Inference Report

Abstract:

Image classification technology has been widely paid attention because of its powerful function. The network structure of the model is determined through the direct selection of the award-winning network or the design of a specific network. At the same time, training parameters are selected according to the actual situation. Due to the existence of noise in the actual situation, we need to train a model to eliminate noise in image recognition. In this project, we use /data/p1_data data to train the Alexnet network to get the network that meets the requirements accurately. We train the GooleNet network using the collected image information, and get a network to classify the collected images. The results of the project show that the models, parameters and data are used to train a model that can eliminate noise and correctly identify the target object.

Introduction:

The /data/p1_data is a data set of three types object that contain candy boxes, bottles, and nothing. There is no obvious difference in the color of all the images in the data set, and the ordinary person can only identify it by the outline of the outer packing and the text information on the package. In data set, some of the data are half an object, so the ability of the model to learn is to be correctly classified for half an object. There are three types of data collected by individuals: apples, jetting kettles, and tea cups. The models hoped to be trained can correctly classify these three types objects. The selected area in the data is each angle of the object, in order to fully collect all the feature information of the object. There are many noises, such as the human foot, the edge of the table, and the many black edges caused by the rotation of the image. The model needs to be properly removed from the noise to avoid misjudging the noise.

Background / Formulation:

In the neural network model, many other parameters and structures are needed to complete the structure and learning of the whole model in addition to the necessary data. For the overall structure of an algorithm, it has many different types of layers. The main layer is the fully connected layer, the convolution layer and the pool layer.

Full connection layer:

The full connection layer connects all the neurons of the current layer to the next layer of neurons and uses different weights. Due to the redundancy of the full connection layer, the efficiency of learning is reduced due to the excessive use of the same parameters.

Convolution layer:

The convolution layer is convoluted through multiple convolution nuclei, and the operation of a convolution kernel extracts the same features of the original map in different positions. Because of the local perception and parameter sharing in the convolution neural network, the parameters of the whole network are greatly reduced.

Pool layer:

The operation of the pool layer is basically the same as that of the convolution layer, but the convolution core of the pool layer takes only the maximum value and average value of the corresponding location (maximum pooling and average pooling).

In the process of network training, the learning rate and the size of epochs need to be designed.

Learning rate

The learning rate for changing the weights after every training network parameter and threshold parameter proportion, the learning rate is small learning will be more precise, because in reality many models are nonlinear, like a curve, with many small linear iterative gradient descent to nonlinear curve, if every step of the long span (learning rate) will lose a lot of information distortion curve, local linear span is too small too serious, you arrive at the far end of the curve will take many steps.

Epoch

Epoch training rounds, the accuracy of the model increases with the increase of training times. The accuracy of the amount of training time can help improve the model, but too many rounds will make the model overfitting lead to lower accuracy.

AlexNet

The number of AlexNet is only 8 layers, but there is a total amount of more than 60M, which proves the effectiveness of CNN under the complex model.

GooleNet

GooleNet proves one thing: with more convolutions, the deeper layers can get better structures. The total number of GooleNet parameters is not large, but the number of calculations is very large.

Both Alexnet and GooleNet are models for image classification, and any type of model can be used for classification training for color images of 256x256.In this project, both Alexnet and GooleNet are applicable, and there is no obvious advantage and disadvantage. So two models are used for two data sets

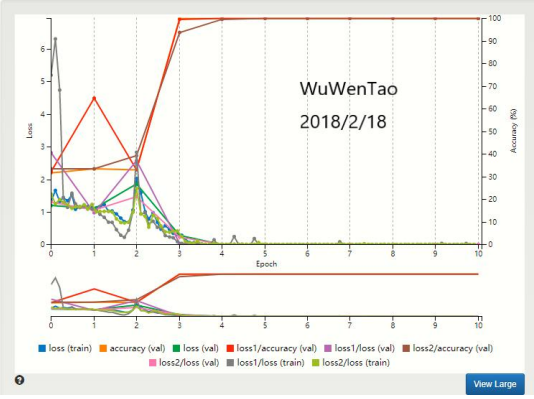
Data Acquisition:

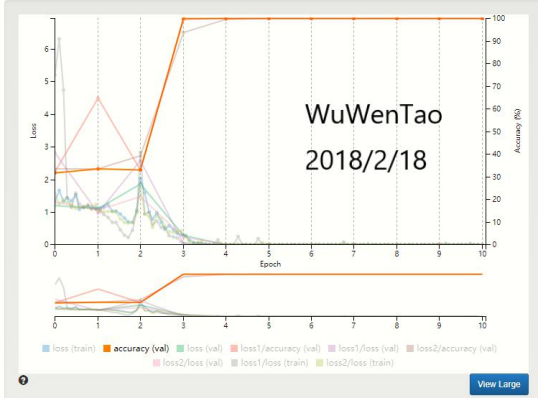
The project mainly uses two sets of data sets, which are three types of data collected by the project's built-in data set /data/p1_data and the individual collected data. The size of each group of data is 256x256, and color images are used. The collected data by individuals are three kinds objects that are captured by video. There are apples, teacups, and watering kettle. The program is used to transform videos into pictures, and convert them into color images of 256x256, and use rotating pictures to get more pictures. Each image contains labeled objects. After learning the characteristics of data sets, model can recognize the rest of the identified images. There are three types of data collected by individuals, and 300 for each class. These three kinds of objects are objects in families. Dataset can continue to collect common objects in families during subsequent work, and the trained models can be used for objects identification in families.

Results:

In project 1, in order to get the result quickly, save GPU time, set up epoch to 5, set up the learning rate of 0.01, select the built-in award-winning network Alexnet. Alexnet uses a color picture of 256x256 in size. After a short period of learning, the precision of the model reaches 75%.

In project 2, the data is 3 types of data, and 290 photos are taken for each class, in which the set is set to 25%, and 9 photos are used as the test set test. Using the GooleNet network, the epoch is set to 10, and the set learning rate is 0.01. GooleNet uses a color picture of 256x256 in size. After a short period of learning, the precision of the model reached 100% and the 9 images used by the test set correctly predicted the category.





All classifications

WuWenTao
2018/2/18

	Path	Top predictions					
1	/home/workspace/test/1/pic_1.jpg	1	100.0%	2	0.0%	3	0.0%
2	/home/workspace/test/1/pic_2.jpg	1	100.0%	2	0.0%	3	0.0%
3	/home/workspace/test/1/pic_3.jpg	1	100.0%	2	0.0%	3	0.0%
4	/home/workspace/test/2/pic_1.jpg	2	99.88%	3	0.11%	1	0.01%
5	/home/workspace/test/2/pic_2.jpg	2	99.93%	3	0.07%	1	0.0%
6	/home/workspace/test/2/pic_3.jpg	2	99.94%	3	0.05%	1	0.01%
7	/home/workspace/test/3/pic_1.jpg	3	100.0%	2	0.0%	1	0.0%
8	/home/workspace/test/3/pic_2.jpg	3	100.0%	2	0.0%	1	0.0%
9	/home/workspace/test/3/pic_3.jpg	3	100.0%	2	0.0%	1	0.0%

Discussion:

Model learn the feature in the image, and the objects used in the second models have obvious features, such as color. So the accuracy of the model in the model can easily reach 100%. For some of the non obvious features, such as the edge trajectory of the image, a deep network is used to fit it. In the data set of handwritten numerals, the characteristics of model learning mainly focus on the characteristic edge trajectories of the numbers. For this model, the accuracy of recognition is very important. When the accuracy of the model is not less than 95%, the prediction speed of the model is not higher than that of 0.2S. Guarantee the accuracy of model prediction is to ensure that the loss caused by wrong prediction can be reduced when executing operation, and the speed of prediction than the accuracy can greatly reduce the requirement.

Conclusion / Future Work:

In two different data sets, two different models, Alexnet and GooleNet, are used to learn two kinds of datasets. The learned models can correctly classify images. In the later work, we plan to collect data on the faces of family members, use TX2's camera to divide in real time, and design the network structure by ourselves, and design a unique model structure. It is not commercially feasible. First, the overall algorithm uses an open network with no unique technology. Secondly, when the product is used in the family, the price problem should be taken into consideration. At this stage, the cost should be high after the product comes out. So the product is not commercially viable.