

CS7643: Deep Learning

Quiz 2 Problem Set Prep

Instructor: Zsolt Kira

January 27, 2024

1. You have an input volume of $32 \times 32 \times 3$. What are the dimensions of the resulting volume after convolving a 5×5 kernel with 0 padding (ie. valid convolution), stride of 1, and 2 filters?
2. How many weights and biases would the layer defined above have?
3. You want to process time-series data with a 1D CONV that has the same configuration as the layer presented in (1) but with a kernel of size 5. The input volume of shape $T \times 3$ models three fluctuating values over time. How many weights and biases does this layer have? Assume the same configuration (padding, stride, number of filters) as in (1).
4. Suppose you have an input volume of dimension $64 \times 64 \times 16$. How many parameters would a single 1×1 convolutional filter have, including the bias?
5. Suppose your input is a 300 by 300 color (RGB) image, and you use a convolutional layer with 100 filters that are each 5×5 . How many parameters does this layer have including the bias parameters?
6. You have an input volume that is $63 \times 63 \times 16$ and convolve it with 32 filters that are each 7×7 , and stride of 1. You want to use a **same** convolution. What is the padding?
7. What is the output volume of a $32 \times 32 \times 16$ input data after applying max pooling with a square kernel of size 2 and stride = 2?
8. What is the resulting volume of padding a $15 \times 15 \times 8$ input volume using pad=2?

Here are a couple of fun 3D convolution problems to solve that will not be in quiz 2.

9. You are working with 3D data. You are building a network layer whose input volume has size $32 \times 32 \times 32 \times 16$ (this volume has 16 channels), and applies convolutions with 32 filters of dimension $3 \times 3 \times 3$, stride=1 and no padding. What is the resulting output volume?
10. You want to process a video with a 3D CONV. The input video can be seen as a sequence of images indexed by time, i.e. a volume of shape $W \times H \times T \times 3$. How many weights and biases does this layer have? Assume the convolution uses 2 filters with kernel of shape $5 \times 5 \times 5$, no padding, and a stride of 1.