IN3310 2024 UiO - Exercise Week 5 - Solution

1 Spatial Output Size

Let O = output dimension, M = input dimension, r = padding, k size = kernel size, and S = stride. Then we have the formula

$$O = \text{ floor } \left(\frac{M + 2r - ksize}{S} \right) + 1$$

- You are given a 2-dimensional convolution with feature map input size (78, 84). When using a kernel of size (5, 5) and stride 3 with padding of 2, what will be the spatial size of the feature map, which is the output of the convolution? Note that the spatial size does not depend on the number of input or output channels.

$$O = \text{floor} \left(\frac{78 + 4 - 5}{3} + 1, \frac{84 + 4 - 5}{3} + 1 \right) = (25, 28)$$

- You are given a 1-dimensional convolution. When using a kernel of size 9 and stride 3 with padding 1, which spatial input size do you need to have so that you have a spatial output size of 16?

$$16 = \text{floor} \left(\frac{M + 2 - 9}{3} + 1 \right).$$

  One possible value of M can be calculated from $3 * 15 = M - 7 \implies M = 52$

- You are given a 2-dimensional convolution. When using a kernel of size (3, 5) and stride 2 with padding of 0, what will be the spatial size of the feature map, which is the output of the convolution, which spatial input size do you need to have, so that you have a spatial output size of (128, 96)?

  a. Let (M_1, M_2) be the input size. Then

$$O = \text{floor} \left(\frac{M_1 - 3}{2} + 1, \frac{M_2 - 5}{2} + 1 \right)$$

  b.

$$(128, 96) = \text{floor} \left(\frac{M_1 - 3}{2} + 1, \frac{M_2 - 5}{2} + 1 \right)$$

Possible values of M_1, M_2 can be calculated by $M_1 = 127 * 2 + 3 = 257, M_2 = 95 * 2 + 5 = 195$.

2 Number of Trainable Parameters

Note that the number of parameters for a convolutional network depends upon the kernel size, and number of input and output channels, but not on the input size or the stride.

How many trainable parameters are in

- a 2-D convolutional layer with input (32, 19, 19), kernel size (7, 7), stride 3, 64 output channels?

  The number of trainable parameters $= (32 * 7 * 7 + 1) * 64 = 1569 * 64 = 100416$. Here $32*7*7$ is the kernel size for a 7×7 window with 32 channels, 1 is for the bias, and there are 64 output channels.

- a 2-D convolutional layer with input (512, 25, 25), kernel size (1, 1), stride 1, 128 output channels?

  $(512 + 1) * 128 = 65664$.

- How many multiplications and how many additions are performed in the first case above?

  The number of multiplications is $32 * 7 * 7 = 1569$ each time an inner product is calculated. The output shape for a 7×7 window over a 19×19 image with stride 3 is floor $\left(\frac{19-7}{3} + 1, \frac{19-7}{3} + 1\right) = (5, 5)$ so there are 25 inner product evaluations per output channel. The total number of multiplications is then $1569 * 25 * 64 = 2510400$.

  The number of additions = 1 each time an inner product is calculated. The total number of additions is, therefore, $1 * 25 * 64 = 1600$.