

CS 466/566 Assignment 1 Report

SEDA NUR DOĞANAY S004299

Network Designs

Meanings:

3x3 convolutional patch creating 32 feature maps with stride 2 = conv 3x3x32-s-2

2x2 Max pooling layer with stride 4 = maxpool 2x2-s-4

Rectified Linear Unit = ReLu

<i>Network Layout</i>	
<i>Network 1</i>	<i>Network 2</i>
<i>conv 5x5x32-s-1</i>	<i>conv 4x4x64-s-2</i>
<i>ReLu</i>	<i>ReLu</i>
<i>maxpool 2x2-s-2</i>	<i>maxpool 4x4-s-2</i>
<i>conv 5x5x64-s-1</i>	<i>conv 3x3x128-s-1</i>
<i>ReLu</i>	<i>ReLu</i>
<i>maxpool 2x2-s-2</i>	<i>maxpool 4x4-s-2</i>
<i>fc</i>	<i>fc</i>
<i>fc</i>	<i>fc</i>

Training Description

<i>Learning rate</i>	<i>0.01</i>
<i>Number of training iterations</i>	<i>20000</i>
<i>Batch size</i>	<i>64</i>

On the side is the training specification of both my networks. Besides, in order to force the system to not to overfit dropout is used in last fully connected layer with a percentage of 25%. In each iteration, a different set of neurons was dropped out therefore the network became more tolerant. As the cost

function softmax cross entropy is chosen because an image can be described by one digit only, i.e. our job was a discrete classification task. Adam is used as the optimization algorithm. It is basically a different version of stochastic gradient descent algorithm, which the gradient used in each iteration, is updated from the previous using a technique based in momenta. Momenta mean the update is a linear combination of the current stochastic gradient and the previous update. In all layers padding was set to

SAME. It means the output has the same dimensions as the input so the algorithm inserted some padding when needed. On the other hand while down-sampling to generate 14x14_dataset, the padding was set to VALID, meaning no padding.

Data Augmentation

My augmentation plans included noise addition up to 5%, brightness change and; rotation between -30° and +30° or affine transformation. Unfortunately, I couldn't finish this part.

Experimental Results

<i>Experimental Results</i>	<i>Performance in %</i>	
Data Set	Network 1	Network 2
28x28_dataset	0.94	0.94
14x14_dataset	0.92	0.91
14x14_augmented_dataset	-	-

Discussion

Surprisingly, I've managed to have nearly the same performance from both networks. This happened because I didn't make a controlled experiment, changed both number of feature maps, patch sizes and strides. Increasing the number of feature maps compensated the effect of decreasing convolution patch size and increasing max pooling patch size.

75% dropout was too much, neurons couldn't figure out anything and 5% dropout was too little that neurons collaborated and overfitted. I believe augmenting some more 14x14 images would result in better performance unless the augmentation is logical.