Machine Learning for Image Analysis, SoSe 2024
Exercise Sheet #7: U-Net & Softmax gradient
Due date: June 18, 2024, before 11 am

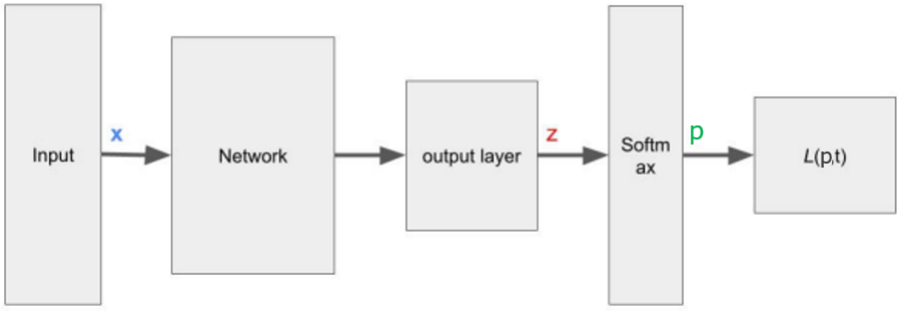
Problem 1 (U-Net - Padding) (10 pt).
You want to apply a U-Net for grayscale images of size 360x360 using **four** down- and upsampling passes. At each level, the U-Net has two consecutive convolutional layers with filter size 3x3, and stride 1. For down- and upsampling we use kernel size 2x2, stride 2 and no padding.
Hint: This problem might be helpful for the current programming exercise.

(a) (2pt) Consider all convolutions to have **same** padding. Assume no additional padding of the input images. What problem does occur during the forward pass?

(b) (4pt) Consider all convolutions to have **same** padding. What would be an admissible padding parameter P for the input image and what would be the output size in the last bottleneck layer? We only care about width/height and ignore channel depth.

(c) (4pt) Consider all convolutions to have **valid** padding. What would be an admissible padding parameter P for the input image and what would be the output size in the last bottleneck layer? We only care about width/height and ignore channel depth.

Problem 2 (Gradient of Softmax Loss) (15 pt).
In one of the previous programming exercises you have already used the simple representation of softmax loss gradient. This week you will have to show that this representation indeed is correct. Consider a Multi-Class classification problem and the following Neural Network with a final softmax layer as shown:



Here, assume that C is the number of different classes. The output layer returns a vector $z = (z_c)_{c=1,\dots,C}$ of output numbers. Let

$$p_c = \frac{e^{z_c}}{\sum_{c'} e^{z_{c'}}}$$

be the softmax function and the loss function be given by

$$L = -\sum_{c}^{C} t_c \ln(p_c),$$

where $t = (t_c)_{c=1,\dots,C}$ is the vector containing the ground truth. Note that t satisfies

$$\sum_{c}^{C} t_c = 1.$$

Show that the gradient of the softmax loss satisfies

$$\frac{\partial L}{\partial z_j}(z) = p_j - t_j$$

Hint: Before you start differentiating, try to formulate the loss as a difference of two terms.