Robotic Inference

Gao Peng

Abstract

In this project, the common CNN model is used to implement the classification of objects and the inference of the specified object category, which mainly includes AlexNet, GoogLeNet and other models. For the project given P1 data set and my own collection of data sets for the network training and testing. The first classification data model is classifying the bottles, candy wrappers and nothing on a moving belt. What's more, the second one is classifying chair, table and nothing.

1 Introduction

In this project, I mainly use AlexNet and GoogLeNet to process 256x256 RGB images. Although LeNet is provided in digits, this model is suitable for handling binary graphs. For Udacity-supplied datasets and my own collected datasets, I Using the traditional AlexNet and GoogLeNet, did not modify the internal structure of the network. First of all, I standardized the dataset; then, select the network model and adjust the hyperparameters for training. Here I only changed the number of iterations to 6. From Loss image, GoogLeNet will reach the loss minimum earlier. Of course, when adjusting the network parameters, you can increase the size of the batch_size to improve the convergence speed, but try not to make the batch_size too large, which is best suited to the size of the epoch. At the same time, in the network training, I use the adam optimization function and the SGD optimization function to compare, it can be found that the adam optimization function will be better, but it is still relatively sensitive to the learning efficiency. In the training network process, the learning efficiency is adjusted to attenuation mode. It will produce better results in the later training of the network and also avoid the occurrence of local minimums.

2 Background

Object recognition is one of the most important issues in the field of computer vision. Many CNN models use this as a basis to measure the pros and cons of the model. The error rate of traditional machine learning for image recognition has been high until the convolution network is in this area. Satisfactory results have been achieved. CNN has characteristics such as weight sharing, local perception, multi-core convolution, etc. AlexNet, as representative of CNN, has improved the accuracy of image recognition through an 8-layer convolutional network. After that VGG is on the basis of this, the network is further deepened and the effect is improved. At the same time, in order to accelerate the training of deep convolutional networks, GoogLeNet has emerged to significantly reduce the parameters required for network training, and even hopes to transplant neural networks to mobile platforms. This article hopes to apply AlexNet, GoogLeNet and other network models to embedded systems to implement robot inference.

3 P1 image classification

3.1 Data processing

P1 dataset is stored in '/data/P1' directory. It include all images of bottles, candy wrappers and no object on a conveyor belt passing under a camera. A swing arm is used to sort all right objects to correct the bins depending on classifying results.

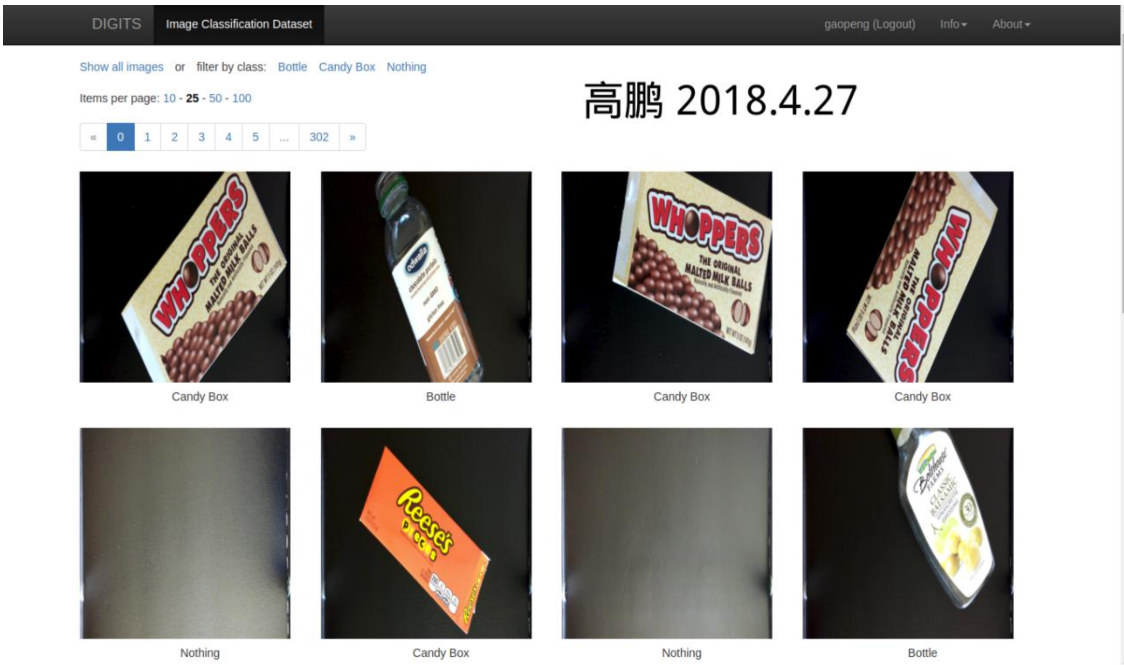


Figure 1. P1 dataset image example

P1 dataset was split two training(75) and validation(25) parts, the color image size is 256 X 256. As shown:

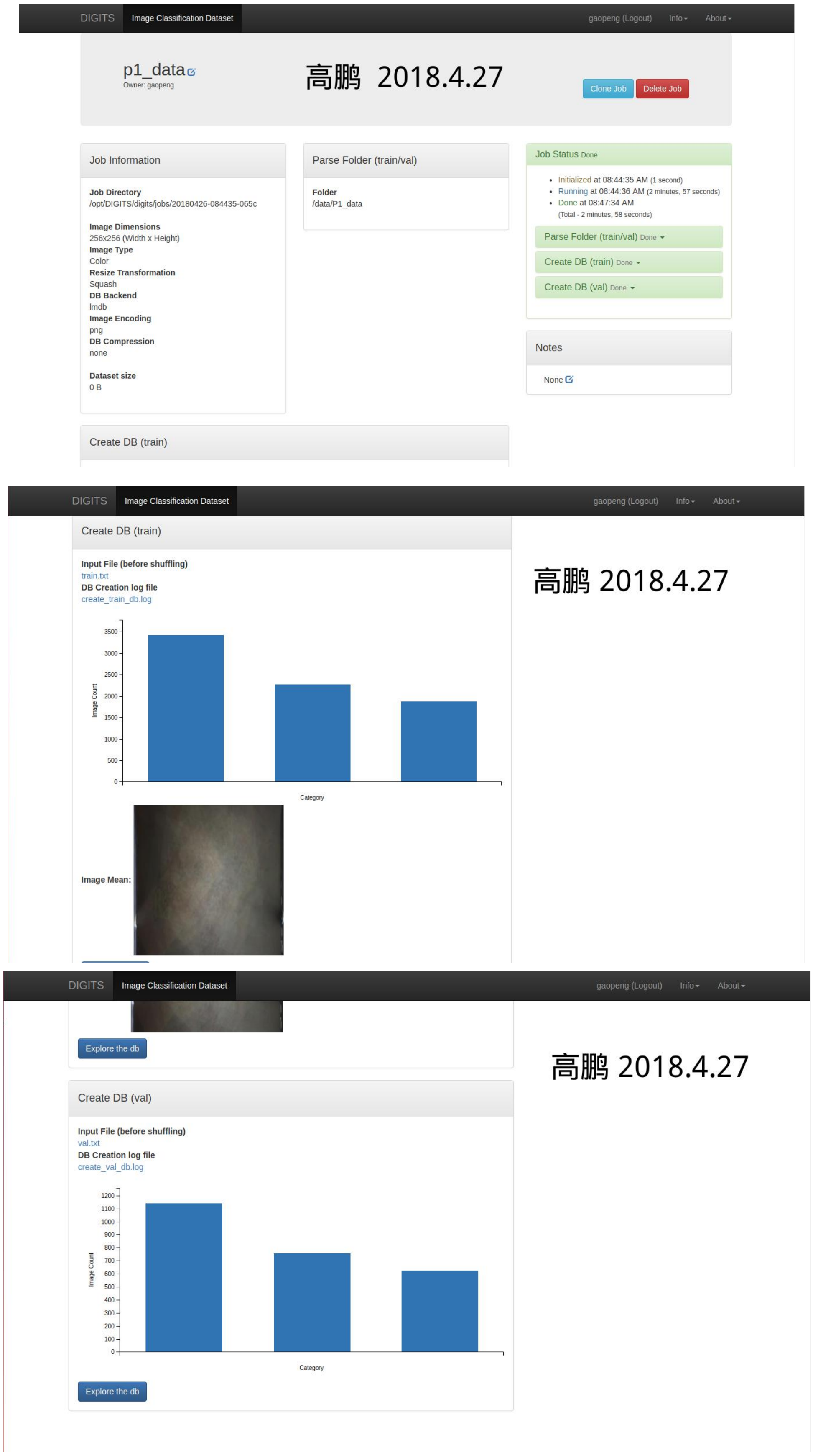


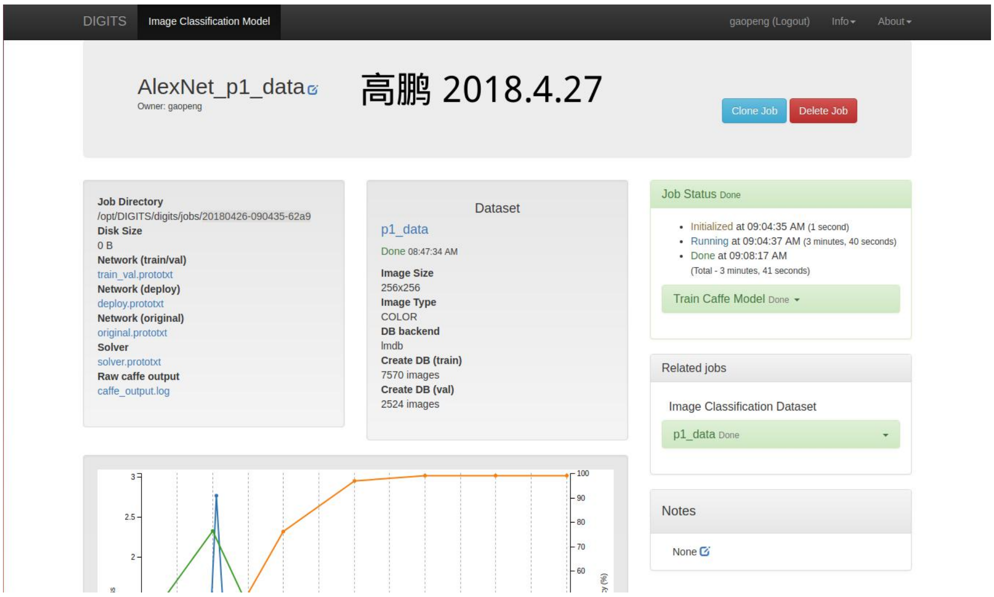
Figure 2. P1 dataset

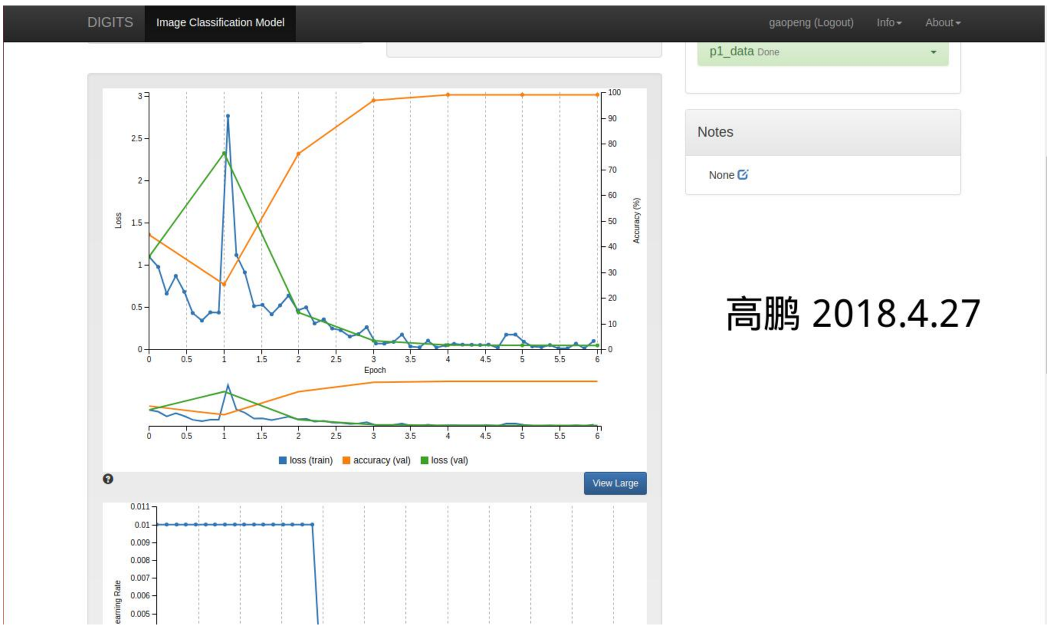
3.2 Data source

The P1 dataset is provided from Udacity Robotics lesson.It were split to two sets: training and validation dataset. They are color image and size is 256 x 256. This dataset is provided from Udacity robotics class.

3.3 Model creation

Based on the above training and validation data sets, AlexNet and GoogLeNet models were established respectively. Both epoch were set to 6. All other parameters used as default. As shown:





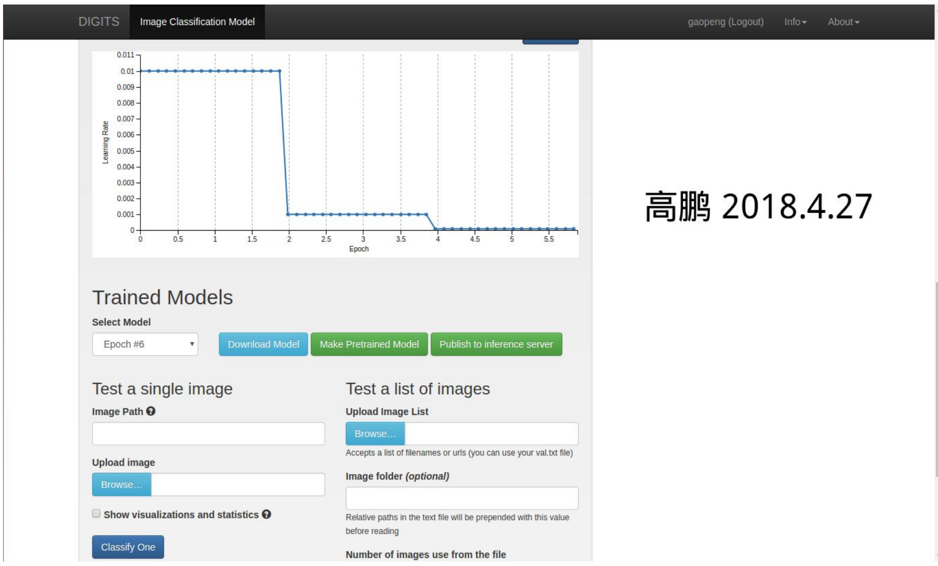
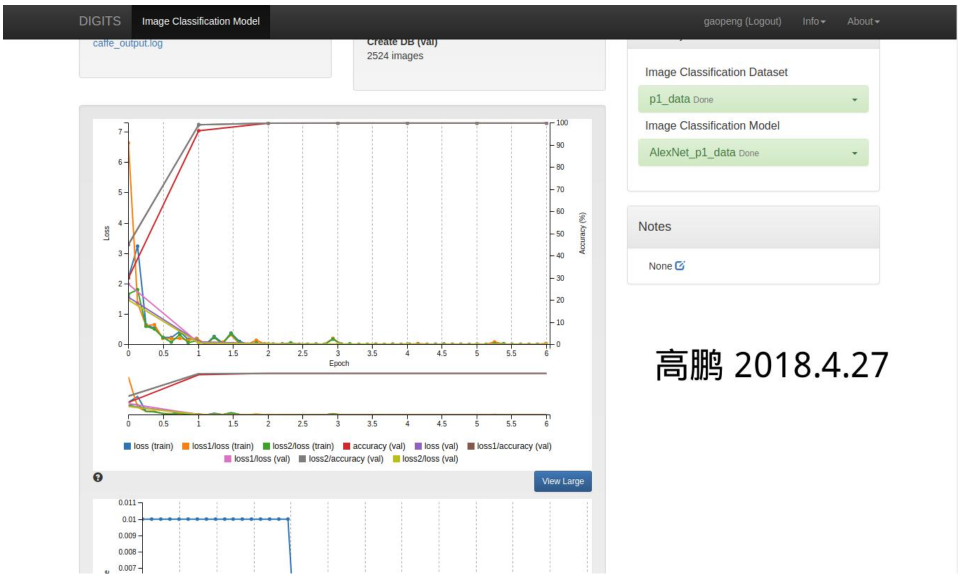


Figure 3. P1 AlexNet Model



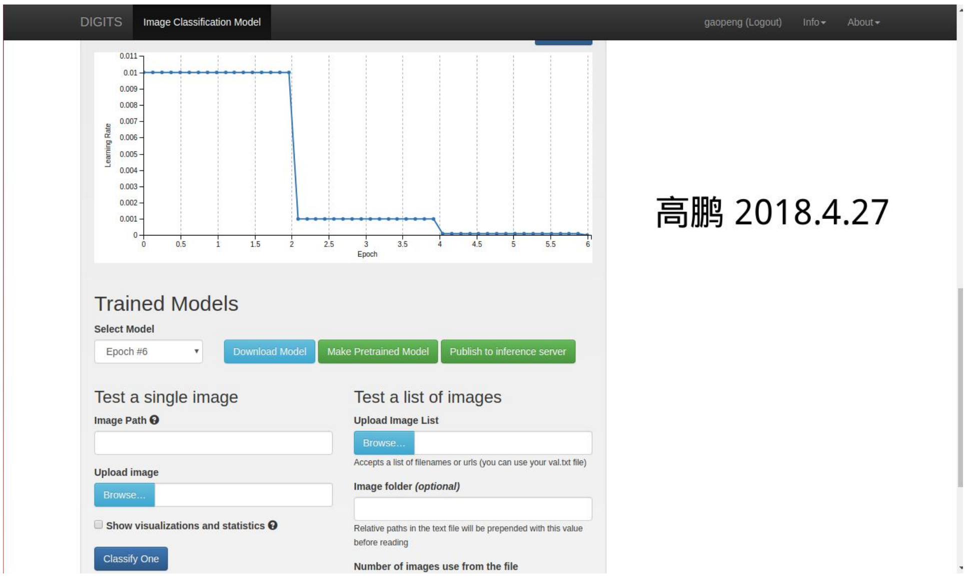


Figure 4. P1 GoogLeNet Model

3.4 Results

Evaluating result for AlexNet Model as:

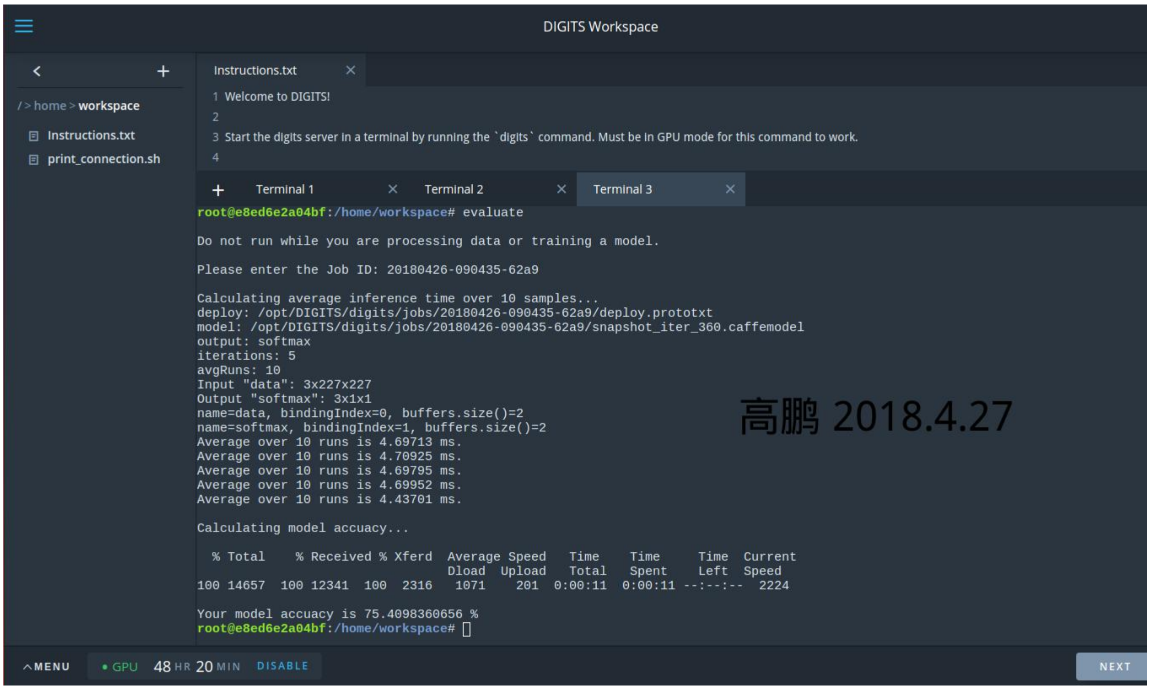


Figure 5. P1 AlexNet Result

Evaluating result for GoogLeNet Model as:

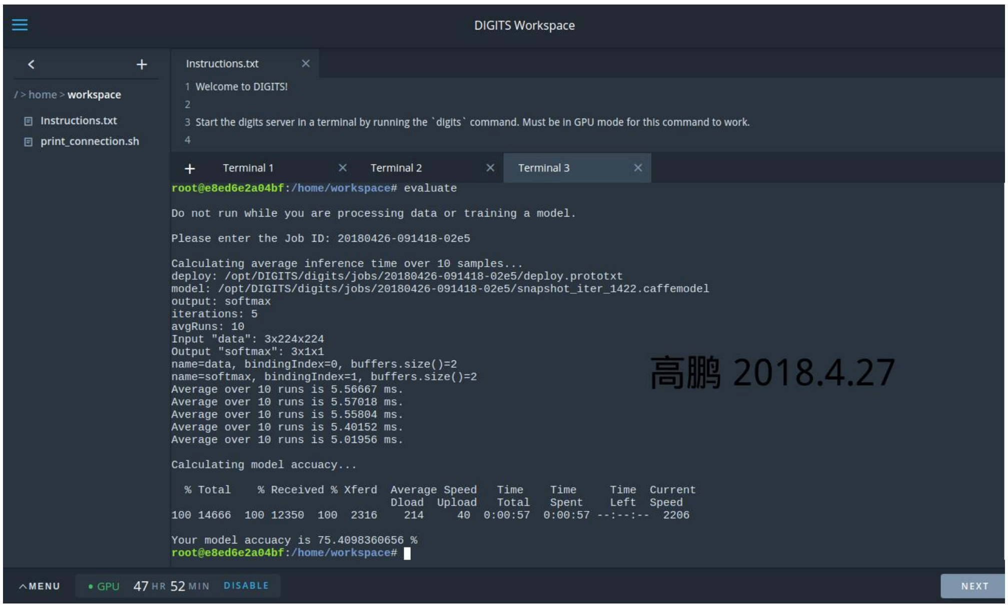


Figure 6. P1 GoogLeNet Result

Both AlexNet and GoogLeNet models are at least 75 percent accuracy and an inference time of less than 10 ms.

4 My image dataset classification

4.1 Data processing

My dataset has 3 types of tables: tables, chairs, empty sets. The total number of data is 3,000, and each type is 1000. I divide the data into 75:24:1 as the training set, verification set, and test set. As shown:

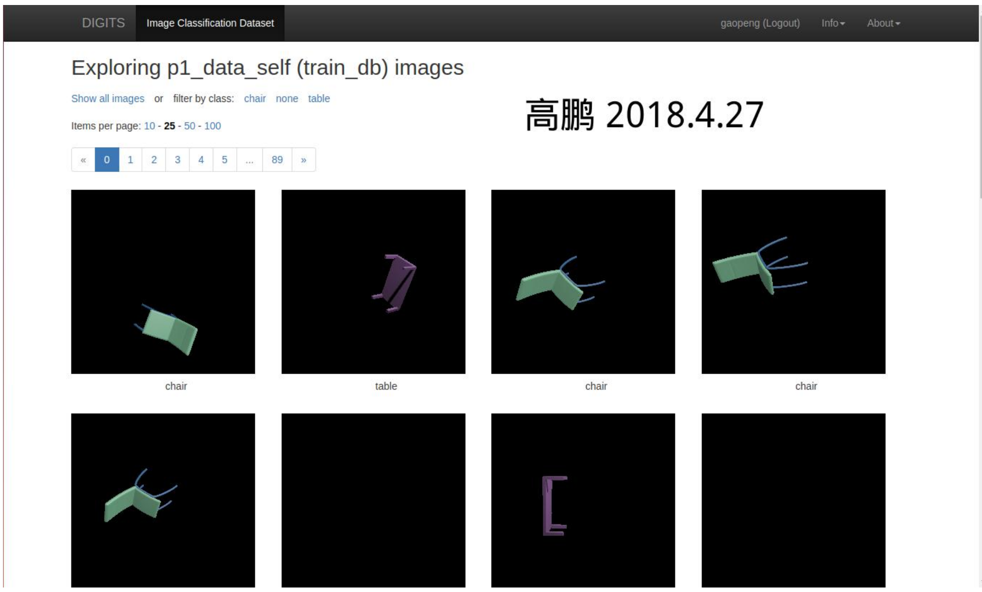
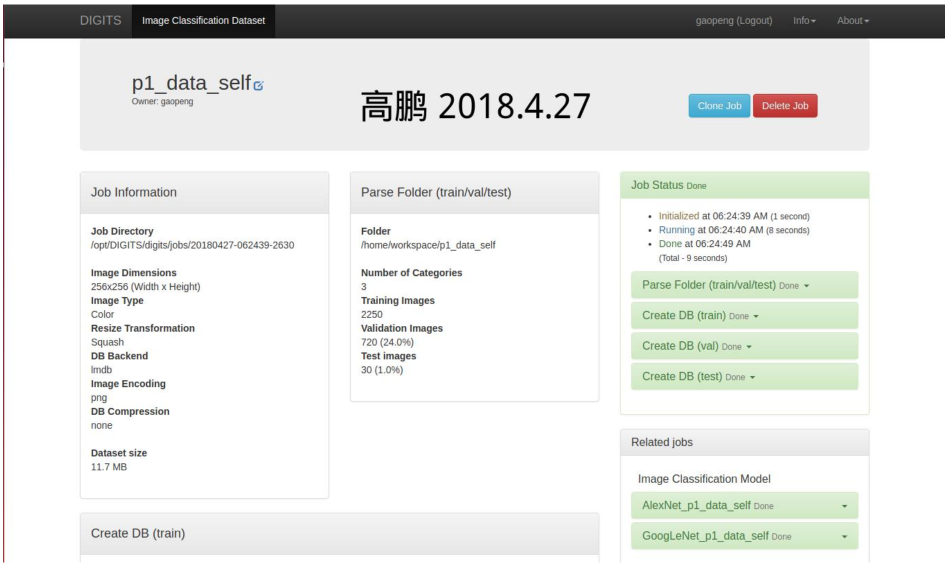
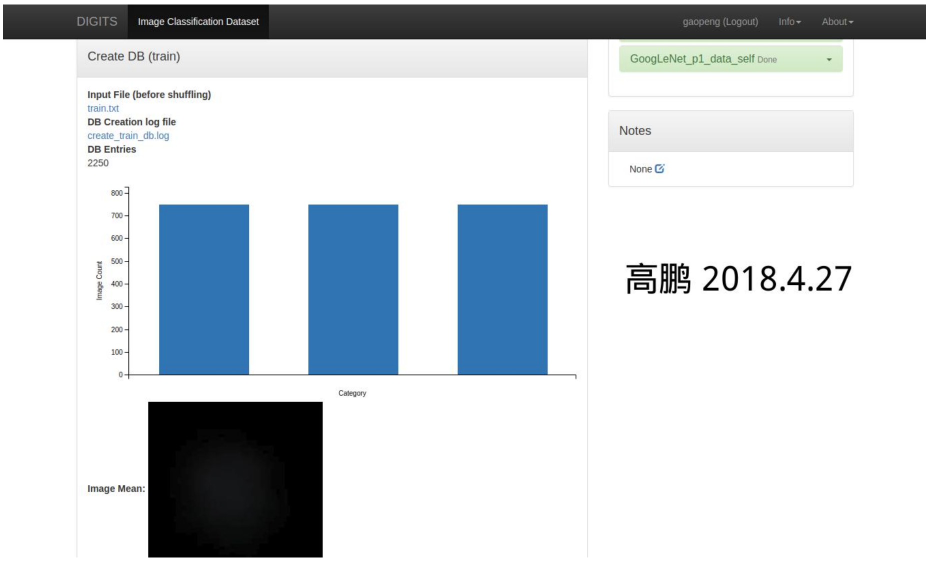


Figure 7. My dataset image example

My dataset is evenly distributed, which will improve the effectiveness of network training. As shown:





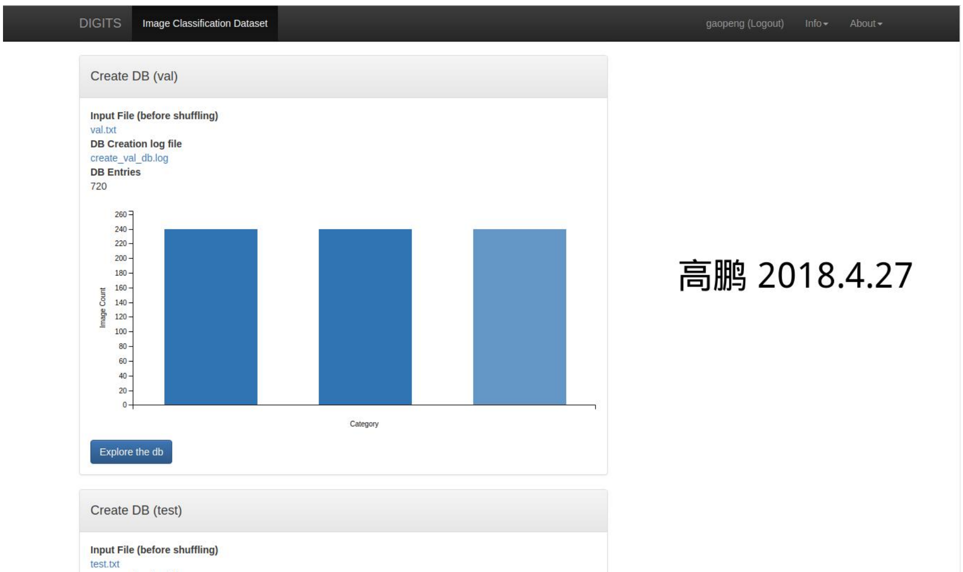


Figure 8. My dataset

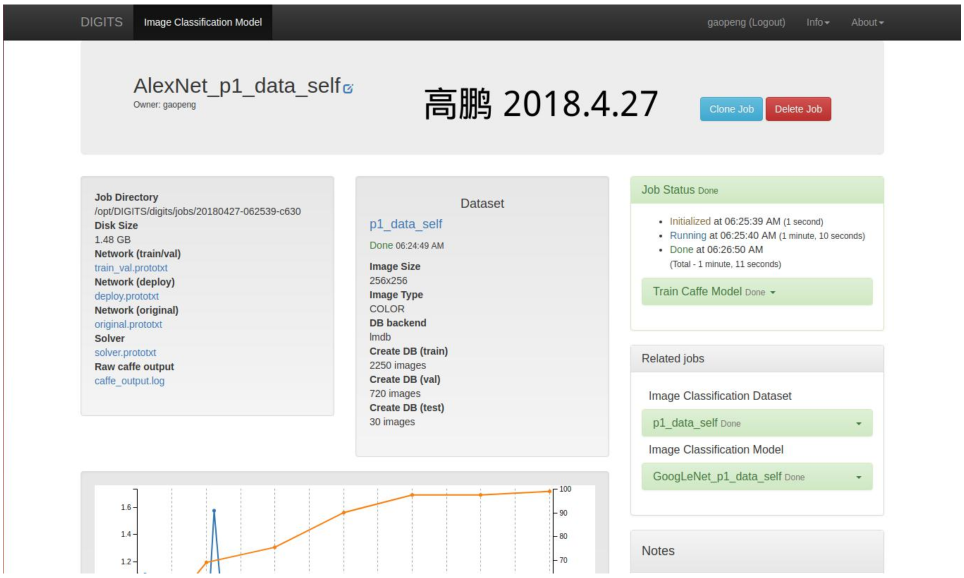
4.2 Data source

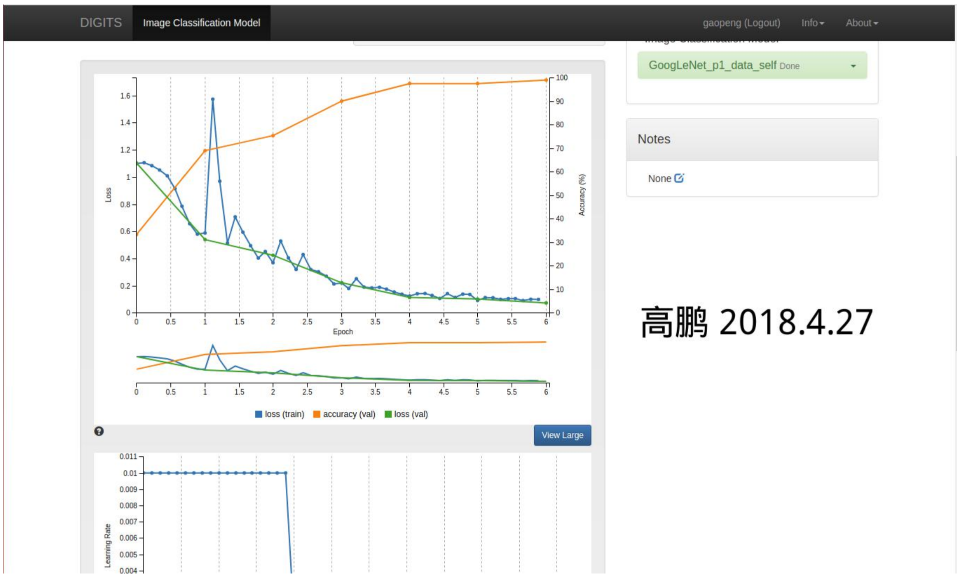
The data is taken by the analog RGB-D camera in blender. The rotation and translation of the 3D model of the chair and chair through the bpy module produce a reasonable data set.

Since the directly acquired dataset contains depth information, I have further processed the image dataset to obtain RGB images.

4.3 Model creation

Based on the above training and validation data sets, AlexNet and GoogLeNet models were established respectively. Both epoch were set to 6. All other parameters used as default. As shown:





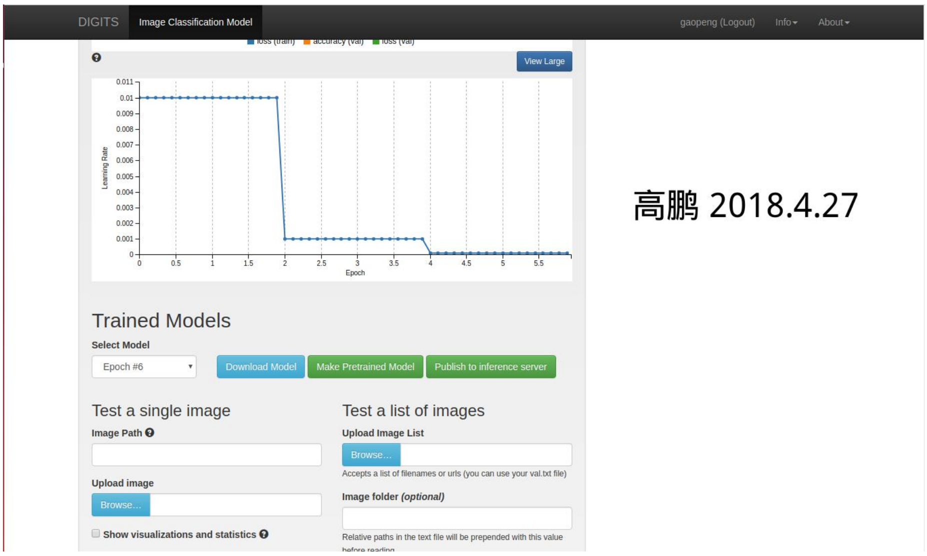
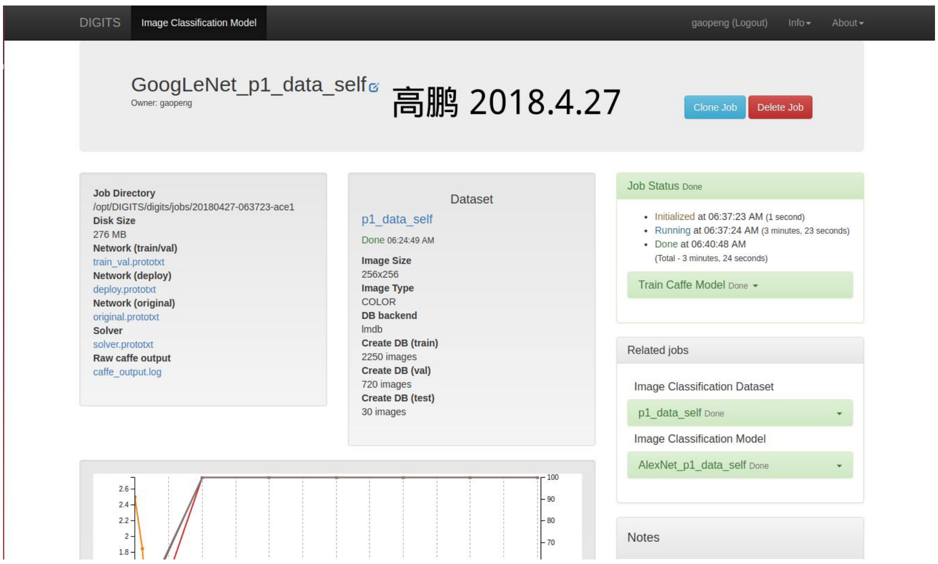
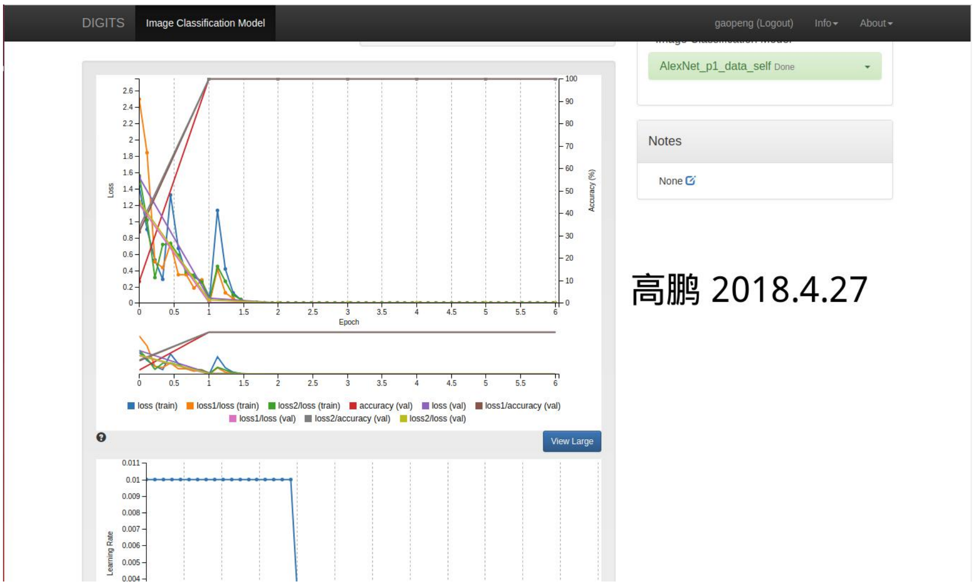


Figure 9. My dataset AlexNet Model





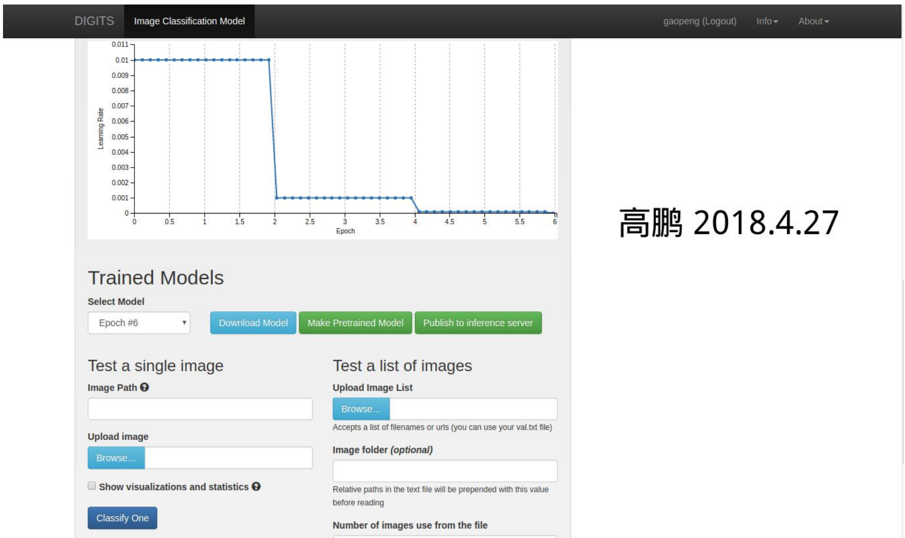


Figure 10. My dataset GoogLeNet Model

4.4 Results

Evaluating result for AlexNet Model as:

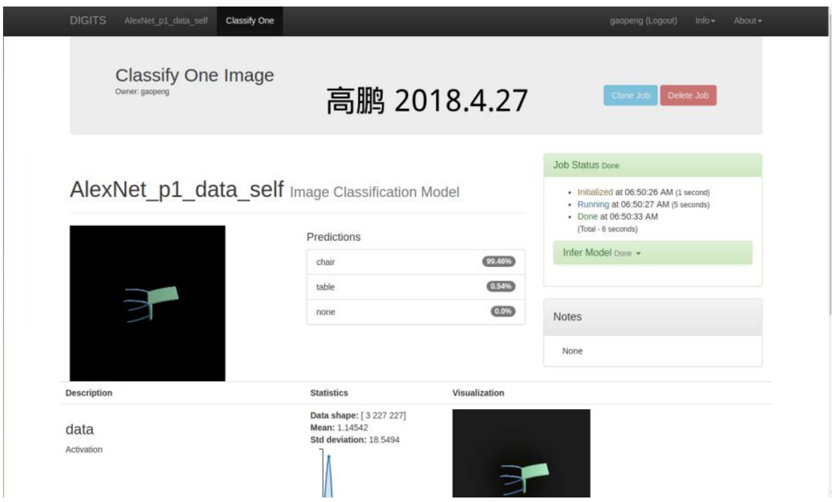


Figure 11. My dataset AlexNet Result

Evaluating result for GoogLeNet Model as:

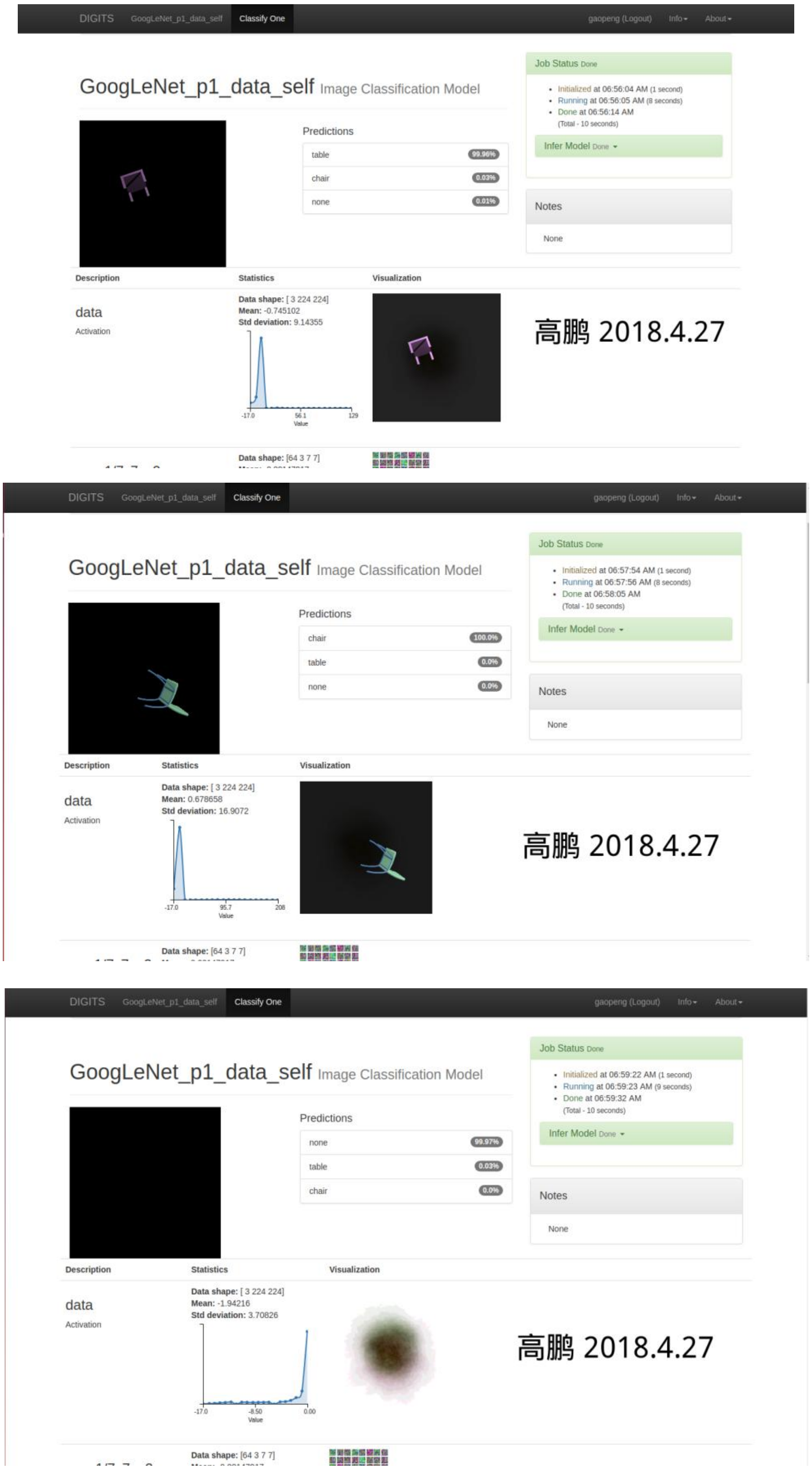


Figure 11. My dataset GoogLeNet Result

In summary, both models can be accurately classified. In comparison, GoogLeNet is better than AlexNet's results. At the same time, since the research in this paper is based on the inference of embedded robots, relatively few parameters of GoogLeNet have a good advantage. This can be clearly seen from the model's capacity. The GoogLeNet model is about 60Mb, and AlexNet is almost 200Mb.

5 Discussion

Based on the above two types of data sets, the advantages and disadvantages of the two types of network models can be intuitively summarized in digits. AlexNet was presented in 2012. Although it has achieved great success in image recognition, it also has many shortcomings. AlexNet The training parameters are too large, and the depth of the network itself does not reach the depth of the new network model. Therefore, compared to GoogLeNet, there are still deficiencies in training duration, network nonlinear expression ability, and classification accuracy. GoogLenNet is designed to speed up Training, easy to transplant, improve the accuracy of results. Its inception structure widens the network in the horizontal direction while providing multiple convolution kernel training. Adding 1x1 convolution in the network structure is used for dimensionality reduction, which is the reduction of network training parameters. In addition, GoogLeNet can Separate convolution structure also reduces network parameters further. Therefore, applying deep learning networks to embedded robot systems, GoogLeNet would be a better choice.

In this model training process, only the number of iterations is adjusted. If you increase the batch_size, it will further increase the speed. However, this process should be moderate. The number of iterations in the current model is 6 times, which is not too large. With a small number of iterations, choosing a small batch_size will allow the model to fully learn from the data.

In summary, the project aims to establish an inferential network model for embedded robot systems. A good inference network model should be weighed against inferred time and accuracy, but because embedded systems have higher real-time performance Requirements, so inferring the time should be more attention. Only in the data refresh time can be quickly inferred in order to further consider the method to improve accuracy. GoogLeNet has better characteristics on this point, not only less network parameters, can reduce inference Time, and its unique inception structure further enhances the accuracy of the network model. Therefore, implementing an inference process on an embedded system, using GoogLeNet or its variants would be a better choice.

6 Conclusion

6.1 P1 dataset

Based on the P1 data set, the training results of the AlexNet and GoogLeNet models are all very satisfactory, the accuracy rate is above 75%, and the test speed is below 10ms.

6.2 My dataset

Based on my data set, under the same parameters, AlexNet and GoogLeNet achieved better results (the number of iterations is set to 6, the rest of the parameters default). However, GoogLeNet has fewer network parameters and shorter training duration. Moreover, From the loss declining statistics graph, it can be seen that GoogLeNet has reached a higher accuracy rate without reaching the sixth iteration, and has not been overfitted afterwards.

7 Future Work

Learn more about the characteristics of GoogLeNet and write your own network for testing, adjusting different network parameters for experimentation.

Perform test experiments on the jetson TX2 module to learn more about the combination of deep learning and embedded robot systems.

Next, we should try to deepen the network model and provide more data to train the model. After all, the data in real life is more complicated. The project has many commercial application values, for example, it can be applied to the control of the robot arm later. Space manipulators that perform missions in space, in particular, need this inferential recognition function because they require higher accuracy. For example, they can be applied to autopilot systems, but this will further achieve semantic segmentation of images. These are the processes that can be continuously explored in the future.