Abstract

In this thesis we implement and train the Stacked hourglass developed by Newell *et al.* [2]. The network is trained, validated and tested on the Microsoft 2017 COCO dataset [1]. This is followed by an interpretation of the developed model. In the interpretation we first off verify Newell *et al.* [2] and Olsen's [3] claim, that the skip-connections are used for recreating details that are lost during the encoder-phase. Then, we analyse the structure of the latent space. In correlation to this we conclude, that the model has learned the difference between moving and stationary people, learned the difference between almost fully-annotated people and not fully-annotated people, as well as use the first principal component to know if a given person is sitting down or standing up. During the interpretation we further conclude, that there are some redundancy and misplaced training samples in the latent space of the model. Finally, we use the obtained knowledge about the model to improve the performance of the model and remove redundancy in the model. This is done by modifying the architecture of the model to include an autoencoder with a reduced information bottleneck.