FlashAvatar: High-fidelity Head Avatar with Efficient Gaussian Embedding

Supplementary Material

A. Additional Ablations and Results

A.1. Additional Ablations

Mouth closure. Since the original FLAME mesh does not model the interior mouth, we add additional faces to close the mouth cavity and find it helpful in modeling the interior mouth. As seen in Fig. 12, if we merely rely on Gaussians in nearby areas like the lips to model the interior mouth, the upper and lower teeth tend to stick together, which leads to blurry results, especially for challenging cases.

Figure 12. Closing the mouth cavity of FLAME mesh with additional faces is useful for modeling the interior mouth like teeth.

Perceptual loss. Besides the pixel-based loss, we adopt the perceptual loss as well. Fig. 13 shows the comparison between the results with/without perceptual loss supervision. As we can see, the perceptual loss helps maintain personalized facial attributes and greatly boosts photo-realism.

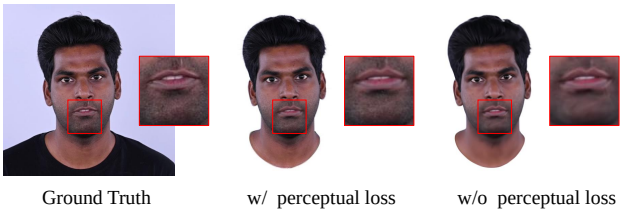

Figure 13. The perceptual loss helps maintain fine-detailed facial attributes of the head avatar.

A.2. Additional Results

Limitation. Our method still relies on a good surface-embedded Gaussian initialization and cannot handle large errors in tracking (see Fig. 14).

B. Implementation Details

B.1. Network Architecture

We show the architecture of the offset network in Fig. 15.

Figure 14. Large errors in tracking lead to wrong results.

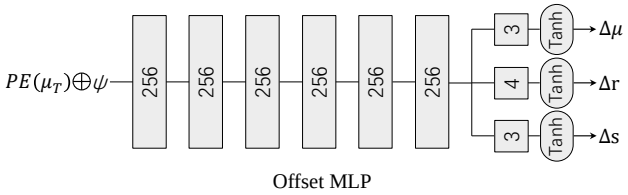

Figure 15. Network architecture of the offset MLP. Except for the last layer, each linear layer is followed by the ReLU activation.

Figure 16. The blue region corresponds to the boundary of the FLAME mesh, which is excluded when sampling Gaussians.

B.2. FLAME Masks

As we only model head regions with neck, we sample Gaussians in the corresponding areas, and we conduct this by adding a flame mask excluding the boundary of FLAME mesh (see Fig. 16).

C. Broader Impact

Our work could reconstruct a digital avatar from a monocular video in minutes and animate it at 300FPS while achieving photo-realistic rendering with full personalized details. This takes an important step towards practical applications of multimodal digital humans, as it provides more space for other interactive tasks to enable real-time interaction. However, there is a risk of misuse, *e.g.* the so-called DeepFakes.

We strongly discourage using our work to generate fake images or videos of individuals with the intent of spreading false information or damaging their reputations. Unfortunately, we may be unable to prevent the nefarious use of our technology. Nevertheless, we believe that performing research in an open and transparent way could raise the public's awareness of nefarious uses, and our work could further enhance forgery detection capabilities.