CS-89.31: Deep Learning Generalization and Robustness

Amittai Siavava

04/27/2023

Problem 1.

Let us denote with:

- $N \times N \times V$ the size of the input volume;
- $n \times n \times V$ the size of the filters;
- Z the number of zeros padded at the top/bottom/left/right of the image;
- S the stride;
- V' the number of filters.

Prove the output volume will have size $M \times M \times V'$, where $M = \frac{N-n+2Z}{S} + 1$.

[Hint] Drawing the convolution operation with Z and S on it can help you get some intuition of proving the formula.

1. Since we pad the input volume with Z zeros on each side, the resulting volume (before convolution) is $N + 2Z$. The resulting volume retains the same depth V.
2. Next, we apply filters of size $n \times n \times V$. Since filters cannot be partially applied(i.e. part of the filter being in the image and part of it being outside), we can only apply the filter from position 0 to position $(N+2Z)-n$ (when the right edge of the filter will be at position $N + 2Z$).
3. A stride S subdivides this interval into $\frac{(N+2Z)-n}{S}$ intervals. However, since both the starting position and the ending position are applied to, we get an extra position for the filter, making the number of intervals $\frac{(N+2Z)-n}{S} + 1$.
4. Therefore, each applied filter will produce a volume of size $M \times M$, where $M = \frac{(N+2Z)-n}{S} + 1$. Since we have V' filters, the resulting volume will be of size $M \times M \times V'$.

See next page for drawing.

For simplicity, I fixed the volume to $9 \times 9 \times 1$, and also fixed the filter to $3 \times 3 \times 1$. In direct contrast to the formula above, $(N + 2Z) = 9$ and $n = 3$.

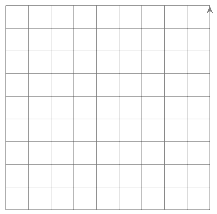

FIGURE 1.
Input-volume
(9×9)

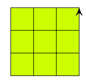

FIGURE 2. Filter

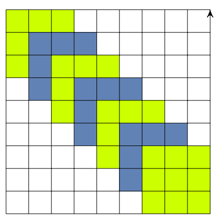

FIGURE 3.
Stride = 1
$M = 7 = \frac{9-3}{1} + 1$

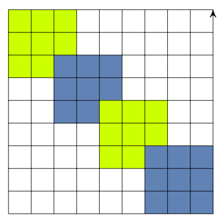

FIGURE 4.
Stride = 2
$M = 4 = \frac{9-3}{2} + 1$

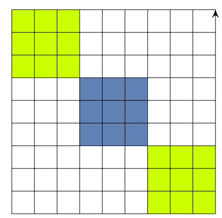

FIGURE 5.
Stride = 3
$M = 3 = \frac{9-3}{3} + 1$