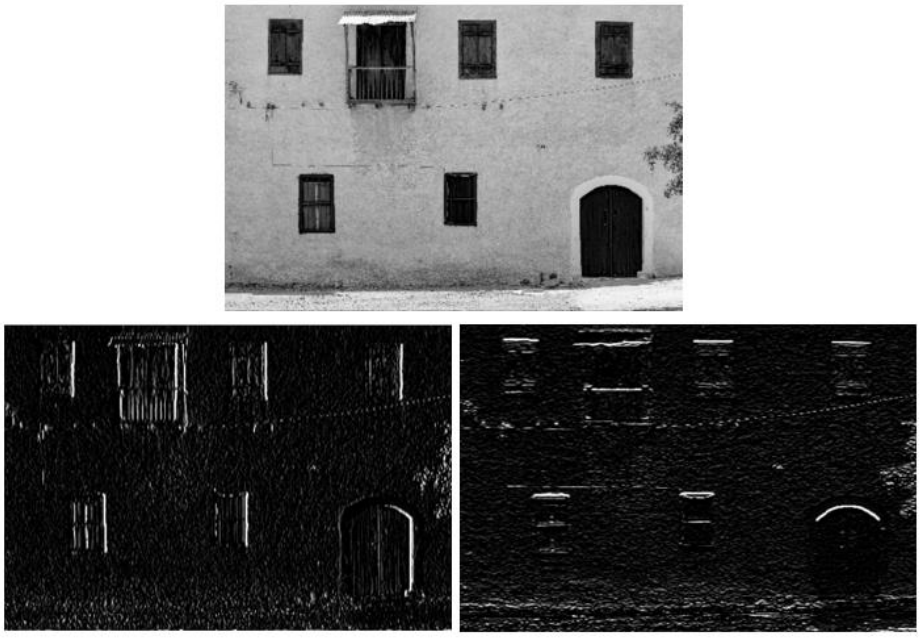
CISC 616

Assignment #2 (Due May 6, 2022 @ 11:59 PM)

1. Consider an input image of shape 500x500x3. The image is flattened and a fully connected layer with 100 hidden units is used. What is the shape of the weight matrix of this layer (without the bias)? What is the shape of the bias.

2. You run this image in a convolutional layer with 10 filters, of kernel size 5x5. How many arameters does this layer have?

3. The top gray image has run through different types of filters and the results are shown in the following images. What type of convolutional filter was used to get each of the resulting images. Explain briefly and include the values of these filters. The filters have a shape of (3,3).



4. In the Adam optimizer. Show what will happen the numbers of steps to compute the exponential moving averages gets large.

5. Given a batch of size *m,* and assume that a batch normalization layer takes an input $z = (z^{(1)}, \ldots, z^{(m)})$ as an input. Write down the output equation(s) of this layer. Give two reasons for using the batch normalization layer.

6. Suppose you have a convolutional network with the following architecture: The input is an RGB image of size 256 x 256. The first layer is a convolution layer with 32 feature maps and filters of size

3x3. It uses a stride of 1, so it has the same width and height as the original image. _ The next layer is a pooling layer with a stride of 2 (so it reduces the size of each dimension by a factor of 2) and pooling groups of size 3 x 3.
Determine the size of the receptive _eld for a single unit in the pooling layer. (i.e., determine the size of the region of the input image which influences the activation of that unit.) You may assume the receptive field lies entirely within the image.

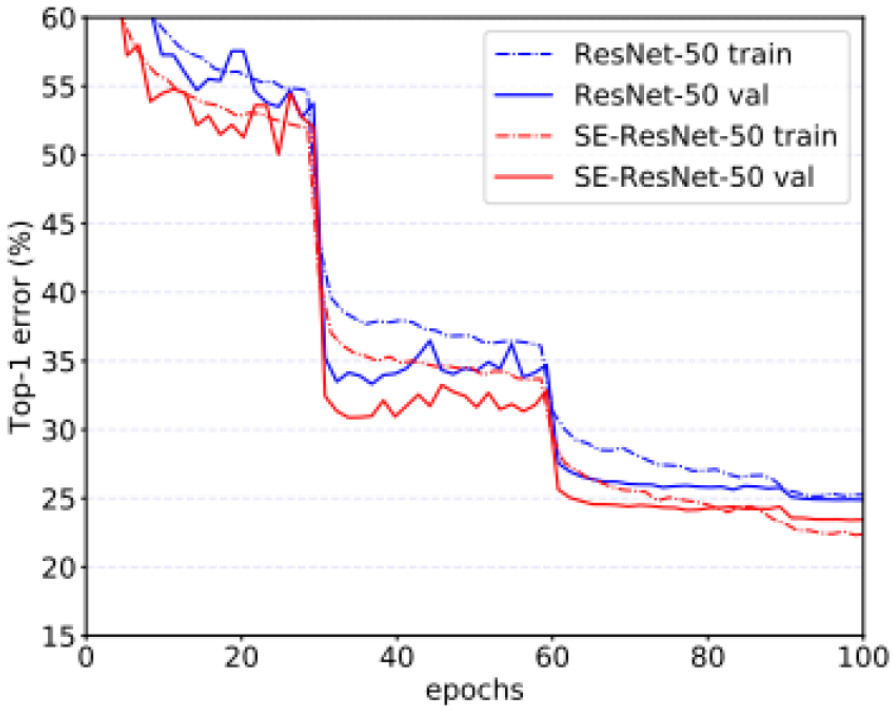
7. If an input data block in a convolutional network has dimension $C \times H \times W = 96 \times 128 \times 128$, (96 channels, spatial dim 128x128) and we apply a convolutional filter to it of dimension $D \times C \times HF \times WF = 128 \times 96 \times 7 \times 7$, (i.e. a block of D=128 filters) with stride 2 and pad 3,what is the dimension of the output data block?

8. What is inverted dropout and what is its advantage?

9. Explain briefly why fully connected neural networks do not work well for image classification.

10. Compute the convolution of the following two arrays: (4 1 -1 3) * (-2 1). Your answer should be an array of length 5. Show your detailed work.

11. Describe what setting change in epochs 25 and 60 could have produced this training curve. Be brief.



12. Why are convolutional layers more commonly used than fully-connected layers for image processing?

13. Dropout layers implement different forward functions at train and test time. Explain what they do. Let p be the probability that node value is *retained.*

14. Explain how standard momentum and Nesterov accelerated gradient differ. How does their performance (convergence as a function of time) differ on convex optimization problems? Use sketches as an aid.

15. How does the actual learning rate of ADAGRAD change with the number of steps?