TexAvatars: Hybrid Texel-3D Representations for Stable Rigging of Photorealistic Gaussian Head Avatars

Jaeseong Lee[1*] Junyeong Ahn[2*] Taewoong Kang[1] Jaegul Choo[1]

[1]KAIST [2]Hanyang University

{webmaster, keh0t0, jchoo}@kaist.ac.kr, hewas1230@hanyang.ac.kr

Figure 1. We propose a high-fidelity head avatar method that combines analytic rigging with texel-space neural regression. Gaussian attributes are predicted in UV space and lifted to 3D via interpolated deformation fields, enabling photorealistic 3D Gaussian Splatting. Decoupling from fixed triangle bindings yields robust extrapolation to extreme expressions and poses. *Left:* unseen expressions show strong muscle activations. *Right:* wrinkles and teeth remain consistent across mirrored views, highlighting robustness and view consistency.

Abstract

Constructing drivable and photorealistic 3D head avatars has become a central task in AR/XR, enabling immersive and expressive user experiences. With the emergence of high-fidelity, efficient representations such as 3D Gaussians, recent works have pushed toward ultra-detailed head avatars. Existing approaches typically fall into two categories: rule-based analytic rigging or neural network-based deformation fields. While effective in constrained settings, both approaches often fail to generalize to unseen expressions and poses—particularly in extreme reenactment scenarios. Other methods constrain Gaussians to the global texel space of 3DMMs to reduce rendering complexity. However, these texel-based avatars tend to underutilize the underlying mesh structure. They apply minimal analytic deformation and rely heavily on neural regressors and heuristic regularization in UV space, which weakens geometric consistency and limits extrapolation to complex, out-of-distribution deformations. To address these limitations, we introduce ***TexAvatars****, a hybrid avatar representation that combines the explicit geometric grounding of analytic rigging with the spatial continuity of texel space. Our approach predicts local geometric attributes in UV space via CNNs, but drives 3D deformation through mesh-aware Jacobians, enabling smooth and semantically meaningful transitions across triangle boundaries. This hybrid design separates semantic modeling from geometric control, resulting in improved generalization, interpretability, and stability. Furthermore, TexAvatars captures fine-grained expression effects—including muscle-induced wrinkles, glabellar lines, and realistic mouth cavity geometry—with high fidelity. Our method achieves state-of-the-art performance under extreme pose and expression vari-*

∗ Equal contribution.

ations, demonstrating strong generalization in challenging head reenactment settings.

1. Introduction

High-fidelity modeling of human head avatars remains a central challenge in graphics, especially with the rising demand for over 4K-quality avatars in telepresence, visual effects, and immersive media [31]. Recent advances in 3D Gaussian Splatting (3DGS) [18] have shown impressive results in photorealistic reconstruction and real-time rendering. Building on this, 3DGS-based head avatars generally follow two paradigms: mesh-driven deformation in 3D space or texel-driven attribute regression.

Mesh-based avatars [23, 25, 35, 39] leverage 3D Morphable Models (3DMMs) [5, 26] for analytic, triangle-wise rigging. This provides stability and interpretability, but linear mesh deformation struggles to capture fine non-linear expressions like wrinkles or asymmetric activations. Furthermore, since each Gaussian is tied to a single triangle, neighboring Gaussians may deform inconsistently across boundaries. ScaffoldAvatar [1] recently addresses this by introducing spatial correlations in 3D, inspired by hierarchical frameworks [30], enabling more coherent local control of facial regions.

Conversely, texel-based avatars [24, 38, 43, 52] leverage the continuous nature of UV space, where attributes are retrieved via interpolative $\mathrm{GridSample}_{\mathrm{lerp}}$. Unlike analytic rigging methods that operate on individually bound Gaussians, this formulation introduces spatial correlation across nearby texels for smoother representations. However, most of these methods discard or minimally use 3DMM geometry and instead use CNNs to regress canonical-to-deformed offsets without explicit mesh-driven modeling. As a result, deformation is fully learned rather than analytically grounded, leading to poor extrapolation under extreme expressions (pose). Many also rely on heuristic offset regularizers, which are difficult to tune reliably.

We observe that mesh- and texel-based approaches offer complementary strengths, and our method is designed to leverage both. Mesh-based models inherently follow the mesh deformation, which provides coarse but, physically plausible initialization. Their use of local-to-global designs [23, 25, 35] enables regularization at the normalized local level, making them well-suited for imposing physically meaningful constraints such as scale bounds or spatial locality in normalized frame. These formulations also allow Gaussians to deform coherently in accordance with the surface geometry. On the other hand, texel-based approaches benefit from operating in a CNN-regressed continuous UV space with a uniform sampling grid. This formulation naturally imposes spatial correlation among neighboring attributes, which introduces a structured inductive bias into learning [44]. As a result, it facilitates better generalization and robustness in deformation prediction, especially when learned attributes are shared across similar spatial contexts.

We propose a hybrid texel–3D rigging strategy that unifies the advantages of mesh-driven and texel-driven paradigms. Rather than predicting canonical-to-deformed offsets or directly regressing in the deformed space, we regress local Gaussian attributes in UV space. A key observation is that such attributes are **not spatially coherent until transformed into global texel space**: standard $\mathrm{GridSample}_{\mathrm{lerp}}$ implicitly assumes local linearity, yet local positions and scales lack correlation in UV coordinates where adjacent texels may correspond to distant mesh regions. To resolve this, we remap triangle-driven Jacobians into texel space and apply deformation analytically prior to sampling, enabling interpolation to occur in a consistent, linearizable domain, called *Quasi-Phong Jacobian Field*. This design preserves geometric fidelity while removing the rigid binding between Gaussians and mesh triangles [23, 35], effectively binding them to a smoothly blended neighborhood. Finally, to compensate for the limitations of 3DMM-based expressions in capturing fine-scale dynamics, we introduce a lightweight latent expression embedding [10], allowing opacity and color to reflect wrinkles, folds, and occlusions without the overhead of an additional shading branch.

In summary, our contributions are threefold. First, **Local Flexible Gaussians in Texel Space**. Instead of binding Gaussians to specific mesh triangles, we regress expression-dependent local attributes directly in texel space using CNNs. This design allows Gaussians to adaptively vary within their local coordinates, while still anchored by analytic rigging for robust generalization. Second, **Texel Space Coupling for Coherent Deformation.** By remapping triangle-wise deformations into texel space and applying them to CNN-predicted local attributes, we form a continuous Gaussian field across the surface. This deformation acts like a structured kernel over the CNN output—injecting mesh-aware gradients that stabilize training and reinforce geometric consistency. Last, **Superior Generalization.** Our approach outperforms prior methods by a significant margin, particularly in cross-identity and cross-expression driving, achieving robust rendering quality.

2. Related Work

2.1. Human Head Modeling

3D head modeling forms the foundation for dynamic avatar construction. Traditional 3D Morphable Models (3DMMs) [5] parameterize facial shape and appearance using PCA, but struggle to capture articulated components such as the jaw, neck, and eyeballs. To address this, FLAME [26] integrates linear blend skinning (LBS) from

body models like SMPL [29], while still relying on PCA for expressions. More recently, NPHM [12, 13] proposed a neural SDF-based model that improves expressiveness and flexibility beyond FLAME, removing fixed topology. With the rise of neural radiance fields (NeRFs) [33], methods such as NerFACE [11], CAFCA [6], and IMAvatar [49] model dynamic heads using 3DMM-conditioned volumetric fields and blendshape-aware skinning. Other work [2, 7, 51] leverage mesh-driven deformation, while neural rendering [4, 15, 21] enhances realism and modularity. Concurrently, UV-space modeling [3, 28] enables convolution-friendly, continuous representations.

2.2. 3DMM Rigging-based Head Avatars

With the advent of 3D Gaussian Splatting (3DGS)[18], the community has shifted toward fast, rasterization-friendly forward deformations. PointAvatar[50] pioneered point-based avatars with isotropic splats driven by blendshape bases, extending IMAvatar into a splatting-friendly regime. Later works [23, 35, 39] tied Gaussians to 3DMM meshes via analytic deformations, avoiding MLPs to preserve real-time efficiency. Fully neural fields [9, 14, 37, 46, 47] regress offsets with MLPs, but incur higher costs and weaker generalization. Hybrids [25, 32] blend 3DMM rigging with neural offsets, yet still deform Gaussians independently, often lacking spatial coherence. ScaffoldAvatar [1] alleviates this by patch-based correlation with anatomical priors. In contrast, our method leverages UV-unwrapped geometry and CNNs to couple Gaussians across texel space while retaining analytic mesh rigging. Unlike mesh-bound Gaussians [23, 35], our attributes vary with expressions but remain anchored in continuous UV coordinates—yielding both flexible expressiveness and stable rigging.

2.3. Texel-Based Head Avatars

Texel-space methods exploit UV-unwrapped facial meshes for structured convolutions in dense attribute regression [22, 31, 38, 43, 45, 52]. Xiang et al. [46] initializes Gaussians in UV space but applies no further deformation, while Lombardi et al. [28] introduces analytic triangle-based rigging with UV-aligned TBN spaces. Zielonka et al. [52] regresses most attributes purely neurally, limiting semantic grounding and robustness under expression extremes. Li et al. [25] combines blendshape-based deformation with texel-space features, and Saito et al. [38], Wang et al. [45] emphasize improved appearance decoding. Kirschstein et al. [22] follows a 3D GAN pipeline, predicting Gaussian attributes via UV-space CNNs. Despite these advances, most approaches rely on neural deformation with heuristic regularization or canonical-to-deformed mappings, which can be unstable in extreme or novel settings.

Concurrent Work. TeGA [24] maps canonical UVD Gaussians to 3D with Adaptive Density Control (ADC) and extra MLPs for deformation and shading. In contrast, we regress Gaussians per frame directly in UV space and lift them analytically to 3D, avoiding ADC and learned offsets, leading to better generalization under expression extrapolation.

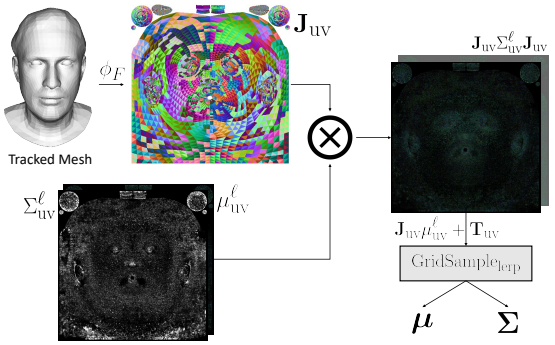

Figure 2. **Overview.** These local attributes are lifted to global space by transforming them with precomputed Jacobians $\mathbf{J}_{\text{uv}}$, remapped from the tracked mesh to texel space by remapping function ϕ_F. This results in globally coherent attributes that are continuous across surface regions via linear grid sampling.

3. Preliminaries

3.1. 3D Gaussian Splatting (3DGS)

3D Gaussian Splatting (3DGS) [18] represents a scene as a set of Gaussian primitives, each defined by a position $\mu \in \mathbb{R}^3$, rotation $r \in \mathbb{R}^4$, scale $s \in \mathbb{R}^3$, SH color $\mathbf{SH} \in \mathbb{R}^{3\times16}$, and opacity $\alpha \in \mathbb{R}$. The rotation and scale are combined into a covariance matrix $\Sigma = RSS^\top R^\top$, and rendering is performed by a tile-based differentiable rasterizer using camera parameters.

3.2. Analytic Rigging-based Gaussian Head Avatars

GaussianAvatars [35] introduces an analytic rigging method that maps local Gaussians to global 3D space via **binding inheritance**, where each Gaussian is associated with a specific mesh triangle. During deformation, each triangle propagates its transformation—using isotropic scaling, barycentric positioning, and TBN-based rotation—to its bound Gaussians. SurFhead [23] extends this idea by adopting a Jacobian-based transformation inspired by the DTF [42] formulation, replacing the scaled rotation model in GaussianAvatars. This enables modeling of both stretching and anisotropic scaling. Throughout this paper, we use superscript ℓ to denote local-space quantities. The global deformation is defined as:

$$\boldsymbol{\Sigma} = \mathbf{J}\Sigma^{\ell}\mathbf{J}^{\top}, \quad \mu = \mathbf{J}\mu^{\ell} + \mathbf{T}, \tag{1}$$

where $\mathbf{T}$ is the triangle centroid and $\mathbf{J}$ is the deformation matrix, realized as a scaled rotation in GaussianAvatars [35] and as a full Jacobian in SurFhead [23].

3.3. 3D Morphable Head Models (3DMMs) and Image Animation

3DMMs is a fundamental component in head avatar reconstruction. In our method, we assume access to per-frame tracked meshes, as obtained from a photometric tracker identical to that used in Qian et al. [35]. Each tracked mesh M is parameterized by an identity-agnostic expression parameter $\psi \in \mathbb{R}^{100}$, an identity parameter $\beta \in \mathbb{R}^{300}$, and a rigid pose parameter $\theta \in \mathbb{R}^{15}$. Since our work focuses on single-identity fitting, we omit the identity parameter β in the remainder of the formulation.

Image Animation, first introduced by Zakharov et al. [48], transfers expression and pose from a driving image to a source image while preserving the source identity. In our setting, we extract an expression-related parameter $\eta \in \mathbb{R}^{128}$ from Drobyshev et al. [10], which serves as an auxiliary signal to the 3DMM expression code for transferring fine-grained expressions, such as wrinkles, that are not well captured by standard 3DMMs.

4. Method

In this work, we introduce a novel parameterization that integrates the strengths of both mesh-based and UV-based rigging to achieve more robust and generalizable deformations. Instead of relying solely on fixed local Gaussians or black-box neural deformations, we regress local Gaussian attributes that are explicitly anchored to mesh triangles within the UV layout. These Gaussians are defined in a local texel space and subsequently mapped into a global (deformed) texel space via triangle's transformations, similar to [35]. However, our formulation enables the model to exploit both the geometric structure of the mesh topology and the spatial continuity of texel-aligned CNN features. We begin in Sec. 4.1 by describing how we regress local attributes, colors, and opacity using CNNs. In Sec. 4.2, we explain how we construct a continuous global texel space using triangle-wise Jacobians derived from the 3DMM mesh. Finally, in Sec. 4.3, we present our regularization strategy, which operates directly in this continuous texel space.

4.1. Local Texel-Attribute Parameterization

Intuition. Qian et al. [35] and Lee et al. [23] achieve efficient rendering by binding Gaussians to 3DMM meshes via analytic rigging from local to global space. However, in their setup, the local Gaussian attributes are fixed with respect to the expression parameters and merely follow the mesh deformation. This rigid coupling restricts expressiveness, as the Gaussians cannot adapt beyond the mesh articulation, leading to limited interpolation capacity.

In this work, since we perform local-to-global transformations without explicitly defining a canonical space, we use the term "global" to denote the deformed space.

In contrast, we regress local Gaussian attributes with a network conditioned on expression, rather than fixing identical attributes across expressions, while still grounding them in analytic rigging. This yields twofold benefits: (i) the learned attributes increase representational flexibility within bounded local coordinates, enhancing interpolation; and (ii) since the predictions remain governed by the smooth global Jacobian, gradients are well-conditioned, acting as a stabilizing kernel. This prevents uncontrolled drift and supports more reliable extrapolation under out-of-distribution expressions.

Analysis. In the texel-based regime, Zielonka et al. [52] and Saito et al. [38] predict canonical-to-deformed offsets or directly regressed deformed quantities. In their formulation, $G_d = G_c + \Delta G$, where G denotes Gaussian attributes and subscripts d and c denote deformed and canonical respectively, the gradient of the image-space training loss $\mathcal{L}$ with respect to the Gaussian texel map decoder f_θ's parameters θ is

$$\nabla_\theta \mathcal{L} \;=\; \frac{\partial \mathcal{L}}{\partial G_d}\,\frac{\partial G_d}{\partial \theta} \;=\; \frac{\partial \mathcal{L}}{\partial G_d}\,\frac{\partial \Delta G}{\partial \theta},$$

which directly scales with the magnitude of global-space displacements ΔG and can therefore become unstable under large deformations. To mitigate this, prior methods often resort to heuristic regularizers that shrink offsets toward zero—e.g., penalizing scale or position displacements—to avoid divergence, at the cost of reduced expressiveness.

In contrast, our local parameterization adopts $G_d = \mathbf{T}(G_\ell)$ with $G_\ell = f_\theta(x)$ denoting local Gaussian attributes regressed in normalized local space. Consequently, $\frac{\partial G_\ell}{\partial \theta}$ remains bounded, independent of the global displacement scale. The gradient then becomes

$$\nabla_\theta \mathcal{L} \;=\; \frac{\partial \mathcal{L}}{\partial G_d}\,\underbrace{\frac{\partial G_d}{\partial G_\ell}}_{\mathbf{J}}\,\frac{\partial G_\ell}{\partial \theta}.$$

Here, $\mathbf{J}$ denotes the Jacobian of the analytic mesh-based transformation $\mathbf{T}$. Since face-level deformations are dominated by rotations and only mild shear or scaling, $\mathbf{J}$ is near-isometric with singular values close to unity, i.e., $\|\mathbf{J}\| \le C$. Thus the gradient magnitude is bounded:

$$\|\nabla_\theta \mathcal{L}\| \;\le\; C \left\| \frac{\partial L}{\partial G_d} \right\| \left\| \frac{\partial G_\ell}{\partial \theta} \right\|,$$

ensuring stable training, **even G_ℓ was predicted from neural networks**. In summary, because errors remain confined to bounded local coordinates and cannot be arbitrarily amplified in global space, our formulation balances interpolation capacity with extrapolation stability.

Design. We employ two expression-dependent CNN decoders inspired by the architecture of Saito et al. [38]:

Figure 3. **Extreme Self- and Cross-Reenactment Scenario.** Our approach demonstrates significantly higher fidelity under extreme facial motions and rigid head rotations (subjects from NeRSemble [20], **FREE** corpus). Notably, our model accurately reconstructs high-frequency features such as nasolabial lines, hair strands, and detailed oral cavity structures. These results highlight the effectiveness of our hybrid texel-rigging framework in preserving semantic details even under highly expressive and challenging settings.

a view-dependent appearance decoder $\mathcal{D}_a$, and a view-independent geometry decoder $\mathcal{D}_g$. Both decoders are conditioned on the FLAME expression parameters ψ, pose parameters θ, and the image-driven expression code η from Drobyshev et al. [10]. While $\mathcal{D}_g$ predicts view-invariant geometry and opacity, $\mathcal{D}_a$ additionally takes the view direction π as input to produce view-dependent appearance. We also replace the spherical-harmonics color representation with a precomputed RGB color $\mathbf{c} \in \mathbb{R}^3$, regressed per texel.

$$G_{\text{uv}} = \mathcal{D}_g(\psi, \theta, \eta), \quad \{\mu^\ell_{\text{uv}}, r^\ell_{\text{uv}}, s^\ell_{\text{uv}}, \alpha_{\text{uv}}\} \subset G_{\text{uv}} \tag{2}$$

$$A_{\text{uv}} = \mathcal{D}_a(\psi, \theta, \eta, \pi), \quad \mathbf{c}_{\text{uv}} \in A_{\text{uv}} \tag{3}$$

Here, the subscript uv indicates texel-space attributes, while the superscript ℓ denotes local-space representations. Unlike Li et al. [24], who introduce an additional shading network to address baked-in effects such as wrinkles or ambient occlusion in Gaussian-based avatars, we find that simply predicting *dynamic* color and opacity with the extra expression code η is already sufficient to capture these appearance variations in practice, without requiring an extra shading stage.

4.2. Rigging from Local Texel to 3D space

Naïve Solution. We begin with a naive approach where local Gaussian attributes—such as rotation r^ℓ, scale s^ℓ, and position μ^ℓ—are predicted in texel space and interpolated via bilinear sampling:

$$[\mu^\ell, r^\ell, s^\ell] = \text{GridSample}_{\text{lerp}}([\mu^\ell_{\text{uv}}, r^\ell_{\text{uv}}, s^\ell_{\text{uv}}]). \tag{4}$$

The interpolated attributes are then lifted to 3D space using the affine transform of the associated triangle F:

$$\boldsymbol{\mu} = \mathbf{J}_F \mu^\ell + \mathbf{T}_F, \quad \boldsymbol{\Sigma} = \mathbf{J}_F \Sigma^\ell \mathbf{J}_F^\top. \tag{5}$$

Here, $\mathbf{J}_F$ and $\mathbf{T}_F$ denote the Jacobian and translation for face F. While texels within the same triangle share a common local frame, discontinuities arise across triangle boundaries, since each triangle defines its frame independently. This leads to ambiguous or inconsistent deformations when interpolating attributes across adjacent texels.

Figure 4. Comparison with RGCA [38]. RGCA adopts a relatively large mesh scale to stabilize training by keeping offsets small, which works well in most cases but can still cause blobby artifacts such as blurred nasal lines under strong stretching.

Prior binding strategies [35] inherit the same issue, producing piecewise-discontinuous fields.

Quasi-Phong Jacobian Field. While our analytic rigging ensures stability, the per-face Jacobians $\mathbf{J}_F$ are inherently piecewise-constant across mesh faces, leading to discontinuities at face boundaries. To address this, we introduce a smooth global texel field by unwrapping triangle Jacobians into UV space. Let $\phi_F : \Delta^2 \to \Omega$ denote the parameterization map from the barycentric simplex Δ^2 of face F to the global UV domain $\Omega \subset \mathbb{R}^2$. We define the texel-space Jacobian field by

$$\mathbf{J}_{\text{uv}}(u) = \sum_{F \in \mathcal{F}} \mathbf{1}_{\phi_F(\Delta^2)}(u)\, \mathbf{J}_F, \tag{6}$$

where $\mathbf{1}_{\phi_F(\Delta^2)}$ indicates texels covered by face F.

Although $\mathbf{J}_{\text{uv}}$ is piecewise-constant, we resample it via bilinear interpolation in UV space, yielding a continuous Jacobian field analogous to Phong surface [40]. We prioritize interpolation behavior, so we set `align_corners=False` in $\text{GridSample}_{\text{lerp}}$ *even when the sampling size matches the source resolution*, ensuring center-based bilinear interpolation rather than corner-aligned identity mapping. The final formulation is

$$\boldsymbol{\mu} = \text{GridSample}_{\text{lerp}}(\mathbf{J}_{\text{uv}}\mu^{\ell}_{\text{uv}} + \mathbf{T}_{\text{uv}}), \tag{7}$$

$$\boldsymbol{\Sigma} = \text{GridSample}_{\text{lerp}}(\mathbf{J}_{\text{uv}}\Sigma^{\ell}_{\text{uv}}\mathbf{J}^{\top}_{\text{uv}}). \tag{8}$$

Crucially, our network-predicted local scale and rotation parameters are merged into covariance form, ensuring that $\boldsymbol{\Sigma}$ still remains positive semi-definite. This guarantees that the resampled covariance lies in a smooth tangent space, enabling spatially coherent interpolation across texels. Fig. 2 provides an overview of our method.

Unlike [35], which operate on piecewise flat surface frames, our formulation blends Jacobians across adjacent faces in UV space. This *Quasi-Phong Jacobian Field* yields smooth, continuous deformations analogous to Phong shading, mitigating discontinuities and providing better-conditioned gradients for optimization.

4.3. Constraint for Local Texel Space

This step is crucial since our local space is not bound to a single triangle. Whereas Qian et al. [35] regularize local parameters within triangle coordinates, our texel-space Jacobian field allows interpolated texels to correspond to blended triangles, yielding a continuous family of global coordinate systems. Thus, local attributes only gain semantic meaning after Jacobian transformation, remaining valid even under blended interpolations. For stability, we follow Qian et al. [35] and impose per-texel lower bounds ϵ_μ and ϵ_s on predicted positions and scales:

$$\mathcal{L}_{\text{reg}_\mu} = \|\max(\mu_{\text{uv}}, \epsilon_\mu)\|, \tag{9}$$

$$\mathcal{L}_{\text{reg}_s} = \|\max(s_{\text{uv}}, \epsilon_s)\|. \tag{10}$$

4.4. Optimization

Our overall training objective integrates photometric reconstruction, perceptual similarity, and stability regularization into a unified loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \lambda_{\text{vgg}} \cdot \mathcal{L}_{\text{VGG}} + \lambda_{\text{reg}_\mu} \cdot \mathcal{L}_{\text{reg}_\mu} + \lambda_{\text{reg}_s} \cdot \mathcal{L}_{\text{reg}_s}. \tag{11}$$

The reconstruction loss $\mathcal{L}_{\text{recon}}$ balances pixel-level accuracy and structural fidelity between the rendered image I and the ground-truth $\hat{I}$:

$$\mathcal{L}_{\text{recon}} = \lambda_{L_1} \cdot \|I - \hat{I}\|_1 + \lambda_{\text{SSIM}} \cdot (1 - \text{SSIM}(I, \hat{I})). \tag{12}$$

Following 3DGS [18], we set $\lambda_{L_1} = 0.8$ and $\lambda_{\text{SSIM}} = 1 - \lambda_{L_1}$. We incorporate a perceptual loss $\mathcal{L}_{\text{VGG}}$ computed using VGG features to enhance high-frequency fidelity. However, introducing it from the beginning destabilizes training and hampers PSNR and SSIM. To mitigate this, we activate the perceptual loss only after 300K iterations (out of 600K total), using a small weight λ_{vgg} thereafter. As introduced in Sec. 4.3, we include two regularization terms to prevent degenerate Gaussians by enforcing lower bounds on predicted local means and scales. The coefficient λ_{reg} controls their influence and remains active throughout training.

5. Experiment

5.1. Baseline Nomination

We compare our method against five recent state-of-the-art baselines: GaussianAvatars [35], SurFhead [23], RGBAvatar [25], GEM [52] and Relightable Gaussian Codec Avatar [38].

Table 1. Simplified comparison across three evaluation settings: Novel Expression (**Held-out**), Novel Expression (**FREE**), and Novel View. Best and second-best scores are highlighted.

Method	Novel Expression (**Held-out**)			Novel Expression (**FREE**)			Novel View		
	LPIPS ↓	SSIM ↑	PSNR ↑	LPIPS ↓	SSIM ↑	PSNR ↑	LPIPS ↓	SSIM ↑	PSNR ↑
GaussianAvatars [35]	0.092 ± 0.018	0.897 ± 0.028	25.00 ± 2.84	0.123 ± 0.024	0.858 ± 0.030	22.01 ± 2.58	0.087 ± 0.021	0.919 ± 0.021	29.23 ± 1.82
SurFhead [23]	0.117 ± 0.040	0.884 ± 0.030	24.47 ± 2.64	0.150 ± 0.036	0.854 ± 0.035	22.06 ± 2.53	0.180 ± 0.122	0.872 ± 0.032	23.80 ± 2.78
RGBAvatar [25]	0.101 ± 0.014	0.889 ± 0.025	24.51 ± 1.58	0.134 ± 0.019	0.860 ± 0.028	21.76 ± 1.68	0.084 ± 0.013	0.924 ± 0.017	28.26 ± 1.01
GEM [52]	0.199 ± 0.025	0.885 ± 0.036	24.12 ± 1.82	0.154 ± 0.027	0.863 ± 0.034	22.01 ± 2.07	0.181 ± 0.021	0.918 ± 0.028	28.16 ± 1.72
RGCA [38]	0.050 ± 0.013	0.890 ± 0.032	25.55 ± 2.07	0.086 ± 0.025	0.854 ± 0.035	22.68 ± 2.59	0.030 ± 0.006	0.943 ± 0.013	34.24 ± 1.23
Ours	0.048 ± 0.013	0.894 ± 0.030	25.61 ± 2.10	0.077 ± 0.017	0.861 ± 0.033	22.84 ± 2.05	0.030 ± 0.005	0.947 ± 0.013	35.15 ± 1.33

Analytic Rigging Methods

- GaussianAvatars (GA) employs explicit triangle-wise rigging using the TBN frame of 3DMM mesh triangles.
- SurFhead (SF) generalizes GA with a Jacobian-based deformation that supports stretching and anisotropic scaling, introducing Jacobian Blend Skinning by blending adjacent triangle Jacobians.

Texel-Neural Hybrid Methods

- RGBAvatar (RGBA) formulates Gaussians in texel space with TBN-based binding, augmented by neural Gaussian blendshapes [32].
- Gaussian Eigen Model (GEM) regresses canonical-to-deformed Gaussian offsets via a style-based CNNs [17].
- Relightable Gaussian Codec Avatar (RGCA) directly predicts deformed Gaussians (positions as mesh-relative offsets) and stabilizes training by enlarging mesh scale, keeping offsets and scales well-conditioned.

5.2. Dataset

We utilize the NeRSemble [20] dataset, following the same train, novel view and novel expression evaluation protocol as GaussianAvatars [35]. Specifically, training uses 10 corpora and 15 cameras, holding out 1 near-frontal camera for novel-view testing, while a single corpus is reserved for novel-expression evaluation. In addition, we introduce an extra test split, denoted as **FREE**, consisting of longer and more unconstrained sequences with arbitrary expressions and head motions, to assess generalization beyond scripted motions [20].

5.3. Comparisons

As shown in Fig. 3, GEM underperforms under extreme expressions and poses, such as strong neck rotations or fine-scale details like glabellar wrinkles and nasal lines, which are tightly linked to subtle muscle activations. While GEM performs reasonably on frontal or mildly expressive frames, it struggles with extreme scenarios due to its reliance on CNN-predicted deltas—from a fixed canonical space to the deformed space—for position, scale, rotation, and opacity. Although position is transformed via a scaled rotation Jacobian, scale and rotation are directly regressed, and color remains pose-invariant, often resulting in the loss of photometric richness. RA adopts a hybrid approach by combining analytic rigging with MLP-predicted blendshapes of Gaussian attributes. However, it still suffers under unseen expressions, as the learned blendshape basis does not extrapolate well—especially when expression parameters fall far from the training manifold. SF proposes a continuous deformation field via learned Jacobian blending weights. Yet, this data-driven blending can be unstable in OOD settings due to its reliance on local examples. GA, leveraging analytic rigging, demonstrates strong generalization across expressions and poses. However, it lacks fine-scale geometric expressiveness and fails to model anisotropic deformation due to its use of isotropic scaled rotation, leading to blob-like artifacts in curved regions—as highlighted in the zoomed-in areas of Fig. 3.

In contrast, our method achieves consistently sharper reconstructions, accurately modeling fine wrinkles, realistic mouth cavities, and sharp boundaries around facial hair. We attribute this to our hybrid rigging mechanism, which combines analytic structure with the adaptability of neural regression. These qualitative improvements are corroborated by quantitative gains (Table 1)

Additional Comparison with RGCA [38]. RGCA [38] was originally proposed as relightable head avatars. Their practical trick—scaling the tracked mesh relatively large so that offsets remain small—makes training stable and interpolation reliable, which is why we selected it as a strong baseline. However, the same design backfires: under strong stretching or out-of-distribution expressions, the tiny offset budget leads to dotted artifacts (Fig. 4). In contrast, our method predicts Gaussians adaptively in normalized local UV space, achieving both stability and expressiveness without relying on such tricks.

More Qualitatives. Additional visual results, including extreme expressions and large head motions, are provided in the supplementary due to space constraints, further highlighting the robustness and fidelity of our approach.

5.4. Ablations

We curate four types of ablations including our contributions. Refer the Table 2 for each ablation. Note that qualitative ablation for VGG loss is curated in supplementary.

Global vs. Local Grid Sampling Interpolating local Gaussian attributes (e.g., position and covariance) directly in local texel space leads to perceptual blurring, especially

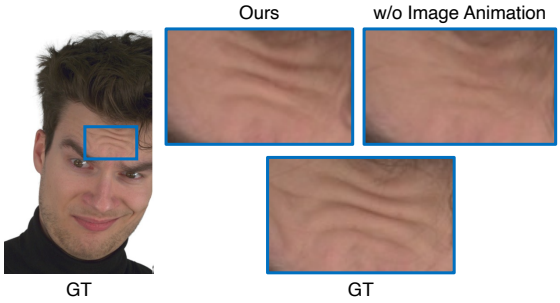

Figure 5. Effect of Image Animation Model. It enables the synthesis of details that are not explicitly represented in 3DMMs, such as wrinkles or subtle skin deformations, thereby enhancing realism under dynamic expressions.

Table 2. Ablation Studies for **FREE** testset. 'w/o' denotes removal of an individual component (not sequential subtraction); mean and standard deviation are computed across identities within a single run. Best score is highlighted.

Method	LPIPS ↓	SSIM ↑	PSNR ↑
Ours	0.077 ± 0.017	0.861 ± 0.033	22.84 ± 2.05
w/o VGG	0.096 ± 0.021	0.859 ± 0.034	22.66 ± 2.24
w/o Global UV	0.096 ± 0.028	0.855 ± 0.035	22.46 ± 2.39
w/o Jacobian	0.080 ± 0.018	0.859 ± 0.035	22.77 ± 2.13
w/o Image Animation (η)	0.078 ± 0.017	0.853 ± 0.034	22.69 ± 1.97

around regions with high geometric variation such as the mouth (Fig. 6). This occurs because these attributes are defined in triangle-specific local frames, and interpolation across triangle boundaries introduces semantic misalignment when UV-adjacent texels map to distant 3D positions. In contrast, we remap mesh-aware Jacobians into global UV space prior to interpolation, enabling geometrically coherent deformation blending. As shown in Fig. 6 and supported by Table 2, our formulation better preserves detail and spatial coherence under challenging articulation.

Jacobian vs. Scaled Rotation Without stretch-friendly Jacobian, models relying solely on scaled rotation deformation—such as in GaussianAvatars [35]—struggle to capture anisotropic changes and are prone to producing blob-like artifacts in highly deformed regions. These limitations arise from the inability of isotropic scaling to account for directional stretch, leading to unnatural shape distortion.

Effect of Image Animation Model We find it crucial for capturing fine-grained details such as wrinkles. As shown in Fig. 5, even in self-driving scenarios, missing η can result in absent wrinkles, which we attribute to the limitations of the FLAME expression space. Since the same FLAME parameters can correspond to different surface details (e.g., with or without muscle activation), we introduce η to disambiguate such cases, and observe noticeable improvements.

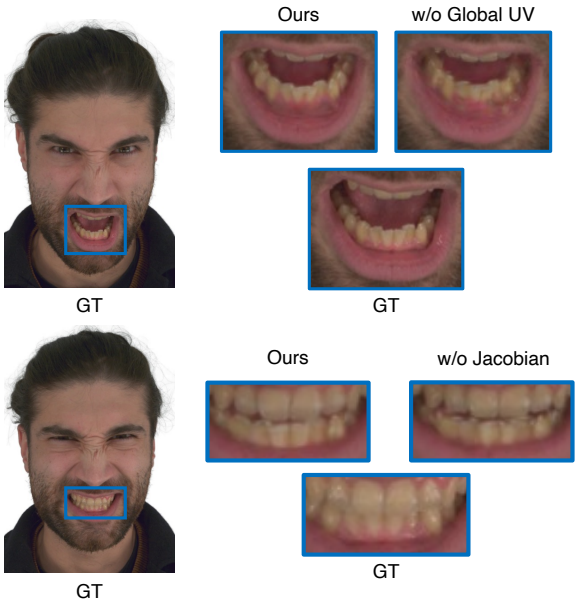

Figure 6. Effect of Global UV Sampling and Jacobian. Remapping mesh Jacobians to texel space enables smooth blending of attributes across triangle boundaries. Jacobian-based deformation effectively models stretch and anisotropic scaling while reducing blob-like artifacts.

6. Conclusion

While TexAvatars demonstrates robust extrapolation of expression and pose across both self- and cross-driving scenarios, several limitations remain. Our pipeline relies on local texel space predictions modulated by a smooth Jacobian field via linear grid sampling in tangent space, which provides spatially coherent and analytically grounded rigging. This allows us to capture high-frequency details such as wrinkles and nasolabial folds, as well as complex cavities like the mouth interior. However, the method is not without its constraints. We provide detailed discussion in the supplementary material and briefly highlight key points: 1) **Hair Motion.** Our mesh-based representation cannot model dynamic hair, leading to artifacts under motion. Future work could incorporate strand-based dynamics [27]. 2) **Tongue Articulation.** FLAME [26] does not include tongue geometry; thus, articulation inside the mouth is not captured. Extending the topology is a promising direction. 3) **Specular Effects.** High-frequency view-dependent effects like eye glints or sebum are insufficiently modeled; explicit specular rendering could address this.

4) **Fixed Number of Primitives.** For the sake of training stability, the number of Gaussians is fixed throughout optimization. While this design choice sacrifices certain high-frequency details (e.g., pores and facial hair), it also opens up avenues for future work to better capture such fine-scale structures.

References

[1] Shivangi Aneja, Sebastian Weiss, Irene Baeza, Prashanth Chandran, Gaspard Zoss, Matthias Niessner, and Derek Bradley. Scaffoldavatar: High-fidelity gaussian avatars with patch expressions. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, pages 1–11, 2025. 2, 3

[2] ShahRukh Athar, Zexiang Xu, Kalyan Sunkavalli, Eli Shechtman, and Zhixin Shu. Rignerf: Fully controllable neural 3d portraits. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pages 20364–20373, 2022. 3

[3] Ziqian Bai, Feitong Tan, Zeng Huang, Kripasindhu Sarkar, Danhang Tang, Di Qiu, Abhimitra Meka, Ruofei Du, Mingsong Dou, Sergio Orts-Escolano, et al. Learning personalized high quality volumetric head avatars from monocular rgb videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16890–16900, 2023. 3

[4] Shrisha Bharadwaj, Yufeng Zheng, Otmar Hilliges, Michael J Black, and Victoria Fernandez-Abrevaya. Flare: Fast learning of animatable and relightable mesh avatars. *arXiv preprint arXiv:2310.17519*, 2023. 3

[5] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, page 187–194, USA, 1999. ACM Press/Addison-Wesley Publishing Co. 2

[6] Marcel C Buehler, Gengyan Li, Erroll Wood, Leonhard Helminger, Xu Chen, Tanmay Shah, Daoye Wang, Stephan Garbin, Sergio Orts-Escolano, Otmar Hilliges, et al. Cafca: High-quality novel view synthesis of expressive faces from casual few-shot captures. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–12, 2024. 3

[7] Chuhan Chen, Matthew O'Toole, Gaurav Bharaj, and Pablo Garrido. Implicit neural head synthesis via controllable local deformation fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 416–426, 2023. 3

[8] Blender Online Community. Blender-a 3d modelling and rendering package. *Blender Foundation*, 2018. 2

[9] Helisa Dhamo, Yinyu Nie, Arthur Moreau, Jifei Song, Richard Shaw, Yiren Zhou, and Eduardo Pérez-Pellitero. Headgas: Real-time animatable head avatars via 3d gaussian splatting. In *European Conference on Computer Vision*, pages 459–476. Springer, 2024. 3

[10] Nikita Drobyshev, Antoni Bigata Casademunt, Konstantinos Vougioukas, Zoe Landgraf, Stavros Petridis, and Maja Pantic. Emoportraits: Emotion-enhanced multimodal one-shot head avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8498–8507, 2024. 2, 4, 5

[11] Guy Gafni, Justus Thies, Michael Zollhofer, and Matthias Nießner. Dynamic neural radiance fields for monocular 4d facial avatar reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8649–8658, 2021. 3

[12] Simon Giebenhain, Tobias Kirschstein, Markos Georgopoulos, Martin Rünz, Lourdes Agapito, and Matthias Nießner. Learning neural parametric head models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21003–21012, 2023. 3

[13] Simon Giebenhain, Tobias Kirschstein, Markos Georgopoulos, Martin Rünz, Lourdes Agapito, and Matthias Nießner. Mononphm: Dynamic head reconstruction from monocular videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10747–10758, 2024. 3

[14] Simon Giebenhain, Tobias Kirschstein, Martin Rünz, Lourdes Agapito, and Matthias Nießner. Npga: Neural parametric gaussian avatars. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–11, 2024. 3

[15] Philip-William Grassal, Malte Prinzler, Titus Leistner, Carsten Rother, Matthias Nießner, and Justus Thies. Neural head avatars from monocular rgb videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18653–18664, 2022. 3

[16] Yan-Bin Jia. Plücker coordinates for lines in the space. *Problem Solver Techniques for Applied Computer Science, Com-S-477/577 Course Handout*, 3, 2020. 2

[17] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020. 7

[18] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 2, 3, 6, 1

[19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 2

[20] Tobias Kirschstein, Shenhan Qian, Simon Giebenhain, Tim Walter, and Matthias Nießner. Nersemble: Multi-view radiance field reconstruction of human heads. *ACM Transactions on Graphics (TOG)*, 42(4):1–14, 2023. 5, 7, 1, 2, 4

[21] Tobias Kirschstein, Simon Giebenhain, and Matthias Nießner. Diffusionavatars: Deferred diffusion for high-fidelity 3d head avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5481–5492, 2024. 3

[22] Tobias Kirschstein, Simon Giebenhain, Jiapeng Tang, Markos Georgopoulos, and Matthias Nießner. Gghead: Fast and generalizable 3d gaussian heads. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–11, 2024. 3

[23] Jaeseong Lee, Taewoong Kang, Marcel Buehler, Min-Jung Kim, Sungwon Hwang, Junha Hyung, Hyojin Jang, and Jaegul Choo. Surfhead: Affine rig blending for geometrically accurate 2d gaussian surfel head avatars. In *The Thirteenth International Conference on Learning Representations*, 2025. 2, 3, 4, 6, 7

[24] Gengyan Li, Paulo Gotardo, Timo Bolkart, Stephan Garbin, Kripasindhu Sarkar, Abhimitra Meka, Alexandros Lattas, and Thabo Beeler. Tega: Texture space gaussian avatars

for high-resolution dynamic head modeling. *arXiv preprint arXiv:2505.05672*, 2025. 2, 3, 5, 1

[25] Linzhou Li, Yumeng Li, Yanlin Weng, Youyi Zheng, and Kun Zhou. Rgbavatar: Reduced gaussian blendshapes for online modeling of head avatars. *arXiv preprint arXiv:2503.12886*, 2025. 2, 3, 6, 7

[26] Tianye Li, Timo Bolkart, Michael. J. Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4D scans. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 36(6):194:1–194:17, 2017. 2, 8, 1

[27] Zhanfeng Liao, Yuelang Xu, Zhe Li, Qijing Li, Boyao Zhou, Ruifeng Bai, Di Xu, Hongwen Zhang, and Yebin Liu. Hhavatar: Gaussian head avatar with dynamic hairs. *arXiv e-prints*, pages arXiv–2312, 2023. 8, 1

[28] Stephen Lombardi, Tomas Simon, Gabriel Schwartz, Michael Zollhoefer, Yaser Sheikh, and Jason Saragih. Mixture of volumetric primitives for efficient neural rendering. *ACM Transactions on Graphics (ToG)*, 40(4):1–13, 2021. 3, 1

[29] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multi-person linear model. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 851–866. 2023. 3

[30] Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, and Bo Dai. Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20654–20664, 2024. 2

[31] Shugao Ma, Tomas Simon, Jason Saragih, Dawei Wang, Yuecheng Li, Fernando De La Torre, and Yaser Sheikh. Pixel codec avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 64–73, 2021. 2, 3

[32] Shengjie Ma, Yanlin Weng, Tianjia Shao, and Kun Zhou. 3d gaussian blendshapes for head avatar animation. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–10, 2024. 3, 7

[33] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 3

[34] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5865–5874, 2021. 1

[35] Shenhan Qian, Tobias Kirschstein, Liam Schoneveld, Davide Davoli, Simon Giebenhain, and Matthias Nießner. Gaussianavatars: Photorealistic head avatars with rigged 3d gaussians. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20299–20309, 2024. 2, 3, 4, 6, 7, 8

[36] Feisal Rasras, Stanislav Pidhorskyi, Tomas Simon, Hallison Paz, He Wen, Jason Saragih, and Javier Romero. The lips, the teeth, the tip of the tongue: Ltt tracking. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–11, 2024. 1

[37] Alfredo Rivero, ShahRukh Athar, Zhixin Shu, and Dimitris Samaras. Rig3dgs: Creating controllable portraits from casual monocular videos. In *International Conference on 3D Vision 2025*, 2024. 3

[38] Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. Relightable gaussian codec avatars. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 130–141, 2024. 2, 3, 4, 6, 7, 1

[39] Zhijing Shao, Zhaolong Wang, Zhuang Li, Duotun Wang, Xiangru Lin, Yu Zhang, Mingming Fan, and Zeyu Wang. Splattingavatar: Realistic real-time human avatars with mesh-embedded gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1606–1616, 2024. 2, 3

[40] Jingjing Shen, Thomas J Cashman, Qi Ye, Tim Hutton, Toby Sharp, Federica Bogo, Andrew Fitzgibbon, and Jamie Shotton. The phong surface: Efficient 3d model fitting using lifted optimization. In *European Conference on Computer Vision*, pages 687–703. Springer, 2020. 6

[41] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1

[42] Robert W Sumner and Jovan Popović. Deformation transfer for triangle meshes. *ACM Transactions on graphics (TOG)*, 23(3):399–405, 2004. 3

[43] Kartik Teotia, Hyeongwoo Kim, Pablo Garrido, Marc Habermann, Mohamed Elgharib, and Christian Theobalt. Gaussianheads: End-to-end learning of drivable gaussian head avatars from coarse-to-fine representations. *ACM Transactions on Graphics (TOG)*, 43(6):1–12, 2024. 2, 3, 1

[44] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018. 2

[45] Cong Wang, Di Kang, He-Yi Sun, Shen-Han Qian, Zi-Xuan Wang, Linchao Bao, and Song-Hai Zhang. Mega: Hybrid mesh-gaussian head avatar for high-fidelity rendering and head editing. *arXiv preprint arXiv:2404.19026*, 2024. 3

[46] Jun Xiang, Xuan Gao, Yudong Guo, and Juyong Zhang. Flashavatar: High-fidelity head avatar with efficient gaussian embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1802–1812, 2024. 3

[47] Yuelang Xu, Benwang Chen, Zhe Li, Hongwen Zhang, Lizhen Wang, Zerong Zheng, and Yebin Liu. Gaussian head avatar: Ultra high-fidelity head avatar via dynamic gaussians. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1931–1941, 2024. 3

[48] Egor Zakharov, Aliaksandra Shysheya, Egor Burkov, and Victor Lempitsky. Few-shot adversarial learning of realistic neural talking head models. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9459–9468, 2019. 4

[49] Yufeng Zheng, Victoria Fernández Abrevaya, Marcel C Bühler, Xu Chen, Michael J Black, and Otmar Hilliges. Im avatar: Implicit morphable head avatars from videos. In *Pro-*

ceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 13545–13555, 2022. 3

[50] Yufeng Zheng, Wang Yifan, Gordon Wetzstein, Michael J Black, and Otmar Hilliges. Pointavatar: Deformable point-based head avatars from videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 21057–21067, 2023. 3

[51] Wojciech Zielonka, Timo Bolkart, and Justus Thies. Instant volumetric head avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4574–4584, 2023. 3

[52] Wojciech Zielonka, Timo Bolkart, Thabo Beeler, and Justus Thies. Gaussian eigen models for human heads. *arXiv preprint arXiv:2407.04545*, 2024. 2, 3, 4, 6, 7, 1