Information page for written examinations at Linköping University

Examination date	2022-10-24
Room (2)	T1(44) U4(3)
Time	14-18
Edu. code	TSBB19
Module	TEN1
Edu. code name **Module name**	Machine Learning for Computer Vision (Maskininlärning för datorseende) Written examiation (Skriftlig tentamen)
Department	ISY
Number of questions in the examination	16
Teacher responsible/contact person during the exam time	Per-Erik Forssén
Contact number during the exam time	013-285654
Visit to the examination room approximately	14:45 and 16:30
Name and contact details to the course administrator (name + phone nr + mail)	Carina Lindström, 013-28 44 23
Equipment permitted	Dictionary Swedish-English-Swedish
Other important information	
Number of exams in the bag	

TSBB19 Machine Learning for Computer Vision
Written Exam 2022-10-24

Instructions: This exam consists of **16** questions that require description of terms, phenomena, relations, etc. Each question gives a score [**0,1,2**], for:

- **0:** Wrong/No Answer
- **1:** Almost Complete/Partially Correct Answer
- **2:** Complete and Correct Answer

Each exam question can give a maximum score of **2p**, and a total of **32p** for the whole exam.

In order to pass with grade **3**, at least **15p** are required.
In order to pass with grade **4**, at least **22p** are required.
In order to pass with grade **5**, at least **27p** are required.

Do not write answers on the exam sheet, instead hand in **separate cross-ruled sheets**. You may answer multiple questions on one sheet, but answer questions in order.

Be brief and to the point. Excessive amounts of irrelevant text indicates lack of understanding and may lead to deduction of points.

Write your AID-number and the date on all paper sheets that you hand in.

Good luck!

Per-Erik Forssén and Michael Felsberg

Question 1: Besides repurposing, there is another way to obtain deep features. Which one? Deep features often improve performance compared to hand-coded features, but also come with a risk. Which one?

Question 2: The DCF for multiple channels and multiple samples has been introduced in the lecture using an equation that approximates the general solution. Under which condition regarding the number of samples is the approximation exact, i.e., how many samples can be processed? The exact solution requires to invert a matrix. What is the rank of that matrix for that number of samples?

Question 3: Two common visual recognition tasks are *classification* and *detection*. What is the difference between these? Activity recognition uses both tasks. Explain how and in what order.

Question 4: What is a colour name descriptor, and how is it different from an RGB vector?

Question 5: What is the vanishing gradient problem? How does ResNet avoid this problem?

Question 6: Region proposals are sometimes used in object detection. Explain what these are, and why they are useful.

Question 7: The logistic activation is often used at the output of a binary classification network. How is it generalized to multiple classes and why is ReLU activation preferred at hidden layers?

Question 8: In descriptor matching, a distance ratio is often used instead of a distance. Explain *which* distance ratio, and why it is useful.

Question 9: The scalar product between two normalised vectors is not a distance. It can however be converted to a distance. Which one, and how?

Questions continue on the next page.

Question 10: Accuracy can be defined using true positives(TP), false positives(FP), false negatives(FN) and true negatives(TN). State this expression, and also define accuracy in the multi-class case.

Question 11: A 1D convolution can be implemented as a matrix multiplication. What is the shape of this matrix? The convolution can also be represented with a matrix in the Fourier domain. What is the shape of this matrix?

Question 12: In segmentation, performance is typically measured using the Jaccard index $\mathcal{J}$ and the F-score $\mathcal{F}$ *of the contour pixels*. Sketch examples for the respective cases: a) $\mathcal{J}$ high and $\mathcal{F}$ low, b) $\mathcal{J}$ low and $\mathcal{F}$ high.

Question 13: Training an ensemble of networks, i.e., multiple instances with different initializations, and implementing a majority voting, improves classification performance in general. What is the major drawback of this approach? How can this approach be approximated, avoiding the drawback of an ensemble?

Question 14: Pooling layers in a CNN are either specified with a sliding window size and a stride (e.g. 3×3, stride $= 2$) or an output size (e.g. $H \times W$). Which variant is used in Spatial Pyramid Pooling? What is a stride?

Question 15: An author trained 100 instances of a network with different initializations, and selected one of these based on performance on the validation data. The paper reports the performance of this network on the test data. Why is this correct methodology? Why is it not OK to report the mean performance of the five best test results?

Question 16: Tracking can make use of features from different layers. Are shallow or deep layers better suited for accurate localization? What other option for improving localization accuracy has been discussed in the lectures?