EN.520.665 Machine Perception Fall 2021

Final

1. (10 points) If P is a 3×4 camera projection matrix and C is the camera center in the world coordinate system (in homogeneous coordinates), show that PC = 0
2. (10 points) We discussed a factorization algorithm for recovering camera information and 3D point locations simultaneously from multiple images. In this algorithm, we assumed that the camera was orthographic and that all points in all views. Now assume that some points are not seen in some views. How would one modify the algorithm to produce reasonable results?
3. (10 points) Suppose that the reflectance map *R* is linear in *p* and *q*, where *p* and *q* are the partial derivatives of depth *z* with respect to *x* and *y* respectively.

Thus R = a*p* + b*q* + c.

We have an image, including the silhouette of a simple convex object of shape *z*= *f* (*x*, *y*). Show that the surface *z*\* = *f* (*x*, *y*) + *g* (b*x* – a*y*),

for an arbitrary differentiable function *g* (s), will give rise to the same image. Does the surface *z* \* have the same silhouette? Assume that the derivative of *g* is bounded.

1. (15 points) Let D and W be the matrices defined in Shi and Malik’s paper on normalized cuts.

Define L=D – W. Prove that

(a) (6 points) For every vector f ∈ Rn we have fTLf = , where T is the transpose operator.

(b) (3 points) L is symmetric and positive semi-definite.

© (3 points) The smallest eigenvalue of L is 0, the corresponding eigenvector is the constant one vector .

(d) (3 points) L has n non-negative, real-valued eigenvalues 0 = λ1 ≤ λ2 ≤ ... ≤ λ

1. (15 points) Suppose you wish to accurately determine the position and attitude of a large object that is constrained to be near some standard position and attitude. You have available a number of camera systems, each of which is capable of accurately locating the image of some notable features on a portion of the surface of the object.
2. (8 points) Determine the number cameras you need to determine the position and attitude of the object and explain how you arrived at this number.
3. (7 points) Suggest and algorithm for recovering the position and attitude and explain how would you determine the required parameters?
4. (20 points) Consider a convolutional neural network block whose input size is 64 × 64 × 8. The block consists of the following layers:

• A convolutional layer 32 filters with height and width 3 and 0 padding which has both a weight and a bias (i.e., CONV3-32)

• A 2 × 2 max-pooling layer with stride 2 and 0 padding (i.e., POOL-2)

• A batch normalization layer (i.e., BATCHNORM) Compute the output activation volume dimensions and number of parameters of the layers. You can write the activation shapes in the format (H, W, C where H, W, C are the height, width, and channel dimensions, respectively.

1. (2 points) What is the output activation volume dimensions and number of parameters for CONV3-32? Neural network problem?
2. (2 points) What is the output activation volume dimensions and number of parameters for POOL2?
3. (2 points) What is the output activation volume dimensions and number of parameters for BATCHNORM?

Now you will design a very small convolutional neural network for the task of digit classification given a 16 × 16 image (with 3 channels i.e., RGB), you want to predict the digit shown in the image. Therefore, this is a 10-class classification problem. Your network will have 4 layers, given in order: a convolutional layer, a max-pooling layer, a flatten layer, and a fully-connected layer. Design the neural network to solve this problem. Of course, there are many, many solutions to this question. To narrow the solution space, here are some restrictions:

• Your convolutional layer must have a stride of 1 and have 4 filters. Because of memory limits, this layer’s activation volume should not have no more than 576 total elements in the tensor per input image (e.g., a 2 × 2 × 8 tensor has 2 ∗ 2 ∗ 8 = 32 total elements in it).

• Your max-pooling layer must have the same stride as pool size e.g., a 2 × 2 pooling layer must have a stride of 2.

• Again because of memory limits, your final fully-connected layer will not have a bias term, and its total number of parameters cannot exceed 1440.

The rest is up to you. Answer the following questions to incrementally build out your small CNN architecture. iv.

1. (4 points) What are the hyperparameters of the convolutional layer you propose (i.e., filter size, padding)? Also, what are the activation volume dimensions for this layer?
2. (4 points) What are the hyperparameters of the max-pooling layer you propose (i.e., pool size)? Also, what are the activation volume dimensions for this layer?
3. (2 points) What are the activation volume dimensions of the final fully-connected layer?
4. (4 points) You start training your model and notice underfitting, so you decide to add data augmentation as part of your preprocessing pipeline. Given that you are working with images of handwritten digits, for each data augmentation technique, state whether or not the technique is appropriate for the task. If not, explain why not.
5. Scaling slightly
6. Flipping vertically or horizontally
7. Rotating by 90 or 180 degrees
8. Shearing slightly