Better Rigs, Not Bigger Networks: A Body Model Ablation for Gaussian Avatars

Derek Austin
dcaustin33@gmail.com

Abstract

Recent 3D Gaussian splatting methods built atop SMPL achieve remarkable visual fidelity while continually increasing the complexity of the overall training architecture. We demonstrate that much of this complexity is unnecessary: by replacing SMPL with the Momentum Human Rig (MHR), estimated via SAM-3D-Body, a minimal pipeline with no learned deformations or pose-dependent corrections achieves the highest reported PSNR and competitive or superior LPIPS and SSIM on PeopleSnapshot and ZJU-MoCap. To disentangle pose estimation quality from body model representational capacity, we perform two controlled ablations: translating SAM-3D-Body meshes to SMPL-X, and translating the original dataset's SMPL poses into MHR both retrained under identical conditions. These ablations confirm that body model expressiveness has been a primary bottleneck in avatar reconstruction, with both mesh representational capacity and pose estimation quality contributing meaningfully to the full pipeline's gains.

1. Introduction

Realistic human avatars have a wide range of impactful use cases, especially in the fields of gaming, simulation and AR/VR. Recently, many works have made substantial leaps and achieved high visual quality, but each new method introduces additional architectural complexity to handle failure modes. Many focus on deformation adjustments, such as custom-built MLPs to compensate for clothing and pose-dependent appearance shifts [3, 14, 16]. What these methods share, however, is a dependence on SMPL [8] as the body model.

We show that upgrading the upstream body model reduces the need for this downstream complexity. Specifically, replacing SMPL with the Momentum Human Rig [2] (127 joints and 18.4k vertices compared to 24 joints and 6,890 vertices in SMPL) estimated via the SAM-3D-Body model [21] achieves the highest reported PSNR values and competitive SSIM and LPIPS values in a minimal Gaussian splatting pipeline. To disentangle pose estimation quality from body model representational capacity, we translate our SAM-3D-Body meshes back to SMPL-X (due to the higher level of detail than SMPL) using the fitting tools of SAM-3D-Body [21] and retrain under identical conditions. We also isolate the contribution that pose quality alone gives us by translating from the given SMPL poses into MHR. Our ablations reveal that both pose estimation quality and mesh representational capacity contribute meaningfully to reconstruction fidelity. Crucially, MHR provides gains along both axes simultaneously. Its richer skeleton and recent pose estimation model yields more accurate pose fitting while its denser mesh better supports Gaussian attachment, explaining why the full pipeline's improvement exceeds either factor in isolation.

Our contributions are as follows:

- A minimal Gaussian avatar pipeline that, by substituting MHR for SMPL and removing all learned deformation modules, achieves the highest reported PSNR on both PeopleSnapshot and ZJU-MoCap.
- A controlled ablation that translates poses between body models under identical training conditions, disentangling the contributions of pose estimation quality and mesh representational capacity.

2. Related Work

2.1. Animatable Avatar Methods

Animatable avatar reconstruction from monocular video has progressed rapidly from NeRF-based to Gaussian-based methods. HumanNeRF [19] established the paradigm of learning a canonical T-pose neural radiance field [9] paired with a skeleton-driven motion field, while also establishing evaluation protocols most commonly used today. InstantAvatar [4] combined hash-based NeRF acceleration with Fast-SNARF articulation, reducing training to under a minute and increasing rendering speeds to 15 FPS+.

The shift to Gaussian splatting allowed for faster rendering times along with a host of optimizations to account for persistent failure modes. 3DGS-Avatar [16] uses a deformation module in concert with a multi-level hash grid to model pose-dependent effects in concert with an 'As-Isometric-

Code: https://github.com/dcaustin33/better_rigs_not_bigger_networks

Table 1. Quantitative comparison on the PeopleSnapshot dataset. We report PSNR↑, SSIM↑, and LPIPS↓ for novel pose synthesis. Best results are in **bold**, second best are underlined. All numbers listed are provided by authors.

	Male-3-casual			Male-4-casual			Female-3-casual			Female-4-casual			Avg		
Method	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Anim-NeRF [12]	29.37	0.9703	0.0168	28.37	0.9605	0.0268	28.91	0.9743	0.0215	28.90	0.9678	0.0174	28.89	0.9682	0.0206
InstantAvatar [4]	29.37	0.970	0.017	30.99	**0.982**	0.029	30.81	0.978	0.028	32.57	**0.981**	0.018	30.94	0.9777	0.0230
GaussianAvatar [3]	30.98	0.9790	**0.0145**	28.78	0.9755	**0.0228**	29.55	0.9762	0.0225	30.84	0.9771	**0.0140**	30.04	0.9769	**0.0185**
SplattingAvatar [18]	33.01	**0.982**	0.020	30.99	**0.982**	0.029	30.81	0.978	0.028	32.57	**0.981**	0.018	31.84	**0.9808**	0.0238
3DGS-Avatar [16]	34.28	0.9724	0.0149	30.22	0.9653	0.0231	30.57	0.9581	**0.0209**	33.16	0.9678	0.0157	32.06	0.9659	0.0186
Ours (SMPL-X)	34.65	0.9679	0.0281	32.32	0.9710	0.0343	35.04	0.9724	0.0333	34.91	0.9726	0.0235	34.23	0.9710	0.0298
Ours (MHR)	**36.94**	0.9776	0.0187	**34.20**	0.9778	0.0265	**37.56**	**0.9787**	0.0267	**37.01**	0.9801	0.0164	**36.43**	0.9786	0.0221

As-Possible‘ loss to ensure geometric consistency of Gaussian offsets from the canonical pose to the rendered pose. Human Gaussian Splats [6] utilizes a feature triplane and optimizes LBS weights during training to account for clothing and human deformation during training. GaussianAvatar [3] uses a pose encoder with an optimizable feature tensor per Gaussian to predict pose-dependent offsets, color and scale.

Two methods adopt a simple strategy close to ours by embedding Gaussians directly on mesh triangles. SplattingAvatar [18] optimizes Gaussians directly on the given mesh while introducing a triangle-walking scheme to continually update the face each Gaussian is assigned to. GaussianAvatars [15] rigs Gaussians to FLAME [7] face mesh triangles without any deformation MLP, demonstrating that triangle-embedded Gaussians can produce high-quality results when the underlying mesh is sufficiently expressive. Our method is most architecturally similar to GaussianAvatars, extending a triangle-embedded Gaussian approach from heads to full bodies.

All of the above methods share a common dependency on SMPL as the upstream body model, and address its limitations exclusively through downstream architectural additions. We instead ask whether upgrading the body model itself can eliminate the need for this complexity.

2.2. Parametric Body Models

SMPL [8] has been the dominant body model in avatar reconstruction. Its 24-joint skeleton with 6,890 vertices provides a compact, well-tooled representation, but its limited joint set lacks twist degrees of freedom in the limbs, producing well-known candy-wrapper artifacts under forearm pronation and volume loss at deeply flexed joints. SMPL-X [11] extends SMPL with articulated hands via MANO [17] and facial detail via FLAME [7], increasing the model to 54 joints and 10,475 vertices, but retains the same LBS formulation and joint-from-surface derivation as SMPL.

Momentum Human Rig offers a substantially richer mesh representation with 127 supplemental twist joints, 18,439 vertices, non-linear pose corrections, decoupled skeleton-mesh architectures and a partitioned identity space [2]. Its twist joints directly address the candy-wrapper artifacts that SMPL-based pipelines must learn to compensate for, and its denser mesh provides finer-grained Gaussian attachment. In order to help estimate MHR parameters, SAM-3D-Body was also released which enables fitting of a mesh from a single image [21]. Despite MHR's representational advantages, no prior avatar method, to our knowledge, has explored replacing SMPL with a higher-capacity body model.

3. Method

3.1. Overview

Similar to GaussianAvatars [15], we embed Gaussians directly on mesh triangles. We start by assigning 30,000 Gaussians to our mesh with at least 1 Gaussian to each triangle followed by a random assignment for the remaining. Each Gaussian stores its properties (position, rotation, scale, RGB color) in the local coordinate frame of the parent triangle. We do not use any learnable feature tensor or spherical harmonics in order to represent color. Per-triangle transforms, constructed from the posed vertices, rigidly map each Gaussian's local parameters to world space without any learned deformation.

A summary of the forward pass:

1. Our body model maps pose θ and shape β to posed vertices via LBS and pose corrections offered in MHR
2. Per-triangle transforms (scaling and position) are then constructed from the posed vertices transforming the Gaussian local parameters to world space via each individual triangle's transforms
3. Gsplat [22] is then used as a differentiable rasterizer to render each image

3.2. Training Details

Our loss is constituted of an L1, SSIM and LPIPS reconstruction loss and small mask regularization, i.e. penalizing Gaussians rendered outside the given segmentation mask. All code and hyperparameters are available in the supplementary materials. We explicitly do not use any spherical

Table 2. Quantitative comparison on the ZJU-MoCap dataset. We report PSNR↑, SSIM↑, and LPIPS↓ for novel view synthesis. Best results are in **bold**, second best are underlined. Fewer baselines report ZJU-MoCap numbers under the single-camera training protocol. All numbers listed are provided by authors.

	377			386			387			392			393			394			Avg		
Method	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
GaussianAvatar [3]	24.86	0.944	0.063	27.10	0.923	0.074	25.63	0.947	0.043	26.18	0.929	0.088	23.90	0.919	0.099	26.11	0.925	0.084	25.63	0.9312	0.0752
SplattingAvatar [18]	32.24	0.977	0.028	30.31	0.953	0.073	**30.69**	**0.967**	0.045	**33.41**	**0.975**	0.040	30.02	0.962	0.047	32.36	0.965	0.044	31.51	0.9665	0.0462
3DGS-Avatar [16]	30.96	0.9778	**0.0198**	33.94	**0.9784**	**0.0247**	28.40	0.9656	**0.0330**	32.10	0.9739	**0.0292**	29.30	**0.9645**	**0.0340**	30.74	**0.9662**	**0.0310**	30.91	**0.9711**	**0.0286**
Ours (SMPL-X)	33.51	0.9799	0.0241	35.43	0.9722	0.0323	29.72	0.9557	0.0492	32.43	0.9635	0.0469	29.63	0.9515	0.0567	32.04	0.9599	0.0456	32.13	0.9638	0.0425
Ours (MHR)	**34.26**	**0.9821**	0.0232	**35.50**	0.9716	0.0340	29.81	0.9603	0.0434	33.00	0.9687	0.0395	**30.21**	0.9596	0.0435	**32.46**	0.9644	0.0388	**32.54**	0.9678	0.0371

harmonics, deformation MLPs, feature triplanes or specialized losses in order to isolate mesh types as the main source of quality.

We run each experiment for 50k iterations with the Adam optimizer [5] with an exponential decay schedule to 10% of the initial learning rate. We start each experiment with 30k initial Gaussians and follow a typical split, prune and densify schedule every 500 iterations. Once the total number of Gaussians has exceeded 100k we do not allow any further splitting or densifying. The only variable we change across all of our experiments is the mesh representation.

3.3. Body Mesh

SAM-3D-Body estimates MHR parameters that we use to train MHR-specific models. Those MHR parameters are then used to fit to SMPL-X via iterative optimization through tools provided in the SAM-3D-Body repository [21]. We choose SMPL-X rather than SMPL as the ablation target due to it being the highest-capacity model in the SMPL family. SMPL-X's articulated hands and expressive face mesh ensure that any remaining quality gap cannot be attributed to SMPL simply lacking geometry in those regions. All code is made available in the supplementary materials and our dataset will be made available to the community.

4. Experiments

4.1. Setup

Datasets. We evaluate on PeopleSnapshot [1] (4 subjects: male-3-casual, male-4-casual, female-3-casual, female-4-casual; monocular video, 540×540) and ZJU-MoCap [13] (6 subjects: 377, 386, 387, 392, 393, 394; multi-view capture, 512×512). For PeopleSnapshot we adopt the temporal train/test split of Anim-NeRF [12]; for ZJU-MoCap we train on a single camera (camera 1) and evaluate on held-out cameras following 3D-Gaussian-Avatars [16].

Metrics. We report PSNR, SSIM, and LPIPS (VGG backbone) to match published baselines. PSNR measures per-pixel similarity whereas SSIM and LPIPS provide structural and textural quality.

4.2. Main Results

Our pipeline achieves the highest average PSNR on both benchmarks while remaining competitive if not exceeding the state of the art on SSIM and LPIPS. The PSNR advantage is notable because PSNR penalizes per-pixel error without spatial tolerance, making it the metric least amenable to compensation by learned texture and appearance corrections. Leading PSNR scores indicate that MHR's posed meshes provide a stable, faithful representation allowing the optimization to place Gaussians closer to the true surface than SMPL-based alternatives, producing geometrically more accurate renderings before any learned correction has a chance to act.

The narrower margins on SSIM and LPIPS are expected as both metrics are more sensitive to perceptual high-frequency texture quality, precisely the phenomena that color MLPs and pose-conditioned decoders in prior methods target. We deliberately disable these components to isolate the body model's contribution and provide a floor for MHR's performance. The remaining gap suggests that lightweight learned corrections on top of MHR would likely close it, a direction we leave to future work.

One possible explanation of our results is that the deformation MLPs that many competing methods use serve as a stand-in for the non-linear pose correctives that MHR employs.

4.3. Qualitative Comparison

We find that MHR provides far higher quality detail on tough-to-model body locations like hands, faces and elbows than even SMPL-X. As shown in Figure 1, our model that utilizes MHR is far more visually realistic in hands and face especially as compared to SMPL-X. We hypothesize that due to the more faithful and finer detail mesh that MHR provides, the optimization process is more stable, allowing the Gaussians to faithfully model hands and facial shape.

4.4. Disentangling Pose Quality and Mesh Capacity

Our main results conflate two factors: SAM-3D-Body may produce more accurate poses than the Anim-NeRF fits used by prior methods, and MHR's mesh may better support Gaussian attachment than SMPL-X. Figure 2 illustrates the pose estimation gap. The SAM-3D-Body fit (right) captures

Figure 1. Rendering of Male-3-casual from the People Snapshot dataset with MHR mesh on the left and SMPL-X rendering on the right.

Table 3. Impact of pose fitting and body model on rendering quality (PeopleSnapshot). Replacing Anim-NeRF's SMPL fits with SAM-3D-Body's MHR fits yields large gains (rows 1 vs. 3), and retaining the native MHR representation rather than converting to SMPL-X provides further improvement (rows 2 vs. 3). Best results are in **bold**, second best are underlined.

	Pose Fitting	Rendering Model	PSNR ↑	SSIM ↑	LPIPS ↓
male-3-casual	Anim-NeRF (SMPL)	MHR	31.65	0.9538	0.0349
	SAM-3D-Body (MHR)	SMPL-X	34.65	0.9679	0.0281
	SAM-3D-Body (MHR)	MHR	**36.94**	**0.9776**	**0.0187**
female-3-casual	Anim-NeRF (SMPL)	MHR	33.05	0.9620	0.0405
	SAM-3D-Body (MHR)	SMPL-X	35.04	0.9724	0.0333
	SAM-3D-Body (MHR)	MHR	**37.56**	**0.9787**	**0.0267**

head orientation, eye gaze, and foot placement more faithfully than the Anim-NeRF fit (left), and its mesh contours follow the underlying human shape rather than the clothing silhouette. Table 3 isolates each factor on the two PeopleSnapshot subjects for which Anim-NeRF's optimized SMPL parameters are publicly available via links provided by the Anim-NeRF team. Averaged across both subjects, upgrading pose estimation alone accounts for +2.50 PSNR, +0.0123 SSIM, and −0.0070 LPIPS, while upgrading the rendering mesh contributes a further +2.41 PSNR, +0.0080 SSIM, and −0.0080 LPIPS, confirming that both factors provide independent, roughly comparable gains.

5. Conclusion

Body model quality is an underexplored axis of improvement in the Gaussian avatar literature. Our results demonstrate that MHR's representational advantages (twist joints, non-linear pose correctives, and higher mesh density) are sufficient to match or exceed methods that compensate for SMPL's failure modes with learned deformation modules, rectification networks, and pose-dependent MLPs. Our controlled ablation, retaining SAM-3D-Body poses while translating to SMPL-X, confirms that representational capacity contributes independently of pose estimation quality as identical poses consistently yield better results when utilizing MHR.

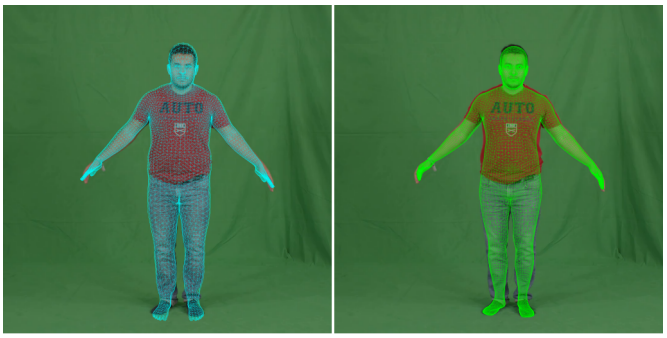

Figure 2. Render of pose provided by Anim-NeRF (left) compared to the pose estimated with SAM-3D-Body (right).

This finding suggests the field's architectural innovations, while effective, have been partially compensating for upstream body model deficiencies. Upgrading the body model may be a simpler and more direct path to quality gains than adding downstream complexity.

Our study has several limitations. Our pipeline deliberately disables all learned corrections to isolate the body model's contribution and consequently falls short of leading methods on SSIM and LPIPS. The remaining perceptual gap indicates these corrections still provide value, particularly for pose-dependent texture detail. We compare only SMPL-family models and MHR; other body models such as STAR [10] and GHUM [20] may offer different trade-offs. MHR's ecosystem is also far less mature than SMPL's, though SAM-3D-Body provides a practical estimation path that requires no custom fitting pipeline.

Three directions follow naturally from this work. First, combining MHR with lightweight pose-dependent corrections should close the remaining perceptual gap while preserving the geometric foundation that MHR provides. Second, ablating individual MHR features would reveal which aspects of body model design matter most for downstream avatar quality, with practical implications for future body model development. Finally, updating existing methods with improved poses and MHR could reveal how much of their learned deformation capacity becomes redundant when the upstream body model already handles twist corrections and fine-grained surface detail, potentially enabling simpler and faster architectures without sacrificing quality.

References

[1] Thiemo Alldieck, Marcus Magnor, Weipeng Xu, Christian Theobalt, and Gerard Pons-Moll. Video based reconstruction of 3d people models. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8387–8397, 2018. 3

[2] Aaron Ferguson, Ahmed A. A. Osman, Berta Bescos, Carsten Stoll, Chris Twigg, Christoph Lassner, David Otte, Eric Vignola, Fabian Prada, Federica Bogo, Igor Santesteban, Javier Romero, Jenna Zarate, Jeongseok Lee, Jinhyung Park, Jinlong Yang, John Doublestein, Kishore Venkateshan, Kris Kitani, Ladislav Kavan, Marco Dal Farra, Matthew Hu, Matthew Cioffi, Michael Fabris, Michael Ranieri, Mohammad Modarres, Petr Kadlecek, Rawal Khirodkar, Rinat Abdrashitov, Romain Prévost, Roman Rajbhandari, Ronald Mallet, Russell Pearsall, Sandy Kao, Sanjeev Kumar, Scott Parrish, Shoou-I Yu, Shunsuke Saito, Takaaki Shiratori, Te-Li Wang, Tony Tung, Yichen Xu, Yuan Dong, Yuhua Chen, Yuanlu Xu, Yuting Ye, and Zhongshi Jiang. Mhr: Momentum human rig, 2025. 1, 2

[3] Liangxiao Hu, Hongwen Zhang, Yuxiang Zhang, Boyao Zhou, Boning Liu, Shengping Zhang, and Liqiang Nie. Gaussianavatar: Towards realistic human avatar modeling from a single video via animatable 3d gaussians. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 1, 2, 3

[4] Tianjian Jiang, Xu Chen, Jie Song, and Otmar Hilliges. Instantavatar: Learning avatars from monocular video in 60 seconds. *arXiv*, 2022. 1, 2

[5] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 2015. 3

[6] Muhammed Kocabas, Jen-Hao Rick Chang, James Gabriel, Oncel Tuzel, and Anurag Ranjan. HUGS: Human gaussian splats. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 505–515, 2024. 2

[7] Tianye Li, Timo Bolkart, Michael. J. Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4D scans. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 36(6):194:1–194:17, 2017. 2

[8] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, 2015. 1, 2

[9] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. 1

[10] Ahmed A A Osman, Timo Bolkart, and Michael J. Black. STAR: A sparse trained articulated human body regressor. In *European Conference on Computer Vision (ECCV)*, pages 598–613, 2020. 4

[11] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. Expressive body capture: 3D hands, face, and body from a single image. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 10975–10985, 2019. 2

[12] Sida Peng, Junting Dong, Qianqian Wang, Shangzhan Zhang, Qing Shuai, Xiaowei Zhou, and Hujun Bao. Animatable neural radiance fields for modeling dynamic human bodies. In *ICCV*, 2021. 2, 3

[13] Sida Peng, Yuanqing Zhang, Yinghao Xu, Qianqian Wang, Qing Shuai, Hujun Bao, and Xiaowei Zhou. Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans. In *CVPR*, 2021. 3

[14] Sen Peng, Weixing Xie, Zilong Wang, Xiaohu Guo, Zhonggui Chen, Baorong Yang, and Xiao Dong. RMAvatar: Photo-realistic human avatar reconstruction from monocular video based on rectified mesh-embedded gaussians. *arXiv preprint arXiv:2501.07104*, 2025. 1

[15] Shenhan Qian, Tobias Kirschstein, Liam Schoneveld, Davide Davoli, Simon Giebenhain, and Matthias Nießner. Gaussianavatars: Photorealistic head avatars with rigged 3d gaussians. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20299–20309, 2024. 2

[16] Zhiyin Qian, Shaofei Wang, Marko Mihajlovic, Andreas Geiger, and Siyu Tang. 3DGS-Avatar: Animatable avatars via deformable 3D gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5020–5030, 2024. 1, 2, 3

[17] Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied hands: Modeling and capturing hands and bodies together. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 36(6), 2017. 2

[18] Zhijing Shao, Zhaolong Wang, Zhuang Li, Duotun Wang, Xiangru Lin, Yu Zhang, Mingming Fan, and Zeyu Wang. SplattingAvatar: Realistic real-time human avatars with mesh-embedded gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1606–1616, 2024. 2, 3

[19] Chung-Yi Weng, Brian Curless, Pratul P. Srinivasan, Jonathan T. Barron, and Ira Kemelmacher-Shlizerman. HumanNeRF: Free-viewpoint rendering of moving people from monocular video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16210–16220, 2022. 1

[20] Hongyi Xu, Eduard Gabriel Bazavan, Andrei Zanfir, Bill Freeman, Rahul Sukthankar, and Cristian Sminchisescu. Ghum & ghuml: Generative 3d human shape and articulated pose models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (Oral)*, pages 6184–6193, 2020. 4

[21] Xitong Yang, Devansh Kukreja, Don Pinkus, Anushka Sagar, Taosha Fan, Jinhyung Park, Soyong Shin, Jinkun Cao, Jiawei Liu, Nicolas Ugrinovic, Matt Feiszli, Jitendra Malik, Piotr Dollar, and Kris Kitani. Sam 3d body: Robust full-body human mesh recovery. *arXiv preprint arXiv:2602.15989*, 2026. 1, 2, 3

[22] Vickie Ye, Ruilong Li, Justin Kerr, Matias Turkulainen, Brent Yi, Zhuoyang Pan, Otto Seiskari, Jianbo Ye, Jeffrey Hu, Matthew Tancik, and Angjoo Kanazawa. gsplat: An open-source library for gaussian splatting. *Journal of Machine Learning Research*, 26(34):1–17, 2025. 2