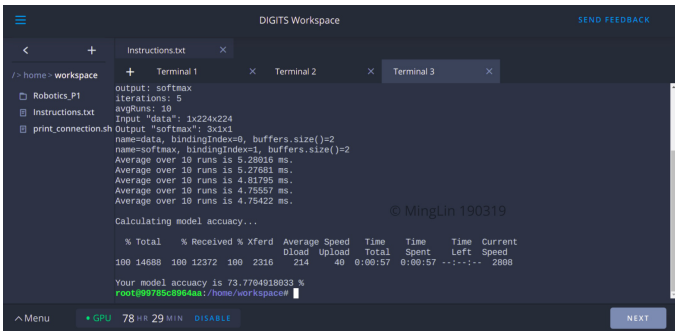

Fig. 8. Evaluate result.

4.2 Tools data from INTEL realsense d435

4.2.1 Loss graph

The loss results that trained with Alexnet and GoogLeNet by using the tools data. From the figures shown as below, one result that can be found is GoogLeNet's result is better than Alexnet. GoogLeNet's training graph convergence quickly than Alexnet. The final accuracy from GoogLeNet is also higher than Alexnet.

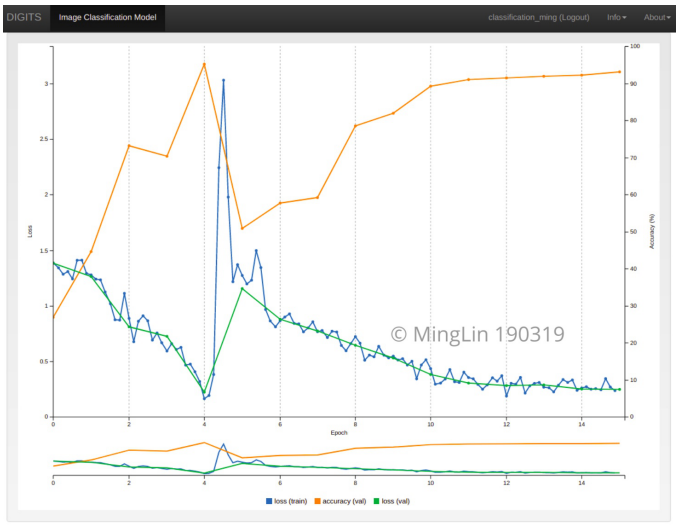

Fig. 9. Alexnet loss graph.

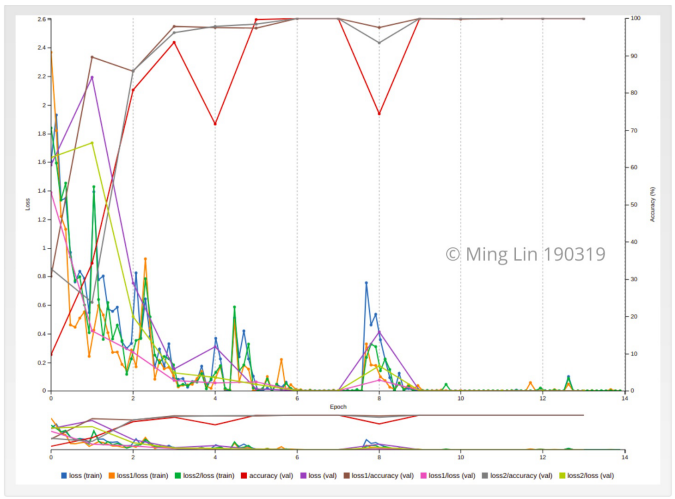

Fig. 10. GoogLeNet loss graph.

4.2.2 *Latency performance*

The training results that trained with Alexnet and GoogLeNet by using the tools data. As the results show below, it can be found that GoogLeNet is faster than the Alexnet. The average processing time of GoogLeNet is 5ms, however Alexnet is 6ms in average.

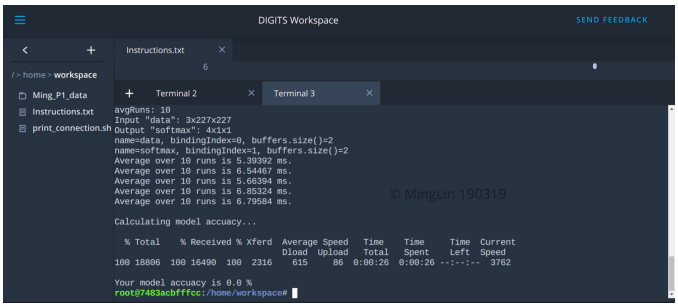

Fig. 11. Alexnet latency.

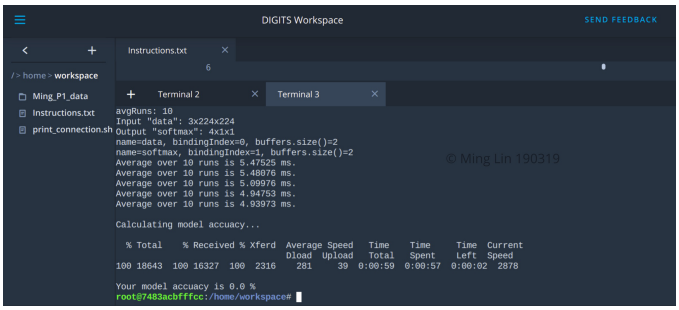

Fig. 12. GoogLeNet latency.

4.2.3 *Classification sample*

The classification results that trained with Alexnet and GoogLeNet by using the tools data. Figures shown below are processing results from GoogLeNet and Alexnet. From the results, it can be found that GoogLeNet got a better inference. The confidence of inference of GoogLeNet is highter than Alexnet.

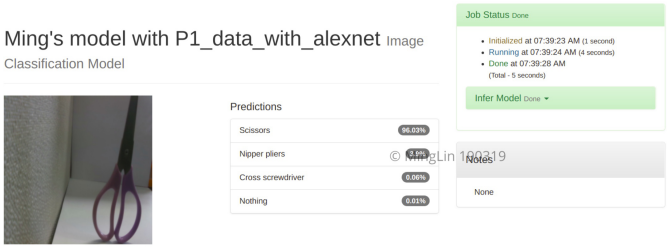

Fig. 13. Alexnet classification result sample.

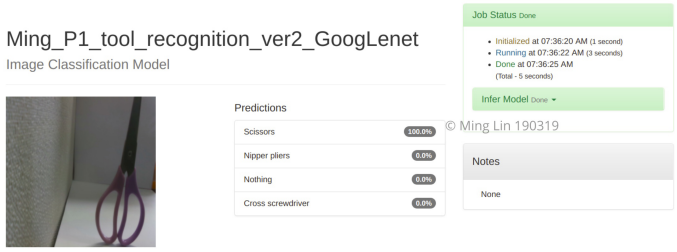

Fig. 14. GoogLeNet classification result sample.

5 Discussion

From the training results it can be found that GoogLeNet is better than Alexnet in this problem. However, one doubt about GoogLeNet is that this model gave a perfect inference result. It means maybe model is not give a inference based on features or patterns, it also has probability that just remember all the data and label. In this case, model will not work as good as now. On the contrast, Alexnet gave a low confidence of inference. Maybe in actual application, Alexnet is better than GoogLeNet. One surprising thing is that the latency of GoogLeNet is shorter than Alexnet. As we known, GoogLeNet has more complex structure than Alexnet.

6 Conclusion

In this paper, a deep convolutional neural network based object classification method is provided. From the results, it can be found that GoogLeNet can gives a good inference about 'P1 data' from udacity and it is also good in 'Tools data' in classification problem. One doubt is GoogLeNet maybe remember all the data rather than gives inference by the image features.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks,"
[2] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions,"
[3] "https://www.udacity.com,"