

Figure 1: Histogram of PCK training accuracies of the model with the skip-connections enabled (left) and the model with the skip connections disabled (right).

## 1 Understanding the Model

In the following section we will be interpreting the model developed in section ??, with the intention of getting an understanding of how the model works.

### 1.1 Motivation

Deep learning models are often complex and work like a *black boxes*. By that it is meant, that when a network is given some input, the model simply just returns some output without any explanation or reasoning behind the output. This can often be a problem, especially in cases where the output of the network can result in a life or death situation of a human. For that reason, understanding and explaining how a network works can be very important - this is what is called *explainable ai* (XAI)

Selvaraju *et al.* [3] argues that there are three cases for using explainable ai:

1. When the network performs worse than humans, an understanding of the network can help us improve the performance of the model.
2. When the network is on par with humans, an understanding of the network is trivial for humans to build trust in the network, as we can understand its strengths and weaknesses
3. When the network performs better than humans, an understanding of the network can teach humans how to perform better.

Throughout section 1 we will be getting an understanding of our model developed in section ??, as we wish to understand its strengths and weaknesses.

### 1.2 Verifying the Effects of Skip-Connections

Olsen [2] and Newell [1] claims, that the skip-connections are used in order to recreate details that are lost during the encoder-phase. Throughout subsection 1.2 we will be verifying or refuting the claim of the effect of the skip-connections. To do so we will be using two models based on the same network:

1. The trained Stacked Hourglass from section ??
2. The trained Stacked Hourglass from section ??, but with the skip-connections disabled.

Thus, the second model has not been retrained and is identical to the first model, however, without its skip-connections.

In Figure 1 the distributions of the PCK training accuracies of the two models have been visualized. We have decided to make use of the training data for computing the PCK accuracies, as we want to look at the data, that the model has been trained on. By looking at the two distributions we can clearly see, how the model without its skip-connections performs much worse, than the model with its skip-connections.

To further understand the decrease of accuracy in the case where the skip-connections are disabled, we have in Figure 2 visualized 20 samples from the training dataset, where the model with skip-connections has an 100% PCK accuracy score. Next to each image the ground truth heatmaps, or the prediction by the model with skip-connections, and the prediction by the model without skip-connections has been visualized.

By looking at Figure 2 we can see, that the model without its skip-connections often struggles with smaller joints, such as the eyes, ears or nose, whereas it performs better, however still not always perfect, on bigger joints, such as the shoulders, hips or knees. This is probably due to the fact, that the details of the smaller joints has a bigger chance of being lost by the max pooling layers in the encoder. Without the skip-connections they are thus not recreatable in the decoder, resulting in bad predictions. Thus, we can verify Olsen's [2] and Newell's [1] claims, that the skip-connections are used for recreating details lost in the encoder.

### **1.3 Shape Analysis of the Latent Space**

#### **1.4 Using Clustering to Separate the Latent Space**

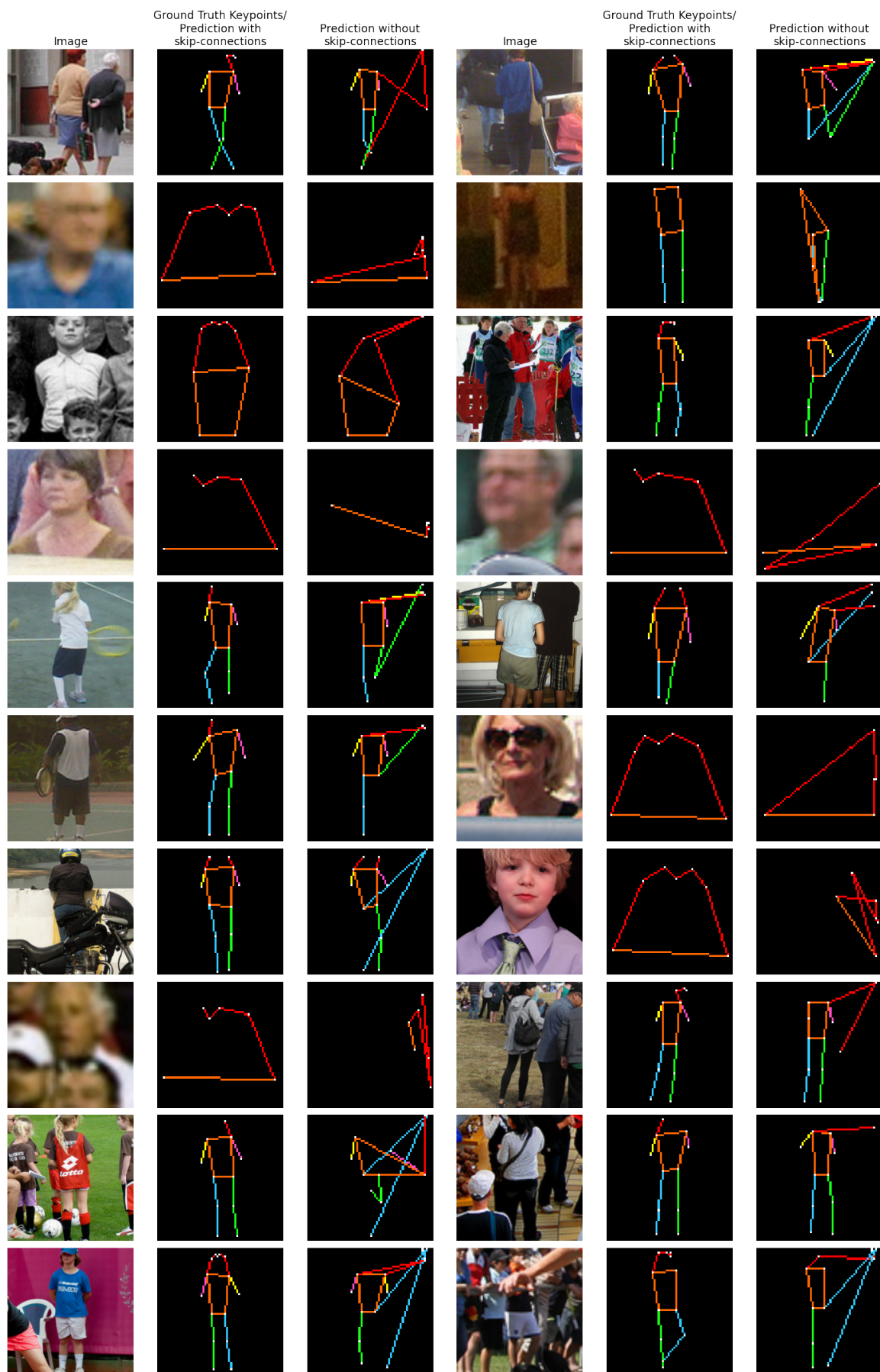


Figure 2: 20 samples of images correctly predicted by the model with skip-connections enabled, the corresponding ground truth heatmap and predictions by the model with skip-connections disabled.