Assignment Two

Matthew C. Scicluna
Département d'Informatique et de Recherche Opérationnelle
Université de Montréal
Montréal, QC H3T 1J4
matthew.scicluna@umontreal.ca

March 5, 2018

1 Convolutions

We compute the full valid and same convolution with kernel flipping for the following matrices: $[1, 2, 3, 4] * [1, 0, 2]$

- The valid convolution is: $[1 \cdot 2 + 2 \cdot 0 + 3 \cdot 1,\ 2 \cdot 2 + 3 \cdot 0 + 4 \cdot 1] = [5, 8]$
- Likewise the same convolution is: $[0, 1, 2, 3, 4, 0] * [1, 0, 2] = [2, 5, 8, 6]$
- Finally the full convolution is: $[0, 0, 1, 2, 3, 4, 0, 0] * [1, 0, 2] = [1, 2, 5, 8, 6, 8]$

2 Convolutional Neural Networks

Consider a 3-layer CNN. We are given an input of size $3 \times 256 \times 256$. The first layer contains 64 8×8 kernels using a stride of 2 and no padding. The shape of its output is $64 \times 125 \times 125$ using relationship 6 from [1]:

$$\text{output length} = \left\lfloor \frac{256 + 2 \cdot 0 - 8}{2} \right\rfloor + 1 = 125$$

The second layer subsamples this using 5×5 non-overlapping max pooling. It is easy to see that the size of its output is $64 \times 25 \times 25$, since $\frac{125}{5} = 25$. The final layer convolves 128 4×4 kernels with a stride of 1 and a zero-padding of size 1 on each border. Using the formula we have that $\left\lfloor \frac{25+2 \cdot 1-4}{1} \right\rfloor + 1 = 24$, and so the output of the last layer has shape $128 \times 24 \times 24$.

(a) The output of the last layer will be of size: $128 \times 24 \times 24 = 73725$

(b) Ignoring biases, we would need $64 \times 25 \times 25 \times 128 = 5120000$ weights

3 Kernel configuration for CNNs

We are given an input shape of $3 \times 64 \times 64$ and the output shape is $64 \times 32 \times 32$ for a convolutional layer.

(a) Assuming no dilation and kernel size of 8×8, we can solve for the stride length s and the padding p by solving the relationship with the given kernel size (setting $s = 2$ for simplicity):

$$\begin{aligned} \left\lfloor \frac{64 + 2 \cdot p - 8}{2} \right\rfloor + 1 &= 32 \\ 32 + p - 4 + 1 &= 32 \\ p &= 3 \end{aligned}$$

Setting 3 padding with 2 stride satisfies the convolution dimensions. Assuming dilatation $d = 6$ and stride of $s = 2$, we can use relationship 15 from [1] to get:

$$\begin{aligned} \left\lfloor \frac{64 + 2 \cdot p - k - (k-1)(6-1)}{2} \right\rfloor + 1 &= 32 \\ \left\lfloor \frac{69 + 2 \cdot p - 6 \cdot k}{2} \right\rfloor &= 31 \end{aligned}$$

This is satisfied when $69 + 2 \cdot p - 6 \cdot k = 63$. We simplify further to get $2p - 6k + 6 = 0$, for which one possible solution is: $p = 3$, $k = 2$. Therefore, setting padding to be 3 and kernel size 2×2 satisfies the convolution dimensions.

(b) Given an input shape of $64 \times 32 \times 32$ and the output shape is $64 \times 8 \times 8$ a configuration assuming no overlapping of pooling windows or padding would have kernel size 4 and stride 1. This is easily seen since $\frac{32}{8} = 4$.

(c) Without any padding and given input shape $64 \times 32 \times 32$ and kernel of size 8×8 and stride 4 we can use the relation to get:

$$\text{output length} = \left\lfloor \frac{32 + 2 \cdot 0 - 8}{4} \right\rfloor + 1 = 7$$

And so the output size would be 7×7.

References

[1] V. Dumoulin and F. Visin, “A guide to convolution arithmetic for deep learning.,” *CoRR*, vol. abs/1603.07285, 2016.