Theoretical Questions

1. What are advantages of CNN over Fully connected DNN for image classification

- The input shape of an image can be anything as long as it is greater than the kernel size in a Conv layer, in a DNN the input shape has to be fixed.
- Once a Conv layer learned a certain pattern in an image, it can be used anywhere, whereas a Dense layer has to find the pattern over and over again.

2. Consider a CNN composed of three convolutional layers

- each with 3×3 kernels, a stride of 2, and "same" padding
- lowest layer outputs 100 feature maps
- middle one outputs 200
- top one outputs 400
- input images are RGB images of 200×300 pixels
- What is the total number of parameters in the CNN?
  - (3x3x3 + 1)x100 = 2800
  - (3x3x100 + 1)x200 = 180_200
  - (3x3x200 + 1)x400 = 720_400
  - total parameters = 903_400
- If we are using 32-bit floats, at least how much RAM will this network require when making a prediction for a single instance?
  - input feature maps of 200x300
  - l1: 100 feature maps of 100x150 (because stride 2, padding same)
  - 200 feature maps of 50x75
  - 400 feature maps of 25x38
  - 32-bit floats = 4 bytes
  - L1: 100x100x150x4 = 6_000_000 bytes = 6 MB per image
  - L2: 200x50x75x4 = 3_000_000 bytes = 3 MB per image
  - L3: 400x25x38x4 = 1_520_000 bytes = 1.5 MB per image
  - CNN Parameters: 903_400 x 4 bytes 3_613_600 bytes = 3.6 MB
  - Total = 14.1 MB per image but only have to take into account the 2 consecutive layers -> 6 + 3 + 3.6 = 12.6 MB
- What about when training on a mini-batch of 50 images?
  - 50x10.5 = 525MB per instance
  - 50x4x200x300x3 = 36_000_000 = 36 MB (#inputimg x #bytes per image x img height x img width x # channels) this is for input images

    - 3.6 MB for parameters
    - Total = 525 + 36 + 3.6 = 564.6 MB per mini-batch (this is a bare minimum)

3. 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 sequence using "valid" padding and stride 1?
    - sequence = 1x8
    - kernel = 1x3
    - output shape = ((1+0-1) / 1) + 1 x ((8+0-3) / 1) + 1 = 1x6
    - 2, 2, 2, 2, 2, 2
- What if we use "same" padding instead?
    - padding = (f-1)/2 = (3-1) / 2 = 1
    - output shape = 1 x ((8+2x1-3) / 1) + 1 = 1x8
    - sequence -> (0, 1, 2, 3, 4, 5, 6, 7, 8, 0)
    - 2, 2, 2, 2, 2, 2, 2, -7

4. Apply max pooling with a kernel size of 2x2 and a stride of 2 to the image

```
[1 2 3 -1]
[4 5 6 -2]
[7 8 9 -3]
[10 0 0 3]
```

Result:

```
[5 6]
[10 9]
```

5. 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?

1+2+3+4+5+6+10+0+3/9 = 34/9 1+0+2+0+1+3+2+0+2/9 = 11/9

Ouput = [34/9, 11/9]

6. Can you name the main innovations in AlexNet, as compared to LeNet-5? What about the main innovations in ResNet?

- AlexNet -> Dropout, Data Augmentation, Local Response Normalization
- ResNet -> Skip connections

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

- What types of layers are used in the network?
    - Conv layers, Pooling layers, Dense layers, Flatten
- How are these layers organized?
    - Conv layers, Pool layers, Conv layers, Pool layers, (...), Flatten, Fully connected layers
- What happens to the spatial dimensions of the image as it passesthrough the network? What happens to the number of channels?
    - The spatial dimensions of the images get smaller as we go deeper into the network
    - The number of channels gets bigger as we go deeper into the network

8. What is a fully convolutional network? How can you convert a dense layerinto a convolutional layer?

- A network that doesn't use Dense layers in the output but used Conv layers
- Amount of filters in Conv same as amount of units in dense
- Kernel size with same dimensions as feature maps
- Valid padding

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

- Shifting, rotating, cropping, changing contrast/brightness of images (data)
- More training data, model becomes more tolerant to variations in data
- Zie eerste antw.

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

- Image Segmentation: Dividing an image into meaningful parts or regions.
- Semantic Segmentation: Classifying each pixel into a specific class without distinguishing between instances of the same class.
- Instance Segmentation: Classifying each pixel and distinguishing between individual instances of the same class.