Homework 3

Please upload your assignments on or before Oct 16, 2020.

- You are encouraged to discuss ideas with each other; but
- you **must acknowledge** your collaborator, and
- you **must compose your own** writeup and/or code independently.
- We **require** answers to theory questions to be written in LaTeX, and answers to coding questions in Python (Jupyter notebooks)
- Upload your answers in the form of a single PDF on Gradescope.

---

1. **(2 points)** *Convolutional layers*. Suppose that a convolutional layer of a neural network has an input 3D tensor $X[i, j, k]$ and computes an output as follows:

$$Z[i, j, m] = \sum_{k_1} \sum_{k_2} \sum_{n} W[k_1, k_2, n, m] X[i + k_1, j + k_2, n] + b[m]$$
$$Y[i, j, m] = \text{ReLU}(Z[i, j, m])$$

for some convolutional kernel W and bias b. Suppose X and W have shapes $(48, 64, 10)$ and $(3, 3, 10, 20)$ respectively. Assume no padding (i.e., neither zero- or circular-padding is applied), and stride 1.

   a. What are the shapes of Z and Y?

   b. What are the number of input and output channels?

   c. What are the total number of trainable parameters in this layer?

   d. How many arithmetic multiplications are needed to perform the convolution operation for this layer?

2. **(2 points)** *Overfitting in neural nets*. This is a three-part question.

   a. Consider neural network learning, and sketch *typical* curves of loss versus iteration count for the training and validation sets using stochastic gradient descent. (Assume that overfitting to the training set happens at some point.)

   b. We discussed data augmentation as a possible approach for combat overfitting in neural net training. List any two other strategies.

   c. For image classification, explain why using data augmentation to may work for cat-vs-dog classifiers, but is **not** the right thing to do for distinguishing between English handwritten characters (a,b,c,d,...).

3. **(2 points)** *The IoU metric*. Recall the definition of the IoU metric (or the Jaccard similarity index) for comparing bounding boxes.

   a. Using elementary properties of sets, prove that the IoU metric between any two pair of bounding boxes is always a non-negative real number in $[0, 1]$.

b. If we represent each bounding box as a function of the top-left and bottom-right coordinates (assume all coordinates are real numbers) then argue that the IoU metric is *non-differentiable* and hence cannot be directly optimized by gradient descent.

4. **(4 points)** In this programming exercise, we will explore the performance of three different object detection networks. We will be using Detectron2, Facebook AI's object detector library; here is the repository. It will be helpful to go through the excellent tutorial (with example code) here.

   a. Download the following test image (a picture of pedestrians in Central Park). We will run two different detectors on this image.

   b. First, consider the COCO Keypoint Person Detector model with a ResNet50-FPN backbone, which is trained to detect human silhouettes. This can be found in the Detectron2 Model Zoo in the "COCO Keypoint" table. Use this model to detect as many silhouttes of people in the test image as you can. You may have to play around with the thresholds to optimize performance.

   c. Second, repeat the above procedure, but with the Mask R-CNN model with ResNet50-FPN backbone, available in the Model Zoo in the "COCO Instance Segmentation" table. This time, you should be able to detect both people as well as other objects in the scene. Comment on your findings.

   d. It appears that the balloons in the test image are not being properly detected in either model. This is because the COCO dataset used to train the above models does not contain balloons! Following the tutorial code above, start with the above pre-trained Mask R-CNN model and train a balloon detector using the (fine-tuning) balloon image dataset provided here. Test it on the original test image and show that you are now able to identify all the balloons.