Deep Learning Lab: Exercise 1 (CV)

Neeratyoy Mallik (4774378)

May 8, 2019

1 Task 1

Lowest **MPJPE** without pretrained ImageNet weights: **21.67254 px** after *15 epochs*.
Lowest **MPJPE** with pretrained ImageNet weights: **12.83504 px** after *19 epochs*.

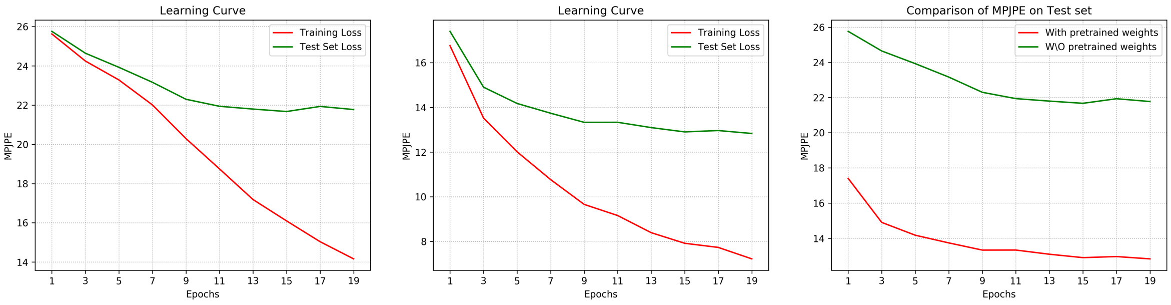

Figure 1: (left) w/o pretrained weights; (middle) with pretrained weights; (right) comparison on test set

The ImageNet weights allow kind of a warmstart wherein a good representation of features or keypoints already exist to begin with. Hence, the model with ImageNet weights show a lower MPJPE by almost 10 pixels in this case, after just 1 epoch. It can therefore focus on learning a generalized human pose rather than learn to identify features first.

2 Task 2

Lowest **MPJPE** without pretrained ImageNet weights: **15.66201 px** after *15 epochs*.
Lowest **MPJPE** with pretrained ImageNet weights: **10.96847 px** after *15 epochs*.

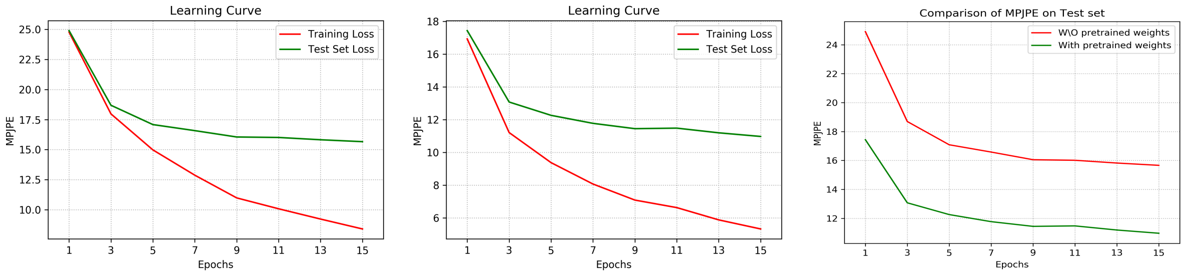

Figure 2: (left) w/o pretrained weights; (middle) with pretrained weights; (right) comparison on test set

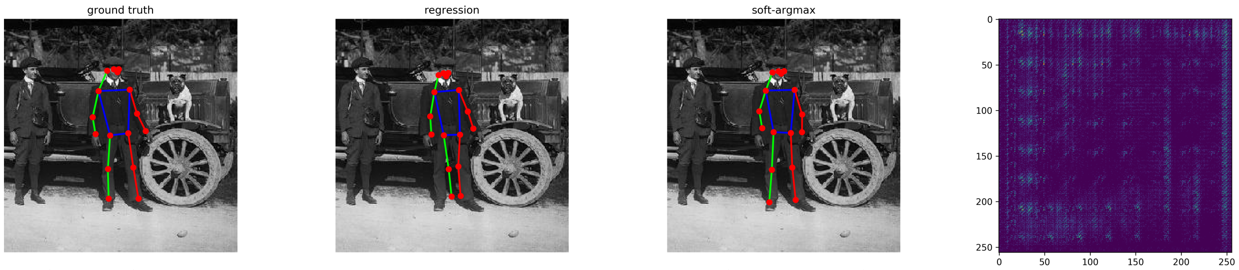

Figure 3: (first 3)Comparison of performance of HPE task; (rightmost)Latent representation of a channel

The pose prediction suffers in cases where parts of the body (keypoints) are occluded, either by other objects or parts of the human body itself due to an unorthodox posture. Also when homogeneity of the clothes worn affect keypoint detection. The network lacks complete understanding of the physiological possibilities of a human pose. These can be seen in Figure 4.

Another anomaly arises in the case of multi-person images, owing to the construction of the dataset. Since the network learns to find the pose from the center of the image, a lack of informative keypoints can throw the prediction completely off. As seen in Figure 5.

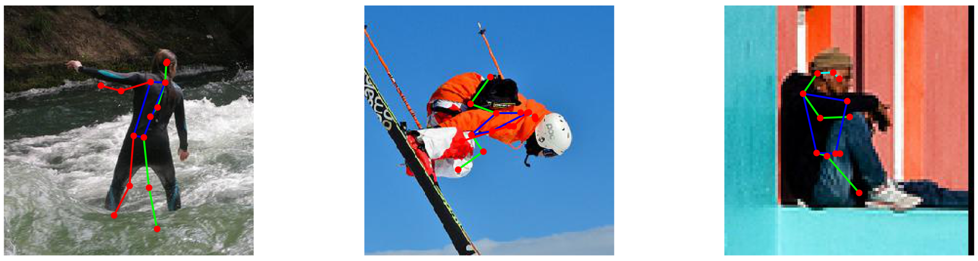

Figure 4: (first) uniform black hinders keypoint detection; (second, third) unnatural pose, occlusion

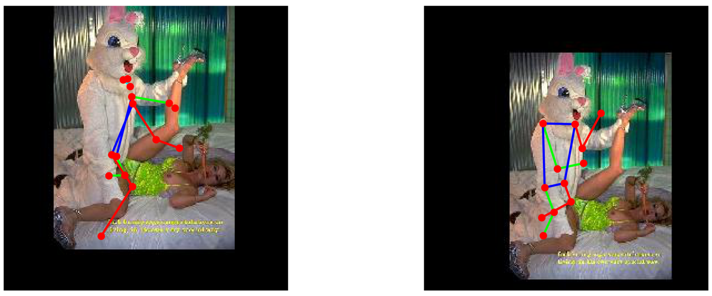

Figure 5: (left) Arms and legs of 2 figures mixed up; (right) same image when shifted does better

3 Task 3

3.1 Upsampling after encoder

# of parameters: *11,177,025*

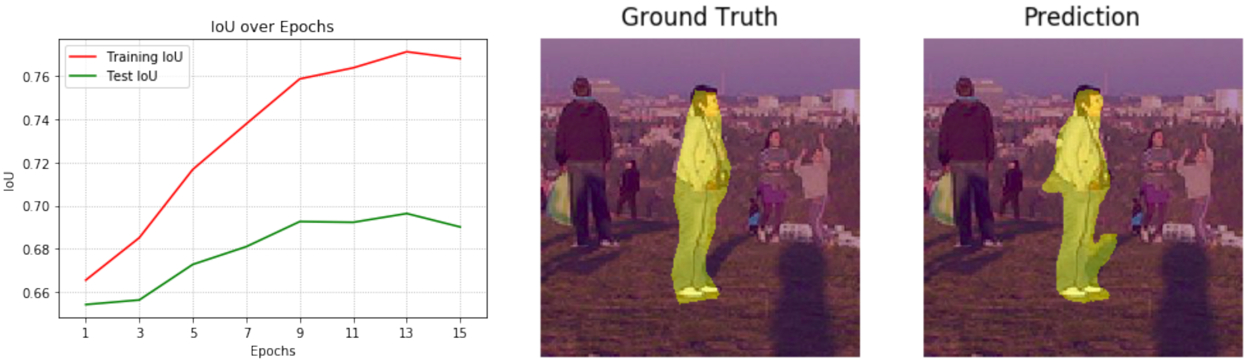

Figure 6: For upsampling:- (left) IoU over epochs; (middle and right) sample output comparison

Highest **IoU**: **0.69257** after *9 epochs*

3.2 Decoders

# of parameters for encoder-decoder: *14,061,505*
# of parameters for encoder-decoder with skip connections: *14,847,937*

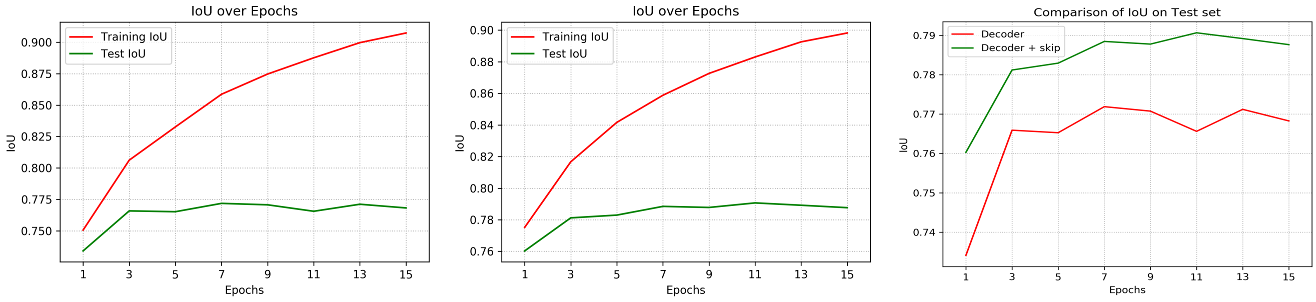

Figure 7: Decoders:- (left) Only upsampling; (middle) Decoder; (right) Decoder with skip connections

Highest **IoU** for encoder-decoder: **0.77188** after *7 epochs*
Highest **IoU** for encoder-decoder with skip connections: **0.79065** after *11 epochs*

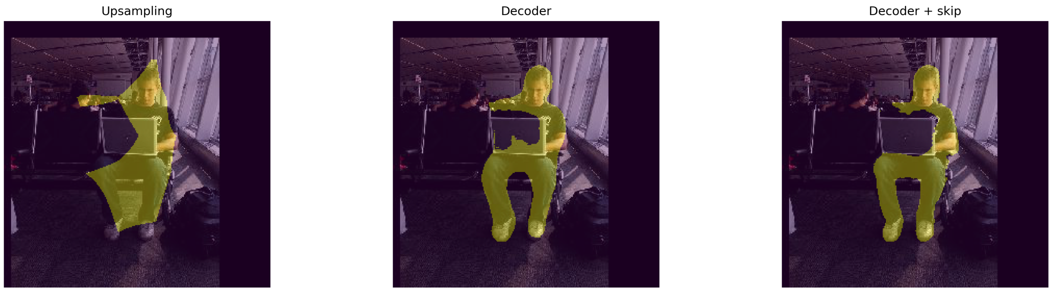

Figure 8: Comparison of the segmentation models

Both quantitatively and qualitatively, the Decoder with skip connections performs the best. As seen from Figure 8, it not only is able to detect a person in a seated position, but also avoids an obstacle occluding part of the body. Moreover, it yields the mask with the smoothest boundaries.