Iqbal et al. - 2018 - Hand Pose Estimation via Latent 2.5D Heatmap Regression

1 Important

Their pose estimation approach works on **single RGB images from the wild** (no camera calibration information given).

They propose a **2.5 dimensional** pose representation (it's scale and translation invariant). Additionally, they propose a way to transfer their 2.5D to a 3D representation. Their 2.5D pose representation looks as follows: The pose consists of several keypoints of the hand. Every keypoint of the hand is represented with the coordinates (x,y,z), where x and y are just the pixel coordinates in the 2D image. The z coodinate is the depth relative to the root keypoint. Additionally, they normalize the scale of the points.

The function that maps the 2D points to 2.5D points is learned via **CNN**. Their loss function has two parts, one for the x,y coordinates of the 2.5D point, and one for the z coordinate. This makes sense since these two kinds of coordinates are aquired differently and need to be treated differently. For each keypoint, it has a 2D **heatmap** and a **depthmap**.

2 Methods

Heatmap regression: is used a lot for 2D pose estimation, but not for 3D estimation since it has high storage cost and uses much more computational resources. The authors propose a more compact 2.5D heatmap.

They are using a **encoder-decoder** network with skip-connections for the heatmap regression.

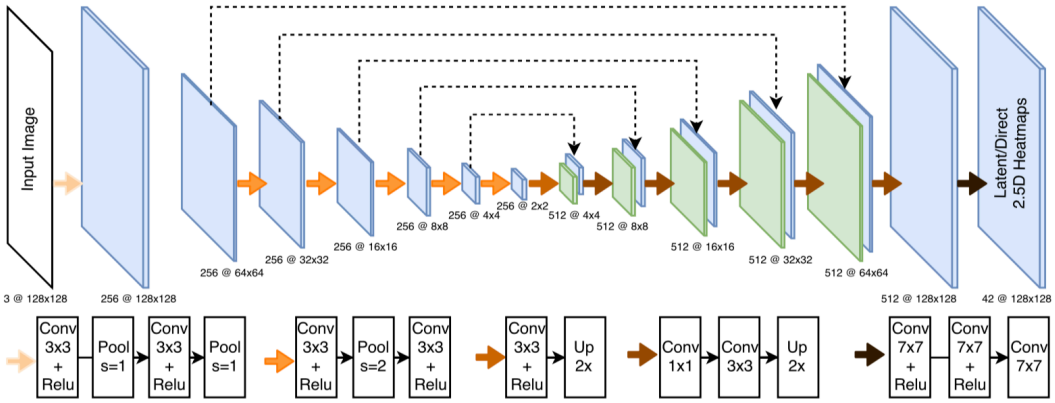

Fig. 5: Backbone network used for 2.5D heatmap regression.

Figure 1

For the holistic regression, they use a **ResNet**.

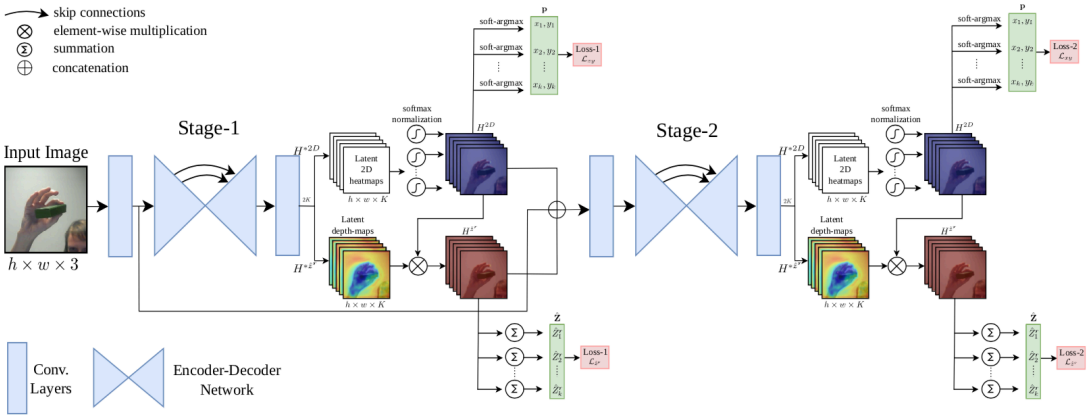

Fig. 6: Overview of the two stage model for latent 2.5D heatmap regression.

Figure 2

3 Additional Processing

Flip left hand training images, which is a common practice for this purpose.

They perform simple **data augmentation**: rotation, translation, scale, color transformations.

They make their model **robust for occlusion**: they randomly add textured ovals and cubes to the training images.