Proceedings of CONF-CDS 2025 Symposium: Data Visualization Methods for Evaluation
DOI: 10.54254/2755-2721/2025.PO25728

High-Fidelity 3D Face Reconstruction from a Single Image via Gaussian Representation and Depth-Guided Refinement

Yudie Cai[1*], Ruoling Qi[1], Yaqiu Zhang[1], Xi Chu[1]

[1]*School of Software, Sichuan University, Chengdu, China*
**Corresponding Author. Email: caiyudie@stu.scu.edu.cn*

Abstract. 3D facial reconstruction has been extensively applied across various domains; however, existing methods continue to face challenges in balancing computational efficiency with geometric accuracy. Traditional techniques depend on expensive hardware and are often time-consuming, while learning-based approaches, despite notable advancements, still struggle with preserving fine details and identity-specific features. To address these limitations, a single-view 3D face reconstruction algorithm is proposed. The method employs a UNet-based architecture to project a monocular facial image into a 3D Gaussian representation, enabling high-quality novel view synthesis through Gaussian Splatting. To mitigate the scarcity of multi-view facial datasets, a large-scale dataset is synthesized using the pre-trained 3D GAN model EG3D, providing high-quality supervision for training. A refinement strategy is further introduced, incorporating face super-resolution and a six-channel image loss function to enhance visual clarity and detail fidelity. Additionally, a depth-guided refinement scheme is proposed, leveraging precise depth maps acquired from structured light scanning systems to optimize the Gaussian field for accurate geometry and fine texture reconstruction. Experimental results show that the proposed method achieves competitive performance in both reconstruction quality and computational efficiency.

Keywords: 3D face reconstruction, single view, depth guidance, novel view synthesis, detail fidelity

1. Introduction

3D facial reconstruction has emerged as a pivotal technology in computer vision, facilitating the generation of high-fidelity 3D facial morphologies from monocular 2D inputs. With the rapid advancement of deep learning, 3D facial reconstruction methods based on the Generative Adversarial Network (GAN) have achieved remarkable progress [1]. Nevertheless, these methods remain computationally expensive and exhibit compromised performance in preserving fine facial details and identity characteristics.

Recent studies demonstrate that the integration of explicit 3D facial geometric priors significantly improves 3DGS-based [2] facial reconstruction [3, 4]. Moreover, several methods directly utilize neural networks to predict 3DGS parameters for each pixel in a single-view image, allowing single-view 3D reconstruction. The Splatter-Image method [5] exemplifies this paradigm through its innovative use of the UNet architecture to directly predict 3DGS parameters from single images.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.PO25728

This approach effectively captures 3D Gaussian information to generate 3D geometric reconstructions and synthesizes multi-view images through novel view rendering.

We propose a single-view 3D facial reconstruction method by adapting 3D Gaussian Splatting (3DGS) to monocular input. A UNet-based network maps a single facial image to a 3D Gaussian field, enabling high-quality novel view synthesis via Gaussian splatting. To compensate for the lack of multi-view facial datasets, we synthesize a large-scale dataset with diverse identities using the pre-trained 3D GAN EG3D [6], providing strong supervision. To address initial rendering blur, we introduce a two-stage training strategy: the Initialization Stage produces coarse reconstruction, while the Refinement Stage enhances detail via the GFPGAN [7] face super-resolution model and a six-channel image loss balancing sharpness and fidelity.

To further enhance geometry and texture quality, we incorporate high-precision depth maps from structured light scanning as supervision, leveraging relative depth to refine the Gaussian field. For broader applicability, our framework also supports monocular depth maps [8, 9], though these are limited by geometric sparsity and reduced reconstruction accuracy compared to structured light-based depth [10].

We conduct comprehensive experiments to evaluate the efficiency and effectiveness of our method, and the results show that it achieves high-quality results comparable to state-of-the-art (SOTA) approaches.

In summary, our contributions are follows:

• We apply the Splatter-image method to complex 3D facial reconstruction and use EG3D to generate a large-scale multi-view face dataset for training and evaluation.

• We design a training strategy comprising two stages: the Initialization Stage that establishes the base geometry, followed by the Refinement Stage that incorporates GFPGAN and a novel six-channel loss function.

• We present a depth-guided refinement method that integrates surface-scanned data as high-fidelity supervision to improve depth estimation accuracy in 3DGS parameters.

2. Related works

2.1. Splatter image

Splatter-Image [5] introduces an ultra-efficient framework for monocular 3D reconstruction by extending the Gaussian Splatting paradigm to single-view settings. It predicts per-pixel 3D Gaussians using a lightweight UNet network composed entirely of 2D operators, resulting in a dense and structured Gaussian field referred to as the "Splatter Image." However, due to the limited geometric constraints in single-view input, it struggles to recover fine-grained structures and accurate depth information, particularly for objects with complex geometry or self-occlusion. The per-pixel prediction strategy may also result in ambiguity in depth estimation, leading to blurred geometry or distorted surface details.

2.2. 3D face reconstruction

Single-view 3D face reconstruction has advanced notably with the rise of 3D-aware GANs. EG3D [6] introduces a hybrid tri-plane architecture with NeRF-based rendering, achieving high-quality synthesis with reduced memory and training costs. However, it is limited to generating random identities from latent codes. To address this, 3D GAN inversion methods [11, 12] have been proposed, but often suffer from identity inconsistency and loss of fine details. Recently, 3DGS [2]

DOI: 10.54254/2755-2721/2025.PO25728

enables direct prediction of 3D representations from images, supporting efficient and geometry-free reconstruction, and offering a promising alternative for high-fidelity face modeling.

2.3. Fringe projection profilometry

Fringe Projection Profilometry (FPP [13]), also known as Digital Fringe Projection Profilometry, is a classical and widely used technique in structured light-based 3D reconstruction. It is favored for its low cost, fast measurement, and wide field of view. The core principle of FPP is to recover 3D surface information by analyzing the phase changes caused by the deformation of projected fringe patterns on the object surface. In our study, we employ a facial scanning device based on FPP to obtain high-precision facial depth maps, as shown in Figure 1.

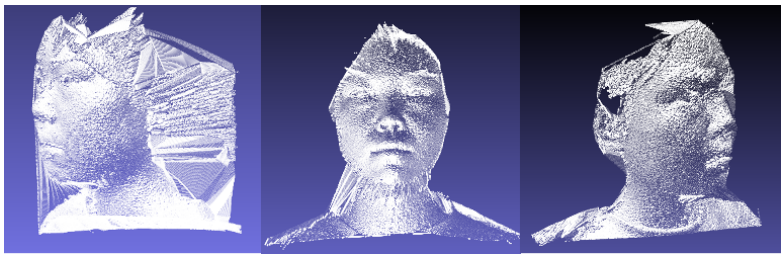

Figure 1. Facial scan depth map

3. Methods

As shown in Figure 2, we present a unified training strategy comprising a two-stage reconstruction framework and a depth-guided fine-tuning module to enhance the accuracy and visual quality of monocular 3D face reconstruction.

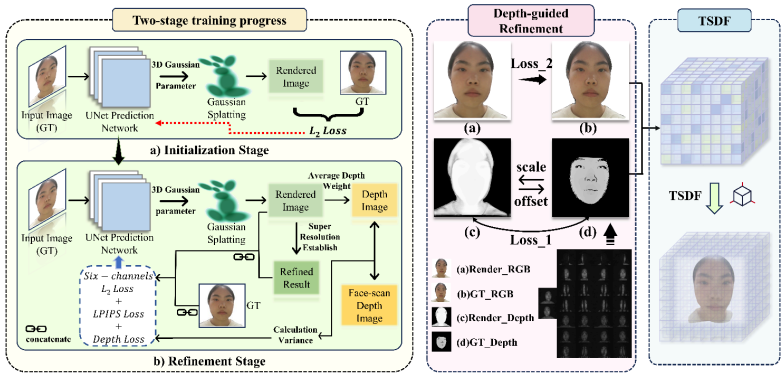

Figure 2. Overview of the proposed framework. A UNet-based network predicts 3D Gaussians from a single-view facial image. The two-stage training includes coarse supervision with EG3D-generated data and refinement with super-resolution and six-channel loss. Depth-guided optimization improves geometry and texture. TSDF is used to reconstruct the 3D facial point cloud

In the Initialization Stage, a UNet-based network predicts per-pixel 3D Gaussian parameters, which are optimized via an L_2 rendering loss to ensure consistency with the input view (Section 3.1). While effective, this stage may yield blurred results under large viewpoint changes. The Refinement Stage addresses this by introducing a refinement module that applies super-resolution via GFPGAN and further optimizes perceptual quality using a six-channel L_2 loss and $LPIPS$ loss (Section 3.2). Subsequently, we perform fast depth-guided fine-tuning of the model using a

DOI: 10.54254/2755-2721/2025.PO25728

single frontal depth map captured by the device (which may have missing facial structures due to hair occlusion and other factors), as described in Section 3.3. This enables the model to quickly adapt to the individual's true facial features. Finally, we employ the TSDF [14] module (Section 3.4) to reconstruct the 3D face mesh from the multi-view images and depth maps generated by the optimally refined model.

3.1. The initialization stage

In this stage, a UNet-based network predicts 3D Gaussian parameters from a 2D image, forming an initial 3D face. These parameters are rendered using the 3DGS pipeline under different camera poses to synthesize novel views. The model is trained by minimizing the difference between rendered and target images, refining geometry through iterative optimization.

We use a synthetic multi-view dataset $D = \{(I, J, \pi)\}$ generated by EG3D, where I is the source image, J the target image, and π the viewpoint transformation. The training minimizes the L_2 loss between rendered and target images:

$$L_{\text{initialization}} = \frac{1}{|D|} \sum_{(I,J,\pi)\in D} \|J - R(P(I), \pi)\|^2 \tag{1}$$

Here, $P(I)$ predicts 3D Gaussians from I, and R renders them under transformation π. This initial stage lays the foundation for later refinement.

3.2. The refinement stage

While the Initialization Stage provides a coarse 3D reconstruction, the results often degrade under large viewpoint changes. To enhance image sharpness and facial detail, we introduce a Refinement Stage.

Face Super-Resolution. We apply GFPGAN, a pre-trained face restoration model, to enhance the outputs from the Initialization Stage. This improves clarity and restores fine details, especially under challenging poses.

Six-Channel Image Loss. Since GFPGAN's weights are fixed, it may cause slight detail loss. To address this, we concatenate the target image J with the super-resolved output $SR(R')$, forming six-channel inputs for computing the loss:

$$R' = R(P(I), \pi) \tag{2}$$

$$L_{channel} = \|K(J, J) - K(R', SR(R'))\|^2$$

where $K(A, B) = \text{concat}(A, B) \in \mathrm{R}^{H\times W\times 6}$.

$LPIPS$ Loss. To improve perceptual quality, we use $LPIPS$ loss between $SR(R')$ and target J:

DOI: 10.54254/2755-2721/2025.PO25728

$$L_{\text{LPIPS}} = LPIPS(J,\ SR(R')) \tag{3}$$

Total Loss. The overall loss combines both terms with weigh $\lambda_1 = 0.99$ and $\lambda_2 = 0.01$:

$$L_{\text{refinement}} = \frac{1}{|D|} \sum_{(I,J,\pi)\in D} \left(\lambda_1 L_{\text{channel}} + \lambda_2 L_{\text{LPIPS}}\right) \tag{4}$$

This stage improves reconstruction quality by refining textures and preserving detailed features.

3.3. Fine-tuning with depth refinement

In this fine-tuning stage, we optimize the parameters of the 3D Gaussian Splatting (3DGS) representation — including splat positions, sizes, and opacities — while keeping the neural network weights fixed. This allows precise geometric refinement without retraining the network. Although TSDF-based fusion methods can combine depth and color, they lack the ability to fine-tune expressive 3DGS parameters. Moreover, synthetic depth images rendered from EG3D often contain artifacts, making real single-view depth data essential for accurate geometry correction.In the following section, we detail how partial depth priors and corresponding loss functions are employed to optimize geometry.

(1) Building Partial Depth Map Priors: Structured-light depth cameras provide accurate but incomplete facial depth maps due to occlusions like hair and shadows. These partial maps serve as geometric priors to correct errors in Gaussian-based reconstructions.

We back-project raw depth D_{raw} to 3D points using the depth camera intrinsic matrix K_d :

$$K_d = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

$$P_d = D_{\text{raw}}\,(u,v) \cdot K_d^{-1} \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{5}$$

Then transform these points to the RGB camera coordinate system: $\mathbf{P}_{\text{rgb}} = R\mathbf{P}_d + T$, where R and T are extrinsic rotation and translation.

We interpolate the projected points to create a dense partial depth map D_{partial} in the RGB view. To resolve scale ambiguity with the rendered depth D_{rendered} , we normalize both using median and percentile statistics:

$$t\,(D) = \text{median}\,(D)$$

DOI: 10.54254/2755-2721/2025.PO25728

$$s\left(D\right)=P_{95}\left(D\right)-P_{10}\left(D\right) \tag{6}$$

Here, $P_{95}\left(D\right)$ and $P_{10}\left(D\right)$ are the 95th and 10th percentiles of D. Percentile $P_{p}\left(D\right)$ means the depth value below which $p\%$ of data falls, helping reduce outlier effects.

$$\text{scale}=\frac{s(D_{\text{partial}})+\epsilon}{s(D_{\text{rendered}})+\epsilon}$$

$$offset=t\left(D_{\text{rendered}}\right)-scale\times t\left(D_{\text{partial}}\right) \tag{7}$$

ϵ is a small constant (e.g., 10^{-6}) to avoid division by zero.
The aligned depth map is:

$$D_{\text{partial}}^{*}=scale\times D_{\text{partial}}+offset \tag{8}$$

This aligned depth map is used as supervision in optimization.

(2) Depth Rendering via Extended Rasterization: 3D Gaussian Splatting (3DGS) uses a rasterization pipeline to project unstructured Gaussian points onto 2D screens efficiently. Each splat contributes color c_i and opacity α_i, with alpha blending:

$$C=\sum_{i\in N}c_i\alpha_i T_i$$

$$T_i=\prod_{j=1}^{i-1}\left(1-\alpha_j\right) \tag{9}$$

To render depth maps, we substitute color c_i with splat depth d_i:

$$D=\sum_{i\in N}d_i\alpha_i T_i \tag{10}$$

This reuse of α_i and T_i incurs minimal overhead. The depth map D is supervised by an L_1 loss with the aligned ground-truth depth:

$$\mathscr{L}_{\text{depth}}=\left\|D_{\text{rendered}}-D_{\text{partial}}^{*}\right\|_1 \tag{11}$$

DOI: 10.54254/2755-2721/2025.PO25728

(3) Hyperparameter Optimization and Loss Construction: We apply dynamic weighting to balance depth and RGB losses during optimization. Depth loss weight decays exponentially to prioritize early geometry correction, while RGB loss weight grows to enhance texture refinement later.

$$\lambda_{\text{depth}} = \exp(-k \cdot \text{iter})^{p_d}$$

$$\lambda_{\text{RGB}} = 1 - \exp(-\alpha \cdot iter)^{p_r} \tag{12}$$

With $k = 0.05$, $p_d = 0.7$, $\alpha = 0.1$, and $p_r = 0.8$, this strategy gradually shifts emphasis from geometry to appearance. The total loss is:

$$\mathcal{L} = \lambda_{\text{RGB}} \| I_{\text{rendered}} - I_{\text{gt}} \|_2 + \lambda_{\text{depth}} \mathcal{L}_{\text{depth}} \tag{13}$$

3.4. Truncated Signed Distance Function (TSDF) module

TSDF is a volumetric representation that encodes the signed distance from each voxel to the nearest surface, truncated to a fixed range. It enables smooth 3D surface reconstruction by fusing multi-view depth data.

After training and optional depth-guided fine-tuning, we obtain multi-view RGB images and corresponding depth maps. These are fused via TSDF to generate a complete 3D face mesh, enhancing geometric consistency and detail completeness.

4. Experiments

4.1. Datasets

We generate a multi-view facial dataset using EG3D, removing background artifacts. The dataset contains 2000 identities with 120 views each, split into training (1400), validation, and test sets (400). Additionally, we capture RGB images and depth maps of individuals from five views using a structured light scanner. The frontal depth maps guide fine-tuning, and RGB images evaluate performance.

DOI: 10.54254/2755-2721/2025.PO25728

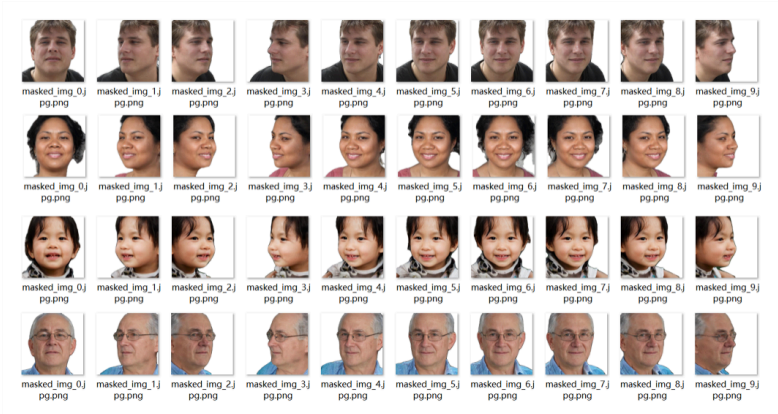

Figure 3. EG3D-generated facial dataset

4.2. Comparison on dataset

We compare the model trained through the Initialization and Refinement Stages with the e4e [15], pSp [16], ReStyle_e4e [17], ReStyle_pSp [17], PTI [18], SPI [19], TriPlaneNetv2 [20], and GOAE [21] methods. For fairness, depth-guided fine-tuning is excluded from our model during this evaluation. We do not include comparisons with recent general 3D reconstruction methods such as GAGAvatar [22], Avat3r [23] and Portrait4Dv2 [24], as they typically rely on multi-view or non-frontal inputs, or target dynamic 4D reconstruction rather than static single-image facial reconstruction.

Quantitative analysis: We evaluate reconstruction quality using multiple metrics. Pixel accuracy is assessed by MSE, $PSNR$, and $SSIM$, while perceptual similarity is measured with $LPIPS$. Identity preservation is verified via ArcFace-based ID score. Additionally, FID evaluates image realism and diversity. Table 1 presents comparisons across five viewpoints on the EG3D dataset.

Table 1. Quantitative comparison of publicly available 3D face reconstruction methods across five viewpoints on the EG3D synthetic dataset

Level	PSNR↑	SSIM↑	LPIPS↓	MSE↓	ID Loss↑	FID↓
e4e	15.6996	0.6273	0.4006	0.034105	0.7595	77.7426
pSp	15.2969	0.6280	0.3967	0.032451	0.7619	67.7643
Restyle_e4e	16.5787	0.6497	0.3856	0.031046	0.7916	65.3841
Restyle_pSp	16.0658	0.6470	0.3964	0.031555	0.7653	67.4145
PTI	16.2533	0.6568	0.3507	0.026559	0.8233	58.9627
SPI	18.0736	0.6903	0.3492	0.022610	0.8346	60.1195
TriPlaneNetv2	17.3300	0.6982	0.3361	0.011741	0.8295	78.6360
GOAE	20.6941	0.6932	0.3389	0.010238	0.8586	64.3422
Ours	21.2589	0.7789	0.2124	0.008094	0.9403	58.7468

Qualitative analysis: We compare reconstruction results on the EG3D dataset from five viewpoints: front, left/right profile (30°), side (90°), and top-down. Figure 4 shows visual comparisons with other methods.

DOI: 10.54254/2755-2721/2025.PO25728



Figure 4. Qualitative comparison of existing 3D face reconstruction methods on the EG3D synthetic dataset. All models are evaluated under five consistent viewpoints for visual comparison

4.3. Ablation study

To evaluate the effectiveness of our two-stage training strategy, we compare results with and without the Refinement Stage using data from the facial scanning device.

Figure 5 shows that the full pipeline produces better geometric results. We also report quantitative metrics for each stage in Table 2, confirming the benefits of our progressive design.

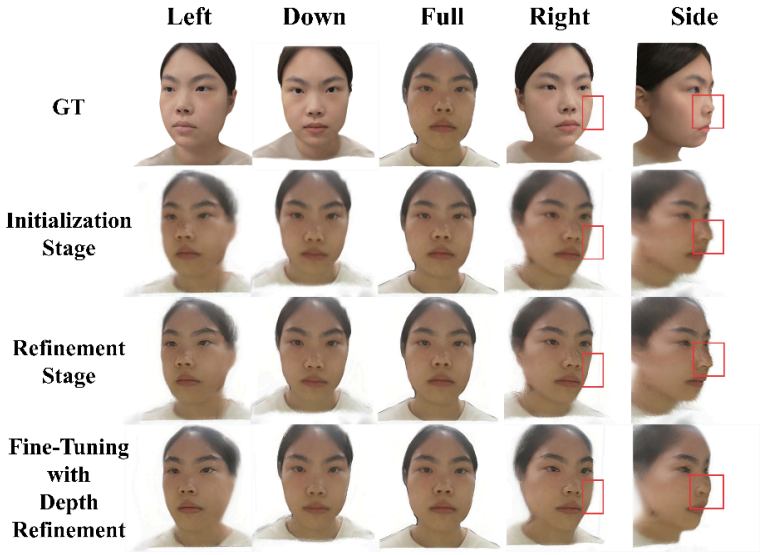

Figure 5. Comparison of rendered images among Initialization Stage, Refinement Stage and fine-tuning with depth refinement on the dataset obtained from the facial scanning device

Table 2. Ablation study on three progressive training stages using the dataset obtained from the facial scanning device

Level	PSNR↑	SSIM↑	LPIPS↓	MSE↓	ID Loss↑	FID↓
Initialization Stage	19.3086	0.8377	0.3044	0.018493	0.8707	77.3584
Refinement Stage	19.4367	0.8532	0.2992	0.018157	0.8924	61.5318
Fine-Tuning with Depth Refinement	20.9637	0.8441	0.2866	0.017977	0.9028	60.3452

DOI: 10.54254/2755-2721/2025.PO25728

5. Conclusion and limitation

We propose a novel 3D facial reconstruction strategy integrating 3DGS, Splatter-Image, and depth-guided refinement to improve accuracy and quality from single-view input. Leveraging EG3D-generated multi-view data and a two-stage training strategy, our method achieves strong performance across multiple metrics.

However, limitations remain. Under extreme viewpoints (e.g., near-profile or top-down), reconstructed faces can show artifacts and unrealistic details due to monocular input ambiguity, incomplete noisy depth from occlusions, and limited extreme-pose data in training.

Future work includes integrating multi-view data, improving depth priors with robust sensing, and applying stronger geometric constraints to enhance reconstruction under challenging views. Advances in neural rendering and geometry-aware models offer promising avenues to mitigate these issues.

References

[1] Diao, H., Jiang, X., Fan, Y., Li, M. and Wu, H. (2024) 3D Face Reconstruction Based on a Single Image: A Review. IEEE Access.
[2] Kerbl, B., Kopanas, G., Leimkühler, T. and Drettakis, G. (2023) 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM Transactions on Graphics, 42(4), 139–1.
[3] Zhao, Z., Bao, Z., Li, Q., Qiu, G. and Liu, K. (2024) PS-Avatar: A Point-Based Morphable Shape Model for Real-Time Head Avatar Creation with 3D Gaussian Splatting. arXiv preprint, arXiv: 2401.12900.
[4] Lan, Y., Tan, F., Qiu, D., Xu, Q., Genova, K., Huang, Z. et al. (2023) Gaussian3Diff: 3D Gaussian Diffusion for 3D Full Head Synthesis and Editing. arXiv preprint, arXiv: 2305.17119.
[5] Szymanowicz, S., Rupprecht, C. and Vedaldi, A. (2024) Splatter Image: Ultra-Fast Single-View 3D Reconstruction. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10208–10217.
[6] Chan, E.R., Lin, C.Z., Nagano, M.A., Pan, B., De Mello, S. et al. (2022) Efficient Geometry-Aware 3D Generative Adversarial Networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 16123–16133.
[7] Wang, X., Li, Y., Zhang, H. and Shan, Y. (2021) Towards Real-World Blind Face Restoration with Generative Facial Prior. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9168–9178.
[8] Bhat, S.F., Birkl, R., Wofk, D., Wonka, P. and Müller, M. (2023) ZoeDepth: Zero-Shot Transfer by Combining Relative and Metric Depth. arXiv preprint, arXiv: 2302.12288.
[9] Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V. et al. (2023) DINOv2: Learning Robust Visual Features without Supervision. arXiv preprint, arXiv: 2304.07193.
[10] Zhan, C., Zhang, Y., Lin, Y., Wang, G. and Wang, H. (2025) RDG-GS: Relative Depth Guidance with Gaussian Splatting for Real-Time Sparse-View 3D Rendering. arXiv preprint, arXiv: 2501.02568..
[11] Ko, J., Cho, K., Choi, D., Ryoo, K. and Kim, S. (2023) 3D GAN Inversion with Pose Optimization. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2967–2976.
[12] Yuan, Z., Zhu, Y., Li, Y., Liu, H. and Yuan, C. (2023) Make Encoder Great Again in 3D GAN Inversion through Geometry and Occlusion-Aware Encoding. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2437–2447.
[13] Xu, J. and Zhang, S. (2020) Status, Challenges, and Future Perspectives of Fringe Projection Profilometry. Optics and Lasers in Engineering, 135, 106193.
[14] Zeng, A., Song, S., Nießner, M., Fisher, M., Xiao, J. and Funkhouser, T. (2017) 3DMatch: Learning Local Geometric Descriptors from RGB-D Reconstructions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
[15] Tov, O., Alaluf, Y., Nitzan, Y., Patashnik, O. and Cohen-Or, D. (2021) Designing an Encoder for StyleGAN Image Manipulation. arXiv preprint, arXiv: 2102.02766.
[16] Richardson, E., Alaluf, Y., Patashnik, O., Nitzan, Y., Azar, Y., Shapiro, S. and Cohen-Or, D. (2021) Encoding in Style: A StyleGAN Encoder for Image-to-Image Translation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June.

DOI: 10.54254/2755-2721/2025.PO25728

[17] Alaluf, Y., Patashnik, O. and Cohen-Or, D. (2021) ReStyle: A Residual-Based StyleGAN Encoder via Iterative Refinement. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October.
[18] Roich, D., Mokady, R., Bermano, A.H. and Cohen-Or, D. (2021) Pivotal Tuning for Latent-Based Editing of Real Images. ACM Transactions on Graphics.
[19] Yin, Y., Zhang, Y., Wang, X., Wang, T., Li, X., Gong, Y., Fan, Y., Cun, X., Shan, Y., Oztireli, C., & Yang, Y. (2022). 3D GAN inversion with facial symmetry prior. arXiv preprint arXiv: 2211.16927.
[20] Bhattarai, A. R., Nießner, M., & Sevastopolsky, A. (2023). TriplaneNet: An encoder for EG3D inversion. arXiv preprint arXiv: 2303.13497.
[21] Yuan, Z., Zhu, Y., Li, Y., Liu, H., & Yuan, C. (2023). Make encoder great again in 3D GAN inversion through geometry and occlusion-aware encoding. arXiv preprint arXiv: 2303.12326.
[22] Chu, X., & Harada, T. (2024). Generalizable and animatable Gaussian head avatar. In *Proceedings of the Thirty-eighth Annual Conference on Neural Information Processing Systems (NeurIPS)*.
[23] Kirschstein, T., Romero, J., Sevastopolsky, A., Nießner, M., & Saito, S. (2025). AVAT3R: Large animatable Gaussian reconstruction model for high-fidelity 3D head avatars. arXiv preprint arXiv: 2405.20717.
[24] Deng, Y., Wang, D., & Wang, B. (2024). Portrait4D-v2: Pseudo multi-view data creates better 4D head synthesizer. arXiv preprint arXiv: 2406.04494.