KeypointNeRF: Generalizing Image-based Volumetric Avatars using Relative Spatial Encoding of Keypoints — Supplementary Material —

A Overview

In this document we provide additional implementation details (Sec. B), information about the baseline methods (Sec. C), more qualitative and quantitative results (Sec. D), and reflect on the limitations of KeypointNeRF and future work (Sec. E).

B Implementation Details

Image Encoders. We employ a single HourGlass [43] network to learn a geometric prior of humans and condition the density estimation network. The input image is normalized to $[-1, 1]$ range and processed by four convolutional blocks (256 filters) interleaved with group normalization. We then employ an HourGlass block (down-sampling rate of four) with group normalization layers and refine the final output with four convolutional layers to produce the deep feature map $F_n^{gl} \in \mathbb{R}^{H/8 \times W/8 \times 64}$. Additionally, after the second convolutional block, we employ the transposed convolutional layer to produce the shallow high-resolution feature map $F_n^{gh} \in \mathbb{R}^{H/2 \times W/2 \times 8}$. As activation function we use ReLU for all layers. We implemented a second convolutional encoder that is independent of the density prediction branch to produce an alternative pathway for the appearance information $F_n^{a} \in \mathbb{R}^{H/4 \times W/4 \times 8}$ as in DoubleField [54]. We follow the design of [25] and implement this encoder as a 15-layer convolutional network with residual connections and ReLU activations.
Multi-view Feature Fusion. The feature fusion network is implemented as a four-layer MLP (128, 136, 120, and 64 neurons with Softplus activations) that aggregates features from multiple views. Its output is aggregated via mean-variance pooling [62] to produce the geometry feature vector $G_X \in \mathbb{R}^{128}$.
Density Fields. The geometry feature vector is decoded as density value σ via a four-layer MLP (64 neurons with Softplus activations).
View-dependent Color Fields. To produce the final color prediction c for a query point X, we implemented an additional MLP that predicts blending weights as an intermediate step which are used to blend the input pixel colors. This network follows the design proposed in IBRNet to communicate information among multi-view features by using the mean-variance pooling operator. The per-view input feature vectors (described in Sec. 4.4) are first fused into a

Fig. B.1. Studio Capture Results. Reconstruction results on held-out subjects from only two input views. Our method produces much sharper results with fewer artifacts compared to prior work. Best viewed in electronic format.

global feature vector via the mean-variance pooling operator. Then this feature is attached to the pixel-aligned feature vectors $\Phi(X|F_n^a)$ and propagated through a nine-layer MLP with residual connections and an exponential linear unit as activation to predict the blending weights (Eq. 4).

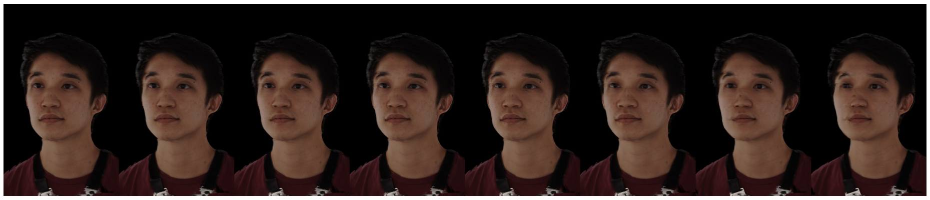

Fig. D.2. Keypoint perturbation via different noise levels (from left to right: 1mm, 2mm, 3mm, 4mm, 5mm, 10mm, and 20mm). The rendered images tend to become blurry around the keypoints (e.g. eyes) for large noise levels (> 10mm).

C Baseline Methods

We used the publicly released code of MVSNeRF [13] and IBRNet [62] with their default parameters. We re-implemented PVA [48] since their code is not public and we directly used the public results of NHP [29] for the experiments on the ZJU-MoCap dataset [45].

D Additional Results

Multi-view studio Capture Results. We further provide qualitative results for two more baseline methods (MVSNeRF [13] and PVA [48]) for the experimental setup described in Sec. 6.1. The results in Fig. B.1 demonstrate that the best performing baseline (IBRNet) produces incomplete images with lots of blur and foggy artifacts. PVA yields consistent, but overly smoothed renderings, while MVSNeRF does not work well for the widely spread-out input views. For more qualitative results, we refer the reader to the supplementary video.

Keypoint perturbation. To evaluate the sensitivity of our method on a less accurate estimation of keypoints, we perturb them with different Gaussian noise levels (ranging from 1 to 20mm) for unseen subjects from Sec. 6.1 and observe that the rendered images (Fig. D.2) occasionally tend to become blurry around the keypoints (e.g. eyes) for large noise levels (> 10mm).

The impact of the iPhone calibration for the in-the-wild capture. We evaluate the robustness of KeypointNeRF to a nosier camera calibration by estimating the iPhone camera parameters without the depth term for the experimental setup presented in Sec. 6.2. We observe (Tab. D.1) a negligible drop (PSNR/SSIM by -0.04/-0.5) in performance for our method, demonstrating the robustness of our method under noisy camera calibration.

Convolutional feature encoders. We further measure the impact of the HourGlass feature extractor and compare it with the U-Net encoder that is used by the other baseline methods [48, 62]. We follow the experimental setup from subsections 6.1 and 6.2 and report quantitative results in Tab. D.2 and D.3 respectively. We observe that HourGlass encoder consistently improves the reconstruction quality.

Table D.1. In-the-wild Captures. Quantitative comparison of IBRNet [62], our method without any spatial encoding, and our method with the proposed keypoint encoding; visual results are provided in Fig. 4 for the iPhone calibration with the depth term

	RGB calibration		RGB-D calibration	
	SSIM↑	PSNR↑	SSIM↑	PSNR↑
IBRNet [62]	81.72	18.41	81.74	18.45
Ours (no keypoints)	79.36	19.85	79.50	19.79
KeypointNeRF	**86.22**	**25.25**	**86.73**	**25.29**

Table D.2. Studio Capture Results. HourGlass [43] vs U-Net [48, 62] encoder for the experiment conducted in Sec. 6.1.

	SSIM↑	PSNR↑
PVA [48]	81.95	25.87
IBRNet [62]	82.39	27.14
KeypointNeRF (w. U-Net encoder [48, 62])	84.34	26.23
KeypointNeRF (w. HourGlass encoder [43])	**85.19**	**27.64**

Table D.3. In-the-wild Captures. HourGlass [43] vs U-Net [48, 62] encoder for the experiment conducted in Sec. 6.2

	SSIM↑	PSNR↑
IBRNet [62]	81.72	18.41
KeypointNeRF (w. U-Net encoder [48, 62])	84.20	**25.67**
KeypointNeRF (w. HourGlass encoder [43])	**86.22**	25.25

E Limitations and Future Work

While our method offers an efficient way of reconstructing volumetric avatars from as few as two input images, it still has several difficulties. The image-based rendering formulation of our method parametrizes the color prediction as blending of available pixels, which ensures good color generalization at inference time, however it makes the method sensitive to occlusions. The method itself has also difficulties reconstructing challenging thin geometries (e.g. glasses) and is less robust to highly articulated human motions (see Fig. E.3). As future work we consider addressing these challenges and additionally integrating learnable 3D lifting methods [20,24] with the proposed relative spatial encoding for more optimal end-to-end network training.

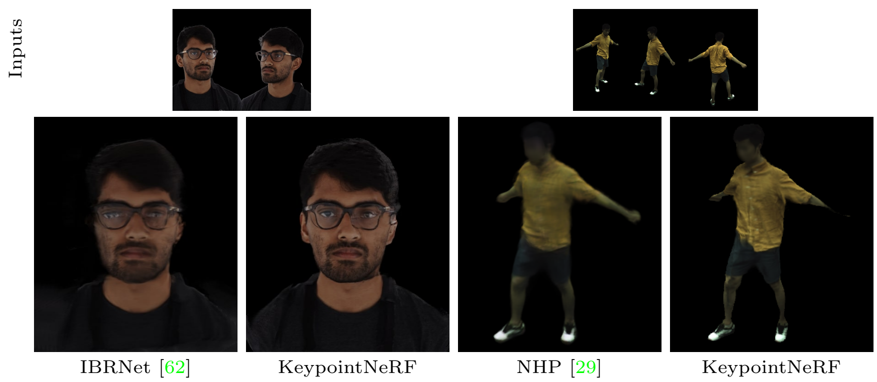

Fig. E.3. Limitations. Our method struggles to reconstruct the thin frames of the glasses (left) and has difficulties reconstructing human articulations that are outside of the training distribution.

References

1. Alldieck, T., Magnor, M., Xu, W., Theobalt, C., Pons-Moll, G.: Video based reconstruction of 3d people models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018) 3
2. Alldieck, T., Xu, H., Sminchisescu, C.: imghum: Implicit generative models of 3d human shape and articulated pose. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021) 4
3. Alldieck, T., Zanfir, M., Sminchisescu, C.: Photorealistic monocular 3d reconstruction of humans wearing clothing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1506–1515 (2022) 8
4. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: Scape: shape completion and animation of people. In: ACM SIGGRAPH 2005 Papers, pp. 408–416 (2005) 4
5. Athar, S., Shu, Z., Samaras, D.: Flame-in-nerf : Neural control of radiance fields for free view face animation. arXiv preprint arXiv:2108.04913 (2021) 4
6. Bansal, A., Chen, X., Russell, B., Gupta, A., Ramanan, D.: Pixelnet: Representation of the pixels, by the pixels, and for the pixels. arXiv preprint arXiv:1702.06506 (2017) 2
7. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3d faces. In: Proceedings of the 26th annual conference on Computer graphics and interactive techniques. pp. 187–194 (1999) 1
8. Buehler, M.C., Meka, A., Li, G., Beeler, T., Hilliges, O.: Varitex: Variational neural face textures. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2021) 4
9. Cao, C., Simon, T., Kim, J.K., Schwartz, G., Zollhoefer, M., Saito, S., Lombardi, S., en Wei, S., Belko, D., i Yu, S., Sheikh, Y., Saragih, J.: Authentic volumetric avatars from a phone scan. ACM Transactions on Graphics (TOG) (2022) 4, 11
10. Cao, C., Wu, H., Weng, Y., Shao, T., Zhou, K.: Real-time facial animation with image-based dynamic avatars. ACM Transactions on Graphics **35**(4) (2016) 3
11. Cao, Z., Simon, T., Wei, S.E., Sheikh, Y.: Realtime multi-person 2d pose estimation using part affinity fields. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7291–7299 (2017) 2, 3, 6
12. Chatziagapi, A., Athar, S., Moreno-Noguer, F., Samaras, D.: Sider: Single-image neural optimization for facial geometric detail recovery. arXiv preprint arXiv:2108.05465 (2021) 4
13. Chen, A., Xu, Z., Zhao, F., Zhang, X., Xiang, F., Yu, J., Su, H.: Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2021) 4, 2, 3
14. Chibane, J., Bansal, A., Lazova, V., Pons-Moll, G.: Stereo radiance fields (srf): Learning view synthesis from sparse views of novel scenes. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (jun 2021) 4
15. Gafni, G., Thies, J., Zollhofer, M., Niessner, M.: Dynamic neural radiance fields for monocular 4d facial avatar reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021) 2, 4
16. Gao, C., Shih, Y., Lai, W.S., Liang, C.K., Huang, J.B.: Portrait neural radiance fields from a single image. arXiv preprint arXiv:2012.05903 (2020) 2, 4
17. Grassal, P.W., Prinzler, M., Leistner, T., Rother, C., Nießner, M., Thies, J.: Neural head avatars from monocular rgb videos. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18653–18664 (2022) 4

18. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge university press (2003) 3, 6
19. He, T., Xu, Y., Saito, S., Soatto, S., Tung, T.: Arch++: Animation-ready clothed human reconstruction revisited. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2021) 2, 4
20. He, Y., Yan, R., Fragkiadaki, K., Yu, S.I.: Epipolar transformers. In: Proceedings of the ieee/cvf conference on computer vision and pattern recognition. pp. 7779–7788 (2020) 5
21. Hu, L., Saito, S., Wei, L., Nagano, K., Seo, J., Fursund, J., Sadeghi, I., Sun, C., Chen, Y.C., Li, H.: Avatar digitization from a single image for real-time rendering. ACM Transactions on Graphics (ToG) **36**(6), 1–14 (2017) 3
22. Huang, Z., Xu, Y., Lassner, C., Li, H., Tung, T.: Arch: Animatable reconstruction of clothed humans. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020) 2, 4
23. Ichim, A.E., Bouaziz, S., Pauly, M.: Dynamic 3d avatar creation from hand-held video input. ACM Transactions on Graphics (ToG) **34**(4), 1–14 (2015) 3
24. Iskakov, K., Burkov, E., Lempitsky, V., Malkov, Y.: Learnable triangulation of human pose. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7718–7727 (2019) 5
25. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision. pp. 694–711. Springer (2016) 7, 1
26. Ke, Z., Sun, J., Li, K., Yan, Q., Lau, R.W.: Modnet: Real-time trimap-free portrait matting via objective decomposition. In: AAAI (2022) 8
27. Kim, H., Garrido, P., Tewari, A., Xu, W., Thies, J., Nießner, M., Pérez, P., Richardt, C., Zollöfer, M., Theobalt, C.: Deep video portraits. ACM Transactions on Graphics (TOG) **37**(4), 163 (2018) 4
28. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Proceedings of International Conference on Learning Representations (2015) 8
29. Kwon, Y., Kim, D., Ceylan, D., Fuchs, H.: Neural human performer: Learning generalizable radiance fields for human performance rendering. In: Advances in Neural Information Processing Systems (NeurIPS). Curran Associates, Inc. (2021) 2, 3, 4, 5, 8, 9, 13, 14
30. Lombardi, S., Saragih, J., Simon, T., Sheikh, Y.: Deep appearance models for face rendering. ACM Transactions on Graphics (TOG) **37**(4), 1–13 (2018) 4
31. Lombardi, S., Simon, T., Saragih, J., Schwartz, G., Lehrmann, A., Sheikh, Y.: Neural volumes: Learning dynamic renderable volumes from images. ACM Transactions on Graphics (TOG) (2019) 1, 4
32. Lombardi, S., Simon, T., Schwartz, G., Zollhoefer, M., Sheikh, Y., Saragih, J.: Mixture of volumetric primitives for efficient neural rendering. arXiv preprint arXiv:2103.01954 (2021) 4
33. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015) 2
34. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: Smpl: A skinned multi-person linear model. ACM transactions on graphics (TOG) **34**(6), 1–16 (2015) 1, 4, 13
35. Martin-Brualla, R., Pandey, R., Yang, S., Pidlypenskyi, P., Taylor, J., Valentin, J., Khamis, S., Davidson, P., Tkach, A., Lincoln, P., et al.: Lookingood: Enhancing performance capture with real-time neural re-rendering. arXiv preprint arXiv:1811.05029 (2018) 4

36. Matusik, W., Buehler, C., Raskar, R., Gortler, S.J., McMillan, L.: Image-based visual hulls. In: ACM SIGGRAPH. pp. 369–374 (2000) 1
37. Meka, A., Pandey, R., Haene, C., Orts-Escolano, S., Barnum, P., David-Son, P., Erickson, D., Zhang, Y., Taylor, J., Bouaziz, S., et al.: Deep relightable textures: volumetric performance capture with neural rendering. ACM Transactions on Graphics (TOG) **39**(6), 1–21 (2020) 4
38. Mihajlovic, M., Saito, S., Bansal, A., Zollhoefer, M., Tang, S.: COAP: Compositional articulated occupancy of people. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Jun 2022) 4
39. Mihajlovic, M., Weder, S., Pollefeys, M., Oswald, M.R.: Deepsurfels: Learning online appearance fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14524–14535 (2021) 4
40. Mihajlovic, M., Zhang, Y., Black, M.J., Tang, S.: LEAP: Learning articulated occupancy of people. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021) 4
41. Mildenhall, B., Srinivasan, P.P., Ortiz-Cayon, R., Kalantari, N.K., Ramamoorthi, R., Ng, R., Kar, A.: Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. ACM Transactions on Graphics (TOG) (2019) 6
42. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: European conference on computer vision. pp. 405–421. Springer (2020) 1, 4, 5, 6, 8
43. Newell, A., Yang, K., Deng, J.: Stacked hourglass networks for human pose estimation. In: European conference on computer vision. pp. 483–499. Springer (2016) 2, 7, 1, 4
44. Peng, S., Dong, J., Wang, Q., Zhang, S., Shuai, Q., Zhou, X., Bao, H.: Animatable neural radiance fields for modeling dynamic human bodies. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14314–14323 (2021) 4
45. Peng, S., Zhang, Y., Xu, Y., Wang, Q., Shuai, Q., Bao, H., Zhou, X.: Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021) 2, 4, 8, 9, 13, 14, 3
46. Prokudin, S., Black, M.J., Romero, J.: Smplpix: Neural avatars from 3d human models. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1810–1819 (2021) 4
47. Raj, A., Tanke, J., Hays, J., Vo, M., Stoll, C., Lassner, C.: Anr: Articulated neural rendering for virtual avatars. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3722–3731 (2021) 4
48. Raj, A., Zollhoefer, M., Simon, T., Saragih, J., Saito, S., Hays, J., Lombardi, S.: Pva: Pixel-aligned volumetric avatars. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021) 2, 4, 5, 9, 11, 13, 14, 3
49. Ramon, E., Triginer, G., Escur, J., Pumarola, A., Garcia, J., i Nieto, X.G., Moreno-Noguer, F.: H3d-net: Few-shot high-fidelity 3d head reconstruction. arXiv preprint arXiv:2107.12512 (2021) 4
50. Rebain, D., Matthews, M., Yi, K.M., Lagun, D., Tagliasacchi, A.: Lolnerf: Learn from one look. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1558–1567 (2022) 4

51. Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., Li, H.: Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2019) 2, 3, 4, 5, 9
52. Saito, S., Simon, T., Saragih, J., Joo, H.: Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020) 3, 4, 5, 9
53. Schwarz, K., Liao, Y., Niemeyer, M., Geiger, A.: Graf: Generative radiance fields for 3d-aware image synthesis. In: Advances in Neural Information Processing Systems (NeurIPS). Curran Associates, Inc. (2020) 8
54. Shao, R., Zhang, H., Zhang, H., Cao, Y., Yu, T., Liu, Y.: Doublefield: Bridging the neural surface and radiance fields for high-fidelity human rendering. arXiv preprint arXiv:2106.03798 (2021) 7, 1
55. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014) 8
56. Tewari, A., Fried, O., Thies, J., Sitzmann, V., Lombardi, S., Sunkavalli, K., Martin-Brualla, R., Simon, T., Saragih, J., Nießner, M., et al.: State of the art on neural rendering. In: Computer Graphics Forum. vol. 39, pp. 701–727. Wiley Online Library (2020) 3
57. Tewari, A., Thies, J., Mildenhall, B., Srinivasan, P., Tretschk, E., Wang, Y., Lassner, C., Sitzmann, V., Martin-Brualla, R., Lombardi, S., et al.: Advances in neural rendering. arXiv preprint arXiv:2111.05849 (2021) 3
58. Torralba, A., Efros, A.A.: Unbiased look at dataset bias. In: CVPR 2011. pp. 1521–1528. IEEE (2011) 3
59. Vlasic, D., Baran, I., Matusik, W., Popović, J.: Articulated mesh animation from multi-view silhouettes. ACM Transactions on Graphics **27**(3), 97 (2008) 1
60. Vlasic, D., Peers, P., Baran, I., Debevec, P., Popović, J., Rusinkiewicz, S., Matusik, W.: Dynamic shape capture using multi-view photometric stereo. ACM Transactions on Graphics **28**(5), 174 (2009) 1
61. Wang, L., Chen, Z., Yu, T., Ma, C., Li, L., Liu, Y.: Faceverse: a fine-grained and detail-controllable 3d face morphable model from a hybrid dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20333–20342 (2022) 4
62. Wang, Q., Wang, Z., Genova, K., Srinivasan, P.P., Zhou, H., Barron, J.T., Martin-Brualla, R., Snavely, N., Funkhouser, T.: Ibrnet: Learning multi-view image-based rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4690–4699 (2021) 4, 5, 7, 8, 9, 10, 11, 12, 13, 1, 2, 3
63. Wang, S., Mihajlovic, M., Ma, Q., Geiger, A., Tang, S.: Metaavatar: Learning animatable clothed human models from few depth images. Advances in Neural Information Processing Systems **34** (2021) 4
64. Wang, S., Schwartz, K., Geiger, A., Tang, S.: ARAH: Animatable volume rendering of articulated human sdfs. In: European conference on computer vision (2022) 4
65. Wang, Z., Bagautdinov, T., Lombardi, S., Simon, T., Saragih, J., Hodgins, J., Zollhofer, M.: Learning compositional radiance fields of dynamic human heads. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021) 4
66. Weng, C.Y., Curless, B., Srinivasan, P.P., Barron, J.T., Kemelmacher-Shlizerman, I.: Humannerf: Free-viewpoint rendering of moving people from monocular video. arXiv preprint arXiv:2201.04127 (2022) 4, 8

67. Xu, H., Alldieck, T., Sminchisescu, C.: H-nerf: Neural radiance fields for rendering and temporal reconstruction of humans in motion. Advances in Neural Information Processing Systems **34** (2021) 4
68. Xu, H., Bazavan, E.G., Zanfir, A., Freeman, W.T., Sukthankar, R., Sminchisescu, C.: Ghum & ghuml: Generative 3d human shape and articulated pose models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6184–6193 (2020) 4
69. Yu, A., Ye, V., Tancik, M., Kanazawa, A.: pixelnerf: Neural radiance fields from one or few images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4578–4587 (2021) 2, 3, 4, 5, 13, 14
70. Yu, T., Zheng, Z., Guo, K., Zhao, J., Dai, Q., Li, H., Pons-Moll, G., Liu, Y.: Doublefusion: Real-time capture of human performances with inner body shapes from a single depth sensor. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7287–7296 (2018) 1
71. Zhao, H., Zhang, J., Lai, Y.K., Zheng, Z., Xie, Y., Liu, Y., Li, K.: High-fidelity human avatars from a single rgb camera. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15904–15913 (2022) 4
72. Zheng, Y., Abrevaya, V.F., Bühler, M.C., Chen, X., Black, M.J., Hilliges, O.: I M Avatar: Implicit morphable head avatars from videos. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Jun 2022) 4
73. Zheng, Z., Huang, H., Yu, T., Zhang, H., Guo, Y., Liu, Y.: Structured local radiance fields for human avatar modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15893–15903 (2022) 4
74. Zheng, Z., Yu, T., Liu, Y., Dai, Q.: Pamir: Parametric model-conditioned implicit representation for image-based human reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence (2021) 4