Deep Learning Lab Course

Silvio Galesso and Christian Zimmermann

Due: 8. May, 2019

Outline The goal of this exercise is to understand main concepts and core components of Semantic Segmentation and Human Pose Estimation using Deep Learning methods. We provide you with a reduced and simplified version of the COCO[1] dataset, with according data readers and a simple baseline network architecture. Based on the code and data provided we ask for a series of experiments that will make implementation of additional code necessary. To pass this exercise send us a short report summarizing your findings that can not exceed 2 pages in length as well as the source code used to create the results. Both should be zipped in an archive file and must not exceed 5 MB in size. The report must include the following items:

- Task1: A diagram showing how the MPJPE[2] evolves over training epochs for the training and evaluation set starting from ImageNet initialization.
- Task1: How is final performance influenced by the ImageNet initialization of training?
- Task2: Provide an ablation study of different design choices you made and make a final comparison between the scalar regression and the softargmax variant.
- Task2: Give some qualitative results for the predicted pose overlayed on the input image and the latent heat map representations the network learned. Discuss common failure cases and overfitting.
- Task3: Compare the three listed network architectures and give qualitative examples.

1 Task 1

First, we ask you to train and evaluate the given network architecture on the prepared dataset. In the first task we want you to use the scalar regression approach that is already implemented in the given code base. For this purpose,

[1] `http://cocodataset.org`

[2] Mean per joint position error, see lecture for details.

clone the Github repository and download the dataset version we provide [3]. Next, you need to implement a training procedure which may include, but is not limited to, defining the loss, setting up the optimizer and saving snapshots of the network weights while training. You also have to think of an evaluation strategy that allows you to report how the networks MPJPE performance evolves during training on both the training and the withhold evaluation set.

We recommend the following setting

- Use an L2 loss to formulate the training objective. Make sure missing keypoint information is handeled accordingly.
- ADAM solver with a learning rate in the range of 1e-4 ususally is a good starting point.

You should be able to archieve below 13 pixel MPJPE on the evaluation set with this setup.

2 Task 2

For the second task you should implement the Softargmax loss and adapt the network architecture accordingly. With this loss formulation you can reach below 11 pixel MPJPE making the right architectural changes to the baseline network.

As far as its analysis is concerned we want you to evaluate the following design choices

- How do architectural choices influence the final performance?
- How do the latent heatmaps look like?

3 Task 3

For the last task you should alter the network to train it for semantic segmentation using a cross entropy loss. We want you to experiment with skip connections and decoder architectures and evaluate the networks on a Intersection over Union (IoU) measure.

More precisely we ask for

- IoU results for an encoder only network.
- How does the encoder only network compare to a Encoder Decoder network with and without skip connections?

[3] `https://lmb.informatik.uni-freiburg.de/data/DL_Labcourse/coco_subset.zip`