

PR-RRN: Pairwise-Regularized Residual-Recursive Networks for Non-rigid Structure-from-Motion - *Supplementary Material*

1. Additional Results on Human3.6M

We evaluate PR-RRN on Human3.6M [1] with ground-truth 2D points and orthographic projection. C3DPO is also evaluated on Human3.6M for comparison. We follow [2] to use the MPJPE metric for evaluation. The results are reported in Table 1. As expected, our method outperforms C3DPO on average.

	Dire.	Disc.	Eat.	Greet.	Phon.	Pos.	Purch.	Sit.
C3DPO	56.1	55.6	62.2	66.4	63.2	62	62.9	76.3
PR-RRN	52.8	54.7	65.1	63.6	63.3	58.4	57.0	78.4
	SitD.	Smo.	Pho.	Wait.	Walk.	Dog.	WalkT.	ALL
C3DPO	85.8	59.9	88.7	63.3	71.1	70.7	72.3	67.8
PR-RRN	98.4	61.1	79.2	59.2	55.0	64.1	64.5	64.6

Table 1. Results on H3.6M

2. Visualization of Representation in Training

We visualize the shape representations learned by PR-RRN in the training process. Similar to the Fig 5. in the paper, the grey points are the representations of randomly selected 1000 frames of CMU Mocap Subject 20. From each group of shapes demonstrated in Fig.5, one shape is selected as an anchor frame and colored in green, blue and orange respectively. Next, we add the representation of all frames in the dataset that have $msr < 0.01$ with each anchor frames, and they are colored the same as the anchor frames. In Fig. 1, representations of these frames from Epoch 0, 10, 50 and 200 are shown. Therefore, each row of Fig. 1 demonstrates how the representations of a set of (near) rigid shapes change in the training process. One can see that, in Epoch 0, the representations are dispersed in the latent space, and as the training goes on, they steadily become closer.

3. Visualization of Reconstruction Results on PASCAL3D+

The visualization of recovered shapes from category Aeroplane, Bicycle, Bus, Car, Chair and Diningtable of PASCAL3D+ are shown in Fig. 2. The

first column contains the origin pictures and visible key-points annotated in green. The following two columns show the reconstructed shapes in two different views. From the figure, one can see that the categorical reconstruction is successful.

4. Visualization of Reconstruction Results on CMU Mocap

We visualize the reconstruction results of CMU Mocap Subject 33 with a video. This Subject captures a throw and catch football motion sequence, containing 2704 frames in total. In the visualization video, the left figure shows the ground-truth 3D shapes, the middle figure shows the predictions of PR-RRN after Procrustes alignment, and the right figure shows the unaligned predictions.

References

- [1] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE TPAMI*, 36(7):1325–1339, 2014.
- [2] David Novotny, Nikhila Ravi, Ben Graham, Natalia Neverova, and Andrea Vedaldi. C3dpo: Canonical 3d pose networks for non-rigid structure from motion. In *ICCV*, 2019.

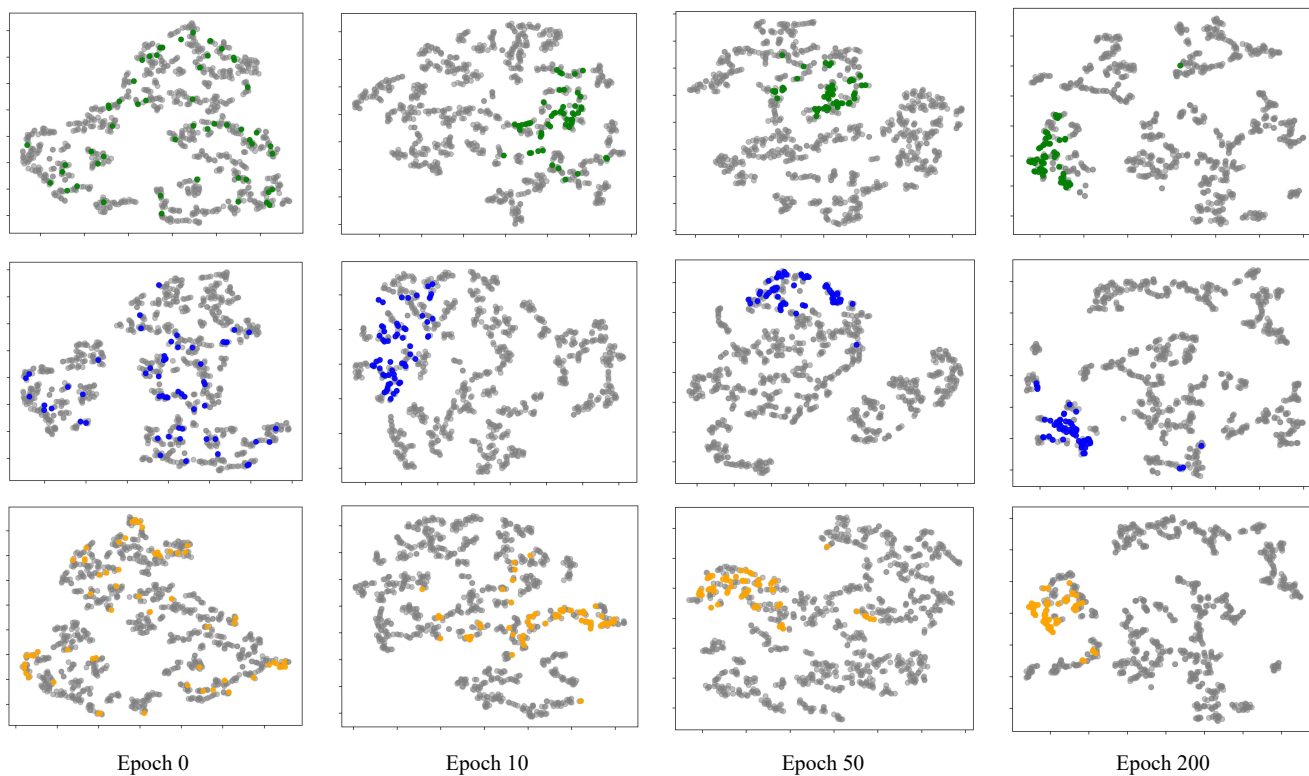


Figure 1. Visualization of representations of (near) rigid shapes in the training process.

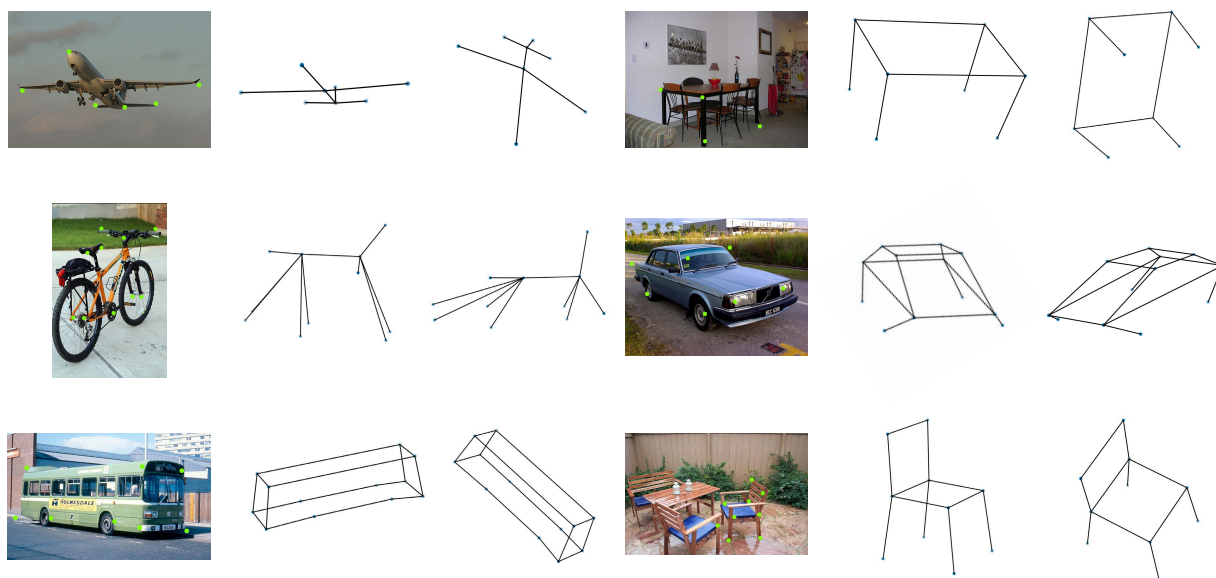


Figure 2. Visualization of reconstruction results on PASCAL3D+.