cs577 Assignment 4: Solution

Yuanxing Cheng, A20453410, CS577-f22
Department of Mathematics
Illinois Institute of Technology

November 9, 2022

Theoretical questions

1

Let I be a 4×4 RGB image where R channel is all 1-s and G channel is all 2-s. The B channel has a value of 1 in its first row, a value of 2 in its second row, a value of 3 in its third row, and a value of 4 in its forth row. Compute the convolution of this image with a 3 by 3 filter hainv all ones without zero padding.

It returns with 3 channels, and we still call them RGB here. The R channel is 2 by 2 where all values are 9; the G channel is 2 by 2 where all values are 18. The B channel is 2 by 2, and it has a value of 18 in its first row and a value of 27 in its second row. Calculations:

$$
\begin{aligned}
9 &= \sum_{i=1}^{9} 1 \times 1 \\
18 &= \sum_{i=1}^{9} 1 \times 2 \\
18 &= 1+1+1+2+2+2+3+3+3 \\
27 &= 2+2+2+3+3+3+4+4+4
\end{aligned}
$$

2

Last question with zero padding

So we extend each channel by 1 more rows and columns full of 0 on all four sides of the image. We end up with the following matrix.

$$
R = \begin{bmatrix} 4 & 6 & 6 & 4 \\ 6 & 9 & 9 & 6 \\ 6 & 9 & 9 & 6 \\ 4 & 6 & 6 & 4 \end{bmatrix} \quad G = 2R = \begin{bmatrix} 8 & 12 & 12 & 8 \\ 12 & 18 & 18 & 12 \\ 12 & 18 & 18 & 12 \\ 8 & 12 & 12 & 8 \end{bmatrix} \quad B = \begin{bmatrix} 6 & 9 & 9 & 6 \\ 12 & 18 & 18 & 12 \\ 18 & 27 & 27 & 18 \\ 14 & 21 & 21 & 14 \end{bmatrix}
$$

3

Last question, but with dialated convolution with a dialation rate of 2

This time we extend the image with 3 more zero rows and columns on every sides. The R channel is 4 by 4 where all values are 4; the G channel is 4 by 4 where all values are 8. The B channel is

expressed in the following matrix.

$$\begin{bmatrix} 8 & 8 & 8 & 8 \\ 12 & 12 & 12 & 12 \\ 8 & 8 & 8 & 8 \\ 12 & 12 & 12 & 12 \end{bmatrix}$$

4

Explain the template matching interpretation of convolution.

As doing the convolution, if the filter shares the same value for some propotional at a local patch of the image, the result would be big. We can use this idea to create vertical edge detection filter, or horizontal edge detection filter.

5

Explain how multiple scale analysis can be achieved with a fixed window size (using a pyramid)

The recptive field is getting larger as the level of convolution increases (closer to the top of the pyramid). Even with fixed window size, for example, 2 by 2 filter, the fields are 4, 9, 16, etc.

6

Explain how to compensate for spatial resolution decrease using depth (number of channels) and the purpose for doing so.

After using filter, the resolution is decreased. so we apply multiple filters obtaining deeper channels to compensate for the loss.

7

Given a 128 by 128 by 32 tensor and 16 convolusion filters of size 3 by 3 by 32, what will be the size of the resulting tensor when convolving without zero padding.

Purpose: depth would be 16, same as the number of filters applied. The size would be (128-3+1) by (128-3+1) thus the resulting tensor would have shape $126 \times 126 \times 16$.

8

Repeat the previous question when using a stride of 2.

$128/2 - 1 = 63$ will be the new width and length.

9

Explain how the number of channels can be reduced using a 1 by 1 convolution.

---

no matter how many channels we had in the last layer, after filted by 1 by 1 single convolution, the depth is 1.

10

Explain the interpretation of convolution layers and the difference between early and deeper convolution layers.

The convolution layers aims to extract certain features layer by layer. In the early layers, we expect to see lines, dots, circles, curves, etc, and in the deeper layer, we might see the shape of eyes, ears, those shapes of certain objects.

11

Let I be an image as in question 1. Write the result obtained using max pooling with a 2 by 2 filter with a stride of 2.

As the image was 4 by 4 originally, the pooling with stride of 2 would do convolution on the 4 separate 2 by 2 squares, resulting in a 2 by 2 output.

12

Explain the purpose of pooling.

To reduce the output dimension on width and length, i.e. the spacial dim.

13

Explain the purpose of data augmentation and when it is mose useful.

We augmente data to increase the generalization and prevent overfitting. Useful when we have relative small dataset.

14

Explain the purpose of transfer learning and when it is most useful.

Save time using pretrained model on feature extraction. Useful when we have relative small dataset. And limited time or resource for training.

15

Explain the need for freezing the coefficients of the pre-trained network.

---

so that the coefficients in these layers will not be modified in a possible bad way.

16

Explain how the coefficients of a pre-trained network can be fine-tuned.

After training the full connected layers, unfreeze the top layers of the pre-trained model then retrained the whole model.

17

Explain the purpose of inception blocks.

We first use parallel conv layers as an increase in dimension, then add 1 by 1 conv layers to these paralleled conv layers as dimension decrease.

18

Explain the advantage of residual blocks.

- if zero weights in the conv part, it becomes an identity block instead of destroying the signal.
- the network can learn to zero out blocks to eliminate unnecessary blocks (if coefficients in a regular networks decay to zero then that part of network is shut down)
- gradients are passed directly through skip connections, thus train fast

19

Explain how intermediate activations of convolution layers can be visualized given an input. What is the purpose for doing so.

We can create a new model with that input and output as the layers of the trained model.
Pusepose: see what each layer extracts

20

Explain how the filter weight of the trained convolution layers can be visualized. What is the purpose for doing so.

If using gradient ascent, we find the input that will maximize the respones of the filter (correlated with the filter); if using gradient descent we find the input that will minimize the loss to the filter. In the case of gradient ascent, we first define a function from the model input to the model loss and grads. Then we start from a gray img with some noise then keep adding the grads value of the model at the input. Finally we convert the result in the desired measure to get the plot.
Purpose: obtain the basis filter.

21

Explain how the heatmap of class activation can be visualized for a specific image and class. Explain how pooled gradients can be used to weight channels in this visualization. Explain the purpoes of this visualizaiton.

In temrs of the multiple channels in one conv layer, we combine their results. The simplest way: equal weight sum; or using the gradients of the loss wrt each channel, which is the grad-cam algorithm where the pooled gradients are the weight for each channel. In the end we superimpose activation on input image.
Purpose: find which parts of the image contributed to the classification and find where's the object.