Intro to Deep Learning. HW3 Yifan Li, yl506

Q1. a) Denote filtersize = $k \times k$, stride = s, current receptive field size = $r \times r$, distance (stride) between 2 adjacent features = j.

- We have the receptive field size of the output feature map is equal to the area covered by k input features $(k-1) \times j_{in}$, plus the extra area covered by the receptive field of the input feature on the border.
- Also, the stride of output receptive field is equal to the stride in the input map times the number of input features that got "jumped over" when applying the convolution.
- Thus, we have

$$\begin{cases} j_{out} = j_{in} \times s \\ r_{out} = r_{in} + (k-1) j_{in} \end{cases}.$$

- For the input image (the actual input to the network), $r = 1$, $j = 1$

b) Pooling operation acts like a convolution given a different stride s' and filter size k'. The above expression still holds.

$$j'_{out} = j_{out} \times s', \quad r'_{out} = r_{out} + (k'-1) j_{out}$$

c) VGG16 architecture: Input → 2 conv3-64 → maxpool1 → 2 conv-128 → maxpool2 → 3 conv-256 → maxpool3 → 3 conv-512 → maxpool4 → 3 conv-512 → maxpool5 → FC-4096 → FC-4096 → FC-1000 → softmax. Also, $k=3, s=1, k'=2, s'=2$

Applying the derived expression we have:

$j_0 = 1, r_0 = 1$ (input) → $j_1 = j_0 s = 1, r_1 = r_0 + (k-1) j_0 = 1+2 = 3$; $j_2 = j_1 s = 1, r_2 = r_1 + (k-1) j_1 = 3+2 = 5$ (conv3-64)

→ $j_3 = j_2 s' = 2, r_3 = r_2 + (k'-1) j_2 = 5+1 = 6$ (maxpool1) → $j_4 = j_3 s = 2, r_4 = r_3 + (k-1) j_3 = 10$; $j_5 = 2, r_5 = 14$; (conv-128) → $j_6 = 4, r_6 = 16$ (maxpool2) → $j_7 = 4, r_7 = 24$; $j_8 = 4, r_8 = 32$; $j_9 = 4, r_9 = 40$; (conv3-256) → $j_{10} = 8, r_{10} = 44$ (maxpool3) → $j_{11} = 8, r_{11} = 60$; $j_{12} = 8, r_{12} = 76$; $j_{13} = 8, r_{13} = 92$ (conv3-512) → $j_{14} = 16, r_{14} = 100$ (maxpool4) → $j_{15} = 16, r_{15} = 132$; $j_{16} = 16, r_{16} = 164$; $j_{17} = 16, r_{17} = 196$ (conv3-512) → $j_{18} = 32, r_{18} = 212$ (maxpool5) → FC.

So the receptive field of VGG16 right before the first fully connected layer has a size of 212×212, and stride 32.