

---

## Computer Vision and Deep Learning

*Final exam: open book, open-notes*

---

**Date** 25.01.2020  
**Name**  
**Time allowed** 2 hours

1. Draw the computational graph for the following expression:

$$f(x, y, z) = e^x + 2*y + z$$

Show how the forward pass and backward pass are computed for the input values:

$$x = 1, y = 3, z = e.$$

2. Compute the result of convolving the grey-scale image  $I$ , with the convolutional kernel  $K$  with a **stride of 2 and no padding**. (Explicitly write the size of the output image and all the values of the resulting feature map).

What kind of features does this filter detect?

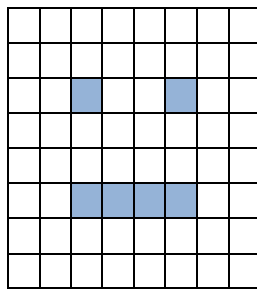
$I$

10	30	20	15	10
5	15	10	40	50
50	50	20	10	0
0	0	0	15	9
5	7	5	5	50
50	30	20	20	10

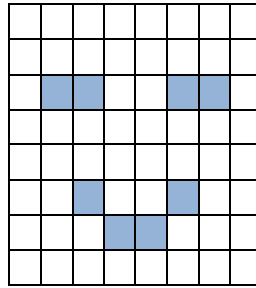
$K$

0	-1	-2
1	0	-1
2	1	0

3. How many **operations** are performed when applying a convolutional layer with kernel size 3x3 and 10 filters, with padding *valid* and stride of 1, on an input volume of size 124x124?
4. You designed an AlexNet-like architecture with 3 dense layers (including the final classification layer). You heard that by replacing the dense layers with convolutional layers can improve the performance of the net. Explain how these dense layers can be replaced with convolutional layers.
5. Given a segmentation task, you obtain the segmentation  $S$  for the ground truth segmentation  $G$ . What is the intersection over union and the average pixel accuracy in this case? You have only two classes (*background* – the with pixels from the drawing and *foreground* – the blue-grey pixels from the drawing).



G



S

6. What is the difference between object localization and object detection?
7. You want to develop a face image manipulation application and for this you first need a CV algorithm that determines the following regions of the face: the hair, the eyebrows, the left and the right eye, the nose, the teeth and the upper and lower lip (just like in the image below).
  - How would you pre-process the input images?
  - What layers would your network use and how would these layers be connected?
  - What would be the output and output size of the network (specify both the spatial size and the depth)?
  - What would be the activation of the last layer in the network?
  - What loss function would you use?
  - What metrics would you use to evaluate your model?
  - Can you run your network on images of different sizes?



8. **Project related question:** What activation did you apply on the last layer from the generator network and why?