6.867 Final Project Writeup

Vickie Ye and Alexandr Wang

Abstract

In this project, we compared different methods for facial expression recognition.

1 Introduction

1.1 PCA

1.2 Facial Landmark Detection

1.3 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a particularly type of artificial neural network that works particularly well for image data. The structure of CNNs exploits strong local correlation in the inputs. This is done by enforcing local connectivity between neurons of adjacent layers. The inputs of a hidden unit at a particular layer n are some locally-connected subset of the units of the previous layer $n-1$, done in such a way that the input to layer n represent some tile of the units of layer $n-1$, where all the tiles overlap.

In addition, CNNs utilize shared parameterizations (weight vector and bias) across layers. By constraining the same types of weights across layers, essentially replicating units across layers, allows for features to be detected regardless to their position in the initial input, making the network much more robust to real world image data. Additionally, by constraining multiple weights to be the same reduces the parameters to be learnt, increasing learning efficiency.

2 Experimental Details

The dataset we used was taken from a Kaggle facial expression dataset.

2.1 Convolutional Neural Network

For implementing our convolutional neural network, we used TensorFlow for the entire implementation, i.e. training, inference, and evaluation.

2.1.1 Minimizing Overfitting

Since the size of our training set was 28,709 labeled examples, and our neural net needed to go through many more than 28,000 examples to achieve convergence, we employed a few techniques to ensure that our convolutional neural net did not overfit to the training set.

First off, we implemented learning rate decay on our convolutional neural net, so that the learning rate decreased as it had been trained through more examples. We used exponential decay, so that that the learning rate decayed by a factor of 0.1 after training through $1,200,000$ examples, and we had it decay in a step change manner as is visible in Figure 2. We found that the step change learning rate decay worked better than a continuous exponential decay. This is probably because maintaining a high learning rate initially could ensure that the CNN was trained towards a good local optimum in fewer steps, whereas a steadily decreasing learning rate could limit the range of the neural network.

In addition, we implemented distortion of the images while training to artificially increase the size of our training set, and make our convolutional neural net more robust to slightly distorted inputs.

3 Results and Analysis

References

[1] Lawrence, S.; Giles, L.; Tsoi, A. C.; Back, A. D. (1997) "Face Recognition: A Convolu-

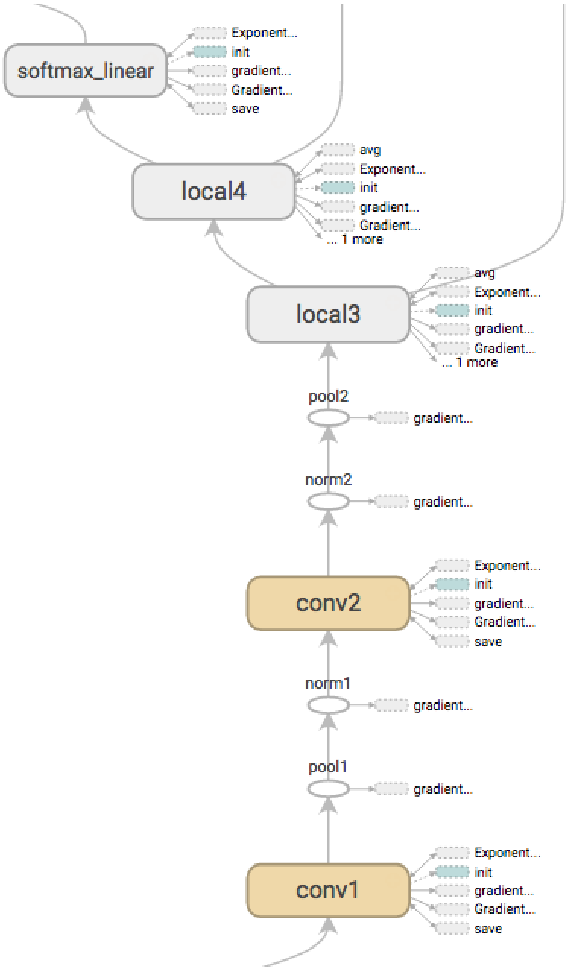

Figure 1: TensorFlow computation graph for the neural network in the CNN we implemented, illustrating each of the layers.

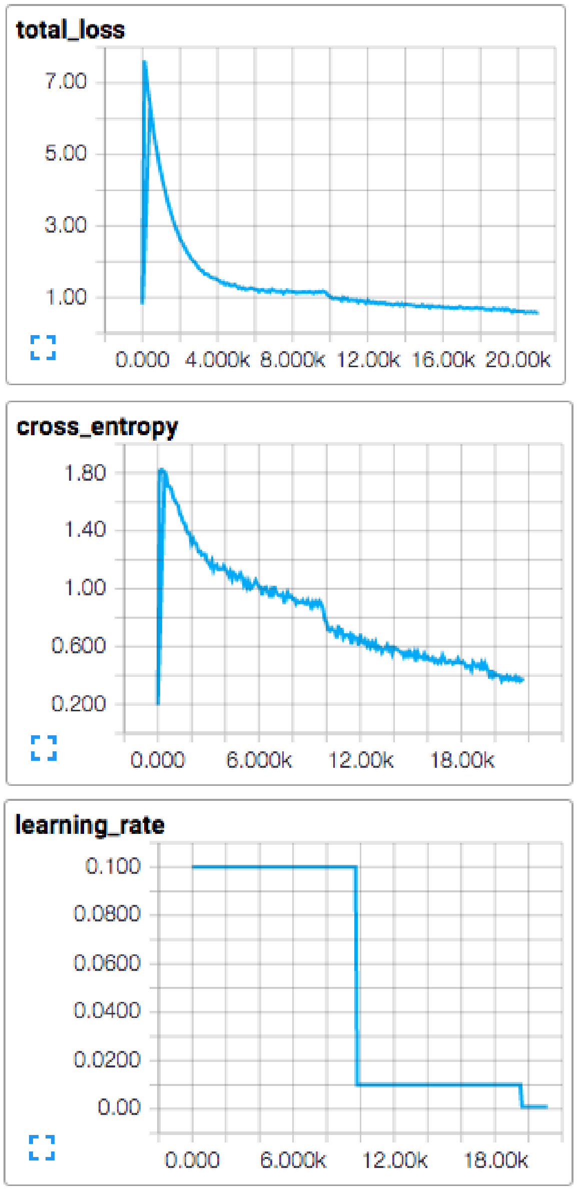

Figure 2: The total loss, cross entropy, and learning rate of the convolutional neural net over time.

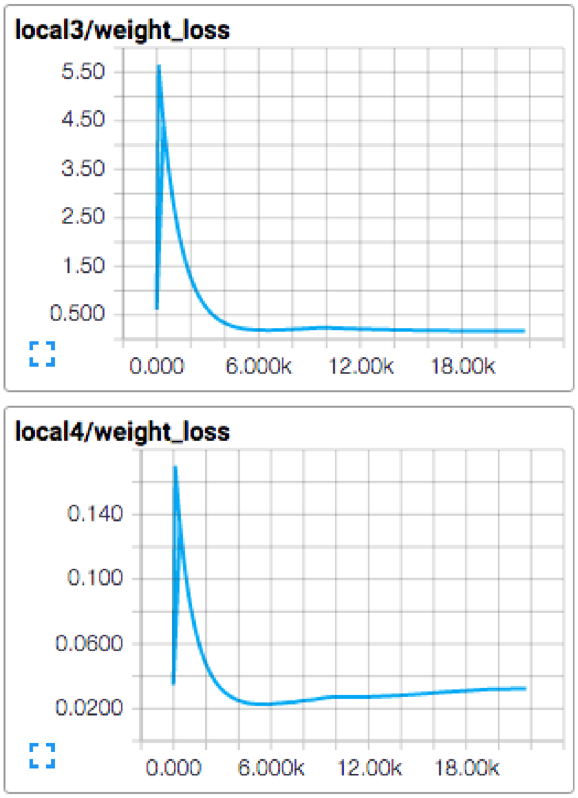

Figure 3: The loss on the local3 layer and local4 layer over time in the convolutional neural net.

tional Neural-Network Approach" *Neural Networks, IEEE transactions on* 8 (1):98-113

[2] Matsugu, M.; Mori, K.; Mitari Y.; Kaneda Y. (2003) "Subject independent facial expression recognition with robust face detection using a convolutional neural network" *Neural Networks* 16 (5):555-559