Machine Learning and Object Recognition
M2R Informatique, GVR

Lecturers: Cordelia Schmid & Jakob Verbeek
Exam, Thursday February 2, 2017, 9:00 – 12:00, room D.109

Before you start...

- **Write clearly** so we can read your answers !
- Indicate your name on each page you hand in, and number the pages.
- You can use printouts of the slides and presented research papers during the exam, no other material.
- Each question is worth 5 points (one for each sub-question) out of 20 total.

1 Features, matching and optical flow

a. Describe how to extract Harris interest points and SIFT descriptors for an image. (Only the SIFT descriptor and not the SIFT region extractor needs to be described).
b. The descriptors presented in (a) are extracted for two images. Describe how to robustly match them.
c. Describe an approach how to extract local features with deep learning.
d. Describe the Lucas-Kanade optical flow approach. Comment on its limitations, weaknesses. Discuss the relation with the Harris interest point detector.
e. Present an alternative state-of-the-art optical flow approach. Briefly state its strong points.

2 Clustering local image features and image representation

a. Given a Gaussian mixture

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(x; \mu_k, \Sigma\right), \tag{1}$$

with fixed uniform mixing weights $\pi_k = 1/K$, and the same fixed isotropic co-variance matrix $\Sigma = \sigma I$ for all mixture components. Show that for $\sigma \to 0$ the soft-assignment $p(k|x)$ is the same as the assignment done in the k-means algorithm.
b. Show that for the above mixture, the EM algorithm to learn the μ_k co-indices with the k-means clustering algorithm (we suppose the mixing weights π_k are fixed, as well as the co-variance matrices Σ that are a small multiple of the identity matrix).

c. Describe (at least 2) differences and (at least 2) similarities between bag-of-word and Fisher vector image representations.
d. Describe how we can visualize the features learned by the convolutional layers in a CNN.
e. Describe what kind of image features are learned across the layers of a CNN. How do they differ across layers, and why is this the case?

3 Classification methods

a. Suppose we use the generative classification method for a problem with two classes, and that for each class we estimate $p(x|y)$ as a Gaussian $\mathcal{N}(x; \mu_y, \Sigma)$, *i.e.* the means are different for each class, but the same co-variance matrix is used. Show that classifier obtained using Bayes' rule for this model is a linear classifier.
b. Describe the multi-class logistic discriminant classifier, and derive the gradient of the objective function.
c. Describe how kernels can be used for classification problems. Give an example of kernel for image classification.
d. Describe max-pooling and strided-convolution in CNNs: what are they, what are they used for, how do they differ.
e. Describe the back-propagation algorithm: how is it used to train deep neural networks, describe the explain the equations in detail.

4 Object category localization

An algorithm for object category localization returns bounding boxes for object categories and associated scores. An example is shown in figure 1.

Figure 1: Example detections for the car category. From left to right the detections with the highest detection scores. The higher the score the more confident the detection.

a. Describe two types of regions used for object category localization and discuss their respective advantages/disadvantages.
b. Describe two types of regions descriptors used for object category localization and discuss their respective advantages/disadvantages.
c. Give an evaluation criterion for an algorithm that performs object category localization. The criterion should be based on precision and recall. Start by defining precision and recall.
d. Evaluate the criterion derived in (d) for the example in figure 1. The dataset contains a total of 5 cars.
e. Discuss how to extend Faster R-CNN to action localization.