STAT 578 Spring 2018 Homework 2

Jifu Zhao Date: 03/25/2018

1 Introduction

Object recognition is a fundamental problem in computer vision. Even in a few years ago, it is still very hard for computers to automatically recognition cat versus dog. But now, with the development of deep learning algorithms, especially Convolutional Neural Networks (CNN), this task is much easier. In this project, using CIFAR10 dataset, I built two simple CNN models and try to recognize 10 different objects among a large amount of images.

CIFAR10 dataset is a collection of images that are commonly used to test computer vision algorithms for object recognition. It consists of 60,000 images in 10 different classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. Each image is $32 \times 32 \times 3$ RGB color image. A subset of 36 images from CIFAR10 dataset are shown in Fig. 1.

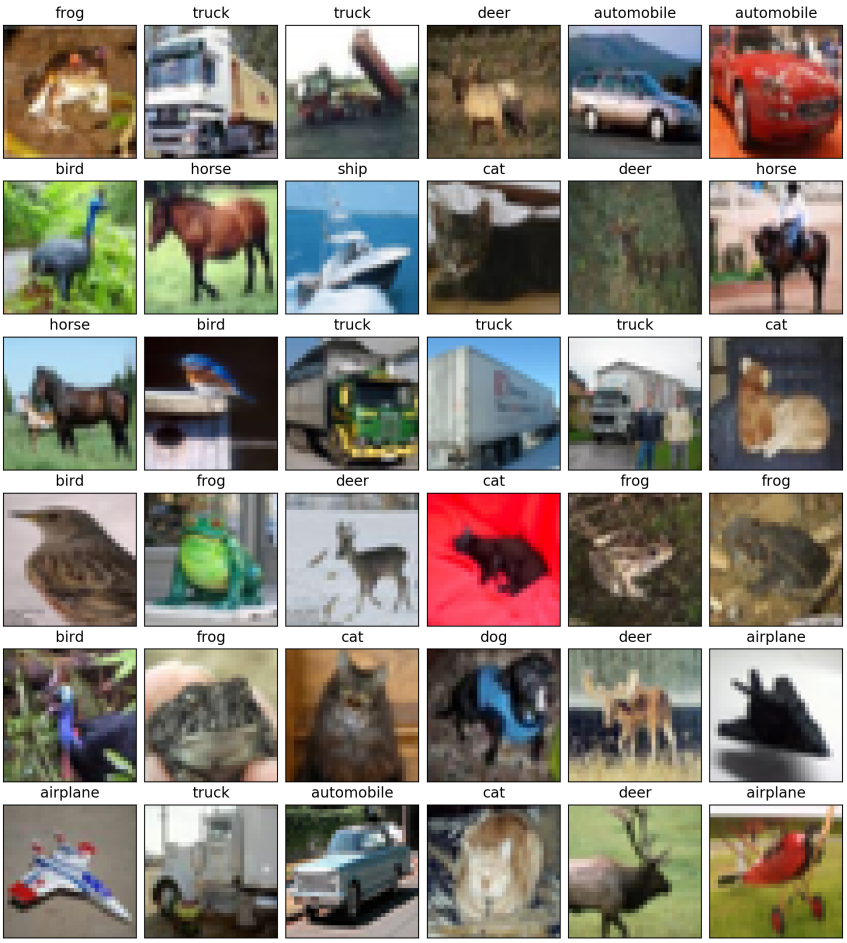

Figure 1: Subset of CIFAR10 dataset.

2 CNN Models

Recent years, there are a lot of different CNN models proposed for object recognition. Their performance has been tested on the famous ImageNet dataset. Fig. 2 contains the comparison of different CNN models on ImageNet dataset. In this project, I implemented two widely used models: VGG and ResNet for object recognition on CIFAR10 dataset using Keras with TensorFlow.

Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
Xception	88 MB	0.790	0.945	22,910,480	126
VGG16	528 MB	0.715	0.901	138,357,544	23
VGG19	549 MB	0.727	0.910	143,667,240	26
ResNet50	99 MB	0.759	0.929	25,636,712	168
InceptionV3	92 MB	0.788	0.944	23,851,784	159
InceptionResNetV2	215 MB	0.804	0.953	55,873,736	572
MobileNet	17 MB	0.665	0.871	4,253,864	88
DenseNet121	33 MB	0.745	0.918	8,062,504	121
DenseNet169	57 MB	0.759	0.928	14,307,880	169
DenseNet201	80 MB	0.770	0.933	20,242,984	201

Figure 2: Comparison of different CNN models (Results come from keras.io).

2.1 VGG

VGG represents a series CNN models for computer vision tasks. It use a series of 3×3 convolution filters with different depths. VGG16 and VGG19 are the widely known models. An illustration of a sample VGG structure is shwon in Fig. 3. In this project, instead of using deep VGG models like VGG16 or VGG19, I implemented a relatively shallow VGG model. The whole model structure is shown in Fig. 4.

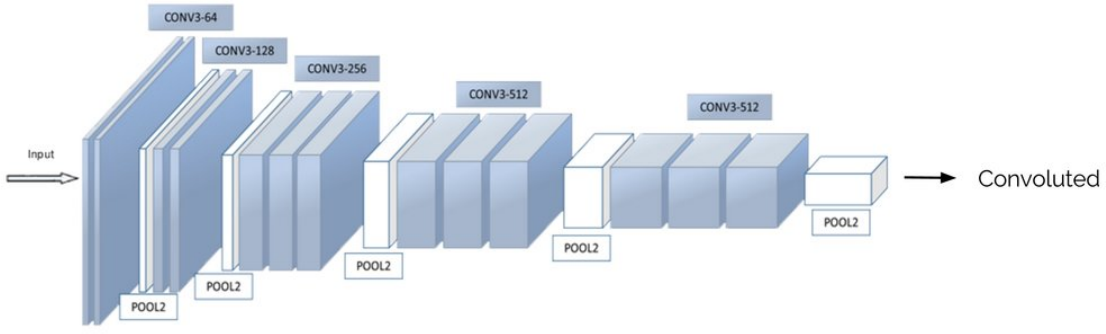

Figure 3: Sample structure of VGG (image downloaded from Internet).

Layer (type)	Output Shape	Param #
block1_conv1 (Conv2D)	(None, 32, 32, 32)	896
block1_conv2 (Conv2D)	(None, 32, 32, 32)	9248
block1_pool (MaxPooling2D)	(None, 16, 16, 32)	0
block1_dropout (Dropout)	(None, 16, 16, 32)	0
block2_conv1 (Conv2D)	(None, 16, 16, 64)	18496
block2_conv2 (Conv2D)	(None, 16, 16, 64)	36928
block2_pool (MaxPooling2D)	(None, 8, 8, 64)	0
block2_dropout (Dropout)	(None, 8, 8, 64)	0
block3_conv1 (Conv2D)	(None, 8, 8, 64)	36928
block3_conv2 (Conv2D)	(None, 8, 8, 64)	36928
block3_pool (MaxPooling2D)	(None, 4, 4, 64)	0
block3_dropout (Dropout)	(None, 4, 4, 64)	0
flatten (Flatten)	(None, 1024)	0
fc1 (Dense)	(None, 256)	262400
fc1_dropout (Dropout)	(None, 256)	0
fc2 (Dense)	(None, 128)	32896
fc2_dropout (Dropout)	(None, 128)	0
prediction (Dense)	(None, 10)	1290

Total params: 436,010
Trainable params: 436,010
Non-trainable params: 0

Figure 4: Structure of VGG.

2.2 ResNet

ResNet refers to the residual neural network structure. A sample ResNet model is shown in Fig. 5. Through residual learning, ResNet can be built very large. For example, the original version on ImageNet has 156 layers. In this work, followed some online instruction, I build a relatively shallow ResNet model.

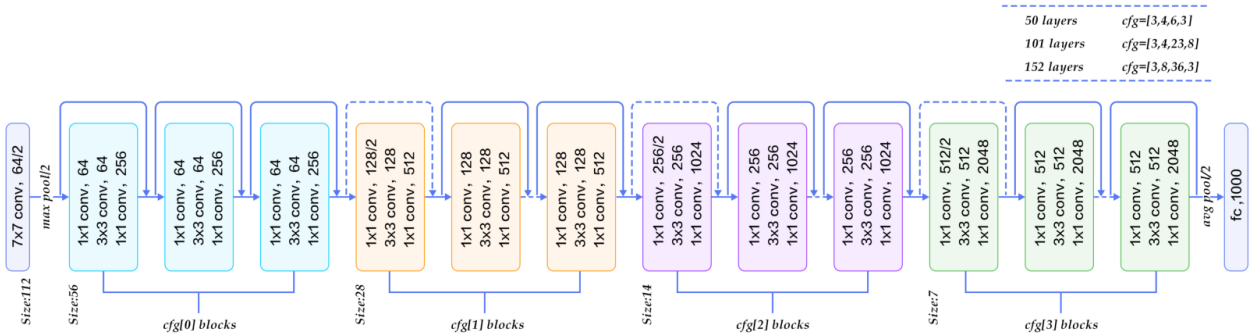

Figure 5: Sample structure of ResNet (image downloaded from Internet).

3 Results

3.1 VGG

The training/validation loss and accuracy versus training epochs are shown in Fig. 6. After 200 epochs, the test accuracy is around 0.85290, which is still underfit.

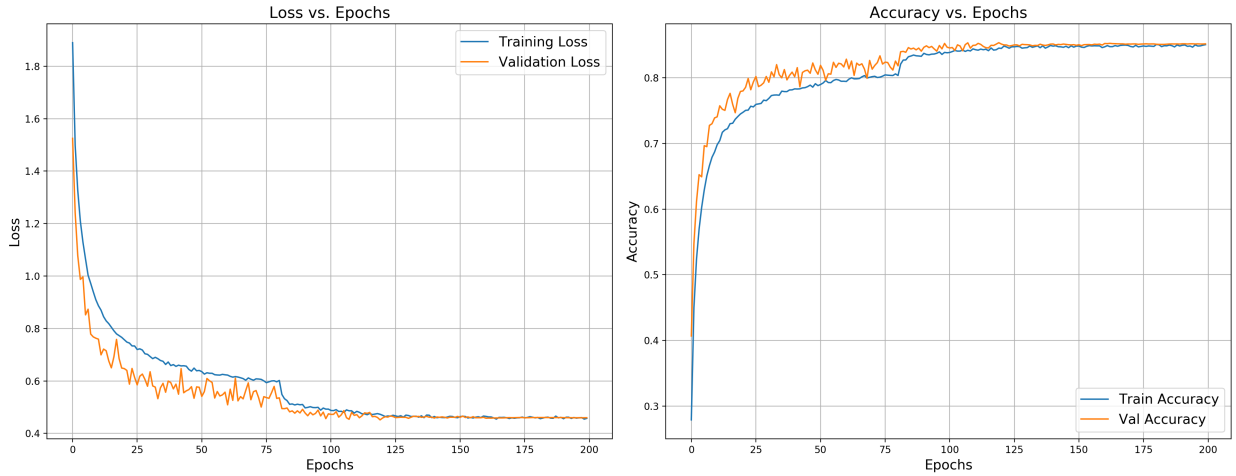

Figure 6: Loss/Accuracy versus training epochs fro VGG.

3.2 ResNet

The training/validation loss and accuracy versus training epochs are shown in Fig. 7. After 200 epochs, the test accuracy is around 0.90840. Not a bad result, but the model seems to overfit a little bit.

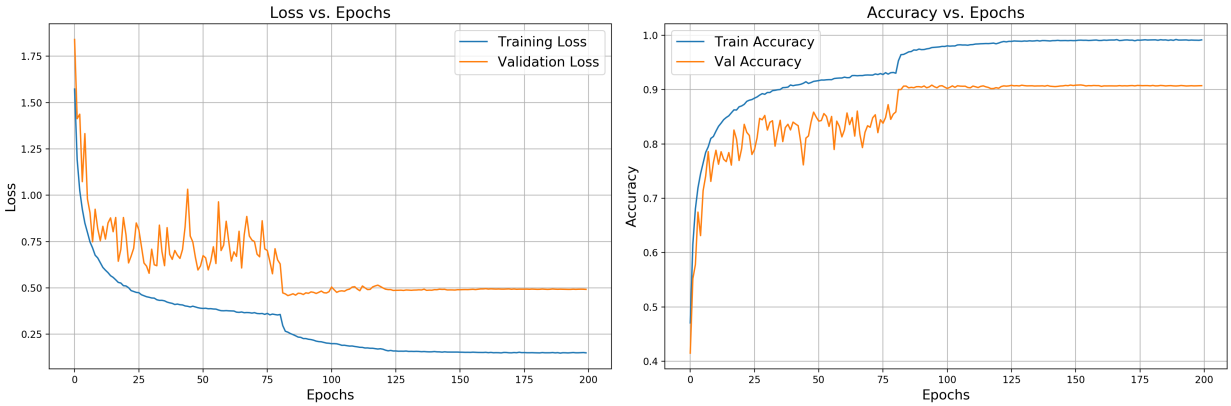

Figure 7: Loss/Accuracy versus training epochs fro ResNet.

Kaggle also holds a competition based on CIFAR10 dataset. The result from ResNet are among the top 20 based on the historical data.