

Yizhou Chen; Xiaotong Chen; Xuanzhen Xu

Abstract

In this project, we built a convolutional neural network which is able to classify 40 breeds of dogs with more than 80% accuracy by transfer learning. To achieve this result, we compared different pretrained model and optimization algorithm, and finally decide to used pretrained VGG19 net and l-bfgs algorithm. And we also investigate the effect of number of categories on test accuracy and found the largest breeds number with tolerable accuracy.

Problem statement

Our target is to classify dog breed from daily life photos with tolerable error rate which is set as 20%. The ideal network should be able to classify all 120 breeds available in the dataset. If the accuracy target cannot be achieved, the maximum number of breeds within the accuracy target should be find. The small inter-class variations and the low-quality samples in the dataset are the main problems of this project.

Dataset

Stanford Dogs Dataset[1] is taken as training and test data in this project. It covers 120 breeds of dogs in 20580 images in total. This dataset is built upon images and annotation from ImageNet. Most of the pictures are life photo instead of close-ups with clean background.

Dataset size:

Number of images: 20,580

Number of categories: 120

Data format:

Image + ID + Bread information

Method

We use the transfer learning on a pretrained VGG-19 network. The VGG-19 network contains five blocks of convolutional layers and each block are separated by max-pooling layer. We implement this model on Keras using ImageNet weights provided by Keras library and ‘lbfgs’ as the solver.

We add our final layer that is a logistic regression that outputs class probabilities.

Results

We select top 40 categories of dogs with highest frequency in the dataset.

There are 4029 training and test set and we get 0.03% training error and 18.59% test error by our model.



Figure 2. Examples of correct prediction by our model.

Categories selected	Accuracy
20	84.73%
40	81.41%
60	79.02%

Table 1. Comparison of accuracy on the number of categories we select

Discussion

- Artificial Neural Network Selection
There are many convolutional neural networks accessible for this task, which makes it hard to select a pre-trained model. Thus we explored different types of CNNs, such as AlexNet, GoogleNet/Inception, VGGNet, ResNet, and pretest the training and testing performance before fine-tuning the post customized components as classifier such as DNN, SVD, K-means. We do this work as stated, and select VGG-19 as the best pretested one.
- Dataset Sizing
There is a long tail in the dataset, which means that the number of dog breeds with small proportion, can accumulate to a large part while each of them is still quantitatively lacking to get a good prediction result. This observes with our common sense that we can not recognize a strange or rare thing especially without adequate knowledge. To maintain a acceptable prediction performance, we reduce the size of dataset and only keep the 60 breeds top in number. But still, the long tail is an important thing in practice, since they have potential value in long run.
- Portable Model Reuse
The pre-trained model is good and handy to use for a good and fast training task, and can output a acceptable prediction in practice. However, this pre-trained model, has little portability to transport onto a embedded platform such as autonomous vehicle and mobiles. So there is a trade-off between pre-trained model selection and customized architecture to balance the task time and portability.

Conclusions

Overall, we are satisfied with the final results which achieves the goals we set in the proposal before. With the classification among the number of dog breeds ranging from 20 to 60 with a step of 20, the testing accuracy drops not so fast and maintain a minimum of 79.02%.

However, we still need higher accuracy for a excellent prediction and lots of options are available. We can preprocess the image, build up the customized CNN networks with fine-tuned layers, add softmax classification layers after CNN layers, train the model, prune the model, or retrain the model to recover the accuracy loss. This is really a difficult task with better performance in the future, and it is needing more knowledge and experiments in the practical work.

VGG 19

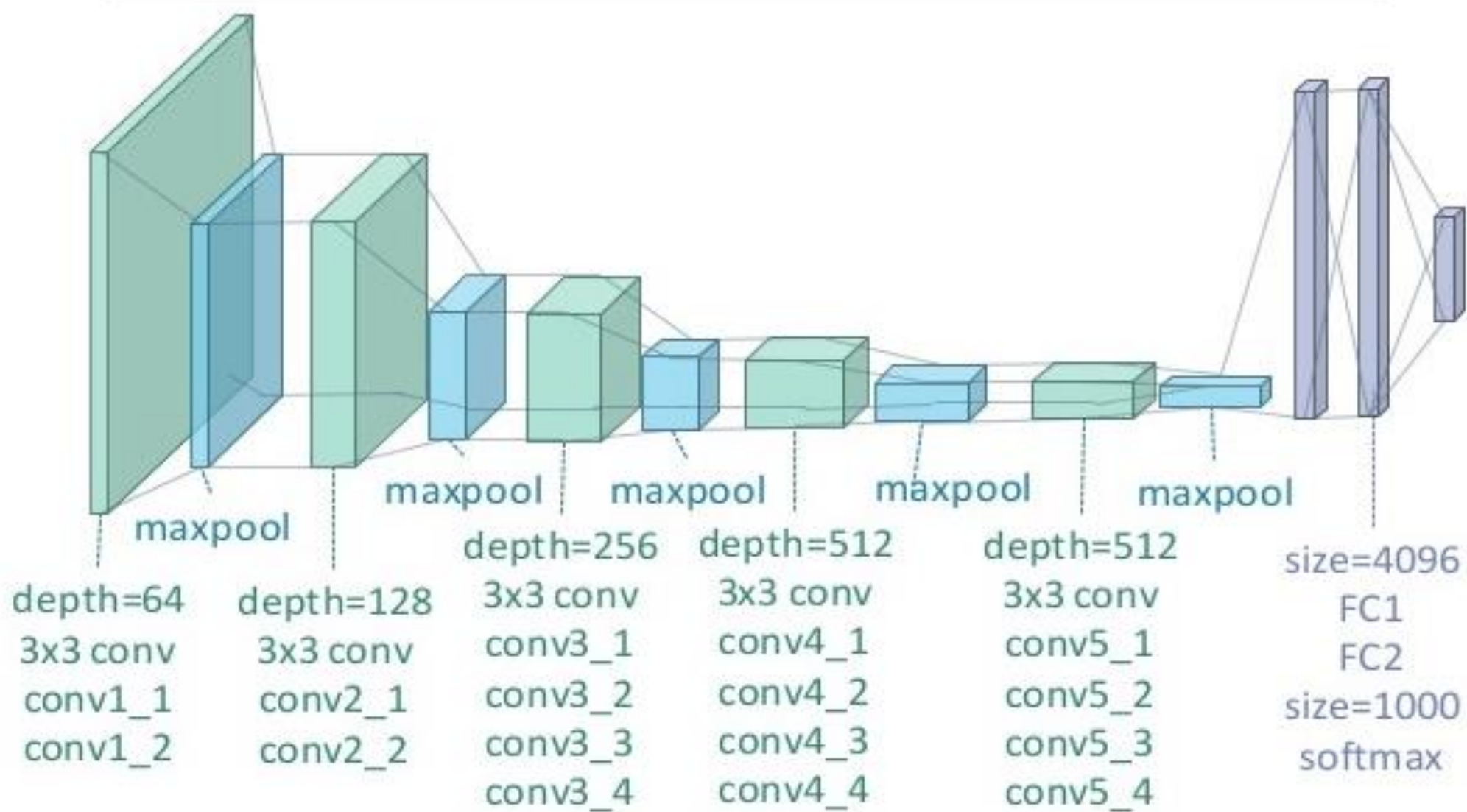


Figure 1. VGG-19 architecture^[2]

Contact

Yizhou Chen yic244@ucsd.edu
Xiaotong Chen xic363@ucsd.edu
Xuanzhen Xu xux021@ucsd.edu

References

1. Aditya Khosla, Nityananda Jayadevaprakash, Bangpeng Yao and Li Fei-Fei. Novel dataset for Fine-Grained Image Categorization. First Workshop on Fine-Grained Visual Categorization (FGVC), IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

2. <http://resizing.info/imgeditor.html>