TSBB17 Visual Object Recognition and Detection
Answers to Written Exam 2018-01-02

Note: The answers given here are not necessarily the only valid answers to the questions. In some cases the answers are also expanded for clarity, and thus longer than absolutely necessary.

Question 1: Describe the triplet loss and what it is used for.

Answer: *The triplet loss is used to learn a descriptor representation. It uses mined groups of samples called anchor(a), negative(n) and positive(p). These are selected such that* $d(a, n) < d(a, p)$ *before learning, and the goal is to have* $d(a, p) < d(a, n)$ *afterwards.*
See lecture on **Feature Descriptors**

Question 2: What is avoided by the approximation scheme in the multi-channel DCF?

Answer: *A matrix inversion.*
See lecture on **DCF for Visual Tracking**

Question 3: Describe at least one similarity and one dissimilarity between object classification and object detection.

Answer: *Both map appearance to a decision. A classifier outputs confidences (label probabilities). A detector outputs locations (bounding box coordinates).*
See lecture on **Visual Object Detection**

Question 4: Give an example of a shape descriptor, and describe when it is useful.

Answer: *Shape descriptors describe coarse details, e.g. contours and depth boundaries. One example is Contour SIFT, others are Shape Context, and Fourier descriptors.*
See lecture on **Feature Descriptors**

Question 5: How many layers are shortcut in ResNet modules?

Answer: *2 or 3.*
See lecture on **Image Classification with CNNs**

Question 6: What is the effect of spatial regularization in the SRDCF tracker?

Answer: *In SRDCF, filter coefficients outside the tracked patch are penalized to mitigate the emphasis on background information in the learned classifier.*
See lecture on **DCF for Visual Tracking**

Question 7: Which activation function fits well with cross-entropy?

Answer: *The logistic, or sigmoid activation.*
See lecture on **Image Classification with CNNs**

Question 8: What are the alternatives to measuring computational speed in EFO-units within the VOT-challenge?

Answer: *Alternatively one could use actual time on a specific hardware platform.*
See lecture on **Visual Object Tracking Introduction**

Question 9: What is represented in a generative model opposed to a discriminative model in object tracking?

Answer: *The patch appearance.*
See lecture on **Visual Object Tracking Introduction**

Question 10: What does causality mean for object tracking?

Answer: *Future frames may not be used.*
See lecture on **Visual Object Tracking Introduction**

Question 11: Briefly explain at least two advantages of Continuous Convolution Operators.

Answer: *1. enable joint fusion of multi-resolution feature maps. 2. avoid explicit resampling of feature maps since it induces artefacts. 3. provide sub-pixel precision.*
See lecture on **DCF for Visual Tracking**

Question 12: What is the computational benefit of strided convolution compared to pooling after convolution?

Answer: *Filter responses need only to be evaluated in a grid downsampled with the stride. This results in faster convolutional layers.*
See lecture on **CNNs, Introduction and Theory**

Question 13: Why should the weights be initialized with non-zero values in CNNs?

Answer: *Different initialisations for different filters are neccessary for learning different filters (otherwise all filters will have identical error gradients).*
See lecture on **Image Classification with CNNs**

Question 14: Describe one advantage and one drawback of including rare cases in tracking evaluation datasets.

Answer: *It encourages development of generic trackers that can handle many different cases. On the other hand, this discourages the use of specific a priori knowledge of the tracking problem.*
See lecture on **Visual Object Tracking Introduction**

Question 15: How does the validation set help to characterize overfitting and underfitting in CNNs?

Answer: *If loss is significantly worse on the validation set than on the training set we have overfitting. If loss is high on both training and validation data we have underfitting. The main purpose of the validation set is the optimization of meta parameters to minimize overfitting.*
See lecture on **CNNs Introduction and Theory**

Question 16: What do deep features from shallow layers and deep layers represent, respectively?

Answer: *Features from shallow layers give accurate localization, while features from deep layers are semantically meaningful and give robust matches with less precise location.*
See lecture on **DCF for Visual Tracking**