This question paper consists of 6 printed pages each of which is identified by the Code COMP5870M01

This is an open book examination. Any written or printed material is permitted.

School of Computing

May/June 2018

COMP5870M01

Image Analysis

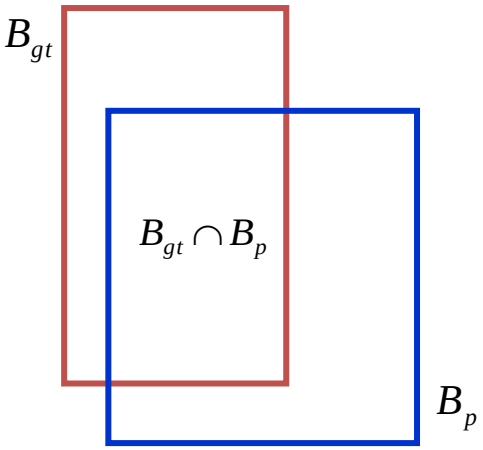
Answer all 3 questions

Time allowed: 2 hours

Question 1

(a) In deciding whether the bounding box for a predicted object is consistent with a ground-truth box, the following measure of overlap is normally used:

$$AO(B_{gt}, B_p) = \frac{|B_{gt} \cap B_p|}{|B_{gt} \cup B_p|}$$



(i) What value would this overlap measure give if the predicted bounding box was exactly the same as the ground-truth box?

[2 marks]

(ii) One might be tempted instead to use the ratio of the overlap area to the area of the ground-truth box, as in:

$$AO'(B_{gt}, B_p) = \frac{|B_{gt} \cap B_p|}{|B_{gt}|}$$

What would be the problem with using this alternative measure in assessing performance? Illustrate your answer with a diagram showing predicted and ground-truth bounding boxes.

[4 marks]

(iii) Bounding boxes are a fairly coarse representation for the position and extent of an object. Give a more precise representation, and explain the implications of using this in collecting a ground-truth dataset and in computing a measure of overlap.

[4 marks]

(b) There are many methods for tracking the objects detected in an image sequence. The problem can be posed as finding an optimal association of detections over the sequence, using a measure of goodness for hypothetical sets of tracks (often expressed as a probability). For each of the following methods, explain why the optimal set of tracks may not be obtained, and what determines the trade-off between performance and compute-time.

(i) MHT

[3 marks]

(ii) Morefield's method

[3 marks]

(iii) Markov Chain Monte Carlo Data Association

[4 marks]

[question 1 total: 20 marks]

Question 2

(a) Consider the following CNN classification stack, with the size of each output on the right.

7	Output of softmax: 10 class probabilities
6	Output of fully-connected (fc) + ReLU: 10
5	Output of reshape to vector (flatten): 4096
4	Output of max-pooling, stride 2: 64x8x8
3	Output of 3x3 convolution + ReLU: 64x16x16
2	Output of max-pooling, stride 2: 32x16x16
1	Output of 5x5 convolution + ReLU: 32x32x32
	Input image: 32x32

(i) For each layer, give the number of parameter values (if any) that must be estimated during optimisation of the CNN on a given dataset. Don't forget that the separate (64 or 32) filters within each of the convolutional layers can be different. Also, don't forgot the bias values. Use the numbering on the left to refer to the layers. Show your working for each layer.

[6 marks]

(ii) What is the effective receptive field size for the outputs at layers 1, 2, 3 and 4? Explain how you arrive at your answers for each layer.

[5 marks]

(iii) One of the variables that must be set prior to optimisation of a CNN is the minibatch size. Give two factors that help in setting this variable.

[4 marks]

(b) Image segmentation can be expressed as an optimisation ('energy' minimisation) problem involving the following cost function:

$$E(\boldsymbol{\alpha}) = \sum_{m} E_r(m) + E_b(m)$$

$$E_r(m) = d\big(\mathbf{z}_m, R(\alpha_m)\big)$$

$$E_b(m) = \textstyle\sum_{m' \in nbrs(m)} [\, \alpha_m \neq \alpha_{m'}] e^{-\beta \|\mathbf{z}_m - \mathbf{z}_{m'}\|}$$

(i) Briefly explain the role of the term $E_b(m)$ in the summation.

[3 marks]

(ii) What is the benefit of expressing the segmentation problem in this form?

[2 marks]

[question 2 total: 20 marks]

Question 3

Repairing holes in the road is a major source of traffic disruption. An unmanned aerial vehicle (UAV) could be used to do this in an autonomous fashion, detecting and repairing cracks in the road surface before they develop into larger holes.

(a) Outline the design of a computer vision system to:

- guide a UAV to survey the roads in a given region, acquiring a set of images that covers the total road surface;

[4 marks]

- segment cracks in the road surface that are visible within the set of images;

[4 marks]

- navigate the UAV to land over each crack;

[4 marks]

- steer a nozzle to deposit filler material into the crack.

[4 marks]

In each part, you should state precisely what techniques you would use and how they fit together. Include any assumptions you make about the positioning of one or more cameras, any additional technology used (e.g. GPS), and any sources of data used for training parts of the system. You are not required to explain how the UAV or nozzle control mechanisms are configured. Treat these as black-boxes that can be told to steer to positions in some coordinate system.

(b) Explain how you would evaluate the system during development.

[4 marks]

[question 3 total: 20 marks]

THE END