CHRONOMESH: TWO-STAGE AGE EDITING FOR 3D HEAD MESH

Anonymous authors
Paper under double-blind review

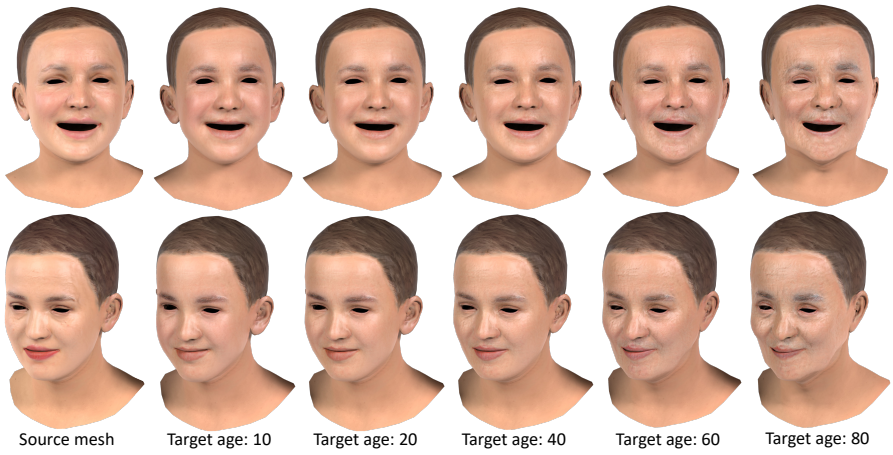

Figure 1: Our age editing results on 3D head meshes. Our method produces age-dependent macro-scale shape deformation and wrinkle details expressed through normal maps.

ABSTRACT

Mesh-based head assets are widely used in production, but many recent aging methods target volumetric 3D representations that do not translate cleanly to UV material maps. We propose *ChronoMesh*, a two-stage framework for age editing of existing 3D head meshes that updates both global head shape and wrinkle-normal details while preserving identity. Because facial aging is inherently multi-scale, jointly optimizing geometry and texture under image-based supervision tends to favor local shading cues, under-expressing macro-scale deformation. Our two-stage design mitigates this issue by separating shape deformation from material editing. In the first stage, we predict an age-conditioned residual in a parametric head model to drive shape deformation toward a target age. During this stage, age-dependent proxy normal maps provide wrinkle cues consistent with the target age. In the second stage, we synthesize age-conditioned edits for UV-space intrinsic maps using a pre-trained StyleGAN with inversion-based latent editing, including wrinkle-normal synthesis. Both stages use image-based supervision through differentiable rendering. Experiments show that our edits preserve identity better than both image-space editors and mesh-reconstruction baselines while moving the rendered appearance toward the target age, with wrinkles represented explicitly in the normal maps rather than baked into diffuse reflectance.

1 INTRODUCTION

Facial age editing is in high demand for reusing 3D character assets in games and films. In current production, creating age variants of the same character often requires costly and time-consuming processes such as prosthetic makeup followed by repeated 3D scanning, or extensive manual revisions by skilled artists. A practical alternative is to directly edit the perceived age of an existing head asset while preserving identity, so that creators can iterate faster, reuse the same asset across multiple scenes and time periods, and reduce the overhead of maintaining large libraries of age-specific

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

variants. This goal is particularly compelling when the editable representation remains compatible with the tools and formats already used in production.

Mesh-based pipelines remain the backbone of digital human production: a head is authored as a polygon mesh with UV-mapped textures and consumed by DCC tools and real-time engines, as reflected in standard asset formats (Khronos Group, 2021). Neural head representations (Chan et al., 2022; Gafni et al., 2021; Hong et al., 2022; Gerogiannis et al., 2025) achieve impressive realism and, in some cases, semantic editing, but deploying them in such a pipeline requires extracting a mesh and converting appearance into UV-space maps, which can introduce artifacts or break compatibility with downstream tools. We therefore target age editing that operates directly on mesh assets and their UV-space maps.

Facial aging is inherently multi-scale: it combines large head shape changes, such as volume loss and altered soft-tissue support, with the emergence of wrinkles and folds (Zoumalan & Larrabee, 2011). Manipulating only high-frequency detail, such as displacement or normal maps, therefore cannot reproduce the macro-scale deformation that contributes to an older appearance, and a mesh-compatible age editor must account for both macro-scale deformation and wrinkle-level detail.

To address these requirements, we introduce *ChronoMesh*, a two-stage mesh aging framework that explicitly decomposes age change into macro-scale mesh deformation and wrinkle-level normal details. We predict age-conditioned updates to the shape parameters of a parametric head model (FLAME) (Li et al., 2017) and represent wrinkles as normal-map variations on the mesh. Training uses differentiable rendering under illumination with an age-estimator-based loss. When shape and normal maps are learned jointly under this loss, the network tends to favor easier local shading changes, so the intended macro-scale deformation is under-expressed even when the age estimator is satisfied.

We therefore adopt a two-stage strategy. Since age cues can be confounded by illumination (Zoss et al., 2022), we pre-optimize a lighting setup that maximizes the measurable age shift and use it as fixed illumination for differentiable-rendering supervision. In Stage 1, we estimate an age-conditioned residual in FLAME shape space to learn macro-scale deformation (Li et al., 2017). Because age supervision still relies on wrinkle cues before predicting normal maps, we attach age-dependent proxy-normal maps during training. In Stage 2, we fix the deformation and learn UV-space intrinsic maps (diffuse reflectance, normal, and specular) via inversion-based latent editing in a pre-trained StyleGAN2 (Karras et al., 2020), enabling wrinkle normal synthesis with consistent map updates. To effectively drive wrinkle-normal learning, we introduce a normal-map age estimator and incorporate its prediction loss as auxiliary supervision.

Our technical contributions are summarized as follows:

- We introduce *ChronoMesh*, a two-stage mesh aging framework that separates macro-scale deformation from wrinkle-level normal synthesis, enabling age edits that remain compatible with UV-mapped head assets.
- We propose a training strategy that explains and mitigates optimization interference between deformation and map learning by (i) using age-dependent proxy-normal maps during the deformation stage and (ii) adding a normal-map age estimator as auxiliary supervision for wrinkle learning.
- We demonstrate, on assets from a public UV-intrinsics dataset (Dib et al., 2024), identity-preserving age edits toward specified target ages, with wrinkles in normal maps rather than baked into diffuse reflectance, supporting consistent relighting (Fig. 1).

We will release our source code and pre-trained models upon acceptance.

2 Related Work

2.1 2D Face Image Aging

Linear latent directions in StyleGAN (Shen et al., 2020; He et al., 2021) entangle age with other attributes, whereas SAM (Alaluf et al., 2021) predicts an age-conditioned latent offset after inversion, AgeTransGAN (Hsu et al., 2022) conditions a GAN on target age groups, Zoss et al. (2022)

learn image-space residuals for production video re-aging, and Gomez-Trenado et al. (2022) control the degree of facial-structure preservation. Diffusion-based editing (Preechakul et al., 2022; Kwon et al., 2023; Chen & Lathuilière, 2023; Li et al., 2023b; Liu et al., 2025b; Gong et al., 2025; Baumann et al., 2025) and personalized aging from several photos of the same person (Ito et al., 2025; Qi et al., 2025; Su et al., 2025) further improve realism, but all operate in image space and cannot be applied to a mesh asset and its UV maps.

2.2 3D Face Model Generation and Editing

Editing age in the shape space of a 3D morphable model dates back to Scherbaum et al. (2007), who learned non-linear aging trajectories of 3DMM coefficients from scans, and age-dependent skin detail has been synthesized from statistics of high-resolution scans (Golovinskiy et al., 2006); our Stage 1 revisits the former with an identity-conditioned residual under image-based supervision, and Stage 2 replaces hand-crafted detail statistics with learned UV-space normal synthesis. 3D-aware GANs (Chan et al., 2022) and their editing, quality, and animation extensions (Li et al., 2023a; Cheng et al., 2024; Kumar et al., 2024; Zhang et al., 2024a; Trevithick et al., 2024; Liu et al., 2025a) support attribute edits, including age (Kumar et al., 2024), with multi-view consistency, but their outputs are neural representations rather than mesh assets with UV maps.

Recent work explicitly addresses facial aging in 3D-aware settings: Teng et al. (2024) disentangle shape and texture factors through 3D reconstruction, Wahid et al. (2024) control pose and target age with a diffusion-based model for multi-view consistent aging, and Li et al. (2024) combine age-driven appearance transformation with age-dependent geometry updates via 3D uplifting. These approaches improve 3D consistency, but they operate on neural 3D representations or reconstructed proxies, so deploying the results still requires converting the output into a mesh asset and reconciling UV-space maps, which can introduce artifacts or break downstream compatibility. Concurrently, ReAge3D (Zeng et al., 2026) re-ages faces represented by 3D Gaussian splatting from multi-view light-stage captures by propagating diffusion-based 2D edits across views; its output remains a radiance-field representation, and geometric change appears only implicitly in normals extracted from the Gaussians.

A few methods address aging and editing for mesh-based 3D faces, but the supported asset components vary widely. MeshWGAN (Zhang et al., 2024b) performs mesh-to-mesh geometric age transformation from real 3D scans, but it focuses on geometry and does not target UV-space map editing. SemUV (Mukherjee et al., 2024) edits age and other attributes of FFHQ-UV diffuse textures along linear latent directions of a StyleGAN trained on UV maps, but leaves the geometry and the normal and specular maps untouched, Canfes et al. (2023) target text-driven avatar editing, and FaceG2E (Wu et al., 2024) edits the geometry and UV texture of text-generated face meshes but not normal or specular maps; none of them models wrinkle-level normal change coupled with macro-scale deformation. In contrast, our method edits existing head meshes while jointly addressing macro-scale deformation and UV-space intrinsic maps, with wrinkle cues represented explicitly in the normal maps.

3 Method: ChronoMesh

3.1 Overview

Our goal is to edit the perceived age of an existing 3D head *mesh asset* while keeping the representation compatible with a standard mesh-and-UV pipeline. Fig. 2 (left) shows the overview of our method. We assume the input head is represented by the FLAME parametric model (Li et al., 2017) with a fixed topology and UV coordinates, and is accompanied by UV-space intrinsic maps: a diffuse reflectance map, a normal map, and a specular map. In addition to the mesh asset, our method takes the source age α_s and a target age α_t as inputs. If α_s is unknown, we estimate it from a rendered image of the input asset using an image-based age estimator (*e.g.*, MiVOLO (Kuprashevich & Tolstykh, 2023)). The output is an age-edited head mesh and its intrinsic maps that match the specified target age.

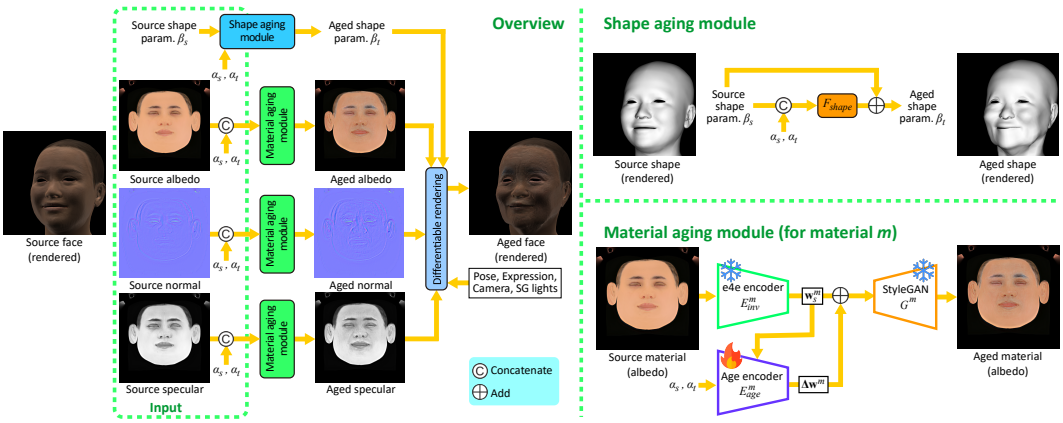

Figure 2: Overview of *ChronoMesh* (left), including the shape aging module (upper right) for macro-scale deformation and the material aging module (lower right) for UV map editing. Here α_s, α_t are the source and target ages, $\boldsymbol{\beta}_s, \boldsymbol{\beta}_t$ the FLAME shape parameters, F_{shape} the shape-residual MLP, and E^m_{inv} (e4e), E^m_{age}, and G^m the inversion encoder, age encoder, and StyleGAN2 generator for map type m, with $\mathbf{w}^m_s$ and $\Delta\mathbf{w}^m$ the inverted latent code and the predicted age offset.

Two-stage learning. Sec. 3.2 introduces the differentiable renderer (Laine et al., 2020) and the losses shared by both stages. Stage 1 (Sec. 3.3) then learns macro-scale deformation and Stage 2 (Sec. 3.4) learns UV-space intrinsic map edits, which avoids the shortcut discussed in Sec. 1.

3.2 Differentiable Rendering, Lighting, and Losses

Rendering and lighting. We render the edited mesh with its UV-space maps under environment lighting using a differentiable rasterizer (Laine et al., 2020), with illumination represented by spherical Gaussians (Wang et al., 2009) and diffuse and specular shading computed from the material maps, so that image-based losses backpropagate to the shape parameters and UV maps. Because wrinkle visibility depends on illumination (Zoss et al., 2022), we pre-optimize a fixed lighting of 16 spherical-Gaussian lobes on the training set so that the age estimator responds maximally to the presence of the normal map, and keep it fixed while training both stages (App. D).

Loss functions. Both stages share the following losses computed on rendered images: (i) an age loss based on MiVOLO (Kuprashevich & Tolstykh, 2023), (ii) an identity loss based on BlendFace (Shiohara et al., 2023) to reduce hairstyle-related bias compared to ArcFace (Deng et al., 2019), (iii) a cycle-consistency loss that edits to α_t and back to α_s, and (iv) regularizers that discourage excessive deviations from the input shape and maps. Stage 1 additionally uses the deformation hinge loss $\mathcal{L}_{hinge}$. Stage 2 additionally uses the normal-map age loss $\mathcal{L}^{norm}_{age}$ (Sec. 3.4). We align rendered images to a canonical crop (similar to the FFHQ protocol) before feeding them into the age and identity networks. Details of the loss functions are provided in Appendix E.

3.3 Stage 1: Learning Macro-scale Shape Deformation

Stage 1 learns an age-conditioned macro-scale deformation of the input head mesh by predicting a residual update in FLAME shape space (Fig. 2, upper right). Let $\boldsymbol{\beta}_s \in \mathbb{R}^{n_\beta}$ (where n_β represents the dimension) denote the FLAME shape parameters of the input mesh. We learn an MLP F_{shape} that predicts an age-conditioned shape residual:

$$\Delta\boldsymbol{\beta} = F_{shape}(\boldsymbol{\beta}_s, \alpha_s, \alpha_t), \qquad \boldsymbol{\beta}_t = \boldsymbol{\beta}_s + \Delta\boldsymbol{\beta}, \tag{1}$$

and obtain the deformed mesh $\mathcal{M}_t$ by instantiating FLAME with $\boldsymbol{\beta}_t$, keeping the remaining FLAME parameters unchanged.

Proxy normal map. Although Stage 1 focuses on learning macro-scale shape deformation, age supervision under shaded rendering still depends strongly on wrinkle cues. However, Stage 1 does

not yet predict age-dependent normal maps, and naively reusing the input normal map can mislead the optimization. For young targets, residual wrinkles in the input normal map encourage the network to "cancel" them through shape changes, resulting in unnatural distortions. For older targets, the lack of increased wrinkles prevents the age estimator from reaching the target age, making deformation learning ineffective. To resolve this chicken-and-egg problem, we attach age-dependent *proxy normal maps* (Fig. 6 in App. B) during Stage 1 training. Specifically, we compute a representative normal map for each age by aggregating training-set normal maps using a weighted average that accounts for variability; details are provided in Appendix B.

Per-vertex deformation hinge loss. To avoid unrealistic facial distortion caused by overly large shape updates, Stage 1 minimizes the common rendering losses (Sec. 3.2) together with a deformation hinge regularizer that softly limits per-vertex displacement:

$$\mathcal{L}_{hinge} = \frac{1}{V}\sum_{v=1}^{V} \max\big(\|\mathbf{x}_s(v) - \hat{\mathbf{x}}_t(v)\|_2^2 - \tau^2,\ 0\big), \tag{2}$$

where $\mathbf{x}_s(v)$ and $\hat{\mathbf{x}}_t(v)$ denote the positions of vertex v before and after deformation, V is the number of vertices, and $\tau = 0.005$. This regularizer is important because identity losses based on face recognition models assume a canonical face alignment; if the deformation changes the facial layout too aggressively, the identity loss becomes less reliable and the face can collapse into implausible shapes. We apply FFHQ-style alignment to the rendered images before feeding them to the identity network, while using $\mathcal{L}_{hinge}$ as a loose global constraint that discourages excessive deformation without hard pinning.

3.4 Stage 2: Learning UV-space Intrinsic Map Edits

Stage 2 learns age-conditioned edits of intrinsic maps attached to the deformed mesh (Fig. 2, lower right). For each map type $m \in \{\text{diffuse}, \text{normal}, \text{specular}\}$, we first train a StyleGAN2 generator G^m on the corresponding dataset maps in an unsupervised manner. Given an input UV map $\mathbf{T}_s^m$, we fine-tune the pre-trained StyleGAN2 generator using pivotal tuning inversion (PTI) (Roich et al., 2022), and invert the input UV map into the latent space using an encoder E_{inv}^m (we use *e4e* (Tov et al., 2021)). We then predict an age-dependent latent offset with an MLP E_{age}^m:

$$\mathbf{w}_s^m = E_{inv}^m(\mathbf{T}_s^m), \quad \Delta\mathbf{w}^m = E_{age}^m(\mathbf{w}_s^m, \alpha_s, \alpha_t), \quad \mathbf{T}_t^m = G^m(\mathbf{w}_s^m + \Delta\mathbf{w}^m). \tag{3}$$

This "invert then offset" design follows latent-based age editing approaches such as SAM (Alaluf et al., 2021), while adapting it to UV-space intrinsic maps. In particular, predicting $\mathbf{T}_t^{norm}$ enables explicit wrinkle normal synthesis.

We edit in the latent space of a pre-trained StyleGAN2 with inversion and pivotal tuning because asset editing must stay close to the input maps, remain consistent across map types, and run in milliseconds; diffusion-based UV-map editing is left for future work (App. F).

Normal-map age supervision. Image-space supervision alone can be insufficient to learn wrinkle normals: depending on illumination, wrinkle shading can be weak and the optimizer may underfit the normal-map change. To provide direct supervision on wrinkle normals, we train an auxiliary age estimator that takes normal maps (or rendered normal images) as input. We implement it with a Transformer-based backbone and define a normal-age loss $\mathcal{L}_{age}^{norm}$ that encourages the predicted normal map to match the target age. This loss is illumination-invariant by construction and substantially strengthens age-dependent wrinkle learning. Architectural details are provided in Appendix C.

3.5 Dataset and Preprocessing

We use FFHQ-UV-Intrinsics (Dib et al., 2024), to our knowledge the only public dataset with diffuse, normal, and specular maps in a common UV layout for thousands of subjects. We fit FLAME to the original FFHQ images with Pixel3DMM (Giebenhain et al., 2025), omit its failure cases, and use 9,237 subjects for training and 236 for testing (App. A).

Table 1: Quantitative evaluation results of ablation for Stages 1 and 2. We report the average metrics obtained by editing each source mesh to every integer age from 0 to 100. "Ours" in Stage 1 uses proxy normal maps. "Single stage" in Stage 2 learns shape and UV maps jointly. The best and second-best scores are shown in **bold** and underlined.

	Setting	ID ↑		Age ↓		
		ArcFace	BlendFace	DEX	MiVOLO	Normal
Stage 1	w/o hinge loss	0.747	0.646	9.40	6.36	23.3
	w/o proxy normal	0.692	0.566	8.78	6.52	**22.4**
	Ours	**0.795**	**0.673**	**7.92**	**5.95**	22.5
Stage 2	Single stage	**0.858**	**0.757**	16.5	3.47	**14.3**
	w/o normal loss	0.806	0.691	15.5	3.90	22.7
	Ours	0.798	0.686	**13.6**	**3.05**	18.9

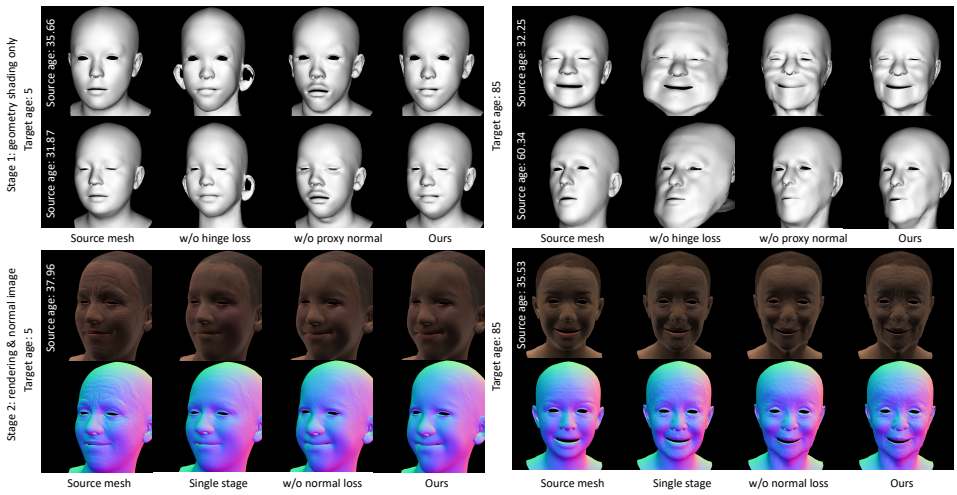

Figure 3: Qualitative evaluation of the ablation study for Stages 1 and 2.

4 Experiments

Implementation details (optimizer, image alignment, eye-region masking, and target-age sampling) are given in Apps. E and F. The source age α_s is estimated by MiVOLO from a rendered source image. Unless noted otherwise, images are rendered with our differentiable rasterizer under the training lighting; Fig. 1 is rendered in Blender under a real-world environment map.

4.1 Ablation Study

Metrics. Identity preservation (ID) is the cosine similarity between face-recognition embeddings of the source and edited renders from BlendFace (Shiohara et al., 2023), used in our training loss, and ArcFace (Deng et al., 2019), which is not. Age fidelity is the mean absolute error in years between the target age and the age predicted by MiVOLO (Kuprashevich & Tolstykh, 2023) or our normal-map estimator, both used in our training, or DEX (Rothe et al., 2015), which is not used in our training but is SAM's age objective (Sec. 4.2).

The ablation removes the hinge loss ("w/o hinge loss") or the proxy normal maps ("w/o proxy normal") in Stage 1 (Sec. 3.3), and in Stage 2 (Sec. 3.4) learns FLAME shape parameters and UV-map edits jointly ("Single stage") or removes $\mathcal{L}_{age}^{norm}$ ("w/o normal loss"); "Ours" uses all components.

In Table 1, Stage-1 "Ours" achieves the best performance on almost all metrics and ranks a close second on "Age (Normal)", indicating that combining the hinge regularizer and proxy normals im-

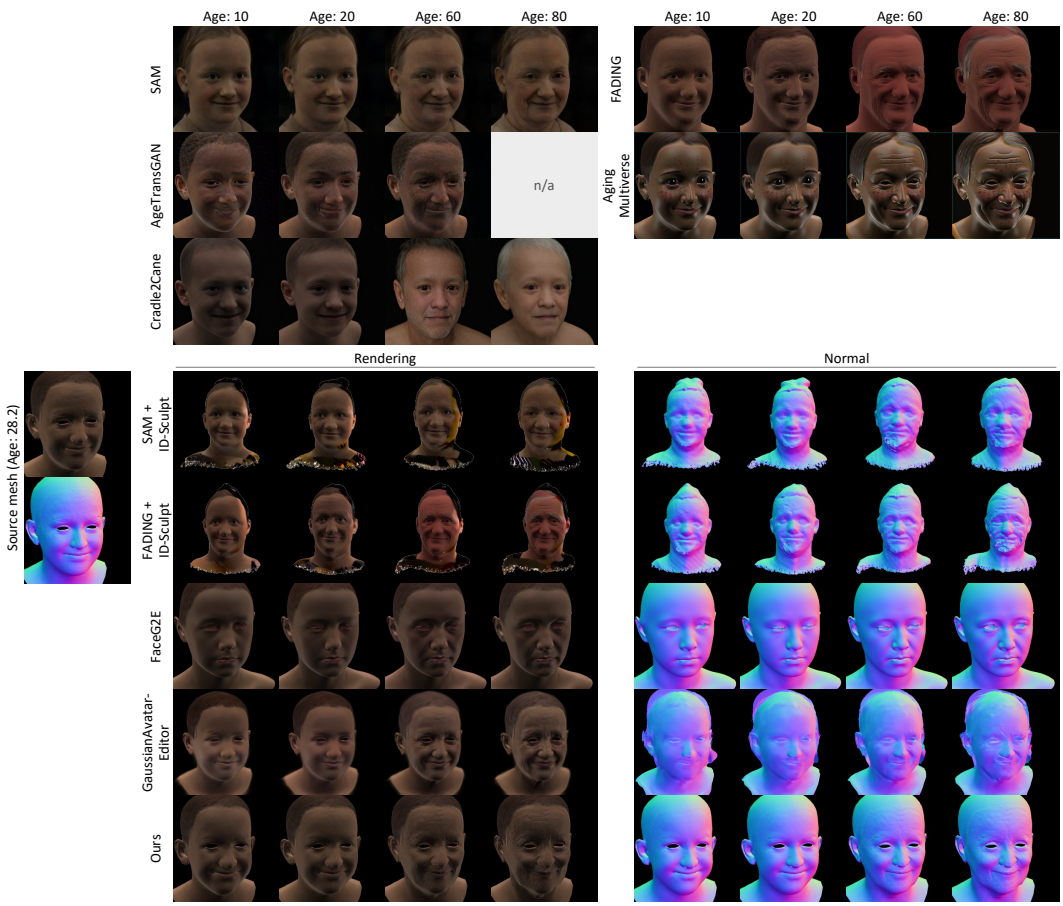

Figure 4: Qualitative comparison with baseline methods. Top: image-based editors, two per row. Bottom: 3D outputs, with renderings (left) and rendered normals (right); ID-Sculpt outputs are in their own reconstruction frame, so their poses differ slightly from the source. n/a: beyond AgeTransGAN's range (up to 70). More subjects are shown in App. G; the output meshes and UV maps and a video are provided as supplementary material.

proves both age control and identity preservation. In Stage 2, "Ours" has lower identity scores than the alternatives, partly because the single-stage variant leaves the geometry nearly unchanged (Fig. 3, bottom), which inflates its identity score; "Ours" attains the lowest DEX and MiVOLO errors in this block, whereas the normal-map estimator favors the single-stage variant, whose maps absorb the age change (the shortcut discussed in Sec. 1).

Fig. 3 shows the failure modes. In Stage 1, "w/o hinge loss" deforms the ears unnaturally for a young target ($\alpha_t = 5$) and inflates the face for an old target ($\alpha_t = 85$), and "w/o proxy normal" distorts the lips and chin, as the optimization compensates for age-inconsistent normal cues by warping the geometry; "Ours" deforms stably. In Stage 2, "Single stage" barely changes the shape while its normal maps still change with age, i.e., joint training explains the age supervision through map updates rather than macro-scale deformation. "w/o normal loss" also produces age-dependent normals, but with less pronounced wrinkles at $\alpha_t = 85$ than "Ours", so $\mathcal{L}_{age}^{norm}$ strengthens wrinkle-normal learning.

4.2 Comparison to Baselines

Baseline selection. To our knowledge, no existing method edits an existing head mesh together with its diffuse, normal, and specular UV maps: FaceG2E (Wu et al., 2024) does not edit normal

Table 2: Quantitative comparison with baseline methods: image-based (2D) age editing and 3D meshes converted from the 2D outputs by ID-Sculpt (Hao et al., 2025). Metrics are averaged over target ages 0 to 100 in 10-year increments; bold: best. [§]: ages 0–70. [†]: used as the age objective when training SAM. [‡]: used in our training loss.

	Method	ID ↑		Age ↓	
		ArcFace	BlendFace	DEX[†]	MiVOLO[‡]
2D	SAM	0.730	0.505	**5.38**	10.6
2D	FADING	0.707	0.532	11.8	12.3
2D	AgeTransGAN[§]	0.686	0.421	5.49	5.29
2D	Aging Multiverse	0.695	0.542	21.4	24.3
2D	Cradle2Cane	0.397	0.348	9.20	5.87
3D	SAM+ID-Sculpt	0.209	0.106	13.7	12.9
3D	FADING+ID-Sculpt	0.262	0.152	16.2	15.1
3D	Ours	**0.786**	**0.672**	14.5	**3.45**

or specular maps, and ReAge3D (Zeng et al., 2026) and GaussianAvatar-Editor (Liu et al., 2025c) output 3D Gaussians, whose conversion into mesh assets would add confounding factors. We therefore compare quantitatively against baselines that operate on renders of the same input mesh under an identical rendering-and-metric protocol.

Baselines. We render the input mesh with the same differentiable renderer and fixed lighting as in training and apply five image-based age editors: SAM (Alaluf et al., 2021), with pivotal tuning inversion (Roich et al., 2022) to adapt it to our renders, FADING (Chen & Lathuilière, 2023), AgeTransGAN (Hsu et al., 2022), Aging Multiverse (Gong et al., 2025), and Cradle2Cane (Liu et al., 2025b). These 2D results indicate how much age change is achievable in image space alone. For a comparison in 3D, we lift the SAM and FADING edits back to head meshes with the single-view reconstruction method ID-Sculpt (Hao et al., 2025) and render them under the same setup ("SAM + ID-Sculpt", "FADING + ID-Sculpt"). FaceG2E and GaussianAvatar-Editor (the latter without mesh conversion) are compared only qualitatively (Figs. 4 and 8–13), as both require a separate optimization for each subject and target age.

Protocol. We use the metrics of Sec. 4.1; the normal-map metric is omitted for the purely image-based results, which have no normal maps. Because ID-Sculpt preprocessing fails on a non-negligible number of edited images, we exclude all samples that fail for ID-Sculpt from the evaluation set.

Quantitative comparison. Table 2 summarizes the quantitative results. Our method obtains the highest identity similarity under both encoders and the lowest MiVOLO error among all methods, but on DEX, the only age estimator unseen by our training, it trails SAM, FADING, AgeTransGAN (ages 0–70), Cradle2Cane, and SAM+ID-Sculpt (14.5 vs. 5.38, 11.8, 5.49, 9.20, and 13.7). SAM is trained with a DEX-style age objective, and PTI further adapts it to our renders; for the other editors, we have not identified the cause. Since SAM and our method each score best on the estimator used in their own training (Table 2, [†] and [‡]), we regard the identity metrics and the mesh-space check in Sec. 4.3 as the more informative evidence. The ID-Sculpt reconstructions are worse than their 2D inputs on all metrics, reflecting degradation introduced by the 2D-to-3D conversion.

Qualitative comparison. Fig. 4 and Figs. 8–13 (App. G) are consistent with these trends. The image-based editors change the apparent age but depart from the original CG rendering: SAM (with PTI) and FADING drift toward a photographic style, AgeTransGAN and Aging Multiverse add skin-texture artifacts, and Cradle2Cane alters identity cues such as face shape. The ID-Sculpt reconstructions have lower geometric quality and are unstable across target ages, as the rendered normals show. FaceG2E changes little with the target age, and GaussianAvatar-Editor adds wrinkles but yields noisy normals near the head boundary. Our method yields more coherent age edits on the mesh while keeping the input asset's rendering style.

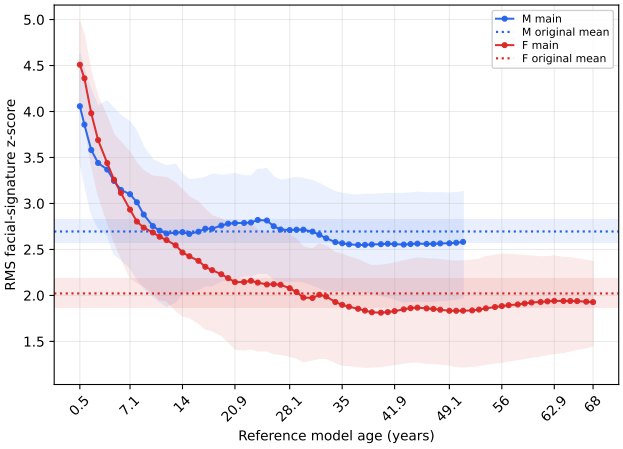

Figure 5: RMS facial-signature z-scores (lower is better) of our edited meshes (solid, "main") against age- and sex-specific statistics of the 3D facial growth curves (Matthews et al., 2021) for male (M) and female (F) subjects, averaged within each age group and shaded with the 25–75% range; dotted: mean score of the unedited source meshes, shaded with its bootstrap 95% confidence interval. Reference statistics exist up to age 51.8 (M) and 68.9 (F).

UV-map behavior and additional results. Fig. 14 (App. G) shows that wrinkle cues appear in the predicted normal maps while the diffuse map stays free of high-frequency wrinkle shading, which supports stable relighting; more examples and a video with temporally smooth age edits are provided in App. G and the supplementary material.

4.3 Mesh-space Plausibility

The metrics above rely on image-based networks, three of which (BlendFace, MiVOLO, and the normal-map estimator) are also used in our training. As a complementary check that uses no rendered image, we register our edited meshes to the open-source 3D facial growth curves of Matthews et al. (2021) and compute the RMS facial-signature z-score (lower is better) with respect to age- and sex-specific 3D statistics. Fig. 5 plots the score against the target age. Unedited meshes already have non-zero scores because of domain gaps: the reference model is built from neutral expressions of a European-ancestry population, whereas our subjects are FFHQ-derived, mixed, and sometimes have open mouths. Scores are high for child targets, consistent with the adult-centric shape space of FLAME (Sec. 5), whereas from late adolescence onward they approach the level of the unedited meshes, indicating that our adult-range edits remain within the plausibility range of the reference statistics rather than merely satisfying the image-based losses. We do not treat the age-specific expected face as an identity-specific ground truth.

5 Conclusion and Future Work

We presented *ChronoMesh*, a mesh-based 3D head aging framework that predicts FLAME shape residuals for macro-scale deformation and StyleGAN latent offsets for UV-space intrinsic maps, representing wrinkles explicitly in normal maps rather than in diffuse reflectance.

Our current system still has several limitations. First, the output quality depends heavily on the StyleGAN generator, and the quality of FFHQ-UV-Intrinsics limits the achievable fidelity; a higher-quality dataset of UV intrinsic maps would directly improve the generator. Second, the input must follow the FLAME topology; fitting FLAME to scanned or artist-made meshes and transferring the predicted deformation back to the original mesh is future work. Third, FLAME's shape space is learned mainly from adult scans (Li et al., 2017), which limits child targets, as Sec. 4.3 shows; age-stratified 3DMMs (Booth et al., 2018; Dai et al., 2020), craniofacial growth models (Ramanathan & Chellappa, 2006), and higher-degree-of-freedom deformation such as neural Jacobian fields (Aigerman et al., 2022) are natural remedies.

AI USE STATEMENT

In this work, we used generative AI tools to design or provide feedback on research methodology or experiments, to implement methods, to assist with translation, and to interpret results. Concretely, we discussed design alternatives for the two-stage formulation and the loss terms with an AI assistant, used AI assistance while writing parts of the training and evaluation code, translated parts of our own draft text into English, and discussed possible readings of the quantitative and qualitative results with an AI assistant. We have not used generative AI tools to generate synthetic data sets, to help develop theoretical models or conceptual frameworks, to formulate mathematical claims, to provide critical ingredients for proving mathematical claims, to assist in the writing of proofs, to propose or refine hypotheses, to clean and reformat data sets, or to support qualitative and thematic data analysis. Additionally, we used generative AI tools to create or edit software code, to lay out figures, to identify relevant literature and summarize existing work, to suggest a structure for this paper, to draft parts of it, to edit it to improve readability, and to format references. In terms of the categories of the submission form, these uses correspond to research ideation or execution (the design feedback, implementation, and interpretation described above), retrieval and discovery (identifying related work), drafting sections of the paper, and aiding or polishing the writing (including translation). We have reviewed all AI-assisted work: all AI-assisted design decisions were evaluated in our own experiments, all AI-assisted code was read and tested by the authors, all translated text was checked against the original drafts, every interpretation reported in this paper is our own and was verified against our results, and every AI-suggested reference was checked against the original publication. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Architectures, loss definitions and weights, and optimizer settings and batch sizes are given in Sec. 3 and Appendices C–F; data preprocessing and the 9,237/236 subject split are described in Sec. 3.5 and App. A; the lighting pre-optimization objective is given in App. D; and the evaluation protocol, including the exclusion of ID-Sculpt failure cases, is described in Sec. 4.2. Source code and pretrained models will be released upon acceptance.

ETHICS STATEMENT

This work uses publicly available face data (FFHQ-UV-Intrinsics and FFHQ-UV, derived from FFHQ photographs) under their respective licenses and the open-source 3D facial growth curves of Matthews et al. (2021); it involves no new data collection from human subjects. The edited 3D assets are reconstructed from photographs of real people, and the method could also be applied to scans of identifiable individuals to misrepresent their age; we will release the code with usage terms that prohibit such deceptive use. The age estimators, the FFHQ-derived data, and the European-ancestry growth-curve reference are demographically imbalanced, so our quantitative results may not transfer equally across populations.

REFERENCES

Noam Aigerman, Kunal Gupta, Vladimir G. Kim, Siddhartha Chaudhuri, Jun Saito, and Thibault Groueix. Neural Jacobian fields: Learning intrinsic mappings of arbitrary meshes. *ACM Trans. Graph. (Proc. of SIGGRAPH 2022)*, 41(4):109:1–109:17, 2022.

Yuval Alaluf, Or Patashnik, and Daniel Cohen-Or. Only a matter of style: Age transformation using a style-based regression model. *ACM Trans. Graph. (Proc. of SIGGRAPH 2021)*, 40(4): 45:1–45:15, 2021.

Haoran Bai, Di Kang, Haoxian Zhang, Jinshan Pan, and Linchao Bao. FFHQ-UV: Normalized facial UV-texture dataset for 3D face reconstruction. In *CVPR*, pp. 362–371, June 2023. doi: 10.1109/CVPR52729.2023.00043.

Stefan Andreas Baumann, Felix Krause, Michael Neumayr, Nick Stracke, Melvin Sevi, Vincent Tao Hu, and Björn Ommer. Continuous, subject-specific attribute control in T2I models by identifying semantic directions. In *CVPR*, 2025.

James Booth, Anastasios Roussos, Allan Ponniah, David Dunaway, and Stefanos Zafeiriou. Large scale 3D morphable models. *International Journal of Computer Vision*, 126(2):233–254, 2018. doi: 10.1007/s11263-017-1009-7.

Zehranaz Canfes, M. Furkan Atasoy, Alara Dirik, and Pinar Yanardag. Text and image guided 3D avatar generation and manipulation. In *WACV*, pp. 4421–4431, January 2023.

Eric R. Chan, Connor Z. Lin, Matthew A. Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J. Guibas, Jonathan Tremblay, Sameh Khamis, Tero Karras, and Gordon Wetzstein. Efficient geometry-aware 3D generative adversarial networks. In *CVPR*, pp. 16123–16133, June 2022.

Xiangyi Chen and Stéphane Lathuilière. Face aging via diffusion-based editing. In *BMVC*, November 2023.

Yuhao Cheng, Zhuo Chen, Xingyu Ren, Wenhan Zhu, Zhengqin Xu, Di Xu, Changpeng Yang, and Yichao Yan. 3D-aware face editing via warping-guided latent direction learning. In *CVPR*, pp. 916–926, June 2024.

Hang Dai, Nick Pears, William Smith, and Christian Duncan. Statistical modeling of craniofacial shape and texture. *International Journal of Computer Vision*, 128(2):547–571, 2020. doi: 10.1007/s11263-019-01260-7.

Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *CVPR*, pp. 4690–4699, June 2019. doi: 10.1109/CVPR.2019.00482.

Abdallah Dib, Luiz Gustavo Hafemann, Emeline Got, Trevor Anderson, Amin Fadaeinejad, Rafael M. O. Cruz, and Marc-André Carbonneau. MoSAR: Monocular semi-supervised model for avatar reconstruction using differentiable shading. In *CVPR*, pp. 1770–1780, June 2024.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2021.

Guy Gafni, Justus Thies, Michael Zollhofer, and Matthias Niessner. Dynamic neural radiance fields for monocular 4D facial avatar reconstruction. In *CVPR*, pp. 8649–8658, June 2021.

Dimitrios Gerogiannis, Foivos Paraperas Papantoniou, Rolandos Alexandros Potamias, Alexandros Lattas, and Stefanos Zafeiriou. Arc2Avatar: Generating expressive 3D avatars from a single image via ID guidance. In *CVPR*, pp. 10770–10782, June 2025.

Simon Giebenhain, Tobias Kirschstein, Martin Rünz, Lourdes Agapito, and Matthias Nießner. Pixel3DMM: Versatile screen-space priors for single-image 3D face reconstruction. arXiv:2505.00615, 2025.

Aleksey Golovinskiy, Wojciech Matusik, Hanspeter Pfister, Szymon Rusinkiewicz, and Thomas Funkhouser. A statistical model for synthesis of detailed facial geometry. *ACM Trans. Graph. (Proc. of SIGGRAPH 2006)*, 25(3):1025–1034, 2006.

Guillermo Gomez-Trenado, Stéphane Lathuilière, Pablo Mesejo, and Oscar Cordón. Custom structure preservation in face aging. In *ECCV*, pp. 565–580, 2022.

Bang Gong, Luchao Qi, Jiaye Wu, Zhicheng Fu, Chunbo Song, David W. Jacobs, John Nicholson, and Roni Sengupta. The Aging Multiverse: Generating condition-aware facial aging tree via training-free diffusion. In *SIGGRAPH Asia 2025 Conference Papers*, 2025.

Jiahong Hao, Jia Tang, Jia Zhang, Rui Yi, Yujia Hong, Min Li, Wei Cao, Yiran Wang, Chen Wang, and Lei Ma. ID-Sculpt: ID-aware 3D head generation from single in-the-wild portrait image. In *AAAI*, pp. 3383–3391, 2025.

Zhenliang He, Meina Kan, and Shiguang Shan. EigenGAN: Layer-wise eigen-learning for GANs. In *ICCV*, pp. 14408–14417, October 2021.

Yang Hong, Bo Peng, Haiyao Xiao, Ligang Liu, and Juyong Zhang. Headnerf: A real-time NeRF-based parametric head model. In *CVPR*, pp. 20374–20384, June 2022.

Gee-Sern Hsu, Rui-Cang Xie, Zhi-Ting Chen, and Yu-Hong Lin. AgeTransGAN for facial age transformation with rectified performance metrics. In *ECCV*, 2022. doi: 10.1007/978-3-031-19775-8_34.

Taishi Ito, Yuki Endo, and Yoshihiro Kanamori. SelfAge: Personalized facial age transformation using self-reference images. *The Visual Computer (Proc. of Computer Graphics International 2025)*, 41(9):6769–6781, 2025.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, pp. 4401–4410, 2019.

Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of StyleGAN. In *CVPR*, pp. 8107–8116, 2020.

Ira Kemelmacher-Shlizerman, Supasorn Suwajanakorn, and Steven M. Seitz. Illumination-aware age progression. In *CVPR*, pp. 3334–3341, June 2014. doi: 10.1109/CVPR.2014.426.

Khronos Group. glTF 2.0 Specification. `https://registry.khronos.org/glTF/specs/2.0/glTF-2.0.html`, 2021. Accessed: 2026-03-01.

Amandeep Kumar, Muhammad Awais, Sanath Narayan, Hisham Cholakkal, Salman Khan, and Rao Muhammad Anwer. Efficient 3D-aware facial image editing via attribute-specific prompt learning. In *ECCV*, pp. 124–141, 2024. doi: 10.1007/978-3-031-72907-2_8.

Maksim Kuprashevich and Irina Tolstykh. MiVOLO: Multi-input transformer for age and gender estimation. *arXiv preprint arXiv:2307.04616*, 2023.

Mingi Kwon, Jaeseok Jeong, and Youngjung Uh. Diffusion models already have a semantic latent space. In *ICLR*, 2023.

Samuli Laine, Janne Hellsten, Tero Karras, Yeongho Seol, Jaakko Lehtinen, and Timo Aila. Modular primitives for high-performance differentiable rendering. *ACM Trans. Graph.*, 39(6):194:1–194:14, 2020.

Jianhui Li, Jianmin Li, Haoji Zhang, Shilong Liu, Zhengyi Wang, Zihao Xiao, Kaiwen Zheng, and Jun Zhu. PREIM3D: 3D consistent precise image attribute editing from a single image. In *CVPR*, pp. 8549–8558, June 2023a.

Peipei Li, Rui Wang, Huaibo Huang, Ran He, and Zhaofeng He. Pluralistic aging diffusion autoencoder. In *ICCV*, pp. 22613–22623, October 2023b.

Tianye Li, Timo Bolkart, Michael J. Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4D scans. *ACM Trans. Graph. (Proc. of SIGGRAPH Asia 2017)*, 36 (6):194:1–194:17, 2017. doi: 10.1145/3130800.3130813.

Xiaohui Li, Giuseppe Claudio Guarnera, Arvin Lin, and Abhijeet Ghosh. Realistic facial age transformation with 3D uplifting. *Comput. Graph. Forum*, 43(4):e15146, 2024.

Hongyu Liu, Xuan Wang, Ziyu Wan, Yue Ma, Jingye Chen, Yanbo Fan, Yujun Shen, Yibing Song, and Qifeng Chen. AvatarArtist: Open-domain 4D avatarization. In *CVPR*, 2025a.

Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. On the variance of the adaptive learning rate and beyond. In *ICLR*, 2020.

Tao Liu, Dafeng Zhang, Gengchen Li, Shizhuo Liu, Yongqi Song, Senmao Li, Shiqi Yang, Boqian Li, Kai Wang, and Yaxing Wang. From cradle to cane: A two-pass framework for high-fidelity lifespan face aging. In *NeurIPS*, 2025b. URL `https://openreview.net/forum?id=E1eVGJ5RYG`.

Xiangyue Liu, Kunming Luo, Heng Li, Qi Zhang, Yuan Liu, Li Yi, and Ping Tan. GaussianAvatar-Editor: Photorealistic animatable Gaussian head avatar editor. *arXiv:2501.09978*, 2025c.

Harold S. Matthews, Richard L. Palmer, Gareth S. Baynam, Oliver W. Quarrell, Ophir D. Klein, Richard A. Spritz, Raoul C. Hennekam, Susan Walsh, Mark Shriver, Seth M. Weinberg, Benedikt Hallgrimsson, Peter Hammond, Anthony J. Penington, Hilde Peeters, and Peter D. Claes. Large-scale open-source three-dimensional growth curves for clinical facial assessment and objective description of facial dysmorphism. *Scientific Reports*, 11(1):12175, 2021.

Anirban Mukherjee, Venkat Suprabath Bitra, Vignesh Bondugula, Tarun Reddy Tallapureddy, and Dinesh Babu Jayagopi. SemUV: Deep learning based semantic manipulation over UV texture map of virtual human heads. arXiv, 2024. arXiv:2407.00229.

Konpat Preechakul, Nattanat Chatthee, Suttisak Wizadwongsa, and Supasorn Suwajanakorn. Diffusion autoencoders: Toward a meaningful and decodable representation. In *CVPR*, 2022.

Luchao Qi, Jiaye Wu, Bang Gong, Annie N. Wang, David W. Jacobs, and Roni Sengupta. MyTimeMachine: Personalized facial age transformation. *ACM Trans. Graph. (Proc. of SIGGRAPH 2025)*, 44(4):140:1–140:16, 2025.

Narayanan Ramanathan and Rama Chellappa. Modeling age progression in young faces. In *CVPR*, pp. 387–394, 2006.

Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. Encoding in style: A StyleGAN encoder for image-to-image translation. In *CVPR*, pp. 2287–2296, June 2021.

Daniel Roich, Ron Mokady, Amit H. Bermano, and Daniel Cohen-Or. Pivotal tuning for latent-based editing of real images. *ACM Trans. Graph.*, 42(1):6:1–6:13, August 2022. doi: 10.1145/3544777.

Rasmus Rothe, Radu Timofte, and Luc Van Gool. DEX: Deep EXpectation of apparent age from a single image. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, pp. 10–15, December 2015. doi: 10.1109/ICCVW.2015.41.

Kristina Scherbaum, Martin Sunkel, Hans-Peter Seidel, and Volker Blanz. Prediction of individual non-linear aging trajectories of faces. *Computer Graphics Forum (Proc. of Eurographics 2007)*, 26(3):285–294, 2007. doi: 10.1111/j.1467-8659.2007.01050.x.

Yujun Shen, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. InterFaceGAN: Interpreting the latent space of GANs for facial attributes. In *CVPR*, pp. 9243–9252, 2020.

Kaede Shiohara, Xingchao Yang, and Takafumi Taketomi. BlendFace: Re-designing identity encoders for face-swapping. In *ICCV*, pp. 7634–7644, October 2023.

Zepeng Su, Zhulin Liu, Zongyan Zhang, Tong Zhang, and C. L. Philip Chen. TimeBooth: Disentangled facial invariant representation for diverse and personalized face aging. In *ICCV*, pp. 12147–12157, October 2025.

Qianrui Teng, Rui Wang, Xing Cui, Peipei Li, and Zhaofeng He. Exploring 3D-aware lifespan face aging via disentangled shape-texture representations. In *ICME*, pp. 1–6, July 2024.

Omer Tov, Yuval Alaluf, Yotam Nitzan, Or Patashnik, and Daniel Cohen-Or. Designing an encoder for StyleGAN image manipulation. *ACM Trans. Graph.*, 40(4), July 2021.

Alex Trevithick, Matthew Chan, Towaki Takikawa, Umar Iqbal, Shalini De Mello, Manmohan Chandraker, Ravi Ramamoorthi, and Koki Nagano. What you see is what you GAN: Rendering every pixel for high-fidelity geometry in 3D GANs. In *CVPR*, pp. 22765–22775, June 2024.

Junaid Wahid, Fangneng Zhan, Pramod Rao, and Christian Theobalt. DiffAge3D: Diffusion-based 3D-aware face aging. arXiv:2408.15922, 2024.

Jiaping Wang, Peiran Ren, Minmin Gong, John Snyder, and Baining Guo. All-frequency rendering of dynamic, spatially-varying reflectance. *ACM Trans. Graph. (Proc. of SIGGRAPH Asia 2009)*, 28(5):133:1–133:10, 2009.

Yunjie Wu, Yapeng Meng, Zhipeng Hu, Lincheng Li, Haoqian Wu, Kun Zhou, Weiwei Xu, and Xin Yu. Text-guided 3D face synthesis - from generation to editing. In *CVPR*, pp. 1260–1269, June 2024.

Libing Zeng, Li Ma, Mingming He, Ning Yu, Paul Debevec, and Nima Khademi Kalantari. ReAge3D: Re-aging 3D faces with view consistency. In *Eurographics Symposium on Rendering (EGSR)*, 2026. arXiv:2606.18156.

Bowen Zhang, Yiji Cheng, Chunyu Wang, Ting Zhang, Jiaolong Yang, Yansong Tang, Feng Zhao, Dong Chen, and Baining Guo. RodinHD: High-fidelity 3D avatar generation with diffusion models. In *ECCV*, pp. 465–483, 2024a.

Jie Zhang, Kangneng Zhou, Yan Luximon, Tong-Yee Lee, and Ping Li. MeshWGAN: Mesh-to-mesh wasserstein GAN with multi-task gradient penalty for 3D facial geometric age transformation. *IEEE Trans. Vis. Comput. Graph.*, 30(8):4927–4940, 2024b. doi: 10.1109/TVCG.2023.3284500.

Michael R. Zhang, James Lucas, Geoffrey Hinton, and Jimmy Ba. Lookahead optimizer: k steps forward, 1 step back. *arXiv preprint arXiv:1907.08610*, 2019. doi: 10.48550/arXiv.1907.08610.

Gaspard Zoss, Prashanth Chandran, Eftychios Sifakis, Markus Gross, Paulo Gotardo, and Derek Bradley. Production-ready face re-aging for visual effects. *ACM Trans. Graph. (Proc. of SIGGRAPH Asia 2022)*, 41(6):237:1–237:12, 2022.

Richard A. Zoumalan and Wayne F. Larrabee. Anatomic considerations in the aging face. *Facial Plastic Surgery*, 27(1):16–22, 2011. doi: 10.1055/s-0030-1270420.

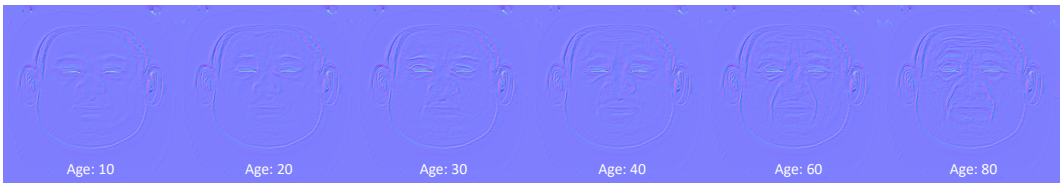

Figure 6: Proxy normal maps used in Stage 1 to represent age-dependent wrinkles.

A Dataset Details

FFHQ-UV-Intrinsics maps are defined only on the facial region, leaving the scalp and neck undefined; to avoid artifacts when computing rendering losses, we blend the diffuse reflectance with the corresponding full-head diffuse map from FFHQ-UV (Bai et al., 2023). While FFHQ-UV-Intrinsics provides 10,000 subjects, we omit failure cases of Pixel3DMM preprocessing, and use 9,237 subjects for training and 236 subjects for testing. We use a constant roughness map because the dataset does not provide one.

B Age-dependent proxy normal maps

Although Stage 1 is designed to learn macro-scale age-dependent shape deformation, age supervision under shaded rendering still depends strongly on wrinkle-related normal cues. Since Stage 1 does not yet predict target-age normal maps, directly reusing the input normal map can provide inconsistent supervision. To alleviate this issue, we attach an age-dependent proxy normal map to each target age during Stage 1 training.

To construct a proxy normal map for a target age α_t, we first precompute two age estimates for each training sample using DEX and MiVOLO. We then use the mean of the two predictions as the representative age of the sample, and use their variance as a reliability indicator. For a given target age, we retrieve the n_{proxy} training samples whose estimated ages are closest to α_t. Their normal maps are then aggregated by a weighted average, where samples closer to the target age receive larger weights, while samples with larger disagreement between DEX and MiVOLO receive smaller weights. The weights are normalized to sum to one.

The selected normal maps are taken from the FFHQ-UV-Intrinsics normal maps and averaged in UV space, followed by per-pixel renormalization. In all experiments, we set $n_{proxy} = 3$. We found that using more samples tends to oversmooth age-related details, especially wrinkle patterns, whereas a small number preserves sharper and more distinctive age cues.

The proxy normal map is used only as auxiliary age-dependent shading guidance in Stage 1. Its purpose is not to provide an exact target normal map, but to supply plausible age-consistent high-frequency cues so that Stage 1 can focus on learning coarse shape deformation without being misled by the source normal map.

C Normal-map age estimator

To provide direct supervision for wrinkle-related age cues in Stage 2, we train an auxiliary age regressor that predicts age from a facial normal image. The estimator takes as input a normal image obtained by applying Pixel3DMM (Giebenhain et al., 2025) to an FFHQ (Karras et al., 2019) face image, together with the annotated FFHQ skin mask, and outputs a scalar age value.

We do not use a convolution-based estimator such as DEX for this purpose, because the visible skin region differs between training and inference, making convolutional features less reliable under varying support regions. Instead, we adopt a Vision Transformer-based regressor (Dosovitskiy et al., 2021) that explicitly focuses on skin regions. Specifically, we concatenate the three-channel normal image and the one-channel skin mask into a four-channel input of resolution 512×512, divide it into 16×16 patches, and convert the patches into tokens. To suppress background and non-skin



Figure 7: Environment map obtained by rasterizing the optimized lighting representation with 16 spherical-Gaussian lobes.

regions, we retain only patches whose average mask value exceeds 0.9, and zero out the remaining patch tokens. The resulting token sequence is augmented with a learnable class token and learnable positional embeddings, processed by a Transformer encoder, and mapped to age by a regression head.

We train this estimator with an L_1 loss on age. The training set contains 68,500 FFHQ images for which Pixel3DMM preprocessing succeeded, and the test set contains 697 images. On the held-out test set, the mean absolute error is approximately four years. We use this network as the normal-based age predictor in the normal-age loss $\mathcal{L}_{age}^{norm}$ in Stage 2.

D LIGHTING PRE-OPTIMIZATION TO AMPLIFY AGE CUES

Perceived age cues are sensitive to illumination in practice (Kemelmacher-Shlizerman et al., 2014; Zoss et al., 2022). To make age supervision more effective during training, we pre-optimize a fixed lighting setup so that the rendered appearance becomes more sensitive to the presence or absence of wrinkle-related normal cues.

Specifically, for each training sample, we render two images under the same lighting condition: one with the normal map and one without it. We then evaluate the absolute difference between their predicted ages using an image-based age estimator. The lighting parameters are optimized so that this age difference becomes large. Since the presence or absence of normal details is expected to produce only a small age change for young subjects, we weight each sample by its FFHQ age annotation, so that older subjects contribute more strongly to the optimization.

Formally, letting α_i^{FFHQ} denote the normalized age annotation of sample i, we minimize

$$\mathcal{L}_{light} = \mathbb{E}_i \left[\alpha_i^{FFHQ} \left(1 - \left| \mathcal{A}(I_i^{w/norm}; \boldsymbol{\ell}) - \mathcal{A}(I_i^{w/o\ norm}; \boldsymbol{\ell}) \right| \right) \right], \tag{4}$$

where $\mathcal{A}(\cdot)$ is the image-based age estimator, $I_i^{w/norm}$ and $I_i^{w/o\ norm}$ are the rendered images with and without the normal map, respectively, and $\boldsymbol{\ell}$ denotes the lighting parameters. In our implementation, the lighting is represented by 16 spherical-Gaussian lobes, arranged symmetrically, with eight on the left and eight on the right, and their intensities, directions, and bandwidths are optimized. After pre-optimization, the resulting lighting is fixed and used throughout training in both stages.

Figure 7 shows the environment map obtained by rasterizing the optimized lighting represented with 16 spherical-Gaussian lobes.

E LOSS FUNCTIONS AND IMAGE ALIGNMENT DETAILS

This section provides the full definitions of the losses used in Stage 1 and Stage 2, together with the image alignment and masking procedure used before computing image-based supervision.

Alignment and masking. Both stages share the same preprocessing pipeline for image-based supervision. Given a rendered image I, we first apply an FFHQ-style 2D alignment before feeding it into the age and identity networks. Following the setup in the main paper, this alignment is implemented by translating the mesh vertices in the image plane before rasterization, while keeping rotation and scale unchanged. We also exclude the eye region from supervision. Let $\mathbf{M}_{eye}$ denote the binary exclusion mask in the aligned image plane, where eye pixels are set to 0 and the remaining pixels are set to 1. We denote the combined alignment-and-masking operation by $\mathcal{P}(\cdot)$, and write $\tilde{I} = \mathcal{P}(I)$. All image-based age and identity losses are computed from $\tilde{I}$.

We use the same age-difference-dependent weighting as SAM:

$$w_{diff}(\alpha_s, \alpha_t) = 0.25\,\cos(\pi\,|\bar{\alpha}_s - \bar{\alpha}_t|) + 0.75, \tag{5}$$

where $\bar{\alpha}_s, \bar{\alpha}_t \in [0, 1]$ denote the normalized source and target ages used for conditioning.

Age loss. For both stages, we use MiVOLO as the image-based age estimator. The rendered target image is first aligned and masked, and the predicted age is then obtained as $\mathcal{A}_{img}(\tilde{I}_t)$. The age loss is

$$\mathcal{L}_{age} = \left(\alpha_t - \mathcal{A}_{img}(\tilde{I}_t)\right)^2, \tag{6}$$

where $\tilde{I}_t = \mathcal{P}(I_t)$.

Identity loss. We use the identity loss adopted in BlendFace. Let

$$f_{id}(x) = \mathrm{normalize}(\mathrm{BlendFace}(x)) \tag{7}$$

denote the normalized identity embedding. The identity loss is defined as

$$\mathcal{L}_{id} = w_{diff}(\alpha_s, \alpha_t)\left(1 - \left\langle f_{id}(\tilde{I}_s), f_{id}(\tilde{I}_t)\right\rangle\right), \tag{8}$$

where $\tilde{I}_s = \mathcal{P}(I_s)$ and $\tilde{I}_t = \mathcal{P}(I_t)$. We prefer this identity loss because it is more suitable when supervising rendered hairless meshes, for which hairstyle information is absent or unreliable.

Cycle consistency. We use a forward-backward cycle in both stages. In Stage 1, after obtaining the forward-edited shape $\hat{\boldsymbol{\beta}}_t$, we edit it back to the source age using a stop-gradient formulation:

$$\hat{\boldsymbol{\beta}}_{s\leftarrow t} = \mathrm{sg}\left(\hat{\boldsymbol{\beta}}_t\right) + F_{shape}\left(\mathrm{sg}\left(\hat{\boldsymbol{\beta}}_t\right), \alpha_t, \alpha_s\right), \tag{9}$$

where $\mathrm{sg}(\cdot)$ denotes stop-gradient. The Stage 1 cycle loss is

$$\mathcal{L}_{cyc}^{shape} = \frac{1}{n_\beta}\left\|\boldsymbol{\beta}_s - \hat{\boldsymbol{\beta}}_{s\leftarrow t}\right\|_2^2. \tag{10}$$

In Stage 2, for each map type m, we first invert the forward-edited map $\hat{\mathbf{T}}_t^m$ to the latent space and then apply the reverse age edit:

$$\hat{\mathbf{w}}_t^m = E_{inv}^m(\mathrm{sg}(\hat{\mathbf{T}}_t^m)), \qquad \Delta\hat{\mathbf{w}}_{t\to s}^m = E_{age}^m(\hat{\mathbf{w}}_t^m, \alpha_t, \alpha_s), \tag{11}$$

$$\hat{\mathbf{T}}_{s\leftarrow t}^m = G^m(\hat{\mathbf{w}}_t^m + \Delta\hat{\mathbf{w}}_{t\to s}^m). \tag{12}$$

The Stage 2 cycle loss is

$$\mathcal{L}_{cyc}^{uv} = \sum_m \mathrm{mean}\left(\left(\mathbf{T}_s^m - \hat{\mathbf{T}}_{s\leftarrow t}^m\right)^2\right). \tag{13}$$

During training, the target age α_t is sampled according to the strategy in App. F, and the backward cycle always maps from α_t back to α_s.

Regularization terms. To keep the edited results close to the source, we use stage-specific regularization.

For Stage 1, we penalize the deviation of the predicted shape parameters from the source:

$$\mathcal{L}_{reg}^{shape} = w_{diff}(\alpha_s, \alpha_t) \frac{1}{n_\beta} \left\| \boldsymbol{\beta}_s - \hat{\boldsymbol{\beta}}_t \right\|_2^2. \tag{14}$$

For Stage 2, we penalize the deviation of each edited UV map from the source UV map:

$$\mathcal{L}_{reg}^{uv} = w_{diff}(\alpha_s, \alpha_t) \sum_m \text{mean}\left(\left(\mathbf{T}_s^m - \hat{\mathbf{T}}_t^m \right)^2 \right). \tag{15}$$

Following SAM, we also regularize the predicted latent code so that it does not move too far from the average latent code of the pretrained generator:

$$\mathcal{L}_{latent} = \sum_m \left\| \mathbf{w}_{avg}^m - \hat{\mathbf{w}}_t^m \right\|_2. \tag{16}$$

Stage 1 hinge loss. Stage 1 additionally uses a per-vertex hinge loss to prevent excessive geometric deformation. Let $\mathbf{x}_s(v)$ and $\hat{\mathbf{x}}_t(v)$ denote the source and predicted target-age positions of vertex v, respectively. We apply this loss to all $V = 5,023$ vertices, including the neck region:

$$\mathcal{L}_{hinge} = \frac{1}{V} \sum_{v=1}^{V} \max\left(\|\mathbf{x}_s(v) - \hat{\mathbf{x}}_t(v)\|_2^2 - \tau^2, 0 \right), \tag{17}$$

where $\tau = 0.005$. We found it important to include the neck vertices; otherwise, the neck region could expand unnaturally in the later stage of training. Since the age and identity losses are computed after 2D alignment, this hinge loss provides an additional loose geometric constraint in the original 3D shape space.

Stage 2 normal-map age loss. Stage 2 additionally uses the normal-map age estimator described in Sec. 3.4. Let $\mathcal{A}_{norm}(\cdot,\cdot)$ denote this estimator, and let $\hat{\mathbf{T}}_t^{normal}$ be the predicted target-age normal map. We evaluate it on the facial skin region only:

$$\mathcal{L}_{age}^{norm} = \left(\alpha_t - \mathcal{A}_{norm}(\hat{\mathbf{T}}_t^{normal}, \mathbf{M}_{skin}) \right)^2. \tag{18}$$

Here, $\mathbf{M}_{skin}$ is the skin mask that excludes the eye region, mouth interior, ears, neck, hair, eyeglasses, and hats.

Final objectives and loss weights. The full Stage 1 objective is

$$\mathcal{L}_{Stage1} = \lambda_{hinge}\mathcal{L}_{hinge} + \lambda_{reg}^{shape}\mathcal{L}_{reg}^{shape} + \lambda_{id}\mathcal{L}_{id} + \lambda_{age}\mathcal{L}_{age} + \lambda_{cyc}\mathcal{L}_{cyc}^{shape}, \tag{19}$$

with $\lambda_{hinge} = 10^8$, $\lambda_{reg}^{shape} = 0.01$, $\lambda_{id} = 0.1$, $\lambda_{age} = 5$, and $\lambda_{cyc} = 1$.

The full Stage 2 objective is

$$\mathcal{L}_{Stage2} = \lambda_{latent}\mathcal{L}_{latent} + \lambda_{reg}^{uv}\mathcal{L}_{reg}^{uv} + \lambda_{id}\mathcal{L}_{id} + \lambda_{age}\mathcal{L}_{age} + \lambda_{norm}\mathcal{L}_{age}^{norm} + \lambda_{cyc}\mathcal{L}_{cyc}^{uv}, \tag{20}$$

with $\lambda_{latent} = 0.005$, $\lambda_{reg}^{uv} = 0.01$, $\lambda_{id} = 0.1$, $\lambda_{age} = 5$, $\lambda_{norm} = 0.1$, and $\lambda_{cyc} = 1$. All scalar weights are fixed throughout training, and we do not use any additional time-varying loss-weight schedule.

F Further implementation details

Training details. We implement our method in PyTorch and train both stages on a single NVIDIA H100 GPU with the Ranger optimizer (RAdam + Lookahead) (Liu et al., 2020; Zhang et al., 2019), as in SAM, with a learning rate of 1×10^{-4}. During training, we apply the FFHQ-style 2D alignment by translating the mesh vertices in the image plane before rasterization (ignoring rotation and scale). Stage-specific settings are as follows. For shape learning, we use a batch size of 32 and predict a residual in FLAME shape space; inference takes 0.346 ms per sample. We sample the target age α_t to cover both reconstruction and extreme ages: $\alpha_t = \alpha_s$ with probability 0.3, otherwise $\alpha_t \in [0, 10]$ (0.2), $[80, 100]$ (0.2), or $[10, 80]$ (0.3), and we use age-dependent proxy normal maps. For UV-map learning, we use a batch size of 7 and output UV intrinsic maps at α_t; we set $\alpha_t = \alpha_s$ with probability $1/3$ (as in SAM) to encourage reconstruction, and inference takes 170 ms per sample.

Stage 1 network architecture. The shape-editing network F_{shape} is implemented as a simple multilayer perceptron that predicts the residual shape parameters $\Delta\boldsymbol{\beta}$ from the source shape parameters and the source/target ages. Specifically, the input is the concatenation of $\boldsymbol{\beta}_s \in \mathbb{R}^{n_\beta}$, α_s, and α_t, and the network outputs $\Delta\boldsymbol{\beta} \in \mathbb{R}^{n_\beta}$. In our implementation, F_{shape} consists of three hidden fully connected layers of width 256 with ReLU activations, followed by a final linear layer that outputs the residual in FLAME shape space.

Stage 2 network architecture. The Stage 2 material-editing module follows the same overall design as SAM, adapted to UV-space intrinsic maps. For each map type $m \in \{diffuse, normal, specular\}$, we prepare an independent StyleGAN2 generator G^m, an inversion encoder E^m_{inv}, and an age encoder E^m_{age}. Given an input UV map $\mathbf{T}^m_s$, we first obtain its latent code using E^m_{inv}, then predict an age-dependent latent offset with E^m_{age}, and finally decode the edited UV map with G^m, as described in the main paper.

The architecture is basically the same as SAM, except that we use e4e (Tov et al., 2021) as the inversion encoder instead of pSp (Richardson et al., 2021) so that it is compatible with pivotal tuning inversion (PTI) (Roich et al., 2022). Thus, Stage 2 can be regarded as a material-wise extension of SAM to UV-space intrinsic maps, with separate networks trained for diffuse, normal, and specular maps.

Why latent editing with a GAN backbone. Stage 2 targets *asset editing* rather than free-form generation: given UV-space intrinsic maps of a specific head, we require age-conditioned updates that (i) stay close to the input maps for identity preservation and pipeline compatibility, (ii) remain consistent across map types, and (iii) support efficient inference for repeated edits. Latent-space editing with StyleGAN2, combined with inversion and pivotal tuning, provides a practical trade-off between reconstruction fidelity and controllability, enabling us to express age change as a compact latent offset while keeping the generated maps aligned with the source. Diffusion models are a promising alternative for high-fidelity synthesis and have recently been applied to face aging and 3D-aware generation, but their iterative sampling and reconstruction constraints make direct UV-map editing less practical in our current setting; we leave diffusion-based UV-map editing as future work.

G Additional qualitative results

With the same setup as the qualitative comparison in Fig. 4 of the main paper, we additionally present qualitative comparisons with all baselines (the five image-based editors, the two ID-Sculpt reconstructions, FaceG2E, and GaussianAvatar-Editor) for six more subjects in Figs. 8, 9, 10, 11, 12, and 13. Furthermore, for the "Ours" results in these examples, we show the corresponding changes in the UV maps in Figs. 15 and 16, in the same manner as Fig. 14. The trends match those in the main paper: the image-based editors depart from the original CG rendering, the ID-Sculpt reconstructions are geometrically unstable across target ages, FaceG2E changes little with the target age, and GaussianAvatar-Editor yields noisy normals in some subjects, whereas our method edits age while preserving the person-specific characteristics of the source mesh, and the age-dependent increase and decrease of wrinkles are reflected in the normal maps. AgeTransGAN has no output at age 80, as it supports target ages up to 70 only.

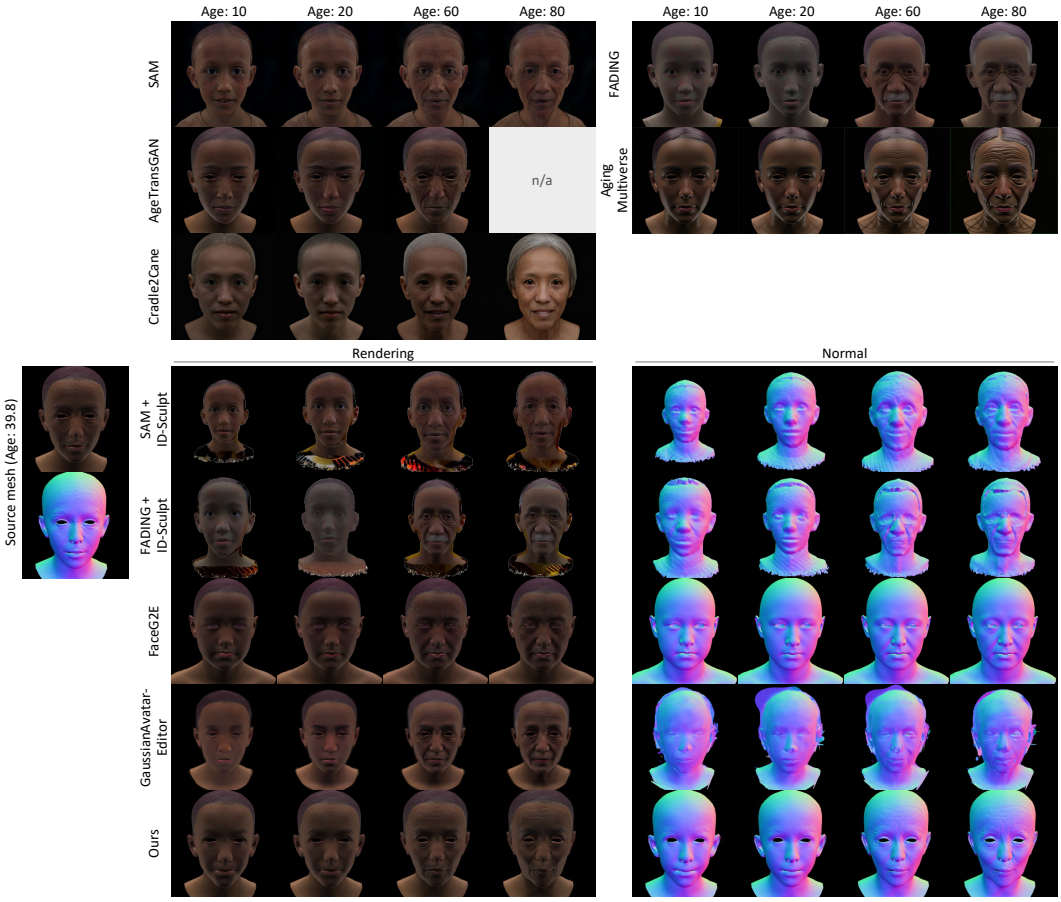

Figure 8: Qualitative comparison with baseline methods for another subject; layout as in Fig. 4. n/a: beyond AgeTransGAN's supported ages (up to 70).

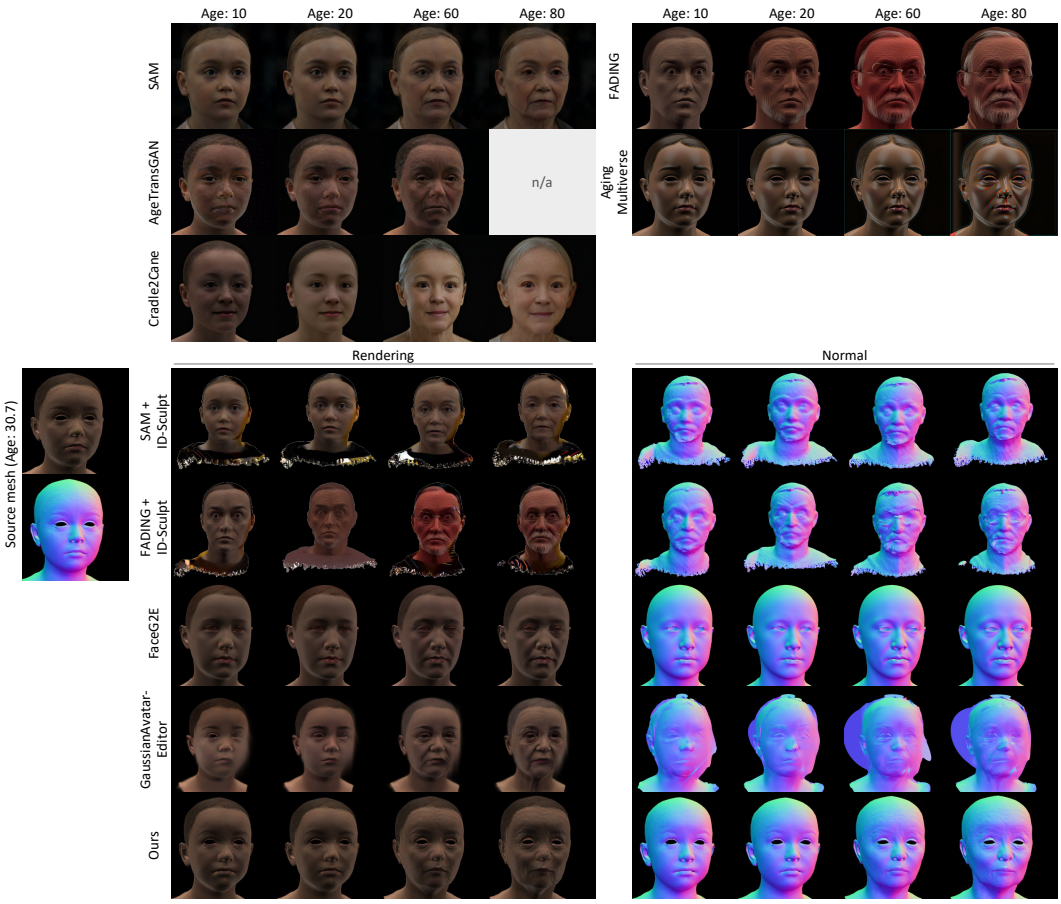

Figure 9: Qualitative comparison with baseline methods for another subject; layout as in Fig. 4. n/a: beyond AgeTransGAN's supported ages (up to 70).

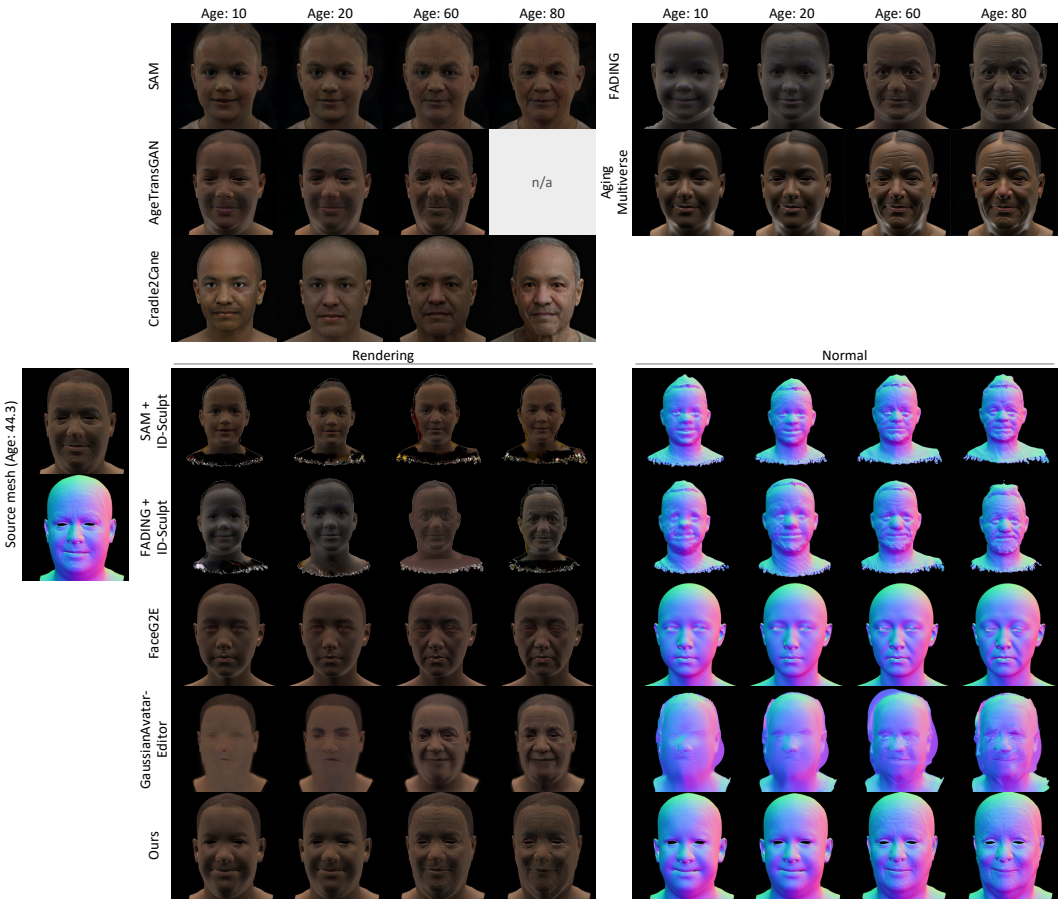

Figure 10: Qualitative comparison with baseline methods for another subject; layout as in Fig. 4. n/a: beyond AgeTransGAN's supported ages (up to 70).

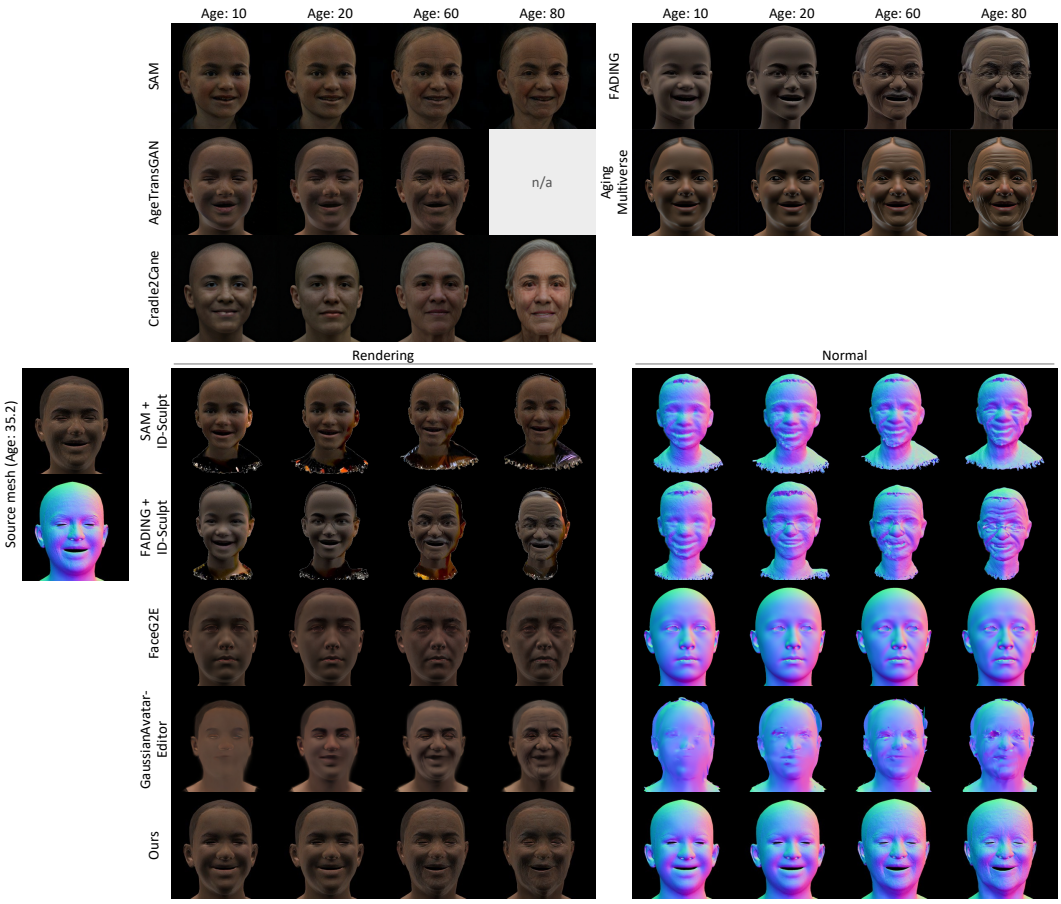

Figure 11: Qualitative comparison with baseline methods for another subject; layout as in Fig. 4. n/a: beyond AgeTransGAN's supported ages (up to 70).

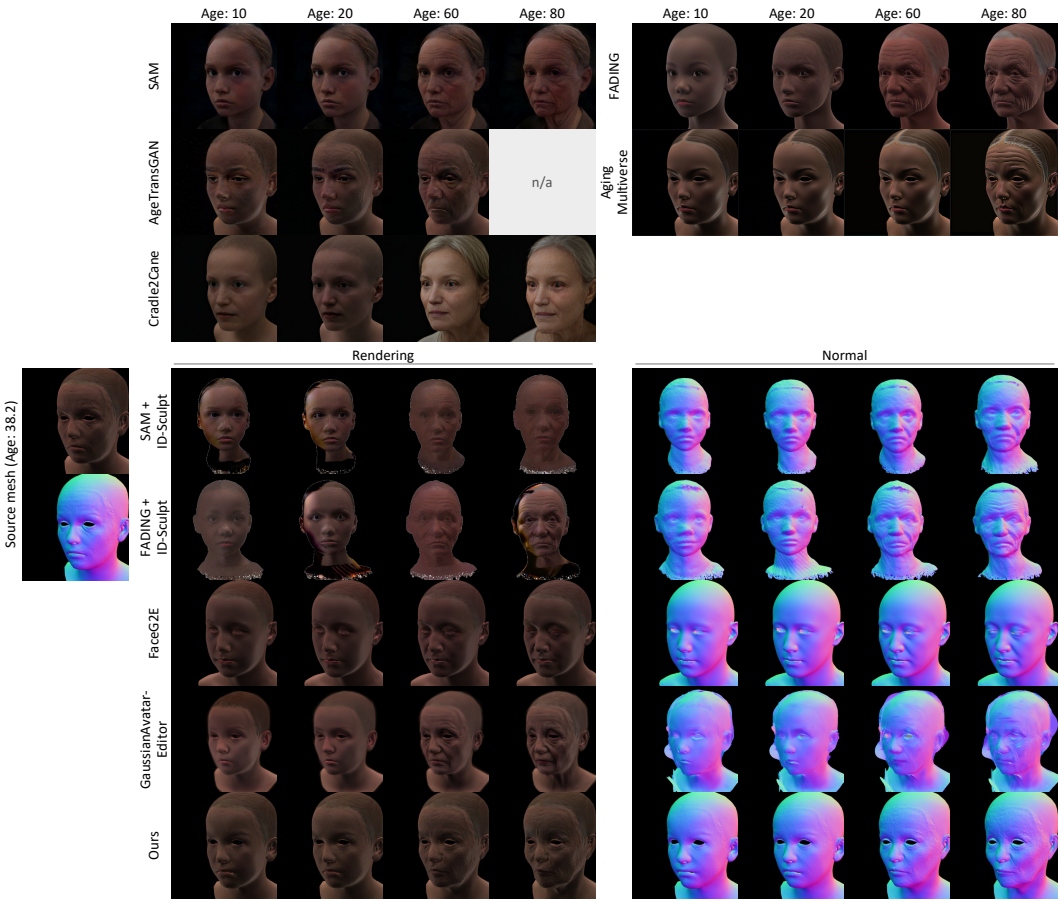

Figure 12: Qualitative comparison with baseline methods for another subject; layout as in Fig. 4. n/a: beyond AgeTransGAN's supported ages (up to 70).

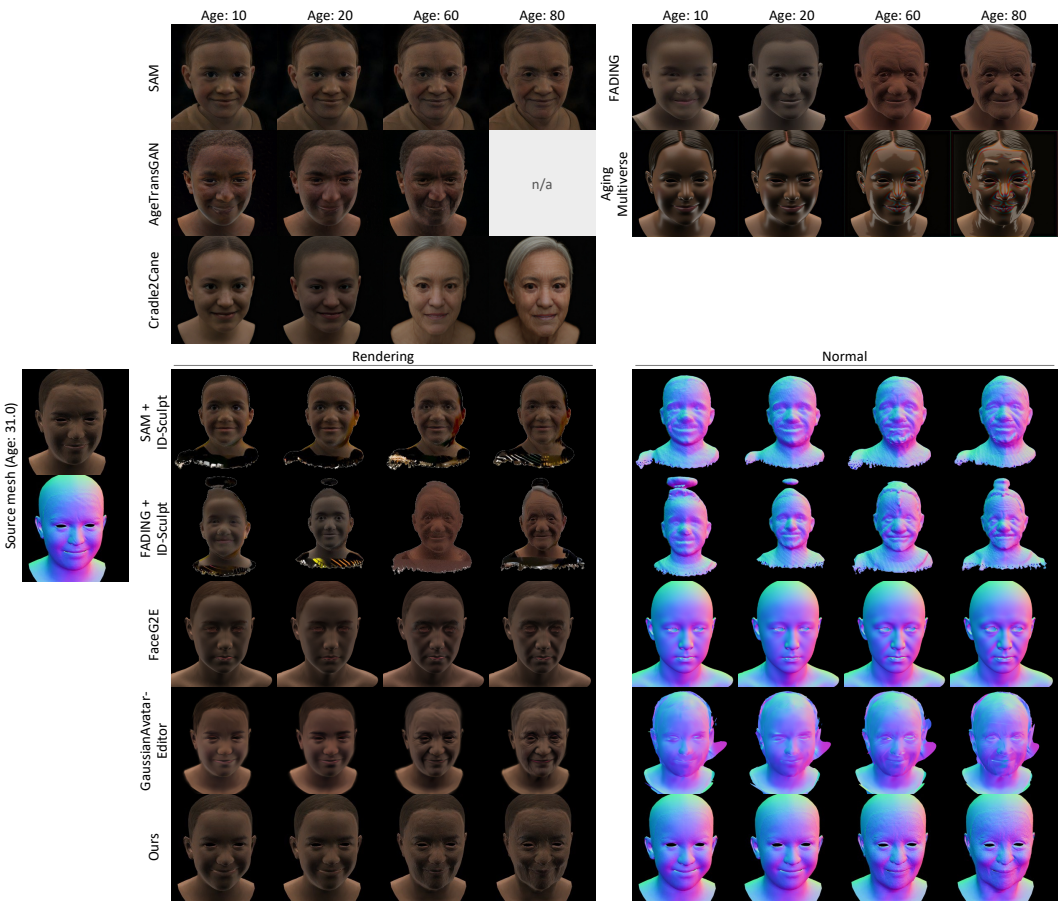

Figure 13: Qualitative comparison with baseline methods for another subject; layout as in Fig. 4. n/a: beyond AgeTransGAN's supported ages (up to 70).

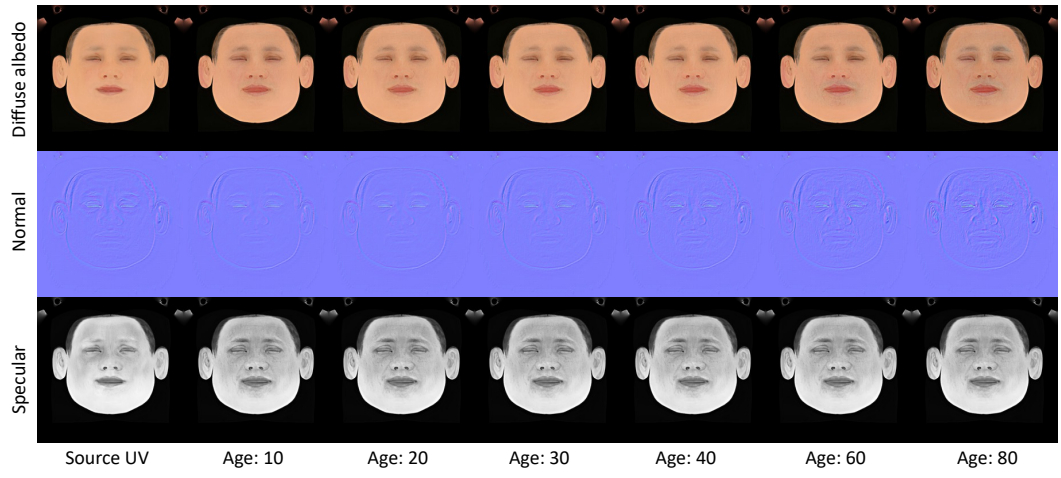

Figure 14: Aging sequence of UV maps of "Ours" used in Fig. 4.

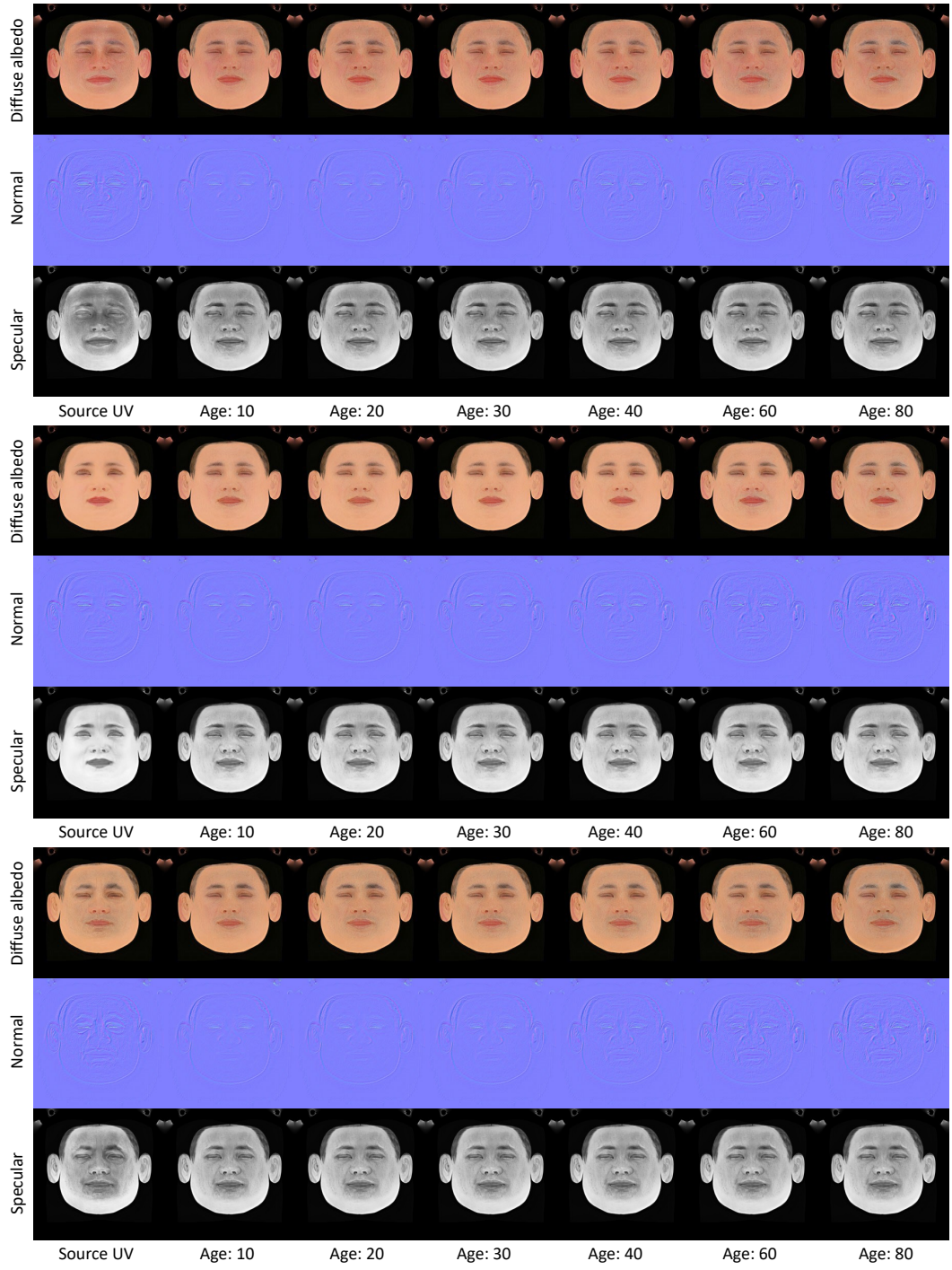

Figure 15: Aging sequences of UV maps of "Ours" used in Figs. 8, 9, and 10.

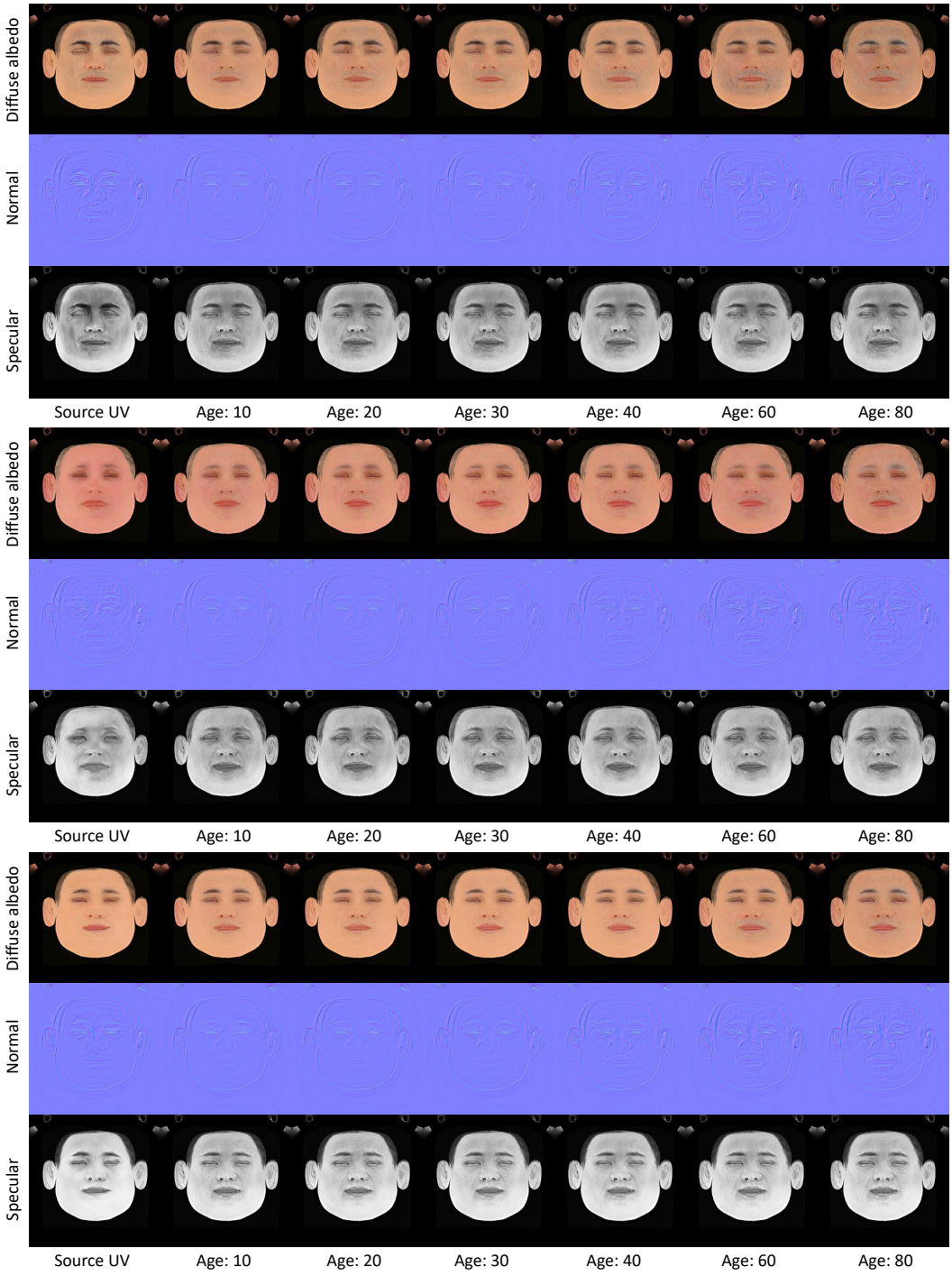

Figure 16: Aging sequences of UV maps of "Ours" used in Figs. 11, 12, and 13.