Computer Vision 1: Homework 10

Important: Mark the homeworks you solved in the homework sheet and bring your solutions with you to the exercise class. For each homework problem, one student will be chosen at random to present their solution.

Programming tasks.

- From the exercise 8, we have two different images of the Hamburg Elbphilharmonie `im` and `im2`. Using `skimage.feature.hog`, generate three HOG feature vectors `v, v2` for each of these images, respectively. The parameters are `orientations = 8, pixels_per_cell =(16, 16), cells_per_block=(1, 1)`.
- Why do the feature vectors have size 9600?
- Compute the Euclidian distances from `v` to `v2`.
- Generate a new augmented image `im3` by adjusting the brightness of the image. We make the image darker by using a technique called `Gamma Correction` with $\gamma = 2$. Using `skimage.exposure.adjust_gamma`, compute and visualize this new image.
- Extract HOG feature vector of `im3` as `v3` (same parameters). What is the distance between `v` and `v3`? How does it compare to the distances between `v` and the feature vectors of the affine transformed image?

Other tasks.

1. We are classifying a 224 x 224 image with the usual three color channels (RGB). In the classification problem, there are 40 different classes an image could belong to. We decide to apply a linear classifier to predict the class scores. The linear classifier evaluates the scores according to

$$y = f(x; W, b) = Wx + b,$$

   where x is the input, W are the weights, b is the bias, and y are the output class scores.

   Write down the sizes of x, W, b, and y, and the total number of parameters the classifier has in case

   (a) we directly represent the input x as the pixels of the image, or

   (b) we use a feature vector of length 512 to represent the input instead.

2. The cross-entropy loss for multi-class classification is defined as:

$$L(\hat{z}, z) = -\sum_{i=1}^{C} z_i \log(\hat{z}_i) \tag{1}$$

   where C is the number of classes, $z_i \in \{0, 1\}$ is the groundtruth probability for class i, $\hat{z}_i \in [0, 1]$ is the probability score of the class i given by a classifier.

   This loss function is also known as categorical cross-entropy loss.

The definition of the categorial cross-entropy loss requires the output $\hat{z}_i$ of the classifier to be in the range $[0, 1]$. Usually this is done by computing the softmax on the output vector as follows:

$$\hat{z}_i = \frac{e^{y_i}}{\sum_{k=1}^{C} e^{y_k}} \tag{2}$$

where y is the output class score vector.

We want to use the linear classifier in Task 1 above together with this categorical cross-entropy loss for classifying whether an image is an cat, a dog or a ship. The indexes for cat, dog and ship are 1, 2 and 3, respectively. For simplicity, assume that the image x is characterized by four features only: $x \in \mathbb{R}^4$.

Suppose we have an image of a cat with $x = \begin{bmatrix} 10 \\ 3 \\ -2 \\ -1 \end{bmatrix}$. The linear classifier has the weight matrix $W = \begin{bmatrix} 1.5 & -2.3 & -1.1 & -0.2 \\ -0.4 & 2.9 & 0.7 & 1.1 \\ 0.5 & -0.1 & 0.1 & 0.3 \end{bmatrix}$ and the bias vector $b = \begin{bmatrix} 1.0 \\ 2.0 \\ -2.0 \end{bmatrix}$. Compute the probability score of each class and the categorical cross-entropy loss.