Deep Learning for Computer Vision
Assignment Sheet 5 - Due 18.02.18, 23:59

Fifth exercise group: 9.2.18, 10:00, A704
Q&A exercise group: 9.2.18, 10:00, A704
Q&A lecture group: 16.2.18, 10:00, A704

Exercise 5.1: action recognition in video sequences with an LSTM (50 + 10 points)
The goal of this exercise is action recognition in a video sequence using a classifier built on Long-Short Term Memory (LSTM) units. We will use the implementation of an LSTM cell provided by Tensorflow.

For this, we prepared a dataset where each datum $\mathbf{x}$ is a record of a person doing one of ten actions, see figure 1 for preview. The dataset is preprocessed and contains cropped sequences with a single person doing only one of the ten actions. Each datum is one video sequence whose length is between 28 and 127 frames using three color channels (it is up to you if want to use color, is is not that meaningful). Each frame has a height of 67 pixels and width of 27 pixels.

We suggest you follow the minimalistic example in Tensorflow which performs classification of the beloved MNIST letters using an LSTM on rows of pixels:

https://github.com/aymericdamien/TensorFlow-Examples/blob/master/examples/3_NeuralNetworks/recurrent_network.py

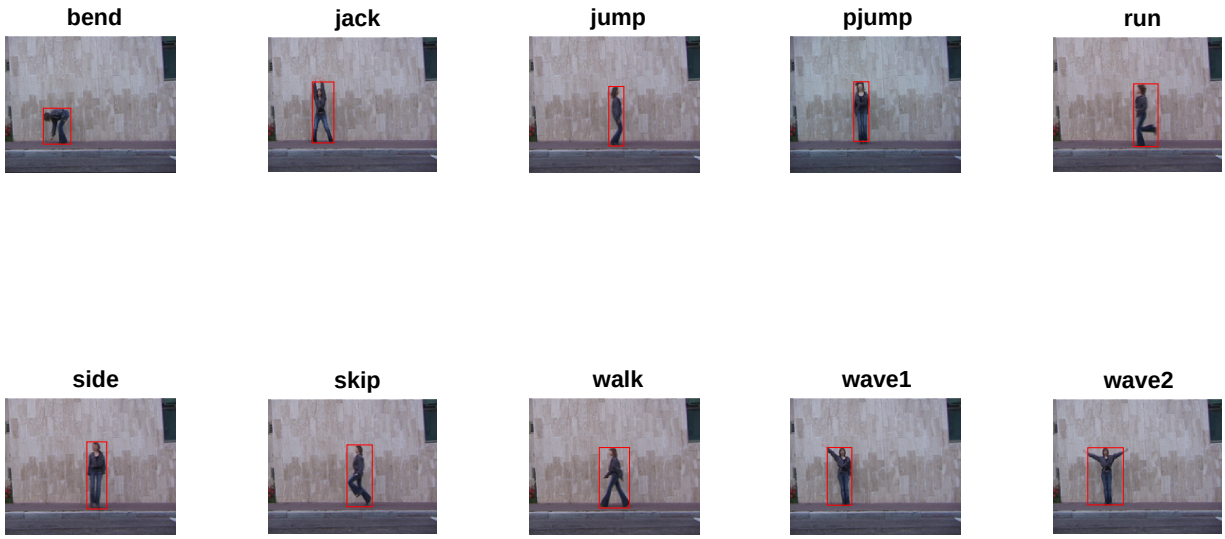

Figure 1: Overview of actions and their labels.

(a) Build an LSTM network based on the architecture of the example, using cross-entropy loss for classification.

(b) Optimize the number of hidden units using 5-fold cross validation.

(c) (Bonus, 10 points) Stitch together video sequences assembled from different parts of the test data, and try to use the network to subdivide these new input sequences into parts where the person is performing different actions.

Exercise 5.2: reading suggestions
As there is formally a week of lectures left during which I am not at university, here are some suggestions for how you could spend the lecture time:

- Work through the proof for the VAE lower bound on slide 6.33 - make sure you understand every step, and how exactly e.g. Bayes rule or the definition of the KL-divergence was applied.
- Work through the slides on generative adversarial networks - understand the formulas, maybe consult Ian Goodfellow et al., "Generative Adversarial Nets", NIPS 2014 for details or watch a video from the Stanford lecture.
- Check out Lecture 12 from Stanford CS231n on "Visualizing and Understanding" neural networks, which I unfortunately did not manage to squeeze in anymore (had some parts covered, but not all of it).
- I would also like to encourage you to take part in the seminar and listen to the talks, even if you are not giving one yourself. We will see interesting GAN architectures, combinations of GANs and autoencoders, applications to tracking and video segmentation, and some more - essentially extensions of the material in the lecture.

  Seminar day is the 2.3.2018, I'll mail around an exact schedule soon. Please sign up on Ilias if you would like to attend so I can reserve an appropriately sized room.