Assignment 3 TDT 4265 Fredrik Almås

1. a) First i add zero padding on I to handle boundary conditions.

0	0	0	0	0	0	0
0	1	0	2	3	1	0
0	3	2	0	7	0	0
0	0	6	1	1	4	0
0	0	0	0	0	0	0

-1	0	1
-2	0	2
-1	0	1

Output:

2	-1	11	-2	-13
10	-4	8	2	-18
14	-1	-5	6	-9

b) iii) Max pooling layers reduce the translational variance.

c) From appendix:

$$W_2 = \frac{W_1 - F_W + 2P_W}{S_W} + 1$$

$$H_2 = \frac{(H_1 - F_H + 2P_H)}{S_H} + 1$$

We want $W_2 = W_1$ and $H_2 = H_1$

$S_W = S_H = 1$

$F_W = F_H = 5$

$$W = \frac{W - 5 + 2P_W}{1} + 1$$

$$2P_W = 5 - 1 + W - W$$

$$P_W = \frac{4}{2} = 2$$

$$H = \frac{H - 5 + 2P_H}{1} + 1$$

$$P_H = 2$$

You should use two layers of padding on each side.

d) $$W_2 = \frac{W_1 - F_W + 2P_W}{S_W} + 1$$

$W_2 = 504$

$W_1 = 512$

$P_W = 0$

$S_W = 1$

$504 = 512 - F_W + 1$

$F_W = 512 - 504 + 1 = 9$

The kernel is square:

$F_H = F_W = 9$

The dimentions of the kernels are 9x9

e) The input to the subsampling is feature maps of size 504×504

The output dimensions from subsampling follow the formulas for output size in convolutional layers and we use no padding when subsampling.

$$W_2 = \frac{504 - 2 + 2 \cdot 0}{2} + 1$$

$$W_2 = \frac{502}{2} + 1$$

$$W_2 = \underline{252}$$

$$H_2 = \frac{504 - 2 + 2 \cdot 0}{2} + 1 = \underline{252}$$

The spatial dimensions of the pooled feature maps are 252×252

f) The input features are of size 252×252

$$W_2 = \frac{252 - 3 + 2 \cdot 0}{1} + 1$$

$$W_2 = 252 - 3 + 1 = \underline{250}$$

$$H_2 = 252 - 3 + 1 = \underline{250}$$

The size of the feature maps in the second layer is 250×250.

g) In layer 1 we have

$F_W = F_H = 5$

$C_1 = 3$ as it is an RGB image

$C_2 = 32$

parameters $= F_H \cdot F_W \cdot C_1 \cdot C_2 + C_2$

$= 5 \cdot 5 \cdot 3 \cdot 32 + 32 = \underline{2432}$

Max pooling uses no parameters

In layer 2 we have

$F_W = F_H = 5$

$C_1 = 32$

$C_2 = 64$

parameters $= 5 \cdot 5 \cdot 32 \cdot 64 + 64 = \underline{51264}$

In layer 3 we have

$F_W = F_H = 5$

$C_1 = 64$

$C_2 = 128$

parameters $= 5 \cdot 5 \cdot 64 \cdot 128 + 128 = \underline{204928}$

Flatten uses no parameters and outputs a vector with size (Height)·(Width)·(N feature maps)

Height and width are unchanged by the convolutional layers as they use the padding found in 1c) with the same size kernels and stride.

Max pooling with a stride of 2 and a kernel size of 2x2 halves the size of

the input, as seen in 1 e).

$H_1 = W_1 = 32$

$H_2 = W_2 = \frac{H_1}{2} = 16$

$H_3 = W_3 = \frac{H_2}{2} = 8$

$H_4 = W_4 = \frac{H_3}{2} = 4$

The vector out of the flatten layer into the fully connected layer has size

$4 \cdot 4 \cdot 128 = 2048$

In layer 4 we have

parameters $= 2048 \cdot 64 + 64 = 131136$

In layer 5 we have

parameters $= 64 \cdot 10 + 10 = 650$

The total parameters for the entire network is

total $= 2432 + 51264 + 204928 + 131136 + 650$

total $= \underline{390410}$

There are 390410 parameters in the network.