SUPPLEMENTARY MATERIALS FOR "ANIMAKE-3D: ANIMATABLE MAKEUP TRANSFER FOR 3D GAUSSIAN HEAD AVATAR"

Anonymous Authors

7. ADDITIONAL RESULTS

Makeup Transfer We present additional results that demonstrate the robustness and superiority of AniMake-3D in makeup transfer. As shown in Fig. 2,3, AniMake-3D excels in identity preservation, even under variations in expression and extreme poses, due to its explicit extraction of the makeup layer. Our framework effectively transfers detailed makeup styles from simple to complex looks while maintaining robustness.

Novel View Synthesis Fig. 1 presents novel views of a makeup-transferred 3D Gaussian head avatar across various viewing angles, demonstrating high consistency across different poses and expressions as it is built upon 3D Gaussian Splatting (3DGS). Our framework offers flexibility in conditioning, allowing the use of either a reference image or a text prompt, making it applicable to a wider range of use cases.

8. PRELIMINARY

3D Gaussian Splatting 3DGS [1] is a point-based volume rendering technique that represents each primitive as a Gaussian kernel, defined as:

$$G(x) = e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}, \tag{1}$$

where μ denotes the Gaussian's position and Σ is the 3D covariance matrix. To ensure Σ remains positive semi-definite, it is factorized into a rotation matrix R and a scaling matrix S as follows:

$$\Sigma = RSS^TR^T. \tag{2}$$

During rendering, 3D Gaussians are projected onto the image plane as 2D Gaussians. The covariance of the projected 2D Gaussians using:

$$\Sigma' = JW\Sigma J^TW^T, \tag{3}$$

where W represents the viewing transformation and J is the Jacobian of the affine approximation of the projective transformation. The final pixel color is then computed through volumetric rendering as:

$$C = \sum_{i\in N}\left(\alpha_i\prod_{j<i}(1-\alpha_j)\right)c_i, \tag{4}$$

where c_i is the color of each Gaussian and α_i is the density determined by the projected 2D Gaussian's covariance Σ' and the Gaussian's opacity o_i.

9. ADDITIONAL DETAILS

Temporal Covariance metric When makeup components shift in position over time on a face, it appears perceptually unnatural. To quantify this effect, we first extract pixel locations corresponding to semantically meaningful facial regions using landmark estimation with SPIGA [2]. Next, we compute the pixel-wise gradient at each landmark. We then calculate the covariance between the gradient values of the method's output and those of the source video to measure their correlation. The formulation is given as follows:

$$\mathrm{Cov}\left(\left.\frac{\partial C_O}{\partial t}\right|_{t-1}^{T}, \left.\frac{\partial C_M}{\partial t}\right|_{t-1}^{T}\right)$$

, where C_O denotes the pixel color of the source image, C_M represents the pixel color of the method's output, T is the number of frames, and Cov denotes covariance.

Since our framework optimizes makeup layer features within a unified UV space representation, the makeup components remain robust to changes in pose and expression. As a result, our framework achieves a higher temporal covariance metric score compared to other image-to-image makeup transfer methods during rendering.

Loss Function The overall loss function for Stage 1 is defined as:

$$\mathcal{L}_{\text{stage1}} = \mathcal{L}_{\text{rgb}} + \lambda_1\mathcal{L}_{\text{vgg}} + \lambda_2\mathcal{L}_{\text{lap}} + \lambda_3\mathcal{L}_{\text{offset}} + \lambda_4\mathcal{L}_{\text{scale}}. \tag{5}$$

The Laplacian loss $\mathcal{L}_{\text{lap}}$ promotes surface smoothness by minimizing the difference between each vertex and the mean of its neighboring vertices. It is defined as:

$$\mathcal{L}_{\text{lap}} = \sum_i \left\| v_i - \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} v_j \right\|_2^2, \tag{6}$$

where v_i represents the position of vertex i, while $\mathcal{N}(i)$ denotes the set of neighboring vertices connected to v_i. The term inside the norm measures the difference between the vertex and the mean of its neighboring vertices. The squared Euclidean norm $\| \cdot \|_2^2$ enforces smoothness by penalizing large deviations from the local mean. This loss encourages each vertex to move toward the centroid of its neighboring vertices, leading to a smoother mesh surface.

$\mathcal{L}_{\text{offset}}$ represents the L1 loss of the offset parameter d_t in 3D Gaussians. $\mathcal{L}_{\text{scale}}$, adopted from [3], prevents the excessive skinning of Gaussians. It is defined as:

$$\mathcal{L}_{\text{scale}} = \frac{1}{N} \sum_{i=0}^{N-1} \max\left(\frac{\max(s_i)}{\min(s_i)} - r, 0 \right). \tag{7}$$

10. ETHICAL ISSUES

Our method performs makeup style transfer from a reference image, enabling re-animation and novel view synthesis. We strongly oppose any misuse of this work for generating deceptive content that spreads misinformation or harms reputations.

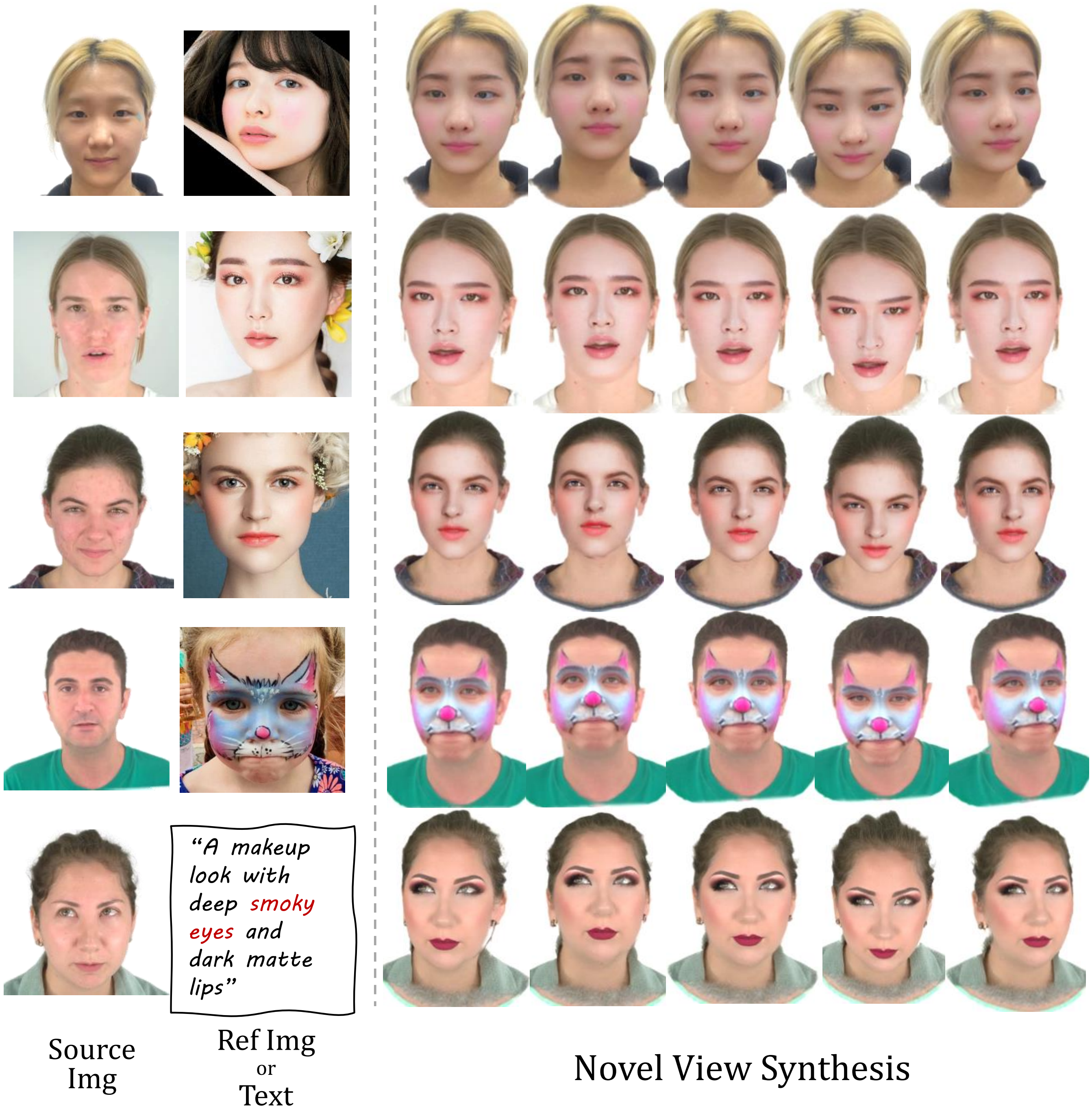

Fig. 1. Makeup transferred 3D Gaussian Avatar in novel view.

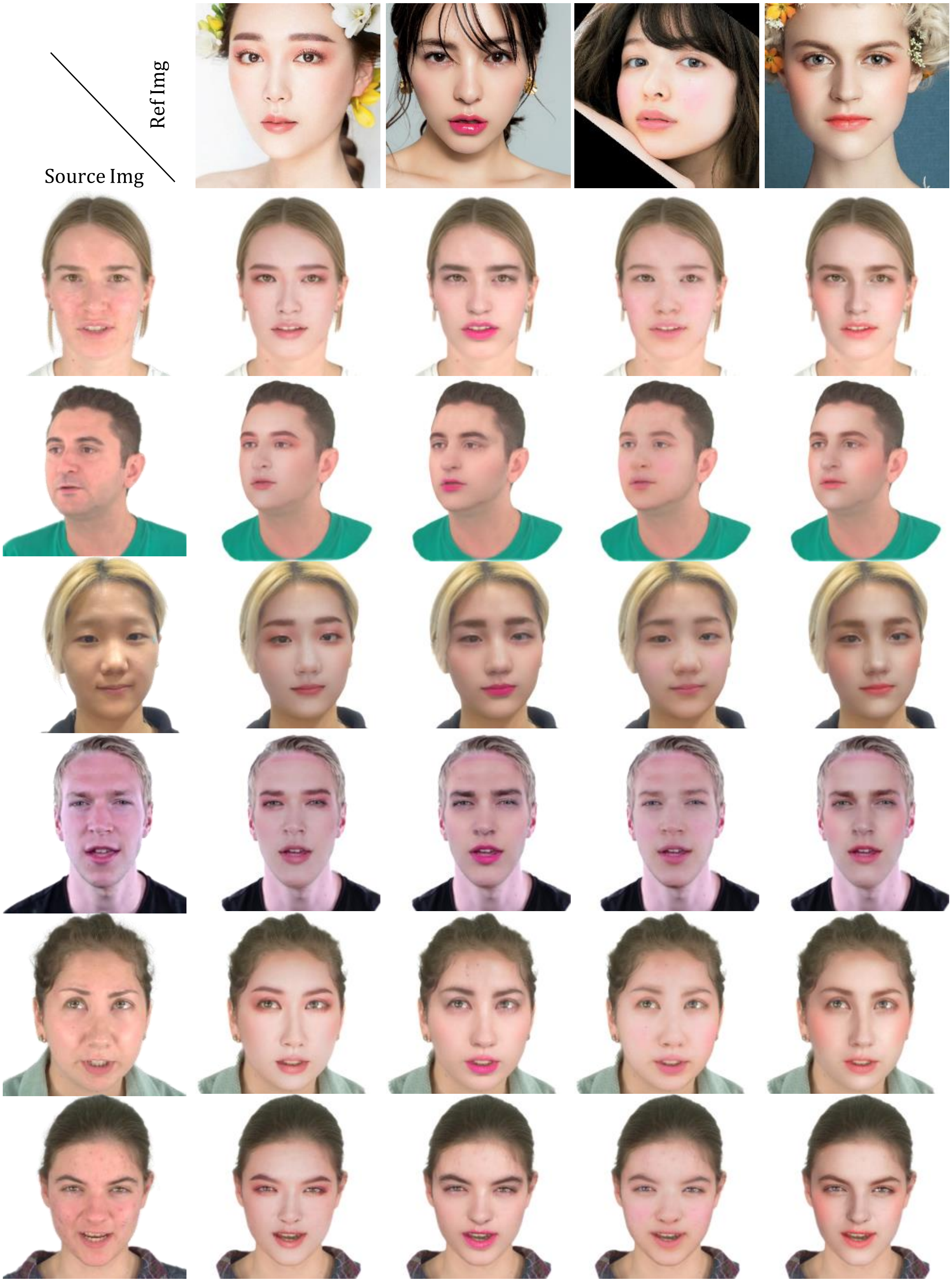

Fig. 2. More results of AniMake-3D in the makeup transfer task.

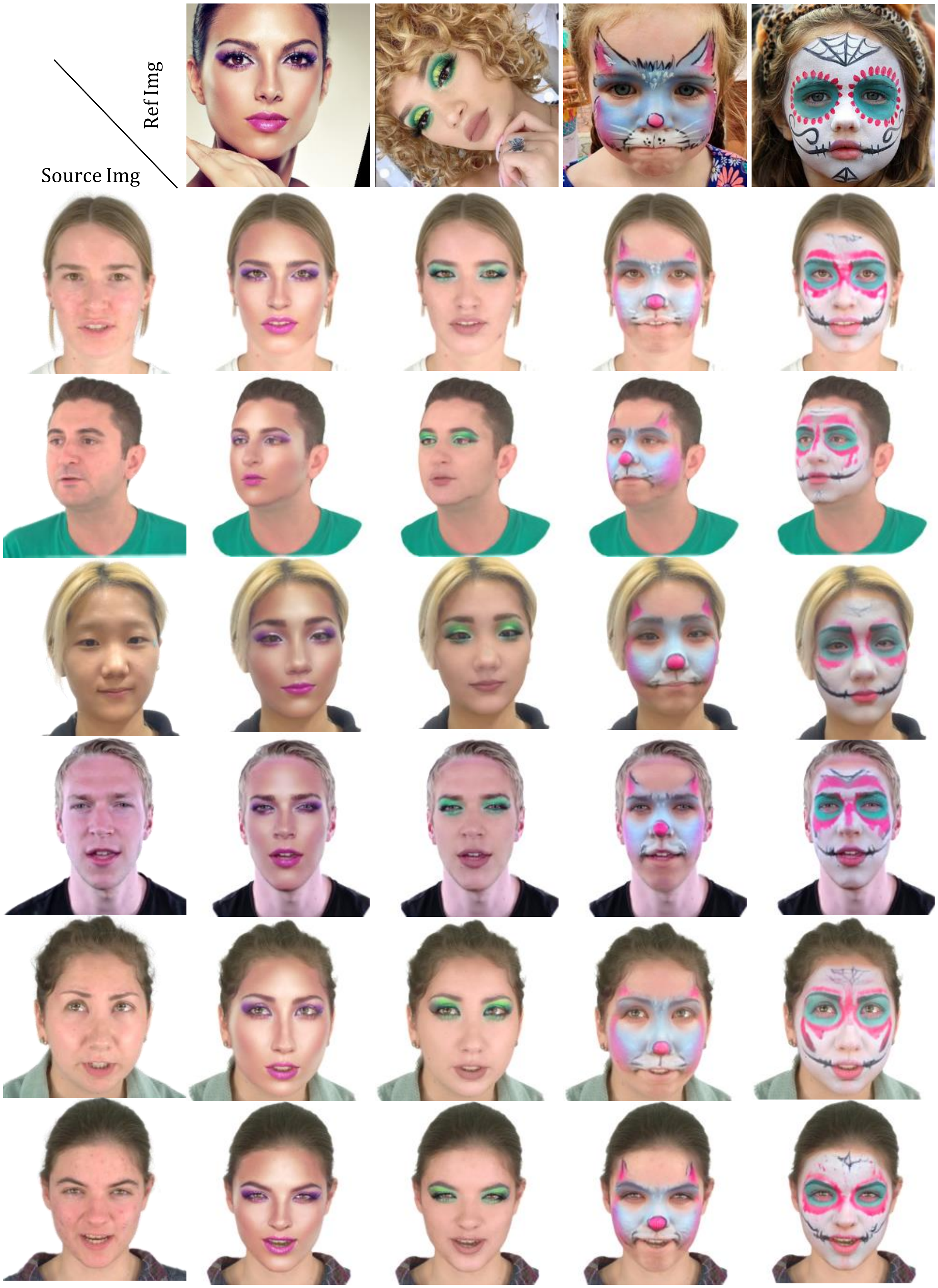

Fig. 3. More results of AniMake-3D in the makeup transfer task.

11. REFERENCES

[1] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis, “3d gaussian splatting for real-time radiance field rendering.,” *ACM Trans. Graph.*, vol. 42, no. 4, pp. 139–1, 2023.

[2] Andrés Prados-Torreblanca, José M Buenaposada, and Luis Baumela, “Shape preserving facial landmarks with graph attention networks,” *arXiv preprint arXiv:2210.07233*, 2022.

[3] Tianyi Xie, Zeshun Zong, Yuxing Qiu, Xuan Li, Yutao Feng, Yin Yang, and Chenfanfu Jiang, “Physgaussian: Physics-integrated 3d gaussians for generative dynamics,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 4389–4398.