DATA.ML.300 Computer Vision Exercise 2

Trung Nguyen

January 2024

1

1a. Output of hidden unit (h) and Output unit (y)

$$h = \sigma(1 * W_1 + b_1) = \frac{1}{1 + exp(-(-2 + 2))} = \frac{1}{2} = 0.5$$
$$y = h * W_2 + b_2 = 0.5 * 4 + 0 = 2$$

1b. Loss value $E = 0.5 * (t - y)^2 = 0.5 * (1 - 2)^2 = 0.5$
1c.

$$\begin{aligned}\frac{\partial E}{\partial W_2} &= (0.5 * (t - y)^2)' = (0.5 * (1 - h * W_2 + b_2)^2)' \\ &= (1 - h * W_2 + b_2) * (-h) \\ &= h * (h * W_2 - b_2 - 1) \\ &= 0.5 * (0.5 * 4 - 0 - 1) = 0.5\end{aligned}$$

1d.

$$\begin{aligned}\frac{\partial E}{\partial W_1} &= (0.5 * (t - y)^2)' = (0.5 * (1 - h * W_2 + b_2)^2)' \\ &= (0.5 * (1 - \sigma(W_1 + b_1) * W_2 + b_2)^2)' \\ &= (1 - \sigma(W_1 + b_1) * W_2 + b_2) * (-W_2) * \sigma(W_1 + b_1) * (1 - \sigma(W_1 + b_1)) \\ &= (1 - \sigma(-2 + 2) * 4 + 0) * (-4) * \sigma(-2 + 2) * (1 - \sigma(-2 + 2)) \\ &= (1 - 0.5 * 4) * (-4) * 0.5 * (1 - 0.5)) = 1\end{aligned}$$

2

2a.

$$
\begin{aligned}
dist(Q, A) &= \sqrt{(2-1)^2 + (1-2)^2 + (6-3)^2 + (4-4)^2 + (2-1)^2} \\
&= \sqrt{1+1+9+0+1} = \sqrt{12} \approx 3.464 \\
dist(Q, B) &= \sqrt{(2-3)^2 + (1-1)^2 + (6-4)^2 + (4-1)^2 + (2-5)^2} \\
&= \sqrt{1+0+4+9+9} = \sqrt{23} \approx 4.796
\end{aligned}
$$

$$
\begin{aligned}
||Q|| &= \sqrt{2^2 + 1^2 + 6^2 + 4^2 + 2^2} = \sqrt{4+1+36+16+4} = \sqrt{61} \\
||A|| &= \sqrt{1^2 + 2^2 + 3^2 + 4^2 + 1^2} = \sqrt{2+4+9+16+1} = \sqrt{32} \\
||B|| &= \sqrt{3^2 + 1^2 + 4^2 + 1^2 + 5^2} = \sqrt{9+1+16+1+25} = \sqrt{52} \\
sim(Q, A) &= \frac{2*1 + 1*2 + 6*3 + 4*4 + 2*1}{||Q|| * ||A||} = \frac{2+2+18+16+2}{\sqrt{61} * \sqrt{32}} \approx 0.905 \\
sim(Q, B) &= \frac{2*3 + 1*1 + 6*4 + 4*1 + 2*5}{||Q|| * ||B||} = \frac{6+1+24+4+10}{\sqrt{61} * \sqrt{52}} \approx 0.8
\end{aligned}
$$

2b. Based on the Euclidean distance and Cosine similarity, A has more similarity to Q than B because A has more Cosine similarity score and closer to Q in term of distance

3

3a. What is the size of third dimension and why?
- Size of x: (384, 512, 3)
- The third dimension of x has the size 3 because x is a RGB color image, each channel in 3 channels contains pixel values for that specific color red, green or blue

Figure 1: Loaded image

3b. Is there any difference between each run? Why? What is the size of the output y, and how is this related to x and w?
- Yes, the result of each filter are different each run.
- Size of output y: (380, 508, 10)
- Size of filter w: (5, 5, 3) with number of filter is 10
- Because y_row and y_col are smaller than x_row and x_col
- The applied filter doesn't calculate at the border image x
y_row = x_row - w_row + 1 = 384 - 5 + 1 = 380
y_col = x_col - w_col + 1 = 512 - 5 + 1 = 508
y_channel = x_channel - w_channel + 1 = 1
- However, there are 10 filters so in the end we will have y_channel = 10

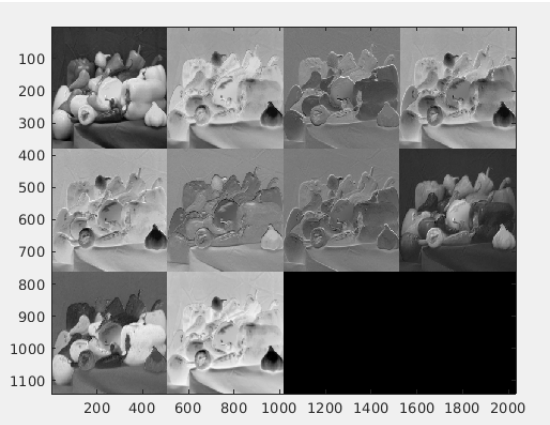

Figure 2: Different convolutional

3c. How does the size of y_pad differ from previous y? Can you explain why?
- Size of y_ds: (24, 32, 10)
- With the stride 16, we have y_ds size calculated as follow:
y_ds_row = (x_row - w_row) // stride + 1 = (384-5)//16+1 = 24
y_ds_col = (x_col - w_col) // stride + 1 = (512-5)//16+1 = 32
Note: // is the integer division operator

- Size of y_pad: (384, 512, 10)
- y_pad now has same number of row and column as x because before applying filter on x, we add 2 zero rows on the top and the bottom, 2 zero columns on the left and the right size of x so that the filtered result will maintain the original size.

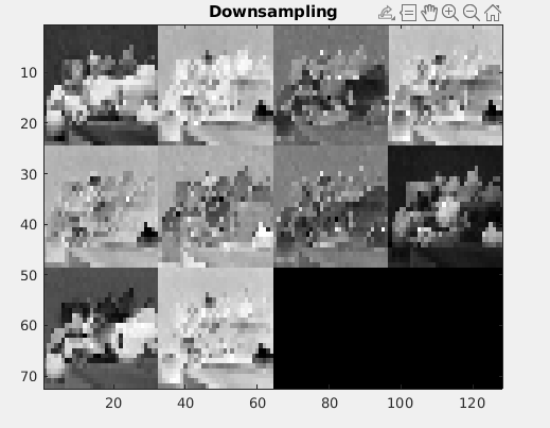

Figure 3: Downsampling result

3d. Why is the repmat function needed here? Take a look at the result. What kind of a features does our filter extract?
- remap function makes 3 copy of w2 and stacks w2 into a (3, 3, 3) filter
- The filter extracts lines, edges in the image

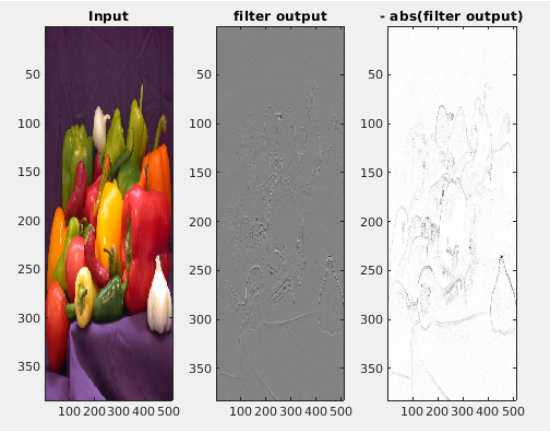

Figure 4: Apply Laplacian filter

3e. Some of the functions in a CNN should be non-linear. Why?
- Convolutional (Conv) filter is a linear filter, so no matter how many Conv layers are stacked on each other, it still acts as a perceptron layer. The non-linear activation function help non-linearize the output of Conv layer so that CNN can estimate complex function in higher order.

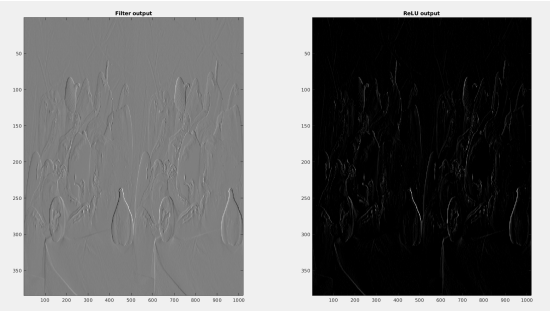

Figure 5: ReLU output

3f. What is the effect of max-pooling? What does the 'Stride' parameter do?
- Size of y: (93, 125, 3) which is smaller than size of x
- Max-pooling layer removes insignificant values in a block which is considered to have small feature representation, and maintains the largest value
- Stride parameter define number of pixels in a row or a column the filter moves each step
Example: current checking pixel is (7, 7), the next one is (7, 11). After finishing filtering on row 7, next pixel will be (11, 7).

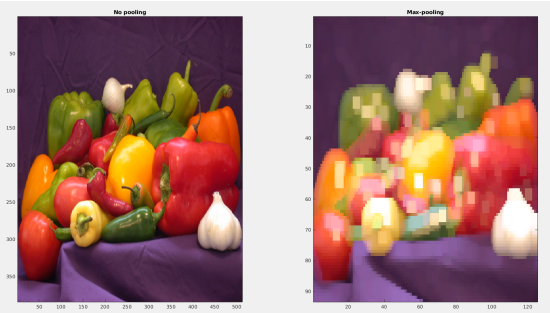

Figure 6: Maxpooling output

3g. How would the histograms set in an ideal case? What is the result here compared to the ideal case?
- In an ideal case, all the positive (green) should be on the right side of the blue lines and negative (red) should be on the left side of the blue lines and both will stays close to the center
- In our case, the green part and red part stay on the correct side but too far from the center

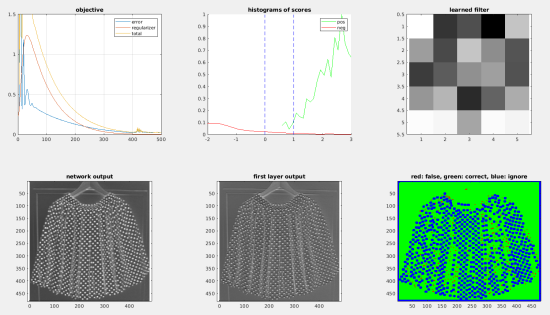

Figure 7: Training result without preprocessing

3h. The learned filter should resemble the discretisation of a well-known differential operator. Which one?
- Laplacian of Gaussian

3i. What is the effect of having too high of a learning rate? Restore the learning rate and set momentum to 0. How does this differ from the previous with the same learning rate? What is the benefit of using momentum?
- Learning rate too high can help the learning faster at first (reduce error quickly), but in the later iteration, because it moves too quick so the learning missed the minimum error
- Without momentum, the learning process become slower (error decrease slower)

and even cannot achieve convergence (minimum error)
- Momentum help learning process converge faster (error decrease faster), and provides better generalization (help avoid overfitting)

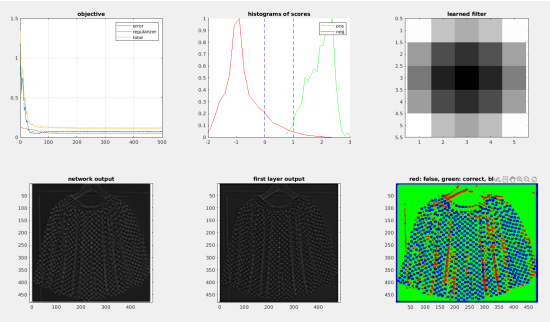

Figure 8: Training result with learning rate = 10

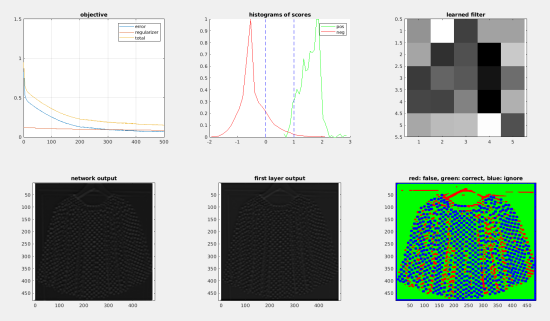

Figure 9: Training result with no momentum

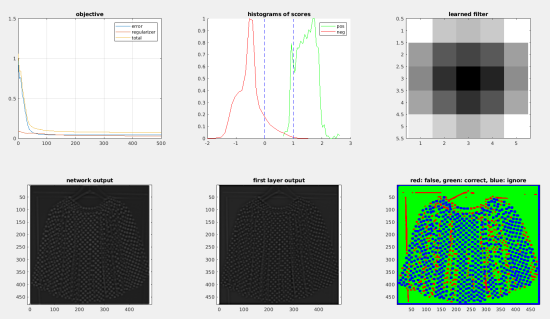

Figure 10: Training result final