

GPAvatar: High-fidelity Head Avatars by Learning Efficient Gaussian Projections

Supplementary Material

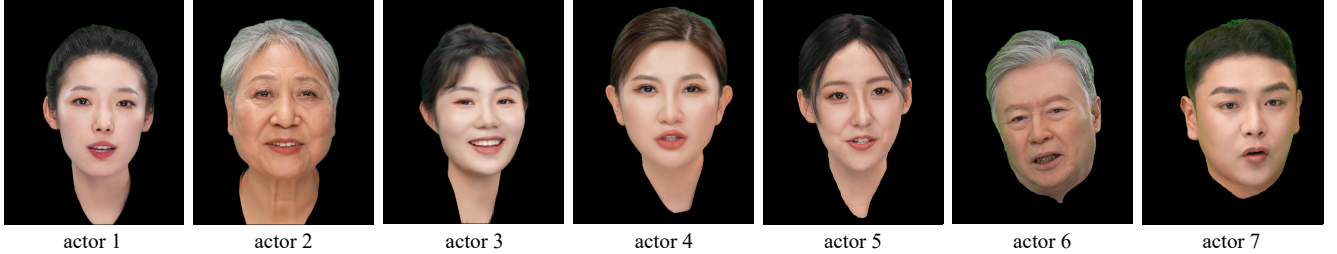


Figure 1. Our collected dataset consists of video data derived from complete segments of streamers’ regular live streams, encompassing seven roles that include male, female, and elderly participants. We present the modeling results for each role utilizing the GPAvatar framework. The dataset captures intricate details such as facial expressions, wrinkles, and distinct facial features observed during live streaming sessions, enhancing its representativeness for real-world applications.

A. Dataset Details

To validate the effectiveness of the algorithm, in addition to conducting experiments on the public datasets INSTA [6] and GBS [1], we also performed evaluations on our self-collected dataset. Our dataset was collected from complete video segments of streamers during their regular live broadcasts, making it more representative of real-world applications. The dataset includes 7 actors, with videos at a resolution of 1024×1024 . For each actor, there are 10,000 frames in the training set and 750 frames in the test set. As shown in 1, we present the modeling results of each actor using the GPAvatar. GPAvatar is capable of modeling facial details, which plays a crucial role in practical applications.

B. Ablation Studies

Table 1. Ablation of perceptual loss (abbreviated as *percep.*) strategy on GBS [2] dataset.

Method	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	Training
<i>percep.</i> every 5 iters	33.75	0.9546	0.0760	2.5h
<i>percep.</i> every 1 iter	33.75	0.9541	0.0725	6.0h
w/o <i>percep.</i>	33.81	0.9553	0.1171	1.5h

We apply perceptual loss [3] (abbreviated as *percep.*) **every 5 iterations** on average during training, resulting in a $2.4\times$ speedup of the training process. Here, we further discuss the effectiveness of this strategy. Tab. 1 and Fig. 2 present a comparison for different configurations: applying perceptual loss every 5 iterations, every 1 iteration, and without perceptual loss supervision. The results show that our strategy effectively balances training time while preserving facial details.

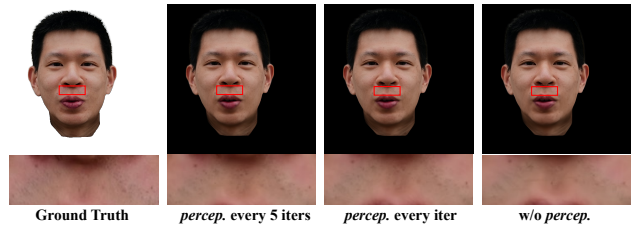


Figure 2. The perceptual loss contributes to preserving fine-detailed facial features of the head avatar. Our “every 5 iterations” strategy shows no significant visual differences compared to the “every 1 iteration” approach.

C. Additional Results

In Fig. 3, we present the photometric error for frames from the main paper, comparing our method to FlashAvatar [4], GaussianBlendShapes [2], INSTA [6], and PointAvatar [5] using RGB-based l_1 distance. The heatmaps make it immediately apparent that our method achieves lower average error, particularly in challenging areas such as hair, eyes, and teeth. This demonstrates our method’s superior ability to accurately capture fine details compared to the other methods.

D. Ethical Consideration

Our work may have potential ethical risks, such as deepfakes and privacy concerns. We acknowledge these risks and emphasize the importance of responsible use of our technology. We encourage researchers and practitioners to consider the ethical implications of their work and to develop safeguards to mitigate potential harm.

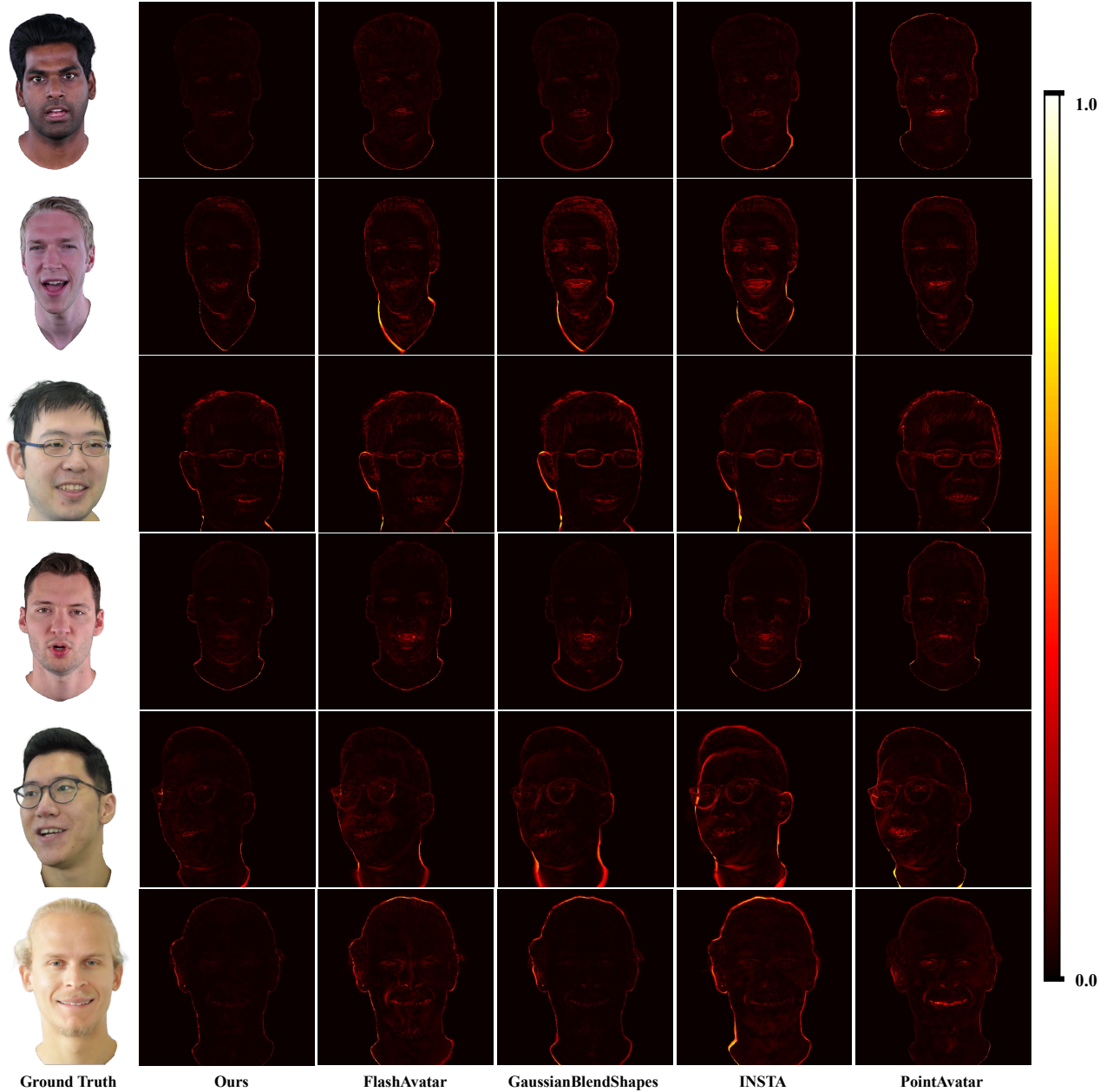


Figure 3. The heatmaps depict photometric errors on the test sequences, based on the l_1 RGB distance.

References

- [1] Xuan Gao, Chenglai Zhong, Jun Xiang, Yang Hong, Yudong Guo, and Juyong Zhang. Reconstructing personalized semantic facial nerf models from monocular video. *ACM Transactions on Graphics (TOG)*, 41(6):1–12, 2022. [1](#)
- [2] Shengjie Ma, Yanlin Weng, Tianjia Shao, and Kun Zhou. 3d gaussian blendshapes for head avatar animation. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–10, 2024. [1](#)
- [3] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. [1](#)
- [4] Jun Xiang, Xuan Gao, Yudong Guo, and Juyong Zhang. Flashavatar: High-fidelity head avatar with efficient gaussian embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1802–1812, 2024. [1](#)
- [5] Yufeng Zheng, Wang Yifan, Gordon Wetzstein, Michael J Black, and Otmar Hilliges. Pointavatar: Deformable point-based head avatars from videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21057–21067, 2023. [1](#)

- [6] Wojciech Zielonka, Timo Bolkart, and Justus Thies. Instant volumetric head avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4574–4584, 2023. [1](#)