Final Report —— Natural Images Classification

Jackson(Shuo Wang)

June 16, 2020

Abstract

The task of image classification is to give an image and its category. For the super powerful human vision system, it's easy to distinguish the category of an image, but for the computer, it can't get the semantic information of an image just like the human. The image set in this report is the natural images: cat, person, fruit, airplane, motorbike, car, dog, flower, a total of eight labels. The CNN neural network is build for image classification and compared with AutoML. Finally, CNN neural network is deployed to IOS device, and the app is made for real-time image classification.

1 Introduction to Image Set

The image set is natural images: cat, person, fruit, airplane, motorbike, car, dog, flower, a total of eight labels. It has 875 images of cats, 968 images of cars, 986 images of person, 702 images of dogs, 1000 images of fruits, 843 images of flowers, 727 images of airplanes, 788 images of motorbikes.

2 CNN Model

Convolutional neural network is a kind of artificial neural network. Its weight sharing network structure makes it more similar to biological neural network, reduces the complexity of network model, and reduces the number of weights. This advantage is more obvious when the network input is multi-dimensional image, which makes the image can be directly used as the network input, avoiding the complex feature extraction and data reconstruction process in the traditional recognition algorithm.
The CNN model has 2 convolutional layers, 2 Maxpooling layers, 2 dropout layers, a flatten layer, 2 dense layer. The loss is 'categorical crossentropy', the optimize is 'adam' and metrics is 'accuracy'. It has 613064 parameters. Figure 1 shows the structure of the model.

```
#structuring the CNN model
from keras import models, layers
model = models.Sequential()
model.add(layers.Conv2D(filters=32, kernel_size=(5,5), activation='relu', input_shape=X_train.shape[1:]))
model.add(layers.MaxPool2D(pool_size=(2, 2)))
model.add(layers.Conv2D(filters=64, kernel_size=(3, 3), activation='relu'))
model.add(layers.MaxPool2D(pool_size=(2, 2)))
model.add(layers.Dropout(rate=0.25))
model.add(layers.Flatten())
model.add(layers.Dense(256, activation='relu'))
model.add(layers.Dropout(rate=0.5))
model.add(layers.Dense(8, activation='softmax'))
```

Figure 1: CNN model

After fitting the model, the training set accuracy is 97.6%. The test set accuracy is 89.63%. Figure 2 shows the changes of loss and accuracy.

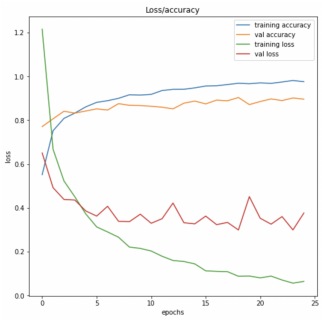

Figure 2: Chances of loss and accuracy

3 AutoML

AutoML can design machine learning model automatically. Through Google's most advanced transmission learning and neural architecture search technology, it helps enterprises or individual users with weak professional knowledge of machine learning build their own high-quality customized model.Cloud automl vision makes it faster and easier to create custom machine learning models for image recognition. With its drag and drop interface, you can easily upload images, train and manage models, and then deploy these training models directly on Google cloud. Users only need to upload the pictures and click training to choose the customized model they want to build or the model provided by Google.
The same data set were used to train the AutoML model. AutoML has a very good result. The precision and recall are nearly 100%. Obviously there is still a big gap between the model I build and it.

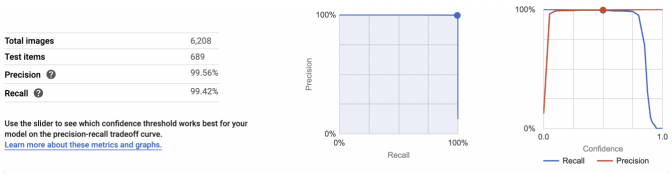

Figure 3: AutoML

4 Deploying to IOS device

According to the official documents of Google, tensorflow Lite is used to deploy the model to IOS devices, and app is obtained and used for real-time image classification.

Figure 4: IOS app

5 Future Improvement Plan

Although the accuracy has reached 90%, there are still many problems in real-time classification. For example, when classifying fruit, my app classifies it as flower, and the accuracy of fruit classification is too low. So I plan to make some changes to CNN model after consulting relevant materials and gradually improve it.

- The images under each category need to be representative, otherwise, during the training process, the model will extract content unrelated to the category, such as some background information. So it's best to use as many pictures as possible from different devices in the scene.
- Increase iterations
- Optimize the network structure
- Hyperparameter tuning
- Fine tune

6 Conclusion

Although this is the first time for me to learn computer vision, I gradually found the fun in it, and also deeply realized the significance of learning to apply.This learning experience is very good. It's not only that we have mastered relevant knowledge, but also that we have learned a thinking mode and practice habits.Thank you very much to the professor, mentor, TA and head teacher for their hard work during the class.I will keep this learning habit and improve the final project in the future.