UNIF: United Neural Implicit Functions for Clothed Human Reconstruction and Animation

Shenhan Qian[1,2] Jiale Xu[2] Ziwei Liu[3] Liqian Ma[1†] Shenghua Gao[2,4,5]

[1] ZMO AI Inc. [2] ShanghaiTech University [3] S-Lab, Nanyang Technological University
[4] Shanghai Engineering Research Center of Intelligent Vision and Imaging [5] Shanghai Engineering Research Center of Energy Efficient and Custom AI IC

ZMO.AI

Code

ECCV TEL AVIV 2022

Introduction

Challenges of Articulated Shape Representation

LBS-based:

- The neural skinning network for neural implicit functions barely generalize to novel poses.

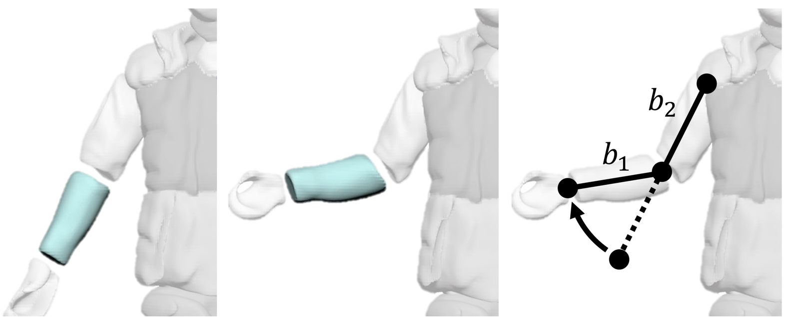
Part-based:

- Relying on ground-truth part labels.
- Lack of explicit modeling of the interaction between parts.

Contributions

We propose Partition-from-Motion, a set of initialization and regularization strategies that helps realize parted-based human body reconstruction with no need for partition labels.

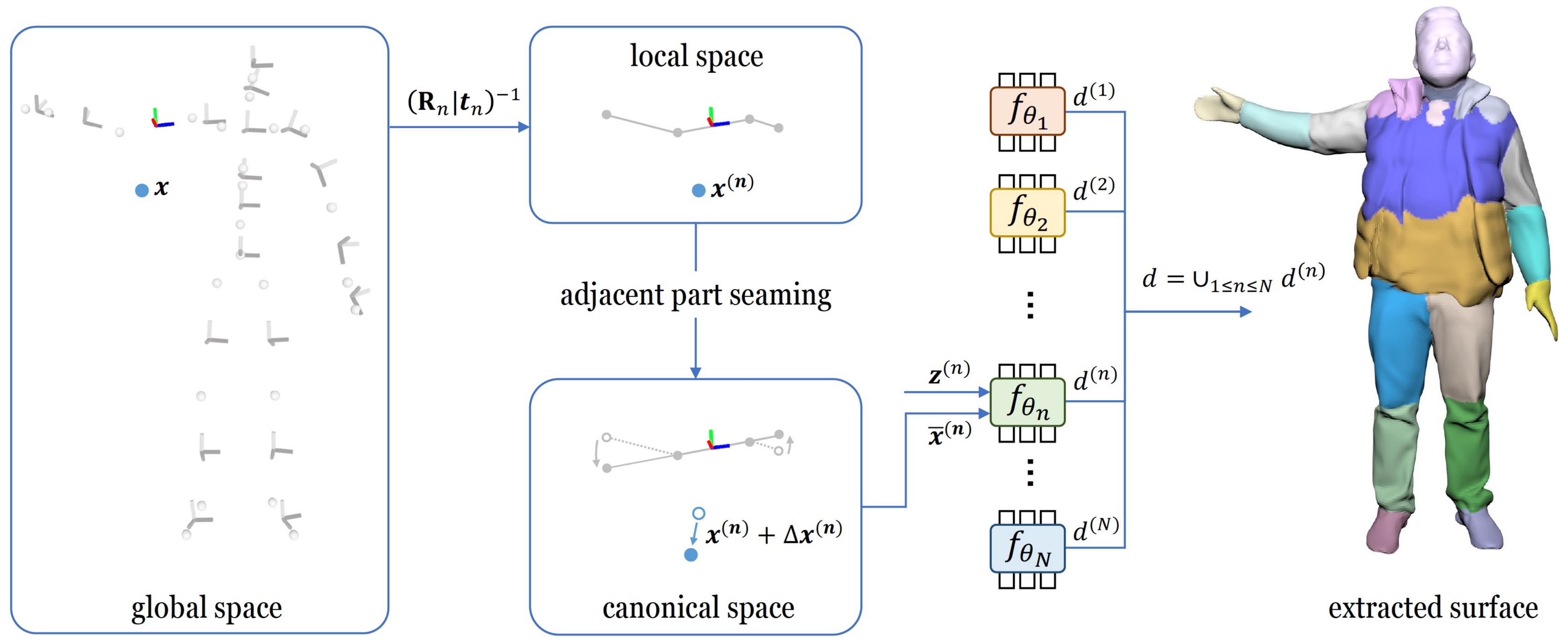
We propose Adjacent Part Seaming, a parametric method that explicitly models the interaction between, improving the pose generalization by a large margin.



Method

Pipeline

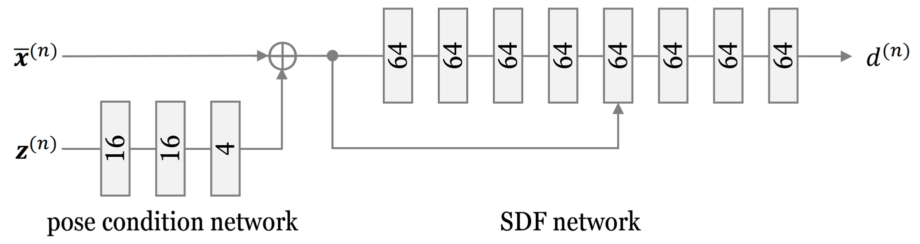
The data we use is a sequence of point clouds, which captures the shapes of a person in varying poses. For each frame, we first fit the body skeleton to the point cloud. Then, we set up local coordinate systems based on the skeleton and define a neural implicit function in each local space. Finally, we optimize the union of the functions to reconstruct the surface.



Input:

$\bar{\boldsymbol{x}}^{(n)}=\boldsymbol{x}^{(n)}+\Delta\boldsymbol{x}^{(n)}$

$\boldsymbol{z}^{(n)}=\oplus_{1\le j\le N}\left(\mathbf{R}_n^T\mathbf{R}_j\oplus\mathbf{R}_n^T(\boldsymbol{t}_j-\boldsymbol{t}_n)\right)$

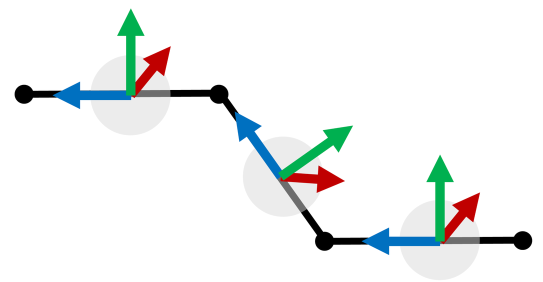


Supervision:

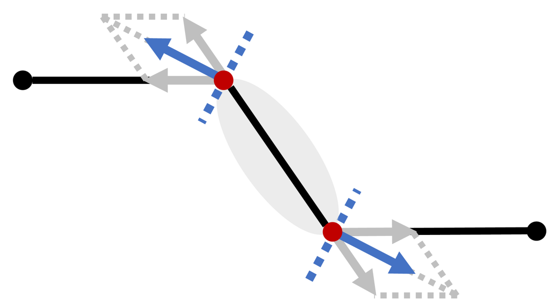
$$\mathcal{L}_{\text{recon}}=\frac{1}{|I|}\sum_{i\in I}\left(|d|+\lambda_{\text{normal}}\left\|\nabla_{\boldsymbol{x}}d-\boldsymbol{n}_i\right\|_2\right)\qquad \mathcal{L}_{\text{unit}}=\mathbb{E}_{\boldsymbol{x}}\left(\left\|\nabla_{\boldsymbol{x}}d\right\|_2-1\right)^2+\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}_{\boldsymbol{x}}\left(\left\|\nabla_{\boldsymbol{x}}d^{(n)}\right\|_2-1\right)^2$$

Partition-from-Motion

Bone-centered initialization: We initialize each part into a small sphere (r = 0.01) at the bone center. Then, parts are not intersected, and the SDF of a part approximately equals the distance to the bone center. This ensures that most points are assigned to the right part when training begins.



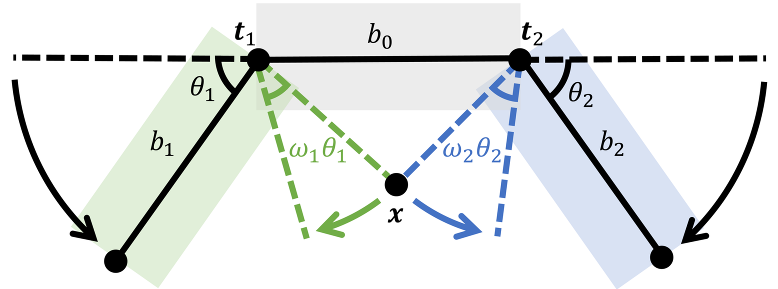
Bone limit loss and section normal loss: When two parts barely have relative motions in the training set, they are at high risk of overlapping. This leads to artifacts when the model is animated under novel poses. Therefore, we propose a bone limit loss and a section normal loss for regularization:



$$\mathcal{L}_{\text{lim}}=\frac{1}{N\cdot|J^{(n)}|}\sum_{n=1}^{N}\sum_{j\in J^{(n)}}\left|d_j^{(n)}\right|,\qquad \mathcal{L}_{\text{sec}}=\frac{1}{N\cdot|J^{(n)}|}\sum_{n=1}^{N}\sum_{j\in J^{(n)}}\left\|\nabla_{\boldsymbol{x}}d_j^{(n)}-\boldsymbol{n}_j^{(n)}\right\|_2$$

Adjacent Part Seaming (APS)

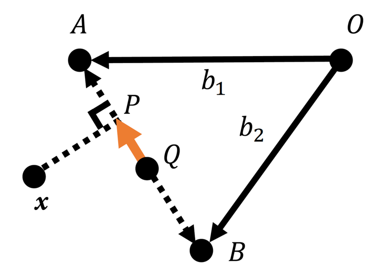
Adjacent part seaming by local rotations: When trying to recover a point $\boldsymbol{x}$ on the bone b_0 to its original position, the point $\boldsymbol{x}$ is expected to go through two different rotations. We blend the offset vectors:



$$\Delta\boldsymbol{x}=\left(\mathbf{R}_{w_1\theta_1}^T(\boldsymbol{x}-\boldsymbol{t}_1)+\boldsymbol{t}_1-\boldsymbol{x}\right)+\left(\mathbf{R}_{w_2\theta_2}^T(\boldsymbol{x}-\boldsymbol{t}_2)+\boldsymbol{t}_2-\boldsymbol{x}\right)$$

Blending weights from "Competing Parts": We define the tendency of a point $\boldsymbol{x}$ to move with either bone b_1 or bone b_2 as

$$r_1=\exp\left(\alpha_1\frac{\overrightarrow{QP}\cdot\overrightarrow{QA}}{\|\overrightarrow{QA}\|^2}+\beta_1\right),\quad r_2=\exp\left(\alpha_2\frac{\overrightarrow{QP}\cdot\overrightarrow{QB}}{\|\overrightarrow{QB}\|^2}+\beta_2\right)$$



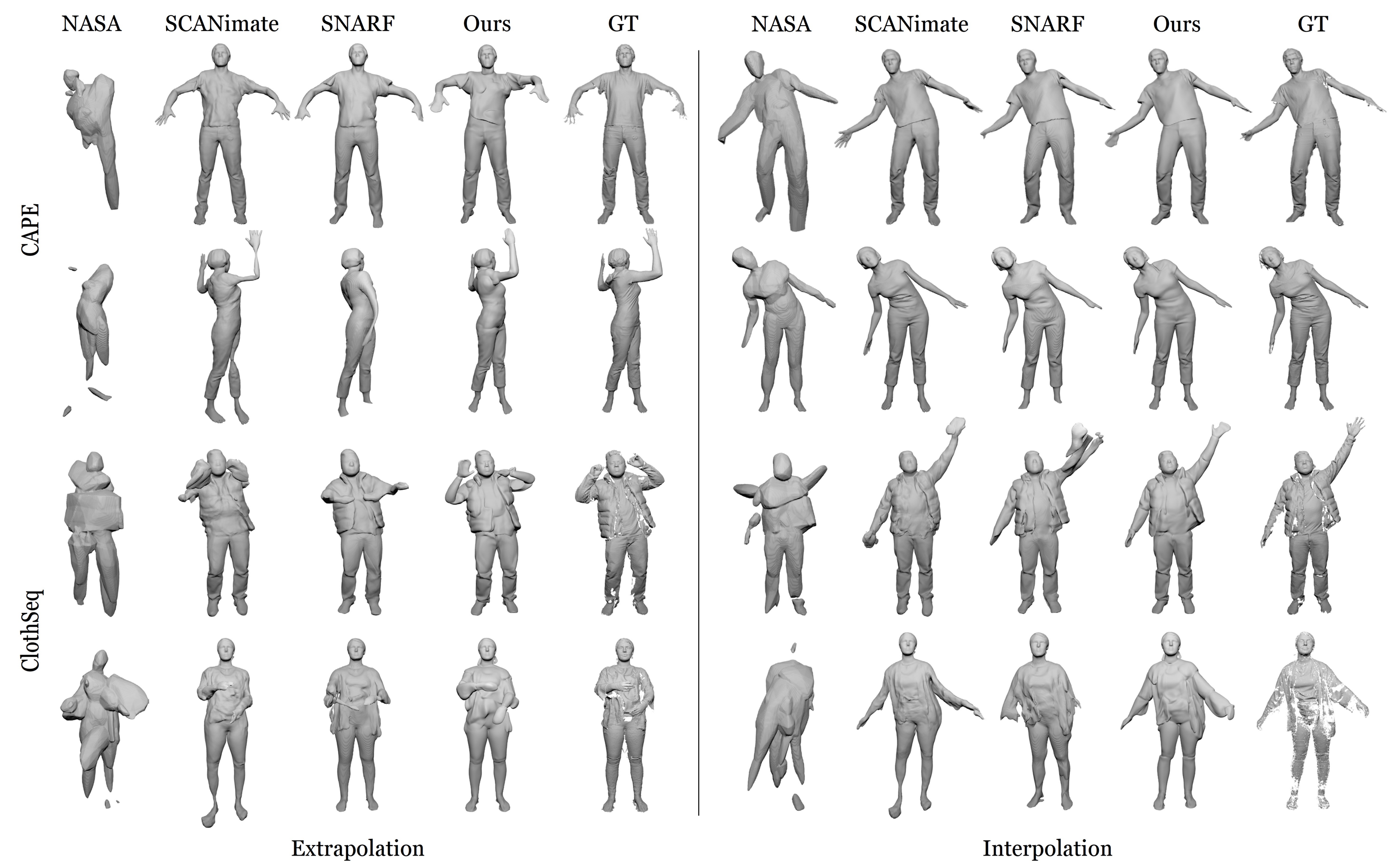
Then we define the blending weights of the point $\boldsymbol{x}$ with respect to bone b_1 and b_2 as $w_1=\frac{r_1}{r_1+r_2}$, $w_2=\frac{r_2}{r_1+r_2}$

Experiments

Comparisons on CAPE and ClothSeq datasets

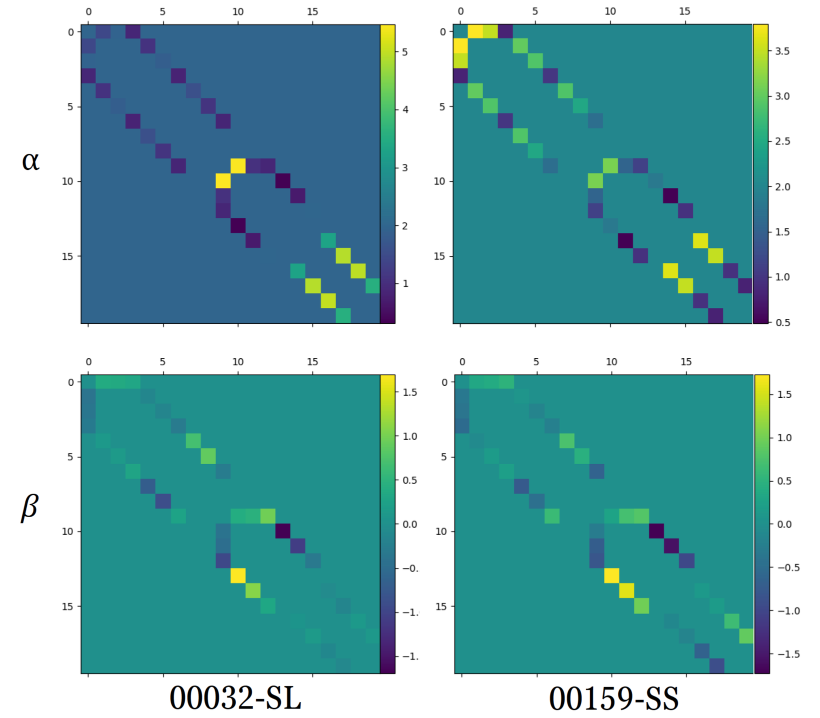
	seq.	SCANimate CD↓	SCANimate F1↑	SCANimate p2s↓	SCANimate Rec.↑	SNARF CD↓	SNARF F1↑	SNARF p2s↓	SNARF Rec.↑	NASA CD↓	NASA F1↑	NASA p2s↓	NASA Rec.↑	Ours CD↓	Ours F1↑	Ours p2s↓	Ours Rec.↑
E	0032-SL	10.19	66.16	10.19	67.31	10.71	65.40	10.68	66.96	98.23	15.78	103.35	15.16	**8.06**	**75.09**	**7.87**	**75.93**
E	0032-SS	9.89	65.56	9.45	66.54	15.49	49.58	15.55	48.58	131.79	9.01	74.52	11.81	**8.37**	**72.55**	**8.18**	**72.86**
E	0096-SL	14.53	56.97	16.89	57.25	12.19	63.70	14.35	63.93	92.74	10.69	93.72	10.53	**10.40**	**64.36**	**10.04**	**65.54**
E	0096-SS	11.25	64.89	11.51	65.50	23.57	**72.47**	24.68	**73.42**	101.51	14.37	86.82	14.77	**8.74**	71.57	**8.50**	72.87
E	0159-SL	7.93	75.24	7.49	76.96	29.34	68.65	33.26	67.22	118.10	8.08	153.04	7.47	**6.64**	**82.42**	**6.28**	**83.05**
E	0159-SS	6.52	84.34	6.15	85.71	20.76	78.39	26.82	77.42	85.46	11.73	81.09	12.37	**5.91**	**86.20**	**5.66**	**87.61**
E	3223-SL	8.12	77.95	8.60	78.28	25.29	68.29	30.17	67.20	66.91	21.31	73.49	20.06	**6.24**	**86.77**	**5.47**	**88.99**
E	3223-SS	9.45	75.08	10.93	74.24	13.90	83.83	16.32	84.41	70.15	22.78	67.47	23.04	**5.61**	**87.88**	**5.31**	**89.61**
I	0032-SL	6.86	85.81	6.80	88.76	4.93	**95.51**	5.06	**97.93**	10.00	74.16	10.01	75.38	**4.14**	95.46	**3.72**	97.60
I	0032-SS	5.70	90.45	5.23	93.39	**4.07**	**96.79**	3.99	**98.23**	10.28	68.45	10.38	69.26	4.17	95.30	**3.83**	97.01
I	0096-SL	8.48	89.50	10.69	91.94	6.48	**96.93**	8.92	98.07	15.47	61.22	18.34	62.08	**4.69**	96.05	**4.47**	**98.33**
I	0096-SS	7.08	82.76	6.47	85.05	4.05	96.41	3.84	97.84	12.73	67.40	11.29	69.12	**3.74**	**97.08**	**3.42**	**98.94**
I	0159-SL	5.18	91.79	4.35	96.01	3.77	96.35	3.18	99.27	11.82	66.37	11.39	69.81	**3.39**	**96.80**	**2.72**	**99.91**
I	0159-SS	4.77	93.75	4.20	96.71	3.42	97.74	3.18	99.19	12.28	65.86	12.04	67.05	**2.94**	**98.00**	**2.69**	**99.81**
I	3223-SL	5.31	93.40	5.26	96.81	5.06	95.70	5.86	95.84	8.17	84.47	7.92	85.61	**3.89**	**96.55**	**3.07**	**99.58**
I	3223-SS	4.89	94.09	4.74	97.15	3.76	97.68	3.88	99.24	7.80	86.61	6.95	87.93	**3.09**	**97.81**	**2.84**	**99.68**
E	JP	14.33	56.25	14.29	58.02	17.72	58.49	21.58	59.16	69.76	16.88	68.24	16.59	**13.04**	**58.60**	**11.24**	**62.06**
E	JS	**11.05**	61.26	10.85	62.58	13.40	57.48	13.91	57.72	116.14	8.51	97.25	9.61	11.84	**65.94**	**9.00**	**70.03**
E	SP	14.32	54.06	14.19	54.80	15.06	60.22	16.47	60.19	65.73	19.93	39.26	21.29	**12.10**	**65.90**	**9.81**	**69.46**
I	JP	10.05	71.78	7.38	79.40	8.43	80.82	8.67	83.92	22.37	44.95	21.97	45.42	**7.87**	**84.66**	**5.47**	**90.01**
I	JS	8.84	74.84	7.77	78.33	**8.81**	80.20	7.86	81.80	33.66	31.61	34.35	31.34	8.89	**81.48**	**5.96**	**86.72**
I	SP	13.20	57.74	12.52	59.28	11.21	73.04	11.08	74.40	48.69	38.02	33.54	40.06	**10.18**	**75.49**	**7.42**	**80.31**

* E: pose extrapolation, I: pose interpolation

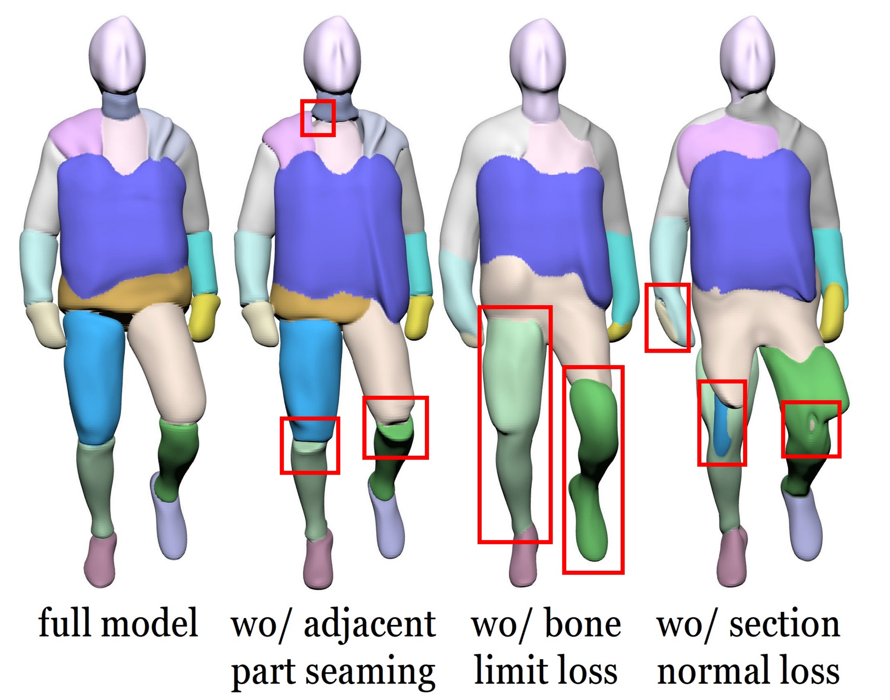


Visualizations

Optimized rigidness coefficients: the matrices of the scaling factor α are symmetric, while the matrices of the bias factor β are skew-symmetric.



Ablation study: We run experiments with each main component disabled and visualize parts in an unseen pose at the early stage of training.



Limitations: Since the #13 and #14 joints of SMPL are too close to the spine, our method learns a small chest and large shoulders. When shoulders move drastically, the model converges to overlapped parts. Therefore, inconsistent part division around the chest can be observed during animation.