

Sameer Bhatti  
HW8

V1b) Yes, this convolutional layer is the same as a 3x3 average pooling layer. They would compute the same function because the kernel contains 9 elements and would therefore look at 9 elements. Since the kernel takes  $1/9^{\text{th}}$  of each element, it is similar to summing up each individual element and dividing by 9. This is exactly as taking the average because there are 9 elements. These would behave similarly during train and test because the average pooling layer would simply take the average of each 3x3 segment while a convolutional layer would utilize the kernel to go through each 3x3 segment with 1 pixel zero padding on the input to produce the same output.

V2a) See Matlab code

V2b) I get about 70% accuracy when training the network. It's not the same every time because there is a lot of randomness associated with training. The weights are randomly initialized, the batches are randomly selected and the stochastic gradient selects a set of random samples to compute the gradient after each update.

V2c) By removing the last convolution layer, the accuracy drops to about 66%, a 4% drop in accuracy compared to the original set of layers.

V2d) Changing average pooling layers to max pooling layers changes the pixels that are selected from the images. Average pooling layers will take the average of the set of values being looked at in the input while max pooling layers will just take the largest value. However, this did not have much of an effect on the accuracy of the layers, coming up as about 71% when all the average pooling layers are changed to max pooling layers.

V3a) The network achieves higher accuracy than last time because it is a much larger network, it is pretrained, and has less classes.

V3b) The softmax layer needs to have 4 outputs because it is classifying 4 categories of images. Therefore, the Alex net is scaled down from outputs of 4096 to 4.

V3c) By removing the ReLu layer and FC64 layer, the outputs go directly from 4096 to 4. Given this, the accuracy actually increased by 1-2% from about 85% to 87%.

V4) It is constantly swinging on descriptions of objects because the code is in a while loop so it is constantly classifying the image but the network was also not possibly trained for the images. For example, I had the webcam on my face and it had classified my face as a "nipple." My face is obviously not an image that would have been trained by the neural network.