Exam

02506
Spring 2024

General information

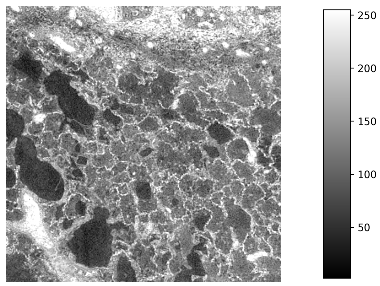
- The exam consists of 20 questions. All questions have equal weight: the correct answer gives 1 point, and the incorrect or missing answer gives 0 points. You only need to submit the answers to the questions. You should not upload any notes or calculations.
- There is one correct answer for each question. Some of the numeric results have been rounded, and might slightly deviate from your result. This should not prevent you from being able to pick the correct answer.
- Each page contains one question. If there are illustrations and images, those refer to the question on that page.
- The notation in the questions is the same as in the course note.
- Data and images are provided for some questions. All filenames typeset in `typewriter font` indicate that you can find files in the data folder. The question text states which data or image should be used.
- To load arrays saved as text files you can use `numpy.loadtxt` in Python and `dlmread` in Matlab.

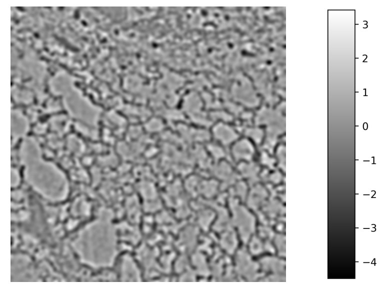
Relevant links

- Course note
  `http://www2.imm.dtu.dk/courses/02506/LECTURE_NOTES_2024.pdf`
- Course page
  `http://www2.imm.dtu.dk/courses/02506`

Question 1

Filtering The image on the left is taken from a CT scan of a lung. The image has been processed by filtering to give the resulting image on the right. The image of the lung is in the file `lung.png`, and the filtering result is in the file `lung_processed.tif`.





How was the image filtered?

(a) Using a 1D first-order Gaussian derivative with a standard deviation of 10 in the horizontal direction and no smoothing in the vertical direction.

(b) Using a 1D Gaussian smoothing kernel in the horizontal direction followed by a 1D first-order Gaussian derivative in the vertical direction. Both filters with a variance of 10.

(c) Using a 1D second-order Gaussian derivative in the one direction followed by 1D Gaussian smoothing kernel in the other direction, repeating the process with directions reversed, and summing the resulting images. This is also known as the Laplacian of Gaussian

$$L = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \, .$$

The kernels have a standard deviation of 4.

(d) Using a 1D first-order Gaussian derivative filter in the horizontal followed by the same in the vertical direction. The filter kernel has a standard deviation of 7.

Question 2

Image smoothness We define image smoothness as

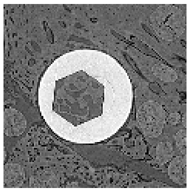
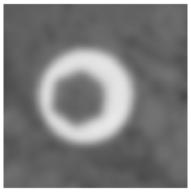
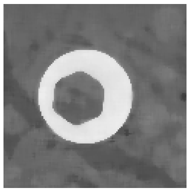
$$S(I) = \sum_{(x,y)\sim(x',y')} d\left(I(x,y), I(x',y')\right)$$

where $(x,y) \sim (x',y')$ indicates neighboring pixel locations (first-order neighborhood), d is a metric

$$d(a,b) = \begin{cases} (a-b)^2 & \text{if } |a-b| < 10 \\ 100 & \text{otherwise} \end{cases},$$

and we assume the pixel intensities a and b to be integers from $\{0, 1, \ldots, 255\}$.

Consider now three images shown below: a slice from a CT scan (available as image `slice.png`), the same slice filtered using a Gaussian filter (image `slice_G.png`), and the slice filtered using a median filter (image `slice_M.png`).

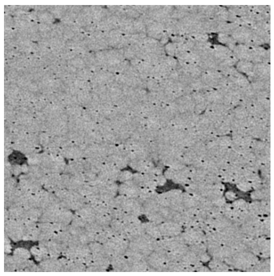




What is the smoothness of these three images?

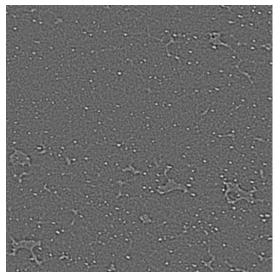
(a) -263473, -1682, and -15608
(b) -16863, 1393, and 1340
(c) 296793, 91229, and 114379
(d) 1201780, 172371, and 158675
(e) 1210511, 177507, and 166288
(f) 1731571, 314862, and 273236
(g) 2115274, 341869, and 291566
(h) 2412291, 349878, and 324963
(i) 24695433, 432640, and 1959738
(j) 47223041, 509213, and 3517167
(k) 47251368, 743104, and 3727908
(l) 48061950, 3410535, and 6443404

Question 3

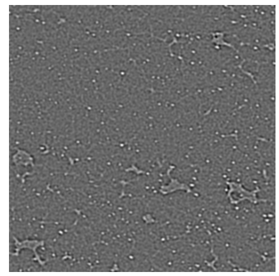
Blob detection We consider blob detection using linear scale space as in the lecture note. We process the image `narwhale.png` shown below in (A). The five images shown in (B) to (F) are Laplacian of the Gaussian at scales $t = \{2, 4, 6, 8, 10\}$ with filenames `narwhale_scale_space_2.tif`, `narwhale_scale_space_4.tif`, ..., `narwhale_scale_space_10.tif`. Scale normalization has not yet been applied.



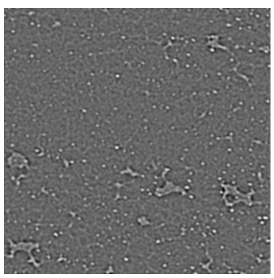
(A)



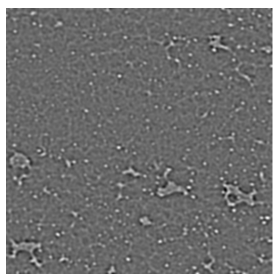
(B)



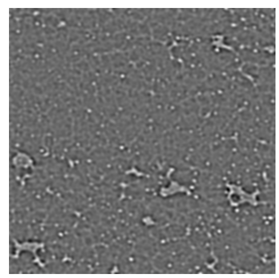
(C)



(D)



(E)



(F)

Consider dark blobs detected at positions (x, y) at a scale $t = 6$ in this scale space.

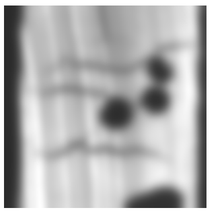
How many of these blobs is located at a pixel where the intensity of the image is less than 100, i.e. $I(x, y) < 100$?

(a) 0
(b) 1
(c) 2
(d) 5
(e) 9
(f) 12
(g) 15
(h) 19
(i) 21
(j) 26
(k) 41
(l) 86
(m) 111
(n) 400

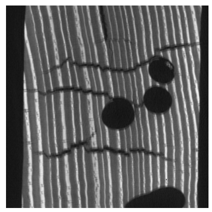
Question 4

Gaussian scale space The Gaussian scale space is computed on the images shown below (intensity normalized). The images filenames are `wood_smooth_A.png` to `wood_smooth_H.png`.

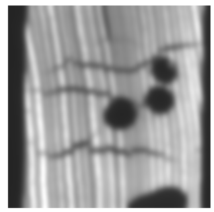
What is the correct order, if the images should be ordered according to increasing scale?



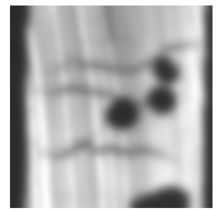
(A)



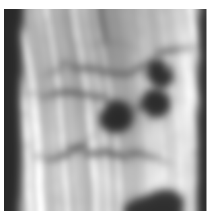
(B)



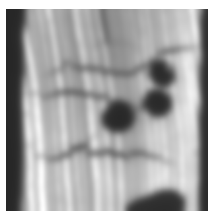
(C)



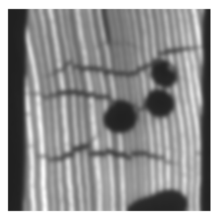
(D)



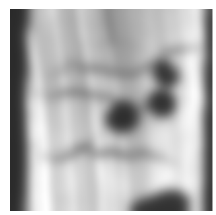
(E)



(F)



(G)



(H)

(a) B, G, C, A, D, E, H, F
(b) B, G, C, F, E, D, A, H
(c) H, A, D, E, F, C, G, B
(d) F, H, E, D, A, C, G, B
(e) B, C, G, E, F, A, D, H
(f) H, D, A, F, E, G, C, B

Question 5

Feature clustering Consider image segmentation by clustering image features as described in the lecture notes. We choose to have 750 clusters for segmenting an image into 3 label classes.

Which of the following arguments for choosing 750 clusters is correct?

(a) 750 clusters are chosen to accurately represent the boundary between segments in the image.

(b) Choosing 750 clusters is a good choice when using mini-batch k-means because it allows well-balanced clusters and makes the algorithm efficient.

(c) Most images contain textures (local appearance) that will make it necessary to choose a number of clusters that is much larger than the number of label classes. Choosing 750 clusters for segmenting an image using 3 labels can be a good choice.

(d) 750 clusters will give redundant cluster probabilities. That does not harm the performance of the segmentation model, but in most cases, it is sufficient to have the same number of clusters as the number of classes.

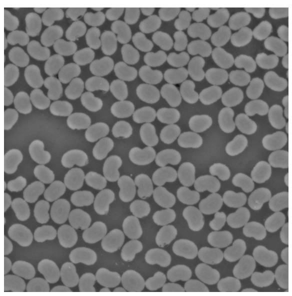
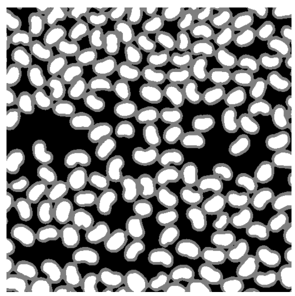
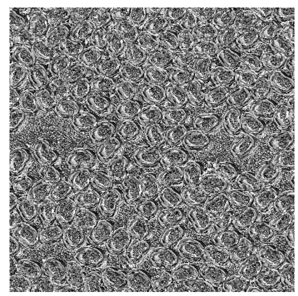
(e) Each pixel must get a label-wise probability. It is only possible to compute a pixel-wise probability if the number of clusters is relatively large and 750 clusters is a safe choice for 3 classes.

Question 6

Feature-based segmentation We want to use feature-based segmentation (as in the lecture notes) to detect bean-shaped fibers. The training image is shown below on the left (file called `bean_fibers.png`), and the associated labels are shown in the middle image (the file `bean_fibers_labels.png`). The labels are:

- Label 1: Black with pixel intensity 0
- Label 2: Grey with pixel intensity 127
- Label 3: White with pixel intensity 255

Image features for all pixels have been extracted and clustered in 256 clusters. The resulting assignment image is shown on the right (available as the image file called `bean_assignment_image.png`). Each pixel of the assignment image has a value between 0 and 255, corresponding to the cluster the pixel belongs to.

What labels will the pixels assigned to the clusters with index values 17, 27, 117, and 212 get?

(a) [1, 1, 2, 1]
(b) [2, 1, 2, 1]
(c) [2, 2, 2, 1]
(d) [2, 3, 1, 2]
(e) [2, 1, 1, 1]
(f) [3, 2, 1, 1]
(g) [3, 2, 2, 2]
(h) [3, 3, 1, 2]

Question 7

Matching blobs Corresponding blobs have been detected in two images and saved in two text files `q_xyd.txt` and `p_xyd.txt`. Each line in the file is one blob with the x-coordinate in the first column, the y-coordinate in the second, and the diameter in the last column.

With $\mathbf{q}$ we denote the coordinates of one set of blobs and with $\mathbf{p}$ the coordinates of another set, both arranged as $2 \times n$ matrix. The transformation from $\mathbf{q}$ to $\mathbf{p}$ is given by

$$\hat{\mathbf{p}} = s\,\mathbf{R}\,\mathbf{q} + \mathbf{t}$$

where s is the scale, $\mathbf{R}$ is the rotation matrix, $\mathbf{t}$ is the translation, and $\hat{\mathbf{p}}$ is the transformed pointset. The transformation parameters are

$$\mathbf{R} = \begin{bmatrix} \cos\theta, & \sin\theta \\ -\sin\theta, & \cos\theta \end{bmatrix}, \quad \theta = 48.3^\circ, \quad s = 0.84, \quad \text{and} \quad \mathbf{t} = [-10.3, 61.78]^T.$$

We should determine how many of blobs match. Blobs match if :

- The Euclidian distance between coordinates from $\mathbf{p}$ and $\hat{\mathbf{p}}$ is less than 2.
- The relative ratio r between the diameters of the blos fulfills $0.8 < r < 1.25$. The relative ratio is
$$r = \frac{d_p}{d_{\hat{p}}},$$
where d_p is the diameter of a blob from $\mathbf{p}$ and $d_{\hat{p}}$ is the diameter of a corresponding blob from $\hat{\mathbf{p}}$.

How many blobs match?

(a) 0
(b) 1
(c) 2
(d) 5
(e) 8
(f) 10
(g) 13
(h) 16
(i) 21
(j) 23
(k) 25
(l) 28

Question 8

SIFT Which statement about SIFT features is correct?

(a) Blob detection using the difference of Gaussians is computed as part of the SIFT feature to ensure invariance to translation with subpixel precision. The gradient magnitude of an area around the SIFT feature gives the feature orientation and scale.

(b) Blob detection using the difference of Gaussians is computed in a linear scale space and is used as part of the SIFT feature to find corresponding points between images.

(c) Blob detection using the difference of Gaussians, which is computed as part of the SIFT feature, ensures that SIFT features become invariant to scale and translation.

(d) Blob detection using the difference of Gaussians is computed at an image location by first computing the orientation of the largest gradients in a 36 bin histogram and at this orientation computing gradient magnitude and orientation histograms in a 4×4 grid resulting in a 128-dimensional descriptor.

Question 9

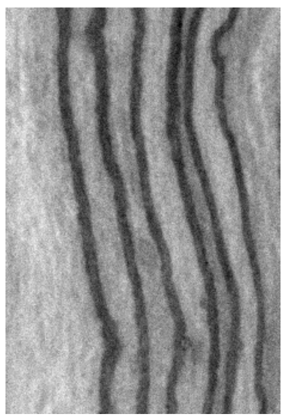
Segmentation energy Consider the image of peripheral nerves, available in the data folder as `nerves.png`. The image values should be normalized to be in the range $[0, 1]$ by dividing the pixel values by 255 after loading.

The segmentation available as `nerves_segmentation.png` (inspect and interpret the pixel values of the image) has been obtained using the MRF approach with

$$\mu_{\text{nerves}} = 0.2, \quad \mu_{\text{background}} = 0.6\,.$$

Since the nerves in the image appear as vertical dark lines, the smoothness parameter in the vertical direction was set to a larger value than the smoothness in the horizontal direction

$$\beta_{\text{vertical}} = 0.1, \quad \beta_{\text{horizontal}} = 0.05\,.$$



What is the total segmentation energy (posterior energy) of this segmentation?

(a) 261.2
(b) 423.4
(c) 436.4
(d) 456.4
(e) 1005.2
(f) 1266.4
(g) 1035.8
(h) 1304.9
(i) 1428.6
(j) 1461.6
(k) 13387.9
(l) 13550.1

Question 10

Parameter range Consider an MRF segmentation as in the lecture notes with

$$\mu_{\text{red}} = 2\,, \quad \mu_{\text{gray}} = 8$$

and first-order neighborhood.

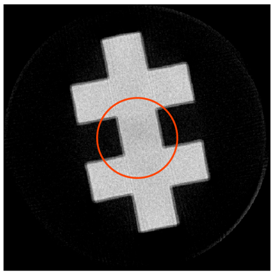
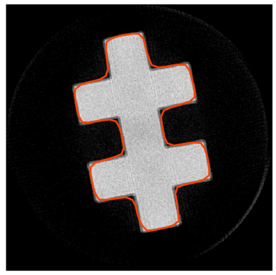
2	2	3	2	7	9	9	8
1	2	2	2	7	8	7	8
2	9	2	4	7	9	2	8
2	3	2	2	7	7	9	7
1	3	3	2	8	8	9	8

For which range of the parameter β do we obtain the segmentation as indicated with the red and the gray color in the illustration?

(a) $(0, 3)$
(b) $(0, 6)$
(c) $(0, 9)$
(d) $(0, 12)$
(e) $(0, \infty)$
(f) $(3, 6)$
(g) $(6, 9)$
(h) $(6, \infty)$
(i) $(9, 12)$
(j) $(12, 15)$
(k) $(12, \infty)$
(l) $(15, \infty)$

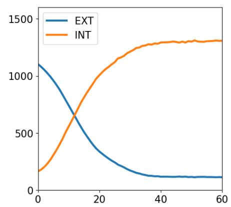
Question 11

Energy evolution A deformable curve has been used to segment an object. The initialization of the curve and the segmentation result after 60 iterations are shown below.

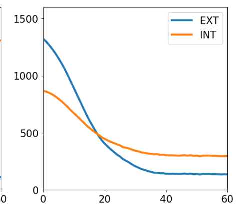
 

We have measured the two contributions to the segmentation energy, the external energy and the internal energy, for every iteration of the curve evolution. The external energy is plotted in blue, and the internal energy in orange.

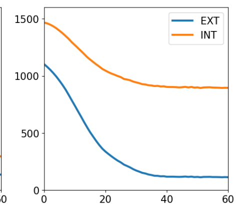
Which of the energy curves below best describes a feasible evolution?



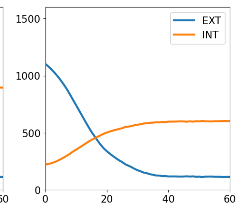
(a)



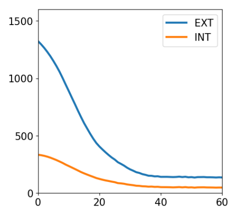
(b)



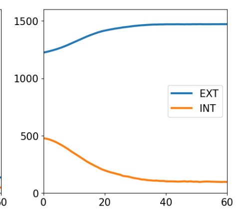
(c)



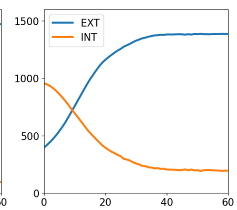
(d)



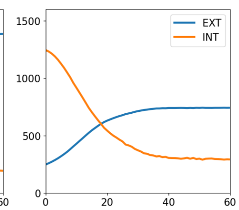
(e)



(f)



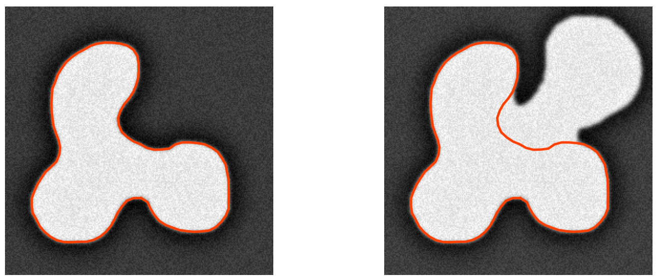
(g)



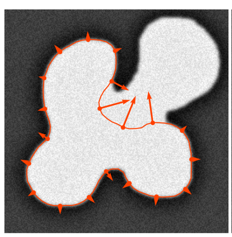
(h)

Question 12

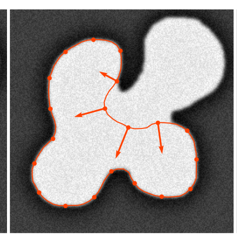
Curve deformation A deformable curve has been used to track an amoeba in a video sequence. The amoeba appears bright on a dark background. In a certain frame, the curve is aligned with the amoeba boundary. Then, between the two frames shown below, the amoeba extends a large 'arm', but does not change the shape elsewhere.



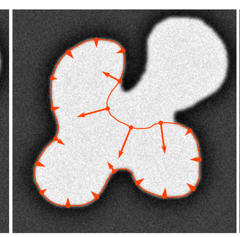
Which of the illustrations best describes the deformation of the curve in the next iteration?



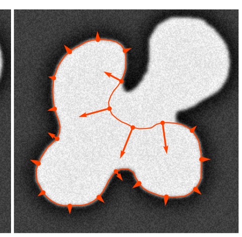
(a)



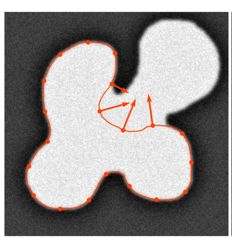
(b)



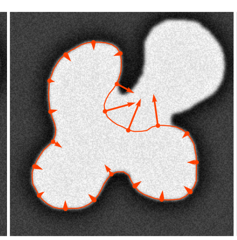
(c)



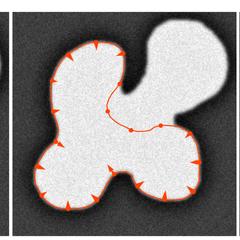
(d)



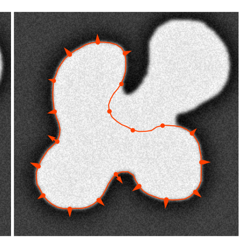
(e)



(f)



(g)



(h)

Question 13

Segmentation cost Consider the color image available in the data folder as `fabric.png` and its segmentation `fabric_segmented.png`.

The segmentation has been obtained using layered surface detection with the in-region cost as in the lecture notes. The red and the green channel of the image were used as c_{below} and c_{above}. Here we use integer pixel values for the color image, while segmentation needs to be appropriately interpreted.

What is the cost of the surface which corresponds to this segmentation?

(a) 2404335
(b) 2562903
(c) 3663433
(d) 4967238
(e) 6067768
(f) 6226336
(g) 6289214
(h) 8693549
(i) 8852117
(j) 9952647

Question 14

Two-line detection The image I, available as `image.txt`, has pixel intensities as given in the illustration.

$z \uparrow$

63	59	59	68	63	69	63	59	67	59
64	26	57	67	22	42	33	22	40	54
25	65	40	33	55	53	51	53	67	31
54	58	69	51	68	60	68	56	70	69
54	15	55	59	70	64	18	68	69	68
65	44	66	33	33	87	59	49	36	33
21	53	32	52	68	45	60	67	66	60
62	68	70	61	57	62	70	53	63	61
25	65	28	33	35	53	61	53	67	31
54	58	69	32	76	60	68	56	26	29

Consider the detection of two terrain-like surfaces with settings for costs, smoothness, and overlap:

$$c_1 = c_2 = I\,,$$

$$\Delta_1 = \Delta_2 = 0\,,$$

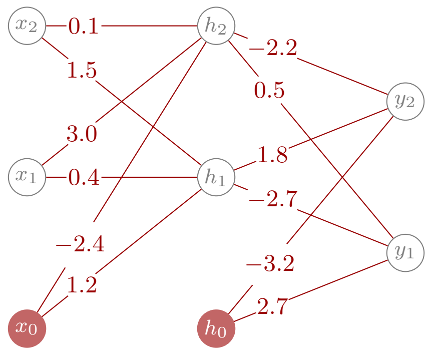
$$\delta_{\text{low}} = \delta_{\text{hight}} = 3\,.$$

What is the cost of the optimal solution?

(a) 417
(b) 427
(c) 451
(d) 878
(e) 900
(f) 912
(g) 924
(h) 932
(i) 936
(j) 956
(k) 967
(l) 975
(m) 978

Question 15

Classification output Consider the classification network



with ReLU activation in the hidden layer and softmax in the last layer. Inputs are given as pairs (x_1, x_2).

What is the classification prediction for the following four points $(-0.1\ ,\ 0.5)$, $(0.1\ ,\ 0.8)$, $(0.2\ ,\ 0)$ and $(0.9\ ,\ -0.1)$?

(a) 1, 1, 1, 1
(b) 1, 1, 1, 2
(c) 1, 1, 2, 1
(d) 1, 1, 2, 2
(e) 1, 2, 2, 1
(f) 2, 1, 1, 2
(g) 2, 1, 2, 1
(h) 2, 1, 2, 2
(i) 2, 2, 1, 1
(j) 2, 2, 2, 2

Question 16

MLP backpropagation Which statement is correct regarding the implementation of the backward pass?

(a) The forward pass is used to compute the predicted values y. The backward pass computes the gradients from the y values and the target values t, and the backward pass requires no additional information from the forward pass.

(b) During the forward pass, for each hidden layer you need to store the activated values h which are used for computing gradients during the backward pass.

(c) During the forward pass, for each hidden layer you have to store both the unactivated neuron values (values z) and the activated values (values h), as both are needed for computing gradients during the backward pass.

(d) During the backward pass, it is important to update the weights of layer l before computing the gradients in layer $l - 1$.

Question 17

Gradients Consider a 5-class classification network with a small modification compared to the lecture notes: in the last layer we use ReLU activation and the squared error loss. That is, we have

$$y_i = \mathrm{ReLU}(\hat{y}_i)\,,$$
$$L = \sum_i (y_i - t_i)^2\,.$$

For a certain input, the values in the last layer before activation are

$$\hat{\mathbf{y}} = [0.1, -0.1, 0.4, 0.7, -0.2]$$

while we know that the target values are

$$\mathbf{t} = [0, 0, 1, 0, 0]\,.$$

What are the values of the partial derivatives

$$\delta_i = \frac{\partial L}{\partial \hat{y}_i}$$

for this specific input?

(a) $[\ -0.2,\quad 0,\quad -0.8,\quad -1.4,\quad 0\ ]$
(b) $[\ -0.2,\quad 0,\quad 1.2,\quad -1.4,\quad 0\ ]$
(c) $[\ -0.2,\quad 0.2,\quad 1.2,\quad -1.4,\quad 0.4\ ]$
(d) $[\ -0.1,\quad 0,\quad 0.6,\quad -0.7,\quad 0\ ]$
(e) $[\ -0.1,\quad 0.1,\quad 0.6,\quad -0.7,\quad 0.2\ ]$
(f) $[\ 0.1,\quad -0.1,\quad -0.6,\quad 0.7,\quad -0.2\ ]$
(g) $[\ 0.1,\quad 0,\quad -0.6,\quad 0.7,\quad 0\ ]$
(h) $[\ 0.2,\quad -0.2,\quad -1.2,\quad 1.4,\quad -0.4\ ]$
(i) $[\ 0.2,\quad 0,\quad -1.2,\quad 1.4,\quad 0\ ]$
(j) $[\ 0.2,\quad 0,\quad 0.8,\quad 1.4,\quad 0\ ]$

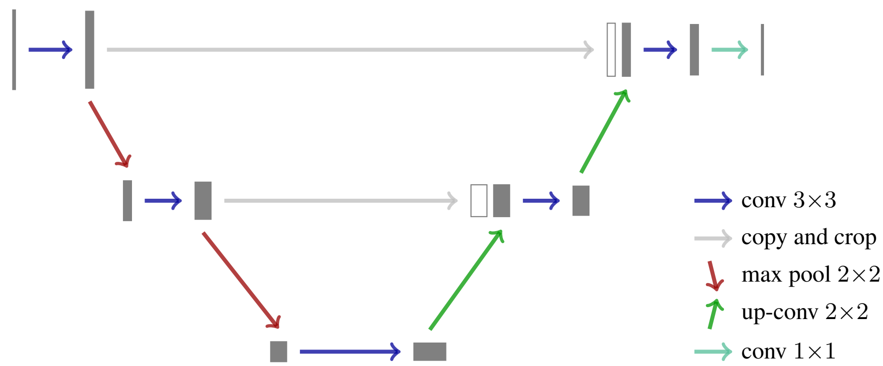
Question 18

Minibatches Which statement about minibatches is correct?

(a) Minibatches are used for ensuring numerical stability when applying adaptive learning rates. If the gradient obtained from a minibatch is small, the algorithm with adaptive learning rates such as the ADAM will ensure that the learning rate increases, and visa versa.
(b) Minibatches are used when training a neural network. By computing an average over the gradients from a minibatch, the obtained gradients will be less noisy and typically allow for a higher learning rate.
(c) Minibatches allow more efficient computation of the forward pass through a neural network, but due to stability reasons, they should only be used for inference.
(d) Minibatches are used for optimizing the structure of data in memory. If a dataset is ordered the same way in memory as it is read or written, then reading and writing with minibatches is much faster. This can be used for faster convergence of neural networks.

Question 19

Network output A neural network similar to U-net is used to segment images. The network architecture is sketched in the illustration.



The elements of the network include convolutions with no spatial padding of the input, max pooling and downscaling over a 2×2 window, and up-convolution (also called transposed convolution). When performed over 2×2 window, up-convolution will double the spatial dimensions (height and width) of the image.

What is the spatial dimension of the output image when the network is given an image of the spatial dimensions 114×102?

(a) 25×22
(b) 42×46
(c) 44×48
(d) 44×50
(e) 48×42
(f) 54×48
(g) 56×50
(h) 92×80
(i) 94×82
(j) 96×84
(k) 112×100
(l) 114×102

Question 20

Learnable parameters Consider a fully convolutional neural network similar to U-net which takes as input RGB images. The following sequence is the first block in the network architecture:

- Convolution using 3×3 kernels with 8 channels, with biases.
- ReLU activation.
- Max pooling and downscaling over a 2×2 window.

Assume that N is the number of learnable parameters in this block.

Which of the following statements is correct?

(a) $N = 80$.
(b) $N = 224$.
(c) $N = 240$.
(d) $N = 256$.
(e) $N = 272$.
(f) N cannot be computed without knowing whether spatial padding of the input has been applied.
(g) N cannot be computed without knowing the size of the input image.
(h) N cannot be computed without knowing both the size of the input image and whether spatial padding of the input has been applied.