CaricatureGS: Exaggerating 3D Gaussian Splatting Faces with Gaussian Curvature

Supplementary Material

8. Implementation considerations

Unless stated otherwise, we optimize each subject's 3D Gaussian Splatting model for 120,000 iterations, adhering to SurFhead's training protocol and evaluation split [21]. All experiments are run on a single NVIDIA RTX 3090 (24 GB VRAM). The optimization time per subject is ≈ 4 hours (this is offline training time, not rendering runtime.)

We used the NeRSemble dataset [34] with 10 subjects, 4 emotions (EMO), and 6 expressions (EXP). Expression `EXP2` is held for testing and Camera 8 serves as the validation view during training.

Caricaturization is performed once at the beginning of the training by solving the unconstrained Poisson equation, deforming the FLAME base template with $\gamma = 0.25$ ($\approx$ 1 min).

Because FLAME uses a shared template across subjects, the deformed surface is saved and reused for all subjects. Unless stated otherwise, we report metrics over 256 frames from the rendered test sequence, aggregated across all camera viewpoints.

CLIP configuration. For text–image alignment, we use OpenAI CLIP with the `ViT-B/32` backbone and the library's default preprocessing.

Prompts are: **Source:** "A realistic neutral head with natural lighting." **Edit:** "A photorealistic caricature of a head with a highly exaggerated nose and large ears, under natural lighting."

Defaults inherited. The optimizer, learning rate schedule, degree of spherical harmonics, and Gaussian growth/pruning follow the SurFhead [21] configuration unless otherwise specified.

9. Linear Model and Error Analysis

Notation. Let $S(u, v)$ be a parametric surface, where $(u, v) \in \mathbb{R}^2$, with a metric G and K denotes the Gaussian curvature at each point of the surface S, and

$$w(\gamma) \;=\; |K|^{\gamma} \;=\; e^{\gamma L}, \qquad L \;\equiv\; \ln|K|. \tag{11}$$

For $\gamma \in [0, \gamma_f]$, denote by S_γ the solution of the weighted Poisson problem with Dirichlet boundary condition x^* on ∂S.

To avoid degeneracies at $K = 0$, we use ϵ to stabilize the magnitude. Note, for convenience we refer to as $|K|_\epsilon = \sqrt{K^2 + \epsilon^2}$ with fixed $\epsilon > 0$. For brevity we write $|K|$ to denote this stabilized quantities.

1) Poisson equation with secant weights. The original family is defined by

$$\Delta_G S_\gamma \;=\; \nabla_G \cdot \big(w(\gamma)\, \nabla_G S\big). \tag{12}$$

Note, that S_0 and S_{γ_f} refer to $\gamma = 0$ and $\gamma = \gamma_f$, respectively. Define the vertex blend,

$$S_{\text{blend}}(\gamma) \;=\; (1-\alpha)\, S_0 \;+\; \alpha\, S_{\gamma_f}, \quad \alpha \;\equiv\; \frac{\gamma}{\gamma_f} \tag{13}$$

By linearity of Δ_G and Equation (13)

$$\begin{aligned} \Delta_G S_{\text{blend}}(\gamma) \;&=\; (1-\alpha)\, \Delta_G S_0 \;+\; \alpha\, \Delta_G S_{\gamma_f} \\ &=\; \nabla_G \cdot \Big(w_{\text{sec}}(\gamma)\, \nabla_G S\Big), \end{aligned} \tag{14}$$

where *secant weight* is

$$w_{\text{sec}}(\gamma) = 1 + \frac{\gamma}{\gamma_f}\big(|K|^{\gamma_f} - 1\big). \tag{15}$$

Thus $S_{\text{blend}}(\gamma)$ solves the exact Poisson equation at level γ with $w(\gamma)$ replaced by $w_{\text{sec}}(\gamma)$, and $S_{\text{interp}}|_{\partial S} = x^*$ (see (4) for x^*).

2) Remainder and properties The secant w_{sec} is the linear interpolant of w in $[0, \gamma_f]$. By the classical interpolation remainder for C^2 functions on a closed interval (e.g., [4, Thm. 3.1], [1, §3.3]), for every $\gamma \in [0, \gamma_f]$ there exists $\xi(\gamma) \in (0, \gamma_f)$ such that

$$w_{\text{sec}}(\gamma) - w(\gamma) \;=\; \frac{w''(\xi)}{2}\, \gamma(\gamma_f - \gamma). \tag{16}$$

Since $w''(\gamma) = L^2 e^{\gamma L}$, we get

$$w_{\text{sec}}(\gamma) - w(\gamma) \;=\; \frac{L^2}{2}\, e^{\xi L}\, \gamma(\gamma_f - \gamma). \tag{17}$$

The secant model is exact at both endpoints (where $\alpha = 0$ and $\alpha = 1$, yielding a analytic expression in $[0, \gamma_f]$ preserving the convexity-induced non-negativity.
Since $w'' \geq 0$, $\gamma \mapsto w(\gamma)$ is convex, hence $w_{\text{sec}} - w$ is non-negative on $[0, \gamma_f]$ and vanishes at the endpoints. In particular, at $\gamma = \gamma_f/2$,

$$\big|w_{\text{sec}}(\tfrac{\gamma_f}{2}) - w(\tfrac{\gamma_f}{2})\big| \;\leq\; \frac{\gamma_f^2}{8}\, L^2\, \max\big(1, e^{\gamma_f L}\,\big). \tag{18}$$

The maximum of this *upper bound* occurs at $\gamma_f/2$ because $\gamma(\gamma_f - \gamma)$ is maximized there.

3) Poincaré and Lax–Milgram for residual bound.

Throughout, we approximate the γ–dependent weight $w(\gamma) = |K|^\gamma$ by its secant $w_{\text{sec}}(\gamma)$ to enable a cheap vertex blend instead of solving a new Poisson problem for each γ. To justify this alternative, we should *quantify* how the weight error propagates to a *geometric residual* $\delta S(\gamma) \equiv S(\gamma) - S_{\text{blend}}(\gamma)$. The goal here is to derive a norm bound on δS that depends only on: (i) ellipticity and Poincaré constants of the domain, (ii) the magnitude of $\nabla_G S_0$, and (iii) the scalar secant remainder from Appendix Sec. 9. This yields a mesh and metric agnostic error budget for the blend.

Setting (frozen operator). Let (S, G) be a compact Riemannian surface with Lipschitz boundary ∂S. We impose Dirichlet conditions $u\big|_{\partial S} = 0$.

We fix the differential operators on the surface S, namely, the gradient and the divergence w.r.t metric G.

Let $V \equiv H_0^1(S)$ and define

$$\begin{aligned} a(u, v) &= \int_S \langle \nabla_G u, \nabla_G v \rangle_G \, dA_G \\ \|u\|_V &\equiv \|\nabla_G u\|_{L^2(S)}. \end{aligned} \tag{19}$$

We also define the *dual norm* by

$$\|F\|_{V'} \equiv \sup_{v \in V \setminus \{0\}} \frac{|F(v)|}{\|v\|_V}. \tag{20}$$

Using *Poincaré inequality*, there exists $C_P > 0$ such that, for all $u \in H_0^1(S)$,

$$\|u\|_{L^2(S)} \le C_P \, \|\nabla_G u\|_{L^2(S)} = C_P \, \|u\|_V. \tag{21}$$

Hence $\|u\|_V$ is a true norm on $H_0^1(S)$ and is equivalent to the standard H^1-norm on $H_0^1(S)$.
By Cauchy–Schwarz,

$$\begin{aligned} |a(u, v)| &\le \|u\|_V \, \|v\|_V \quad \text{(boundedness)}, \\ a(v, v) &= \|v\|_V^2 \qquad \text{(coercivity with } \alpha = 1) \end{aligned} \tag{22}$$

where coercivity means that there exists $\alpha > 0$ such that

$$a(v, v) \ge \alpha \, \|v\|_V^2 \quad \forall \, v \in V.$$

Lax–Milgram. If a is bounded and coercive on the Hilbert space V and $F \in V'$ is bounded, then, there exists a unique solution $u \in V$, solving $a(u, v) = F(v)$ for all $v \in V$, with estimate

$$\|u\|_V \le \frac{1}{\alpha} \, \|F\|_{V'} \overset{(22)}{=} \|F\|_{V'}. \tag{23}$$

For each γ, we solve the weighted Poisson PDE given by

$$\Delta_G S_\gamma = \nabla_{\mathrm{G}}\big(w(\gamma) \, \nabla_G S\big), \qquad S_\gamma\big|_{\partial S} = x^*. \tag{24}$$

Let $S_{\text{blend}}(\gamma) = (1 - \alpha) S_0 + \alpha S_{\gamma_f}$ with $\alpha = \gamma/\gamma_f$, and define

$$\begin{aligned} \psi(\gamma) &\equiv w_{\text{sec}}(\gamma) - w(\gamma) \\ \mathcal{R}_\Delta(\gamma) &\equiv \nabla_{\mathrm{G}}\big(\psi \, \nabla_G S\big). \end{aligned} \tag{25}$$

Define $F \in V'$ (weak residual functional) by

$$\begin{aligned} F(v) &= \langle \mathcal{R}_\Delta, v \rangle \\ &= \int_S \big(\nabla_{\mathrm{G}}(\psi \, \nabla_G S)\big) \, v \, dA_G \\ &= -\int_S \psi \, \langle \nabla_G S, \nabla_G v \rangle_G \, dA_G, \end{aligned} \tag{26}$$

with $v\big|_{\partial S} = 0$.

Using the dual norm and by Cauchy–Schwarz and $\|\psi\|_{L^\infty}$-bound, we readily have

$$\begin{aligned} |F(v)| &\le \|\psi\|_{L^\infty(S)} \, \|\nabla_G S\|_{L^2(S)} \, \|\nabla_G v\|_{L^2(S)} \\ &= \|\psi\|_{L^\infty} \, \|\nabla_G S\|_{L^2(S)} \, \|v\|_V, \end{aligned} \tag{27}$$

and using (20) we get

$$\|F\|_{V'} \le \|\psi\|_{L^\infty} \, \|\nabla_G S\|_{L^2(S)}. \tag{28}$$

Let $\delta S \equiv S_{\text{blend}} - S_\gamma$. Subtract the weak forms for S_{blend} and S_γ to obtain

$$\begin{aligned} a(\delta S, v) &= a(S_{\text{blend}}, v) - a(S_\gamma, v) \\ &= \int_S w_{\text{sec}} \, \langle \nabla_G S, \nabla_G v \rangle_G \, dA_G \\ &\quad - \int_S w(\gamma) \, \langle \nabla_G S, \nabla_G v \rangle_G \, dA_G \\ &= \int_S \psi \, \langle \nabla_G S, \nabla_G v \rangle_G \, dA_G \\ &= -\int_S \nabla_{\mathrm{G}}\big(\psi \, \nabla_G S\big) \, v \, dA_G \quad (*) \\ &\equiv -F(v). \end{aligned} \tag{29}$$

Where in (*) we use integration by parts and Dirichlet boundary conditions on ∂S.

Testing with $v = \delta S$ and using coercivity and duality,

$$\begin{aligned} \|\delta S\|_V^2 &= a(\delta S, \delta S) \\ &= -F(\delta S) \le \|F\|_{V'} \, \|\delta S\|_V \\ \Rightarrow \|\delta S\|_V &\le \|F\|_{V'}. \end{aligned} \tag{30}$$

Combining with the bound on $\|F\|_{V'}$ yields the *energy* estimate

$$\|\delta S\|_V \le \|\psi\|_{L^\infty(S)} \, \|\nabla_G S\|_{L^2(S)}$$

$$\|\delta S\|_V \le \|w_{\text{sec}} - w\|_{L^\infty} \, \|\nabla_G S\|_{L^2(S)}. \tag{31}$$

Optional L^2 bound. By Poincaré on $H^1_0(S)$,

$$\begin{aligned} \|\delta S\|_{L^2(S)} &\leq C_P \, \|\delta S\|_V \\ &\leq C_P \, \|w_{\text{sec}} - w\|_{L^\infty} \, \|\nabla_G S\|_{L^2(S)}. \end{aligned} \tag{32}$$

In summary, the secant error bound yields the energy bound for the residual δS by

$$\begin{aligned} \|\delta S(\gamma)\|_{L^2} \lesssim \; & C_P (\ln \|K\|)^2 e^{\max(0, \gamma_f \ln \|K\|)} \\ & \times \gamma(\gamma_f - \gamma) \, \|\nabla_G S\|_{L^2(S)}. \end{aligned} \tag{33}$$

which depends on geometric constants of the domain (C_P). The curvature in (33) is evaluated at its global maximum

$$\|K\| = K_\infty = \max_{s \in S} |K(s)| \tag{34}$$

We note that $S_0 = S$ (for $\gamma = 0$ by definition since there is no deformation done to S), hence (33) can be written using either terms.

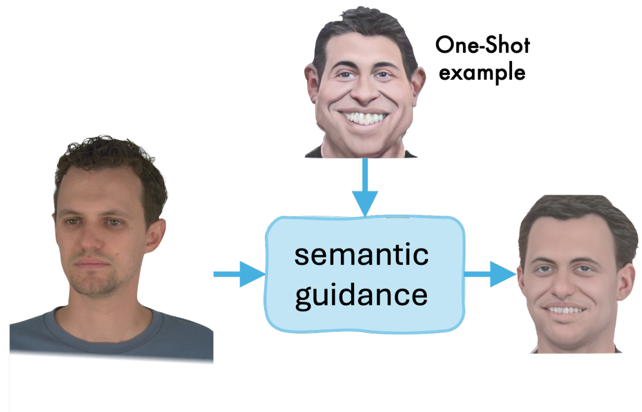
10. Caricature GT*via one-shot stylization

As discussed in Sec. 3, one-shot stylization methods (e.g., Deformable StyleGAN [46]) address the natural-caricature domain gap by aligning DINO features and adapting a pre-trained GAN to a single caricature exemplar. Given a target style image (Fig. 8a), they synthesize stylized outputs for arbitrary inputs. In practice, we observe pronounced identity–expression entanglement, which degrades both identity fidelity and expression accuracy (Fig. 8). Moreover, the outputs are not consistent across viewpoints or expressions: under view changes or when transferring expressions from the source, the method exhibits structural drift and a collapse toward the reference style (Figs. 8b and 8c), limiting its suitability for our 3DGS reconstruction setting.

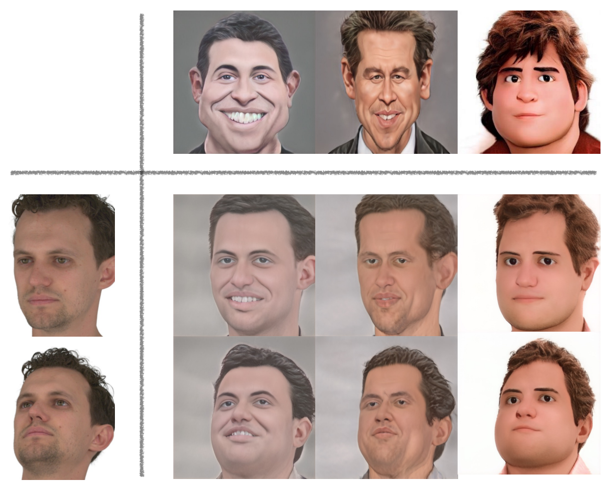
Protocol. We ran [46] using the official implementation, employing `Style1`, `Style2`, and `Style3` as target style exemplars and `EMO3`, `EMO4` for expression prompts.

11. Masking and GT*

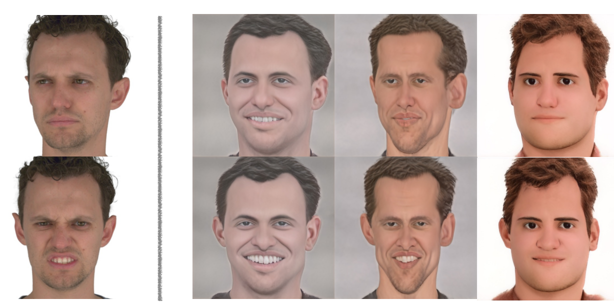
As noted in Sec. 3.2, GT* supervision is constructed by projecting the FLAME mesh, fitted to each original frame, onto the image. Consequently, the quality of GT* inherits any mesh–image misregistration. In practice, small fitting errors that are negligible at γ=0 are amplified as the caricature strength increases, with the most visible drift around delicate geometry such as the eyelids and eyeballs; see Fig. 9. In addition, the deformation can reveal triangles that were occluded in the original projection (e.g., along the eyelid crease), creating pixels with no reliable photometric support.



(a) Deformable StyleGAN [46]: stylization conditioned on a target style exemplar.



(b) View variation induces identity drift and structural artifacts (*e.g.* neck geometry).



(c) Expressions are not preserved, outputs bias toward the style exemplar (*e.g.* persistent smile, forward gaze).

Figure 8. Limitations of one-shot stylization for caricature. Identity–expression entanglement and lack of view/expression consistency hinder 3DGS supervision.

To prevent these failure modes, we build a visibility-aware GT* mask. We (i) suppress supervision on triangles that become newly visible at nonzero γ relative to the original projection, and (ii) mask anatomically fragile regions prone to amplified alignment error (eyelids, ear tips).

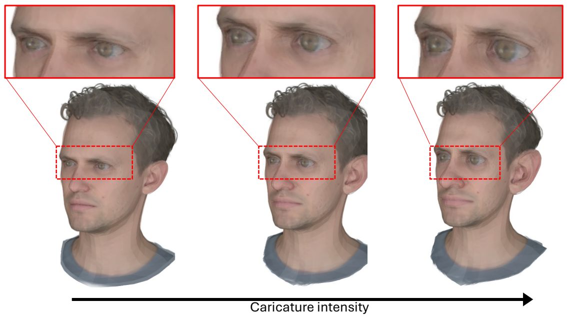


Figure 9. FLAME–image misregistration under increasing caricature strength γ. Projection drift concentrates on thin, high-curvature structures (eyelids/iris rim) and grows with γ, introducing erroneous supervision if used unfiltered.

This filtering removes inconsistent labels before they reach Gaussians anchored to those areas, yielding cleaner gradients and more stable appearance/geometry during training. The resulting GT* thus preserves the benefits of deformation-aware supervision while avoiding artifacts introduced by projection drift and occlusions.

2. put eyelids iris break and combine it with zoom on mesh eyeballs-related to small FLAME alignment errors

12. Ablation: Alternating Supervision

Setup. As motivated in Sec. 5.1, we seek a *single* 3DGS model that renders both the original avatar (γ=0) and its caricatured counterpart (γ=γ_f). We compare three training schedules using identical budgets: (i) *Original-only*: supervision from original frames only. (ii) *GT*-only*: supervision from caricatured (GT*) frames only. (iii) *Alternating (ours)*: alternating mini-batches from both sources. We set the target exaggeration to γ_f=0.25 and evaluate along the interpolation path $\gamma \in \{0, 0.10, 0.15, 0.20, 0.25\}$.

Findings. Original-only (i) fits the undeformed scene well but fails to generalize to caricatured geometry Fig. 10, yielding visible distortions under nonzero γ. Conversely, GT*-only (ii) represents the caricatured avatar but degrades markedly at γ=0. In addition, GT*-only exhibits systematic artifacts around hair and other structures that extend beyond the tracked mesh support (*e.g*. holes or under-coverage), because those pixels are never directly supervised in the warped domain, see Fig. 11.

Our alternate schedule (iii) maintains high fidelity at both endpoints and produces smooth interpolation across γ (see Fig. 12), avoiding the hair/occlusion failures seen in (ii). Practically, alternating acts as a simple multi-domain regularizer, as it preserves appearance outside the mesh support (from original frames) while learning the exaggerated geometry and view-dependent effects required by GT*.

Conclusions. Alternating supervision is necessary to obtain a *single* 3DGS that is faithful at γ=0 and γ=γ_f and stable along the interpolation path, while training on either domain alone leads to domain-specific overfitting and characteristic failure modes.

Figure 10. Training on original frames only

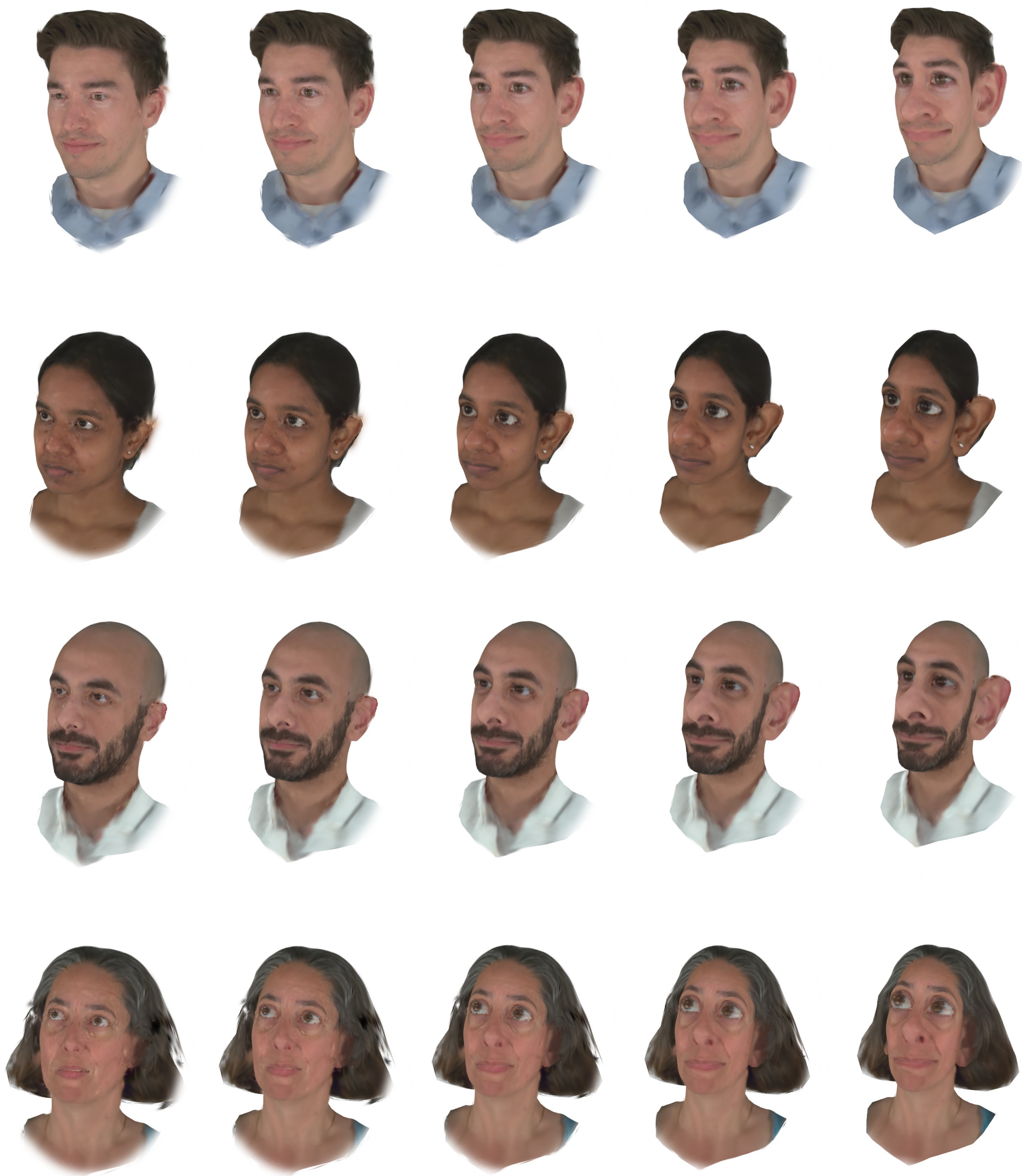


Figure 11. Training on GT* frames only.

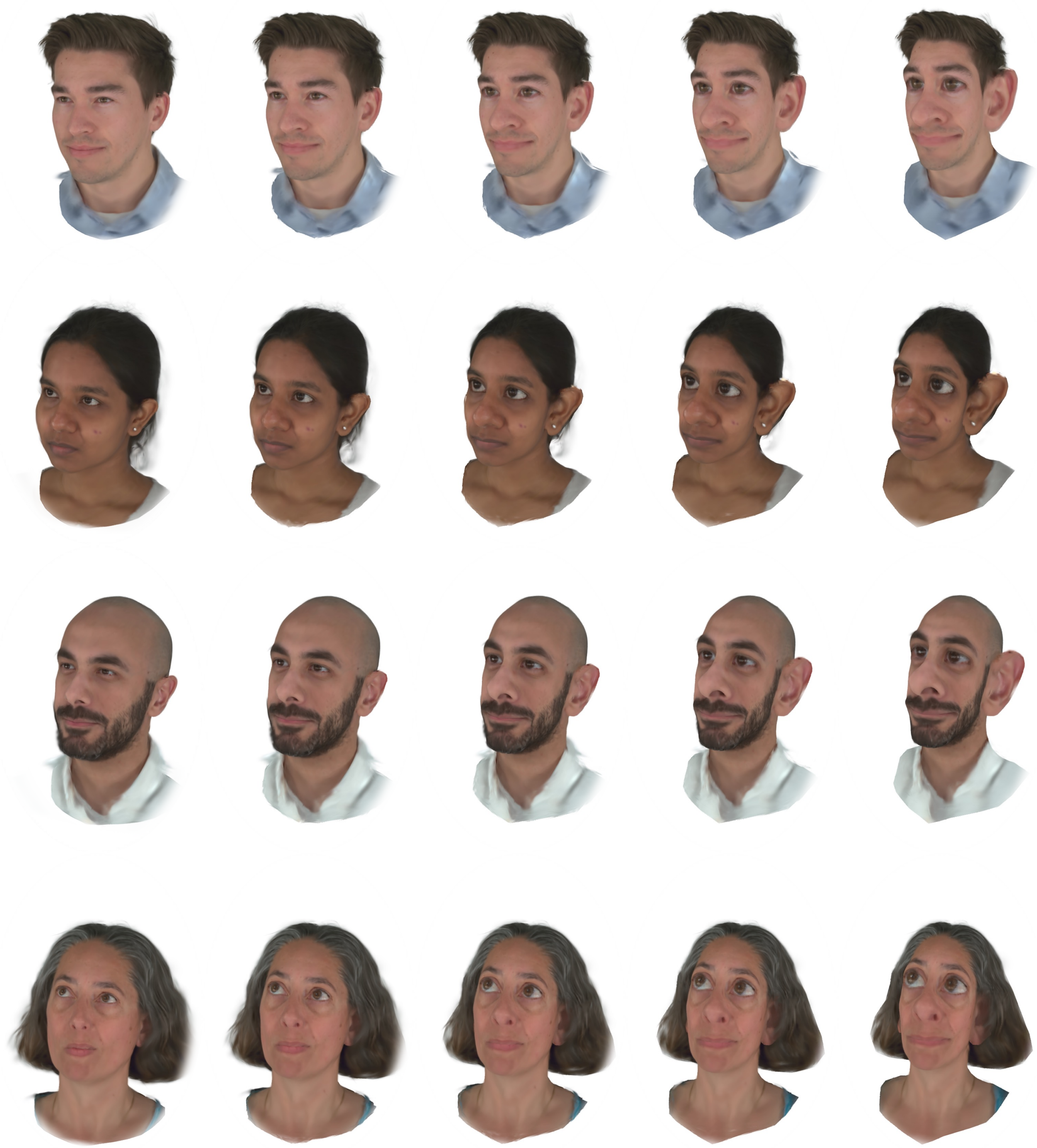


Figure 12. Training on both original and GT* frames interleaved