# CS4851/6851 IDL: Homework 3 February 28, 2022

Note: All coding problems to be submited with Github Link. Do not Upload the files/folder. Use git commands only.

Note: this is the distribution of questions:

1. Question 1 to Question2: Required for everyone.
2. Question 3: Required only for Graduate Students
3. Question 4: Bonus question for both Graduate Students and Undergraduate Students

# Problem 1 (20 points)

You have a convolutional neural network that takes as an input image of size 512 *×* 512 *×* 3 and passes it through a layer that convolves the image using 3 filters of dimensions 5 *×* 5 *×* 3 with a valid padding.

1. List all learnable parameters of this convolution layer.

512 x 512 – width and height of images, 3 – number of channels

Number of weights, Number of Bias, Input layer – images of size 512 x 512 x 3 = 786,432, number of padding (however valid padding means no requirement for padding), number of stride

* Input Layer has no learnable parameters since it just contains the input data
* Convo Layer- We have 3 from our input layer. # of outputs = # of filters times filter size = 3 x 5 x 5 = 75. Multiplying by our three inputs 75 x 3 = 225 + 3 ( # of biases = # of filters) = 228.

1. What if you want to replicate the behavior of this convolutional layer using a fully connected layer? How many parameters would that fully connected layer have?

* A fully connected layer has each input have effects on each output.
* Input Layer has no learnable parameters since it just contains the input data
* # of connections between the first and second layer: 3 x 3 = 9, # of connections between the second and third layer: 3 x 2 = 6, # of bias between first and neurons of the second layer is 1 x 3 = 3, # of biases between second layer and neurons of the third layer: 1 x 2 = 2.
* # of learnable parameters – (3 x 3) + (3 x 2) + (1 x 2) + 1 x 2 = 38

# Problem 2 (20 points)

Given a binary input image of diagonal streaks (see example in Figure [2](#_bookmark2)) and two filters (see Figure [1a](#_bookmark0)) describe how would you build a detector for finding the location of pattern shown in Figure [1b](#_bookmark1) on the input image. Allowed operations are convolution, summation, and argmax.

A binary image has only two values, pixel intensity of 0 or 1. Conducting a convolution on the inputted image, the argmax is then calculated In each part of the feature map of the image, a convolution layer is then conducted again with argmax layer following right after, summation is then conducted to provide an outputted image.

Bonus for undergraduates beyond this line

# Problem 3 (20 points)

Demonstrate that convolution is translation invariant for 1D convolution (Note: this can be extended to N-D convolutions as well).

Translation invariance – if a model’s output is unaffected by shifts of the input. It can be achieved by combining convolutions with spatial pooling operators.

It can

Maxpool-Convolution-Maxpool

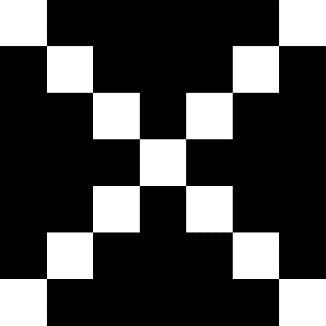
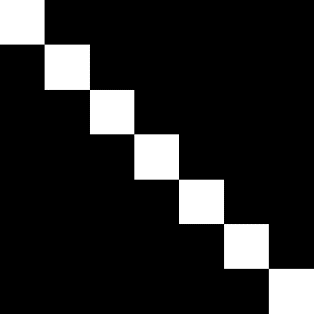
F1

F2

F3

Bonus for both undergraduates and graduates beyond this line.

0 0 0



1 1 1

2 2 2

3 3 3

4 4

4

5 5

5

6 6

6

0 2 4 6

0 2 4 6

0 1 2 3 4 5 6

* 1. Available fixed filters (b) Target shape to detect

Figure 1: Filters to use ([1a](#_bookmark0)) and pattern to find ([1b](#_bookmark1)).

# Problem 4 (20 points)

You have to choose between two papers given below:

1. Paper 1: [High-Performance Neural Networks for Visual Object Classification-](https://arxiv.org/pdf/1102.0183.pdf):
   1. Give a short summary of the paper.
2. This paper provides an understanding as to how CNNs work with GPUs over CPUs. It starts by discussing the parts that create a CNN ranging from the pre-processing layer of image processing, convolution, max pooling, and classification layers. It then gets into depth as the GPU implementation of NVIDIA series 400 & 500 with forward propagation and back propagation. Experiments are then conducted on different datasets MNIST, CIFAR-10, NORB.
   1. What were the parameter sizes for CIFAR-10 and MNIST? why do you think the parameter size differed for CIFAR-10 vs MNIST?
3. CIFAR-10 contains a collection of natural color images that are 32x 32, 10 classes, 5000 samples in the training set, 1000 samples in the test set, small kernels are used, skipping factor of 0, a learning rate of 0.001, a 0.993 epoch size, number of maps. The parameters for MNIST is the number of epochs, learning rate, deformed images. The images in CIFAR-10 images are already smaller than NORBs, so subsequently the are already smaller than MNIST allowing for more parameters.
4. Paper 2: [ImageNet Classification with Deep Convolutional Neural Networks](https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf):
   1. Give a short summary of the paper.
   2. Why is there a big fluctuation of loss for the last epoch of training?

0



100

200

300

400

500

600

700

0 100 200 300 400 500 600 700

Figure 2: An example image for the architecture