TDT4195 Image processing assignment 2 report

Nils Reed

November 2021

Task 1.a)

We have: $S_W = S_H = 1$, $F_W = F_H = 5$

Want $W_1 = W_2$, $H_2 = H_2$

Using eq.s (1) & (2):

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

Inserting known values:

$$W_2 = W_1 - 5 + 2P_W + 1$$

$$= W_1 + 2P_W - 4$$

$$\Rightarrow W_2 - W_1 = 2P_W - 4$$

$$W_2 = W_1 \Leftrightarrow W_2 - W_1 = 0$$

Thus:

$$2P_W - 4 = 0$$

$$2P_W = 4$$

$$\underline{P_W = 2}$$

Similarly, for H:

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

$$H_2 = H_1 - 5 + 2P_H + 1$$

$$H_2 - H_1 = 2P_H - 4$$

$$H_2 - H_1 = 0 \Rightarrow 2P_H - 4 = 0$$

$$2P_H = 4$$

$$P_H = 2$$

Thus:

Padding of 2 should be used both in height and width direction

Table 1.b)

Know: $W_1 = H_1 = 512$, $W_2 = H_2 = 504$

$S_W = S_H = 1$, $P_W = P_H = 0$

Thus, from (1) & (2):

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

$$504 = 512 - F_W + 1$$

$$\Rightarrow F_W = 512 - 504 + 1 = 9$$

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

$$504 = 504 - F_H + 1$$

$$\Rightarrow F_H = 9$$

Thus:

The dimensions of the kernels are 9×9

Task 1.c)

For subsampling we have:

$$W_2 = \frac{W_1 - F}{S} + 1 \quad , \quad H_2 = \frac{H_1 - F}{S} + 1$$

$$W_2 = \frac{504 - 2}{2} + 1 = \frac{502}{2} + 1 = 251 + 1 = \underline{252}$$

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

Thus

The dimensions of the pooled feature maps

is 252×252

Task 1.d)

From (1) & (2)

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

$$= 252 - 3 + 1 = 250$$

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

$$= 252 - 3 + 1 = 250$$

Thus:

The dimensions of the feature maps in the 2nd layer are 250×250

Task 1. e)

Need to find no. of inputs to the FCNN. Track dimensions through the layers:

Layer 1:

$$W_2 = \frac{W_1 - F_W + 2P_W}{S_W} + 1 = \frac{32 - 5 + 2 \cdot 2}{1} + 1 = 32$$

$$H_2 = \frac{H_1 - F_H + 2P_H}{S_H} + 1 = \frac{32 - 5 + 2 \cdot 2}{1} + 1 = 32$$

Pooling:

$$W_2 = \frac{W_1 - F}{S} + 1 = \frac{32 - 2}{2} + 1 = 15 + 1 = 16$$

$$H_2 = \frac{H_1 - F}{S} + 1 = \frac{32 - 2}{2} + 1 = 16$$

See by now that since everything is quadratic, one only needs to calculate W or H, since they're the same number.

Layer 2

$$W_2 = H_2 = \frac{W_1 - F + 2P}{S} + 1 = \frac{16 - 3 + 2 \cdot 1}{1} + 1 = 15 + 1 = 16$$

Pooling:

$$W_2 = H_2 = \frac{W_1 - F}{S} + 1 = \frac{16 - 2}{2} + 1 = 7 + 1 = 8$$

Layer 3:

$$W_2 = H_2 = \frac{W_1 - F + 2P}{S} + 1 = \frac{8 - 3 + 2 \cdot 1}{1} + 1 = 7 + 1 = 8$$

Pooling:

$$W_2 = H_2 = \frac{W_1 - F}{S} + 1 = \frac{8-2}{2} + 1 = 3 + 1 = \underline{4}$$

Thus, the number of inputs to the 1st FCNN layer is

$$4 \cdot 4 \cdot 128 = 2048$$

Layer	Layer Type	Number of Hidden Units/Filters	Activation
1	Conv2D (kernel size=5, stride=1, padding=2)	32	ReLU
1	MaxPool2D (kernel size=2, stride=2)	–	–
2	Conv2D (kernel size=3, stride=1, padding=1)	64	ReLU
2	MaxPool2D (kernel size=2, stride=2)	–	–
3	Conv2D (kernel size=3, stride=1, padding=1)	128	ReLU
3	MaxPool2D (kernel size=2, stride=2)	–	–
	Flatten	–	–
4	Fully-Connected	64	ReLU
5	Fully-Connected	10	Softmax

Assuming the image is RGB (thus $32 \times 32 \times 3$)

Parameters for layer 1:

$$P_1 = F_W \cdot F_H \cdot C_1 \cdot C_2 + C_2 = 5 \cdot 5 \cdot 3 \cdot 32 + 32 = 2432$$

Layer 2:

$$P_2 = F_W \cdot F_H \cdot C_1 \cdot C_2 + C_2 = 3 \cdot 3 \cdot 32 \cdot 64 + 64 = 18496$$

Layer 3:

$$P_3 = F_W \cdot F_H \cdot C_1 \cdot C_2 + C_2 = 3 \cdot 3 \cdot 64 \cdot 128 + 128 = 73856$$

Layer 4:

$$P_4 = C_1 \cdot C_2 + C_2 = 2048 \cdot 64 + 64 = 131136$$

Layer 5:

$$P_5 = C_1 \cdot C_2 + C_2 = 64 \cdot 10 + 10 = 650$$

Thus, the total no. of parameters is:

$$D = \sum_{n=1}^{5} P_n = 2432 + 18496 + 73856 + 131136 + 650 = \underline{226570}$$

Task 2.a

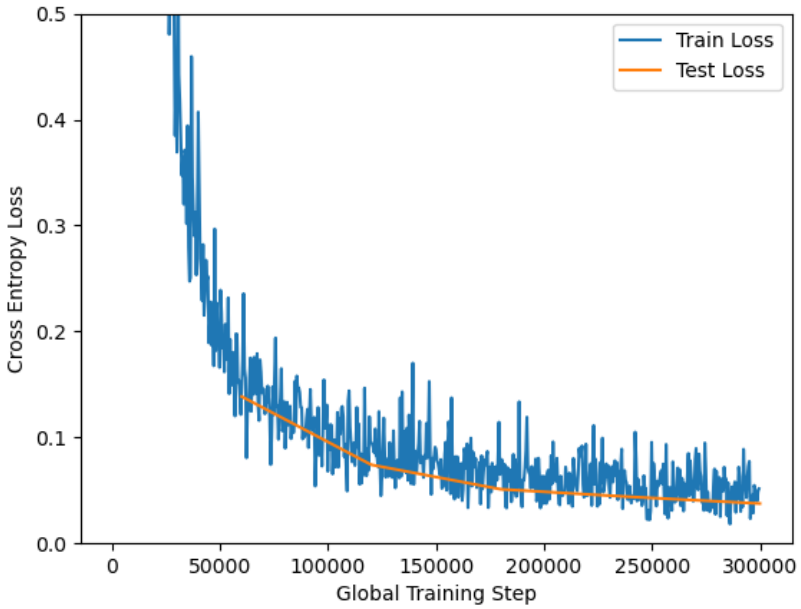

Figure 1: Training and validation loss for 2a

I would say there is no indication of overfitting. This is because the training and validation loss generally decrease at the same rate, and if there was overfitting, their derivatives would have diverged more from each other.

Task 2.b

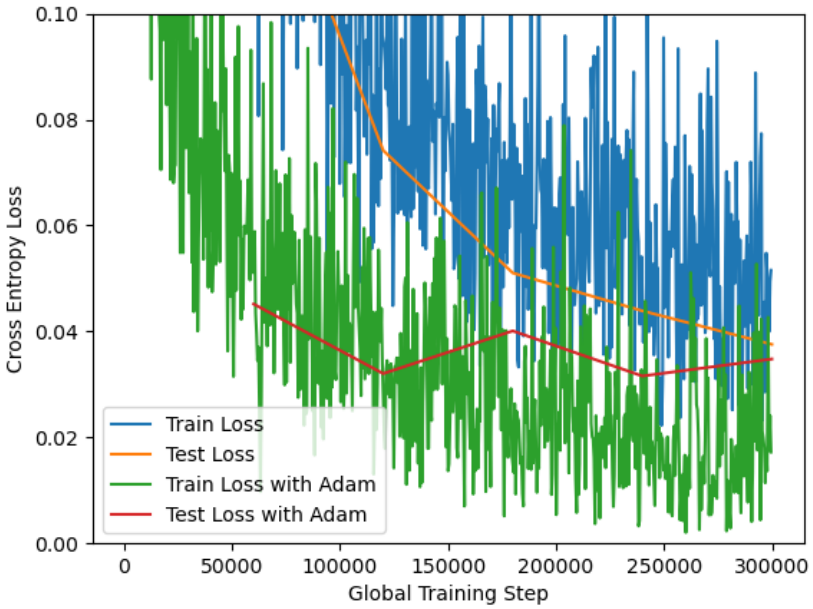

Figure 2: Training and validation loss for SGD and Adam in the same plot

Task 2.c

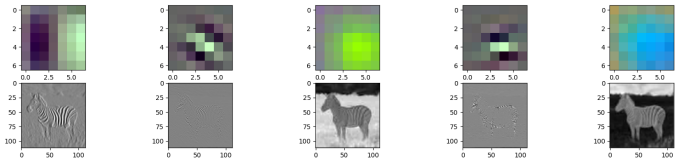

Figure 3: Filters and activations for 2c

Task 2.d

I will address the filters in order from left to right.

The first filter seems to be a filter which highlights differences in colour, specifically in the horizontal direction. This can be seen from the filter weights, since

it's clearly split down the middle.

The second second filter seems to highlight diagonal lines (at a -45 degree angle from horizontal). This can be seen from, the lines on the neck of the horse, and the fact that the filter is divided into three diagonal segments

The third filter is very green, so it will highlight pixels with a high G value in RGB. This explains all the grass (yellow = red + green) being highlighted, and the white stripes still being visible on the zebra.

The fourth second filter seems to highlight horizontal and near horizontal lines. This can be seen from, the lines on the neck of the horse, and the fact that the filter is divided into three diagonal segments

The third filter is very blue, so it will highlight pixels with a high B value in RGB. This explains the sky and the skies being highlighted, and the white stripes still being visible on the zebra.

Task 3.a)

Mapping:	Reasoning:
1a : 2e	1a has the highest freq. in the vertical direction & 2e has dots furthest from the origin in the vertical direction (which signifies higher freq.)
1b : 2c	Middle freq. in vert. dir., dots not furthest out from nor in towards the origin in spectrum.
1c : 2f	Lowest freq. in vert. dir., dots closest to origin in vert dir.
1d : 2b	Lowest freq in horiz. dir., dots closest to origin in horiz. dir.
1e : 2d	Highest freq. in horiz dir, dots furthest from origin in horiz. dir.
1f : 2a	Only ones left + Middle freq. in horiz. dir., dots not furthest from nor closest to origin in horiz. dir.

Task 3.b)

High-pass filters are filters that attenuate the low-frequency components of the signals they're fed

Low-pass filters are filters that attenuate the high-frequency components of the signals they're fed.

Task 3c)

Since filtering in the spatial domain is equal to multiplication in the frequency domain, it's straightforward to see that the black areas in the spectra of the filters are the parts that'll get attenuated

Thus, a is a high-pass filter, because the center is black, meaning that low freq.s are attenuated.

b is a low-pass filter because only the center is white, meaning that only the low freq.s are not attenuated (or, the high freq.s are attenuated).

Task 4.a

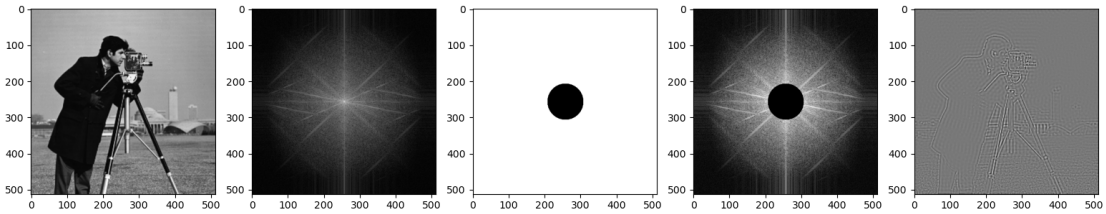

Figure 4: High-pass filter figures 4a

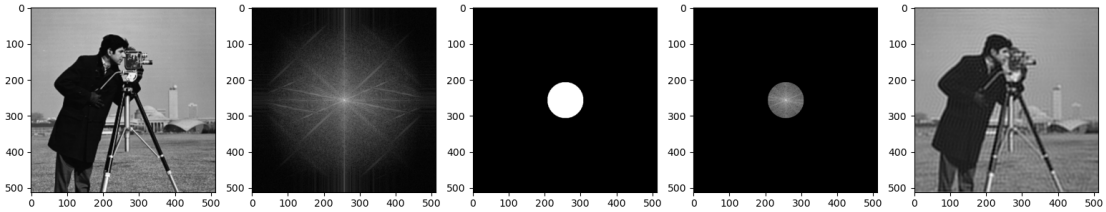

Figure 5: Low-pass filter figures 4a

The ringing effect observed is due to the filters being ideal, which means they are 1 in the area where frequencies pass through and 0 otherwise. Implementing the filters with a bit of smoothing between the two regions should fix this.

4.b

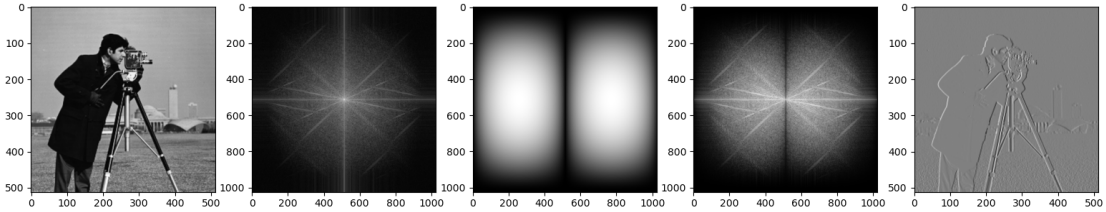

Figure 6: Sobel filter figures 4b

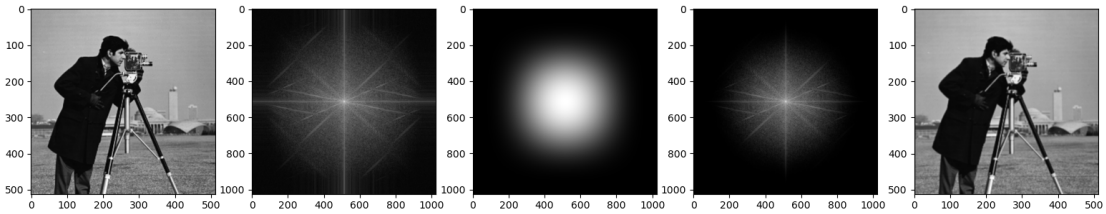

Figure 7: Gaussian filter figures 4b

Task 4.c

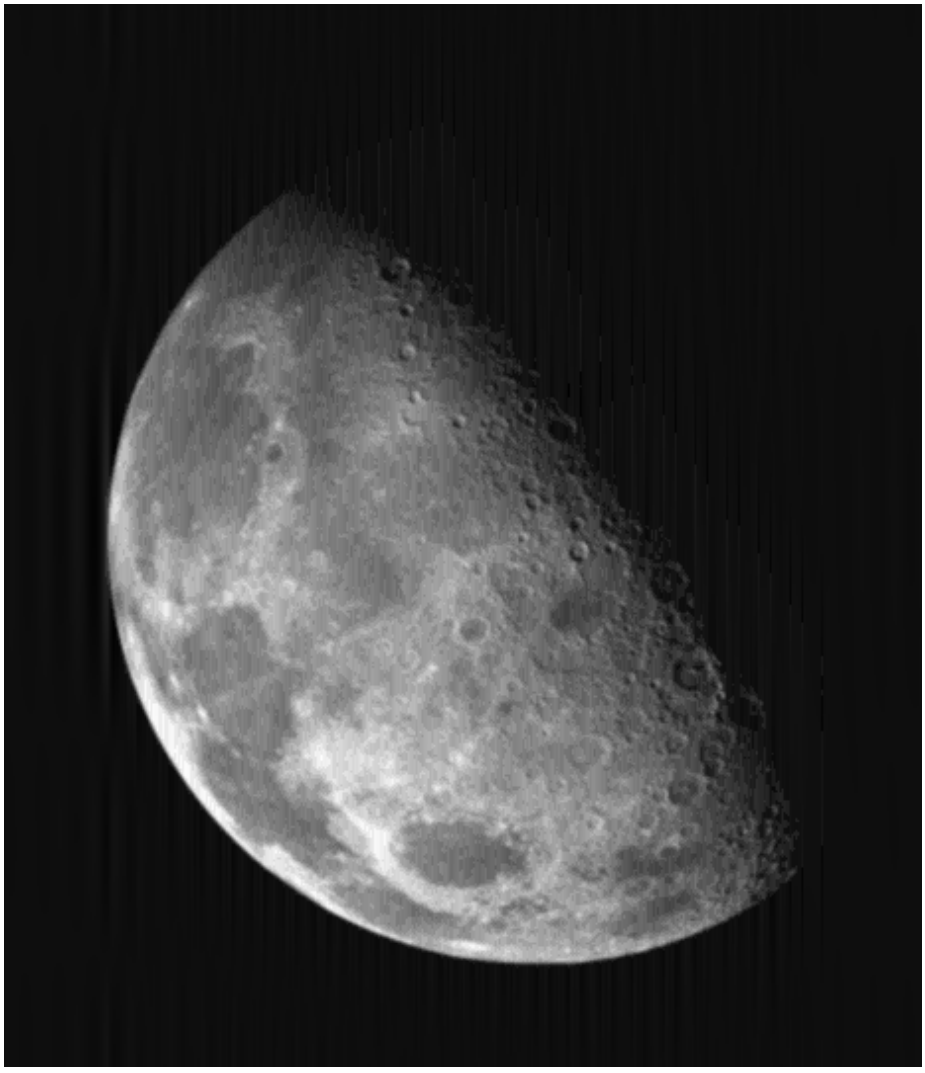

Figure 8: Filtered moon

Task 4.d

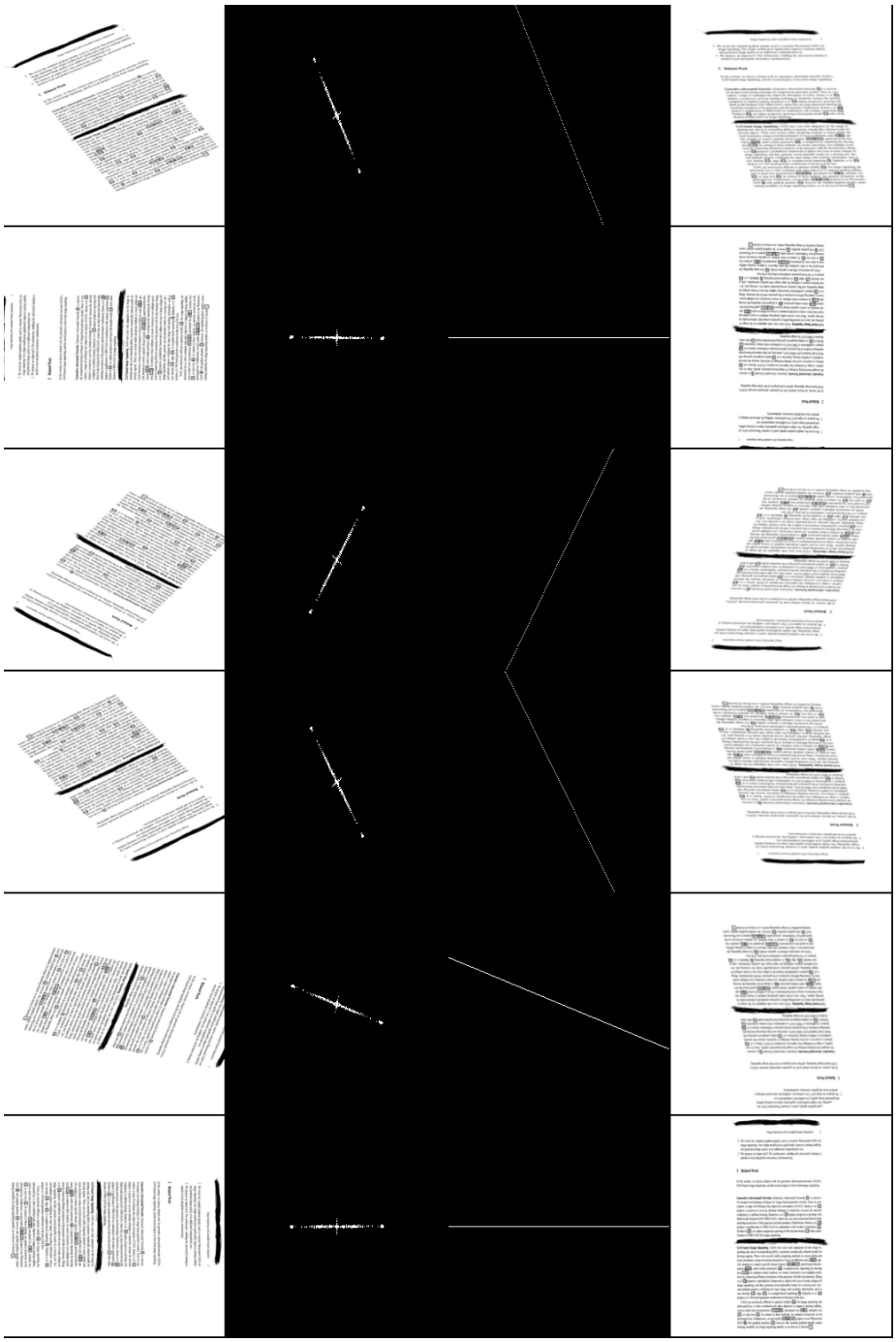

Figure 9: Generated picture from 4d