Supplementary Material
Codec Avatar Studio: Paired Human Captures for Complete, Driveable, and Generalizable Avatars

Julieta Martinez[1], **Emily Kim**[2], **Javier Romero**[1],
Timur Bagautdinov[1], **Shunsuke Saito**[1], **Shoou-I Yu**[1], **Stuart Anderson**[1], **Michael Zollhöfer**[1],
Te-Li Wang[1], **Shaojie Bai**[1], **Shih-En Wei**[1], **Rohan Joshi**[1], **Wyatt Borsos**[1],
Tomas Simon[1], **Jason Saragih**[1], **Paul Theodosis**[1], **Alexander Greene**[1], **Anjani Josyula**[1],
Silvio Mano Maeta[1], **Andrew I. Jewett**[1], **Simon Venshtain**[1], **Christopher Heilman**[1],
Yueh-Tung Chen[1], **Sidi Fu**[1], **Mohamed Ezzeldin A. Elshaer**[1], **Tingfang Du**[1], **Longhua Wu**[1],
Shen-Chi Chen[1], **Kai Kang**[1], **Michael Wu**[1], **Youssef Emad**[1], **Steven Longay**[1], **Ashley Brewer**[1],
Hitesh Shah[1], **James Booth**[1], **Taylor Koska**[1], **Kayla Haidle**[1], **Matt Andromalos**[1],
Joanna Hsu[1], **Thomas Dauer**[1], **Peter Selednik**[1], **Tim Godisart**[1], **Scott Ardisson**[1],
Matthew Cipperly[1], **Ben Humberston**[1], **Lon Farr**[1], **Bob Hansen**[1], **Peihong Guo**[1],
Dave Braun[1], **Steven Krenn**[1], **He Wen**[1], **Lucas Evans**[1], **Natalia Fadeeva**[1], **Matthew Stewart**[1],
Gabriel Schwartz[1], **Divam Gupta**[1], **Gyeongsik Moon**[1], **Kaiwen Guo**[1], **Yuan Dong**[1],
Yichen Xu[1], **Takaaki Shiratori**[1], **Fabian Prada**[1], **Bernardo R. Pires**[1], **Bo Peng**[1],
Julia Buffalini[1], **Autumn Trimble**[1], **Kevyn McPhail**[1], **Melissa Schoeller**[1],
Yaser Sheikh[1]

[1]Codec Avatars Team, Meta Reality Labs Research
[2] Robotics Institute, Carnegie Mellon University

In this Supplementary Material we provide more details on our dataset and open source releases. We start with Ava-256, and continue with Goliath-4.

1 Ava-256

As promised in the main submission, we provide a comprehensive list of previous datasets compared with the Ava-256 in Table 1. We include smaller scale datasets first—those with fewer than 60 ids or less than 20 views, then larger-scale datasets, finally ours for comparison.

Preprint. Under review.

Table 1: Summary of multi-view face datasets. Smaller scale dataset (fewer than 60 ids or 20 views) as shown first, then large-scale datasets, then ours.

Dataset name	IDs	Views	Sentences + expressions	Images	Resolution	Size	Headset captures
D3DFACS[6] [2]	10	6	19-97	-	1 280 × 1 024	-	✗
CMU Multi-PIE [4]	337	15	6	750K	3 072 × 2 048	0.3 TB	✗
4DFAB[6][1]	180	7	15	-	1 600 × 1 200	-	✗
MEAD [8]	48	7	(8)	-	1 920 × 1 080	0.9 TB	✗
Interdigital Light-Field [7]	5	16	-	-	2 048 × 1 088	0.1 TB	✗
Multiface [9]	13	40/150	168	15M	2 048 × 1 334	65.0 TB	✗
NerSemble [5]	222	16	25	31.7M	3 208 × 2 200	1.0 TB	✗
HUMBI Face [12]	617[1]	68	20	17.3M	200 × 150[5]	1.3 TB	✗
Facescape [10]	359[2]	68	20	400K	4 344 × 2 896	0.9 TB	✗
i3DMM[4][11]	64	137	10	-	-	-	✗
RenderMe-360 [6]	500[3]	60	37/54	243M	2 448 × 2 048	5.8 TB	✗
Ava-256 Dome Captures (Ours)	256	80	35	217 M	2 048 × 1 334	32.0 TB	✓

[1] We only found the data of 403 subjects available online. We have contacted the authors to clarify this discrepancy.
[2] Out of 847 total subjects captured, only 359 subjects are available with multi-view image data.
[3] 500 captures are expected to be released later this year, but only 21 subjects are currently available online. 5.8 TB is the size of this smaller subset containing about 4.2 % of the data, so we expect the final release size to be well over 100 TB.
[4] We requested access to the dataset on May 20th but have not heard back at the time of submission, and are thus unable to obtain more data about this dataset.
[5] The head images are cropped from the full-body 1 920 × 1 080 image, and are thus of low resolution.
[6] Dataset was not available on the website.

Dataset documentation and intended uses. We provide a dataset datasheet in the format proposed by Gebru et al. [3] at `https://github.com/facebookresearch/ava-256/blob/main/DATASHEET.md`.

Dataset and code URL. The URL for this dataset and associated code is `https://github.com/facebookresearch/ava-256`. The `README.md` file has instructions on how to download the data. We provide a simple python script (`download.py`) to download the entire dataset. The script can be customized to download only certain assets, or a subset of the captures.

URL to Croissant metadata record. We have chosen not to provide a Croissant metadata record for our dataset.

Author statement. The authors confirm that they bear responsibility in case of violation of rights. The authors also confirm that the code and data are released under the CC-by-NC 4.0 License.

Hosting, licensing and maintenance plan. Our dataset is hosted in a public S3 bucket in AWS. Please refer to `download.py`, our script which downloads all our data. The data, code and model weights are available under the CC-by-NC 4.0 License. Maintenance is done by the Codec Avatars Team at Reality Labs Research within Meta.

2 Goliath-4

Dataset documentation and intended uses. We provide a dataset datasheet in the format proposed by Gebru et al. [3] at `https://github.com/facebookresearch/goliath/blob/main/DATASHEET.md`.

Dataset and code URL. The URL for this dataset and associated code is `https://github.com/facebookresearch/goliath`.

URL to Croissant. We have chosen not to provide a Croissant metadata record for our dataset.

Author statement. The authors confirm that they bear responsibility in case of violation of rights. The authors also confirm that the code and data are released under the CC-by-NC 4.0 License.

Hosting, licensing and maintenance plan. Due to the sensitive nature of the minimally-clothed captures, this dataset is hosted at the moment as zips protected by password, and users should email javierromero1@meta and julietamartinez@meta.com to obtain access to the password.

We have uploaded the links to the full 4 TB dataset at https://fb-baas-f32eacb9-8abb-11eb-b2b8-4857dd089e15.s3.amazonaws.com/goliath-for-review-full/index.html.

To ease the job of reviewers, we provide a small sample with 100 frames of each capture, which can be used to run the baselines that we provide. The size of each small sample is less than 5 GB, and the combined size of all samples is only 39 GB. The links to download the smaller samples are available at https://fb-baas-f32eacb9-8abb-11eb-b2b8-4857dd089e15.s3.amazonaws.com/goliath-for-review-100frames/index.html.

The password for all the zip files is RLR-2015-2024.

The data, code and model weights are available under the CC-by-NC 4.0 License. Maintenance is done by the Codec Avatars Team at Reality Labs Research within Meta.

References

[1] S. Cheng, I. Kotsia, M. Pantic, and S. Zafeiriou. 4DFAB: A large scale 4D database for facial expression analysis and biometric applications. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2018.

[2] D. Cosker, E. Krumhuber, and A. Hilton. A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2296–2303, 11 2011. doi: 10.1109/ICCV.2011.6126510.

[3] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. D. Iii, and K. Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, 2021.

[4] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. In *Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–8, 2008. doi: 10.1109/AFGR.2008.4813399.

[5] T. Kirschstein, S. Qian, S. Giebenhain, T. Walter, and M. Nießner. NeRSemble: Multi-view radiance field reconstruction of human heads. *ACM Transactions on Graphics*, 42(4), jul 2023. ISSN 0730-0301. doi: 10.1145/3592455. URL https://doi.org/10.1145/3592455.

[6] D. Pan, L. Zhuo, J. Piao, H. Luo, W. Cheng, Y. Wang, S. Fan, S. Liu, L. Yang, B. Dai, Z. Liu, C. C. Loy, C. Qian, W. Wu, D. Lin, and K.-Y. Lin. RenderMe-360: Large digital asset library and benchmark towards high-fidelity head avatars. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023.

[7] N. Sabater, G. Boisson, B. Vandame, P. Kerbiriou, F. Babon, M. Hog, T. Langlois, R. Gendrot, O. Bureller, A. Schubert, and V. Allie. Dataset and pipeline for multi-view light-field video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2017.

[8] K. Wang, Q. Wu, L. Song, Z. Yang, W. Wu, C. Qian, R. He, Y. Qiao, and C. C. Loy. Mead: A large-scale audio-visual dataset for emotional talking-face generation. In *Proceedings of the European Conference on Computer Vision*, 2020.

[9] C.-h. Wuu, N. Zheng, S. Ardisson, R. Bali, D. Belko, E. Brockmeyer, L. Evans, T. Godisart, H. Ha, X. Huang, A. Hypes, T. Koska, S. Krenn, S. Lombardi, X. Luo, K. McPhail, L. Millerschoen, M. Perdoch, M. Pitts, A. Richard, J. Saragih, J. Saragih, T. Shiratori, T. Simon, M. Stewart, A. Trimble, X. Weng, D. Whitewolf, C. Wu, S.-I. Yu, and Y. Sheikh. Multiface: A Dataset for Neural Face Rendering. In *arXiv*, 2022. doi: 10.48550/ARXIV.2207.11243. URL https://arxiv.org/abs/2207.11243.

[10] H. Yang, H. Zhu, Y. Wang, M. Huang, Q. Shen, R. Yang, and X. Cao. FaceScape: A large-scale high quality 3D face dataset and detailed riggable 3D face prediction. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2020.

[11] T. Yenamandra, A. Tewari, F. Bernard, H. Seidel, M. Elgharib, D. Cremers, and C. Theobalt. i3DMM: Deep implicit 3D morphable model of human heads. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2021.

[12] Z. Yu, J. S. Yoon, I. K. Lee, P. Venkatesh, J. Park, J. Yu, and H. S. Park. HUMBI: A large multiview dataset of human body expressions. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2020.