EE 569: Homework #4: Image Recognition with Convolutional Neural Network

Issued: 03/26/2017 Due: 11:59PM, 04/23/2017

General Instructions:

1. Read Homework Guidelines for the information about homework programming, write-up and submission. If you make any assumptions about a problem, please clearly state them in your report.
2. Do not copy sentences directly from any listed reference or online source. Written reports and source codes are subject to verification for plagiarism. You need to understand the USC policy on academic integrity and penalties for cheating and plagiarism. These rules will be strictly enforced.
3. In this homework, it is recommended to use *Tensorflow* for deep learning programming. *Torch7* is an alternative choice if you have experience and feel more comfortable with it.

Problem 1: CNN Training and Its Application to the CIFAR-10 Dataset (50 %)

In this problem, you will learn to train one simple convolutional neural network (CNN) derived from the LeNet-5 introduced by LeCun *et al.* [1]. Furthermore, you need to apply it to the CIFAR-10 dataset [2]. The CIFAR-10 dataset consists of 60,000 RGB 32x32 pixel images in 10 classes (with 6000 images per class). It includes a labeled training set of 50,000 images and a test set of 10,000 images. Fig.1 shows some exemplary images from the CIFAR-10 dataset.

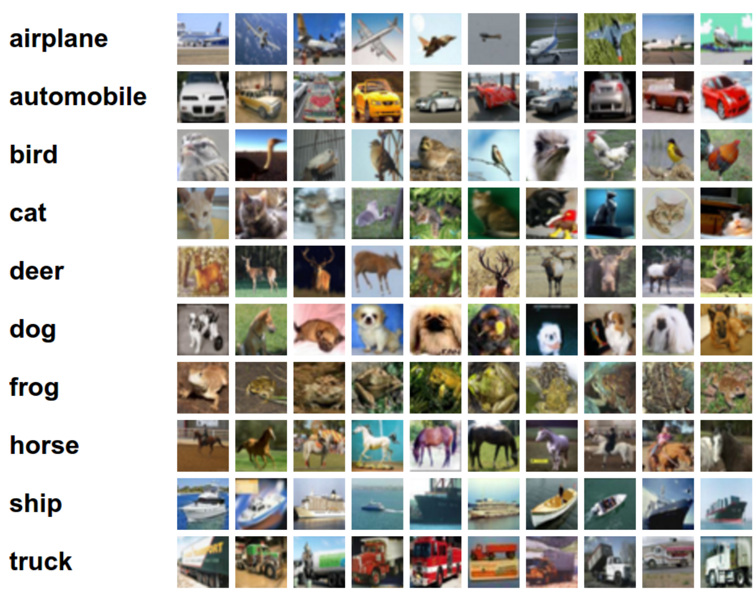

Figure 1: CIFAR-10

The CNN architecture for this assignment is given in Figure 2. This network has two *conv* layers, and three *fc* layers. Each *conv* layer is followed by a *max pooling* layer. Both *conv* layers accept an input receptive field of spatial size *5x5*. The filter numbers of the first and the second *conv* layers are 6 and 16 respectively. The stride parameter is 1 and no padding is used. The two *max pooling* layers take an input window size of *2x2*, reduce the window size to *1x1* by choosing the maximum value of the four responses. The first two *fc* layers have 120 and 80 filters, respectively. The last fc layer, the output layer, has size of 10 to match the number of object classes in CIFAR-10. Use the popular ReLU activation function [3] for all *conv* and all *fc* layers except for the output layer, which uses softmax [4] to compute the probabilities.

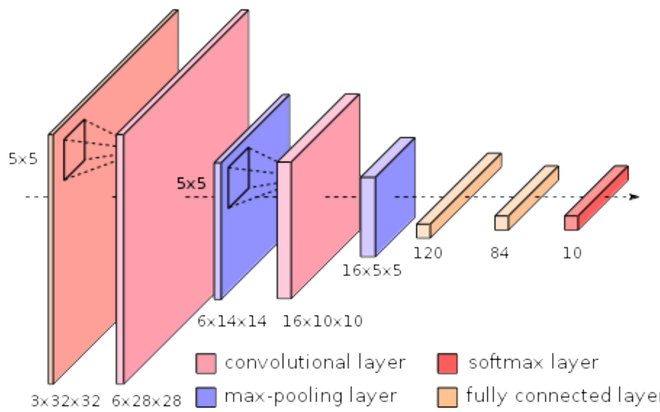

Figure 2: A CNN architecture derived from LeNet-5.

(a) CNN Architecture and Training (15%)

Explain the architecture and operational mechanism of convolutional neural networks by performing the following tasks.

- Describe CNN components in your own words: 1) the fully connected layer, 2) the convolutional layer, 3) the max pooling layer, 4) the activation function, and 5) the softmax function.
- What are the functions of these components?
- What is the major difference between a CNN and the traditional multi-layer perceptron (MLP)?
- Why CNNs work much better than other traditional methods in many computer vision problems? You can use the image classification problem as an example to elaborate your points.
- Explain the loss function and the classical backpropagation (BP) optimization procedure to train such a convolutional neural network.

Show your understanding as much as possible in words in your report.

(b) Application of the CNN to CIFAR-10 Dataset (15%)

Train the CNN given in Fig. 2 using the 50,000 training images from the CIFAR-10 dataset. You can adopt proper preprocessing techniques and the random network initialization to make your training work easy.

- Explain your preprocessing step and the random network initialization scheme. Justify these practices.
- Compute the mAP (mean averaged precision) performance curves using the epoch-accuracy (or iteration-accuracy) plot on training and test datasets separately. Plot the performance curves under 5 different yet representative parameter settings. Discuss your observations.

- Find the best parameter setting to achieve the highest accuracy on the test set. Then, plot the performance curves for the test set and the training set under this setting.

(c) K-means with CNNs (20%)
Professor Kuo mentioned the use of the k-means method to initialize the network in [5]. Explain how it works. Does the k-means initialization work better than the random initialization for the CIFAR-10 dataset? Conduct experiments, report your results and make discussion.

Problem 2: Capability and Limitation of Convolutional Neural Networks (50%)

(a) Improving Your Network for CIFAR-10 Dataset (Advanced: 25%)
Feel free in modifying the baseline CNN in Figure 2 to improve the classification accuracy obtained in Problem 1(b). For example, you can increase the depth of the network by adding more layers, or/and change the number of filters in some layers. You can augment the dataset. You can also try different activation functions or optimization algorithms. They all have a potential to improve the result. You may need to fine-tune the training parameters to get the training job done. Report the best accuracy that you can achieve and describe your network architecture and the training parameter setting to reach this result. Discuss the sources of performance improvement. Your grading in this part will be based on your obtained performance in comparison with other students in the same class. Thus, you can view this as a competition (or a challenge) among students in EE569.

(b) State-of-the-Art CIFAR-10 Implementation (Advanced: 25%)
Check the state-of-art implementation on CIFAR-10 classification in [6]. Select one paper from the list for discussion.

- Describe what the authors did to achieve such a result. You do not have to implement the network.
- Compare the solution with the baseline CNN and your improved CNN and discuss pros and cons of the three methods.

Your answer should fit in two pages. You can add pictures, flowcharts, and diagrams in your report. If you do so, you need to cite their sources.

References

[1] LeNet-5 http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf
[2] CIFAR-10 https://www.cs.toronto.edu/~kriz/cifar.html
[3] ReLU https://en.wikipedia.org/wiki/Rectifier_(neural_networks).
[4] Softmax https://en.wikipedia.org/wiki/Softmax_function
[5] Prof.'s Paper https://arxiv.org/abs/1701.08481
[6][CIFAR-10 Leaderboard
http://rodrigob.github.io/are_we_there_yet/build/classification_datasets_results.html#43494641522d3130