

Supplementary Material for ConVol-E: Continuous Volumetric Embeddings for Human-Centric Dense Correspondence Estimation

Amogh Tiwari Pranav Manu Nakul Rathore Astitva Srivastava Avinash Sharma

Center for Visual Information Technology (CVIT), IIT Hyderabad

{amogh.tiwari, pranav.m, nakul.rathore, astitva.srivastava}@research.iit.ac.in, asharma@iit.ac.in

1. Additional Qualitative Results

Figure 1 and Figure 2 show results of our method on internet images containing single, and multiple humans and occlusions. Our method yield good generalization for pixel-wise embeddings on internet images.

2. Dense Correspondence Matching

Figure 3 shows visualization of dense correspondence matches across multiple images obtained using our method.

3. Segmentation-Label and Garment Transfer

Figure 4 and Figure 5 show results of BodyMap [1] for real world applications of part-segmentation label transfer, and cloth transfer. Corresponding results obtained using our method have been reported in the main paper in Figure 6 and Figure 7. It can be seen that without Neural Rendering, BodyMap representation drastically fails in transferring the garment appearance correctly across a different subject, due to presence of incorrect correspondences.

4. Anchor Points Selection

The positions and ConVol-E values (RGB colors) of 19 anchor points on the SMPL mesh are selected manually (as shown in Figure 6) to ensure that the extrapolation to remaining vertices does not result in any repetitions, which is not guaranteed by a random procedure. We empirically found that manual placement is a good strategy for better efficacy. One can achieve a similar effect by taking a large number of randomly designated anchor points, e.g. designating all 6890 vertices of the SMPL mesh as anchor points. However, neural networks cannot learn such drastically varying embedding values, which will further become dramatic once extrapolated to human scans. On the other hand, placing anchor points symmetrically on the SMPL will ease the learning of the embeddings.

Table 1 shows a quantitative study of anchor point selection. We compare the manual placement of 19 anchor

points with random placement and also with varying numbers of anchor points using the proposed metric Neighborhood Consistency Score on the meshes from 3DHumans [2] dataset.

| No. of Anchor Points | NCS ↓ |
|----------------------|--------------|
| 10 (Random) | 0.975 |
| 19 (Random) | 0.952 |
| 50 (Random) | 0.913 |
| 100 (Random) | 0.847 |
| 19 (Manual) | 0.838 |

Table 1. Quantitative study regarding anchor point selection.

5. Failure Cases of ConVol-E

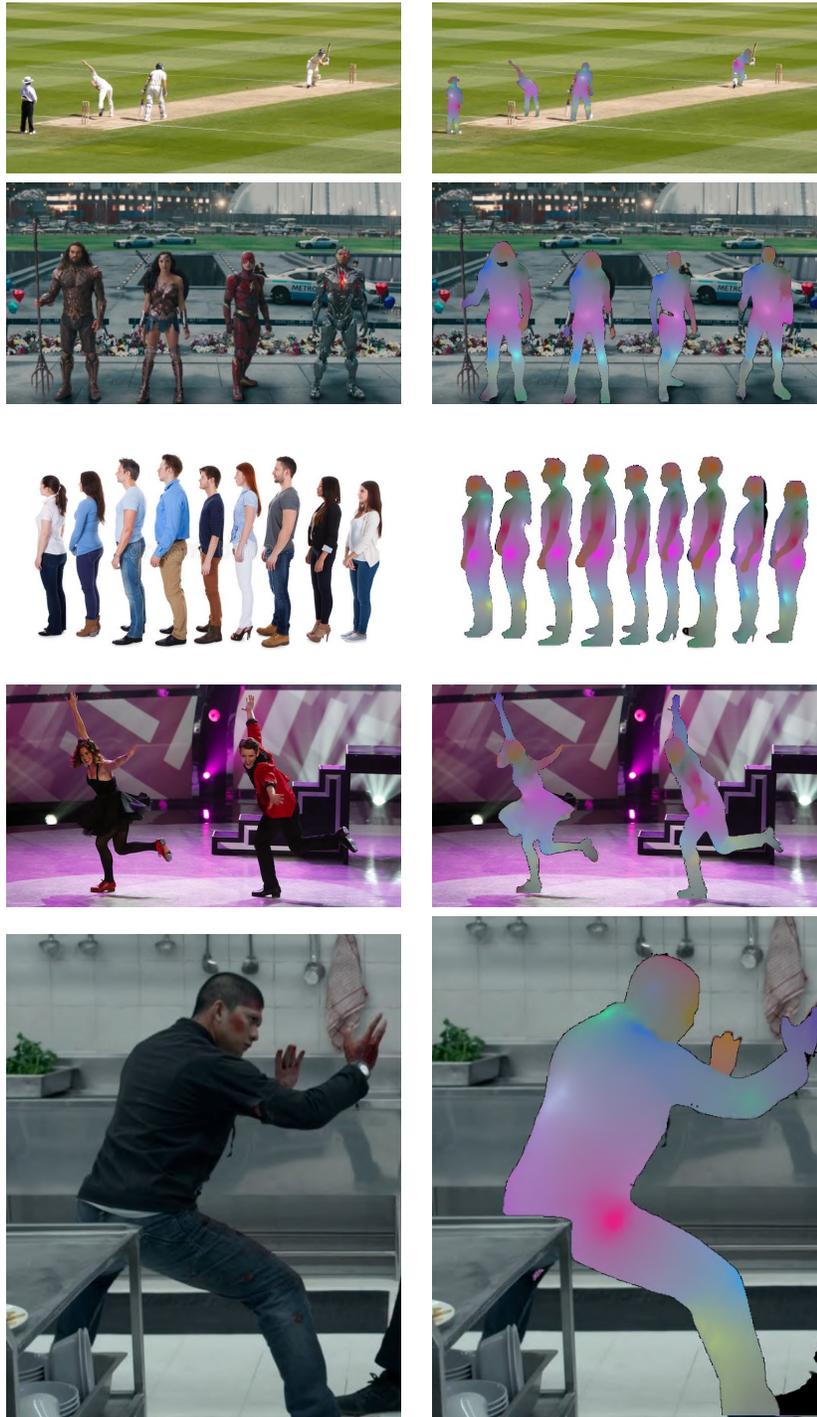
Figure 7 show limitation of our method with first case (top row) involving the garment type (south-asian attire *Saree*) that is very loose, wrapped clothing as well as out of training distribution (from the training set). In the second case (bottom row), the failure is due to severe occlusion caused by another body with a similar appearance of the garment.

References

- [1] Anastasia Ianina, Nikolaos Sarafianos, Yuanlu Xu, Ignacio Rocco, and Tony Tung. Bodymap: Learning full-body dense correspondence map. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13276–13285, 2022. 1, 4, 5
- [2] Sai Sagar Jinka, Astitva Srivastava, Chandradeep Pokhariya, Avinash Sharma, and P. J. Narayanan. Sharp: Shape-aware reconstruction of people in loose clothing. *International Journal of Computer Vision*, Dec. 2022. 1



Figure 1. Visualization of ConVoIE embeddings of our method on internet images.



(a) Input RGB

(b) Output

Figure 2. Results of our method on internet images with multiple humans and occlusions.

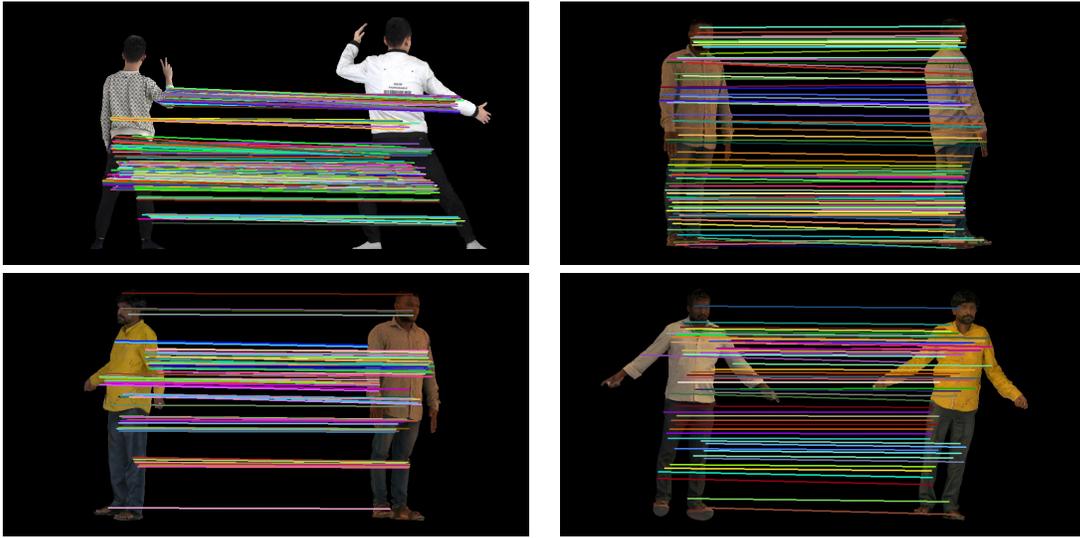


Figure 3. Correspondence matching across images (number of correspondences have been sampled for visualization.)



Figure 4. BodyMap [1] fails to transfer labels correctly across subjects.

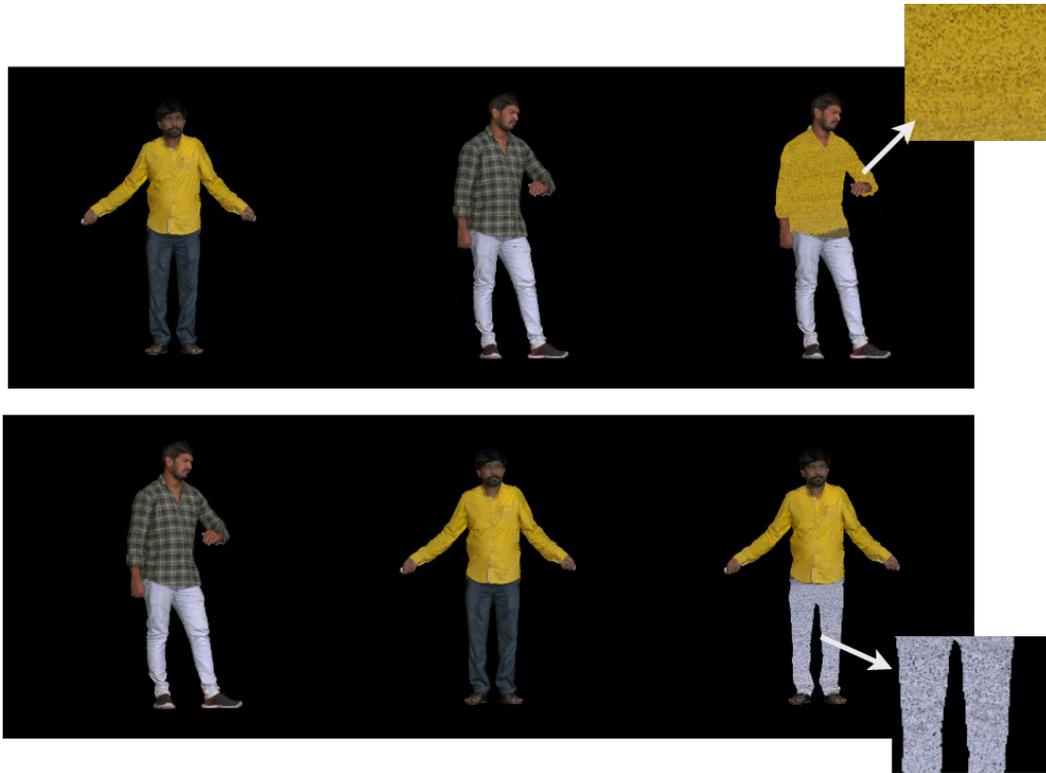


Figure 5. BodyMap [1] representation drastically fails in transferring the garment appearance correctly across different subjects.

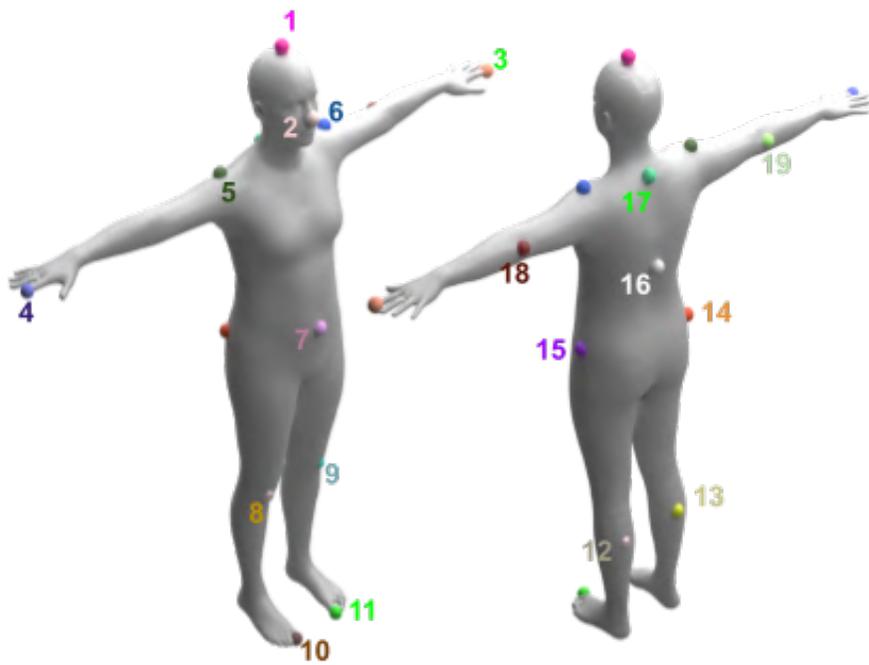


Figure 6. Manually selected anchor points on SMPL



(a) Input RGB.

(b) Output ConVolE Embeddings.

Figure 7. Failure cases of proposed ConVol-E.