

# HW8

Sunday, October 31, 2021 1:32 PM

## Inception Network Motivations:

For one specific layer, you have data 28X28 image by 192 channels, i.e., data frame 28X28X192.

You have two options with filters for the next layer:

Option 1:

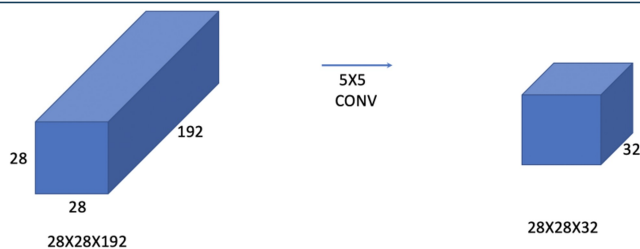
You have, 5X5 filter size with 32 filters, and stride 1 and with padding.

Option 2:

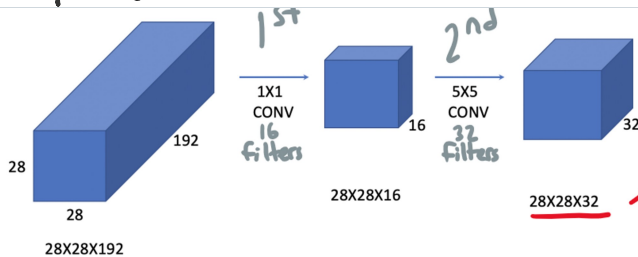
You have, first, 1X1 filter size with 16 filters, and stride 1 and with padding; then you follow up to apply 5X5 filter size with 32 filters, and also stride 1 and with padding.

Compute the related parameters in option 1 vs. option 2. Which option requires more? Explain why.

Option 1



Option 2



Option 1

$$28 \times 28 \times 5 \times 5 \times 192 \times 32 = 120,422,400$$

Input image size      Filter size      Depth      # of filters      # of multiplications needed

## Option 2

$$\text{Part 1: } 28 \times 28 \times 1 \times 1 \times 192 \times 16 = \underline{2,408,448}$$

Input image size      Filter size      Depth      # of filters      # of multiplications needed for 1st part

$$\text{Part 2: } 28 \times 28 \times 5 \times 5 \times 16 \times 32 = \underline{10,035,200}$$

# of multiplications needed for 2nd part

$$2,408,448 + 10,035,200 = \boxed{12,443,648}$$

Total # of multiplications needed

As you could see, option 2 requires far fewer multiplications compared to option 1. As a result, option 2 has a much lower computational cost. By doing the  $1 \times 1$  convolution before the  $5 \times 5$  convolution, we are reducing the feature size or dimensionality of the input image while maintaining the spatial dimension. This makes option 2 much cheaper than option 1.