Chapter 14: Theoretical Questions

- (Exercise 1 from the book) What are the advantages of a CNN over a fully connected DNN for image classification?

- (Exercise 2 from the book) Consider a CNN composed of three convolutional layers, each with 3×3 kernels, a stride of 2, and "same" padding. The lowest layer outputs 100 feature maps, the middle one outputs 200, and the top one outputs 400. The input images are RGB images of 200×300 pixels:

  - What is the total number of parameters in the CNN?
  - If we are using 32-bit floats, at least how much RAM will this network require when making a prediction for a single instance?
  - What about when training on a mini-batch of 50 images?

- Consider a one-dimensional sequence $(1, 2, 3, 4, 5, 6, 7, 8)$ and a kernel $(-1, 0, 1)$.

  - What is the output of applying this kernel to the sequence using "valid" padding and a stride of 1?
  - What if we use "same" padding instead?

- Apply max pooling with a kernel size of 2×2 and a stride of 2 to the image

$$\begin{bmatrix} 1 & 2 & 3 & -1 \\ 4 & 5 & 6 & -2 \\ 7 & 8 & 9 & -3 \\ 10 & 0 & 0 & 3 \end{bmatrix}$$

- Suppose the output of one of the layers in a convolutional neural network consists of the two feature maps (channels) below. What is the result of applying a global average pooling layer to these feature maps?

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 10 & 0 & 3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 2 & 0 & 2 \end{bmatrix}$$

- (Part of exercise 6 from the book) Can you name the main innovations in AlexNet, as compared to LeNet-5? What about the main innovations in ResNet?

- Describe the typical architecture of a convolutional neural network used for image classification.

  - What types of layers are used in the network?
  - How are these layers organized?
  - What happens to the spatial dimensions of the image as it passes through the network? What happens to the number of channels?

- What is a fully convolutional network? How can you convert a dense layer into a convolutional layer?

- What is data augmentation? What are the benefits of using it? Describe some common data augmentation techniques for image data.

- Describe the task of image segmentation. What is the difference between *semantic* segmentation and *instance* segmentation?