

# GEAR: Reconstruction of Classical Paintings via Geometry Grounding and Appearance Restitution

Qinyu Zhang  
zhangqinyu@stumail.nwu.edu.cn  
Northwest University  
Xi'an, China

Xinda Liu\*  
liuxinda@nwu.edu.cn  
Northwest University  
Xi'an, China

Yunchen Li  
liyunchen@stumail.nwu.edu.cn  
Northwest University  
Xi'an, China

Yunzhuo Liu  
liuyunzhuo\_stu@stumail.nwu.edu.cn  
Northwest University  
Xi'an, China

Chenxi Hu  
chenxi\_hu@stumail.nwu.edu.cn  
Northwest University  
Xi'an, China

Kang Li  
likang@nwu.edu.cn  
Northwest University  
Xi'an, China

Guohua Geng  
ghgeng@nwu.edu.cn  
Northwest University  
Xi'an, China

## Abstract

Classical paintings preserve rich spatial, cultural, and historical content, making their reconstruction as explorable 3D scenes valuable for digital preservation, immersive exhibition, and cultural engagement. Yet, unlike photographs, they often depict scenes in a single-view, stylized manner, with weak perspective, lighting, and depth cues. Existing 3D reconstruction methods are largely built on natural-image priors, making it difficult to recover geometrically plausible and visually faithful 3D representations from such inputs. To address this challenge, we introduce **Classical Painting-to-3D (CP3D)**, a new task that aims to recover a 3D representation from a single classical painting while jointly ensuring geometric plausibility, appearance fidelity to the source artwork, and plausible novel-view synthesis. We further propose **GeAR**, a training-free two-stage framework for Geometry Grounding and Appearance Restitution. GeAR first converts the input painting into a geometry-grounded representation with more coherent shading and illumination cues, improving the stability of 3D Gaussian reconstruction. It then restores artwork-faithful appearance across views under spatial constraints and multi-view consistency, recovering the painterly textures and details weakened during grounding. In addition, we construct **HeriArch**, a curated benchmark of 10,160 high-resolution classical artworks for systematic evaluation of CP3D. Extensive experiments and user studies show that GeAR consistently outperforms strong baselines in geometric plausibility, appearance fidelity, and human preference. Code is publicly available at <https://github.com/liuxindazz/GEAR>.

\*Corresponding author.

## CCS Concepts

• **Applied computing** → **Arts and humanities**; • **Computing methodologies** → **Computer vision problems**.

## Keywords

Classical Painting-to-3D, 3D Gaussian Splatting, Cultural Heritage

### ACM Reference Format:

Qinyu Zhang, Xinda Liu, Yunchen Li, Yunzhuo Liu, Chenxi Hu, Kang Li, and Guohua Geng. 2026. GEAR: Reconstruction of Classical Paintings via Geometry Grounding and Appearance Restitution. In *Proceedings of the 34th ACM International Conference on Multimedia (MM '26)*, November 10–14, 2026, Rio de Janeiro, Brazil. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3767308.3835699>

## 1 Introduction

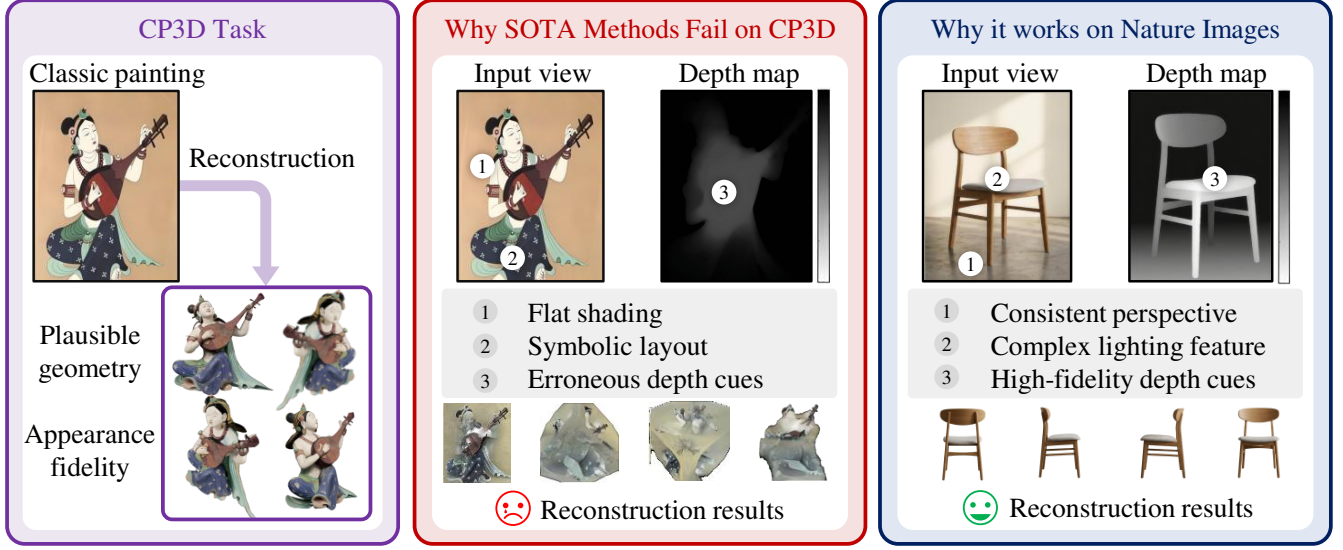
Classical paintings are enduring visual records of lost worlds, preserving not only people and events but also how past cultures organized space, ritual, and everyday life. Reconstructing them as explorable 3D scenes is therefore valuable not only for recovering the spatial experience embedded in the artwork, but also for supporting digital preservation, immersive exhibition, and new forms of cultural engagement [6, 24, 34]. Unlike photographs, however, classical paintings often rely on flattened composition, stylized perspective, and symbolic depiction rather than physically consistent scene geometry [21, 23]. Consequently, their reconstruction cannot be handled by directly applying existing single-image 3D reconstruction methods developed for natural images.

Motivated by this mismatch, we introduce a new task, Classical Painting-to-3D (**CP3D**), for inferring a plausible 3D representation from a single classical painting. CP3D requires a 3D representation that jointly satisfies geometric plausibility and appearance fidelity to the source painting, while enabling plausible novel-view synthesis (Fig. 1 (a)). This dual requirement makes CP3D fundamentally different from standard single-image 3D reconstruction, which assumes to be a camera observation of a real, physically consistent 3D scene rather than a planar depiction shaped by artistic intent and stylized representation.



This work is licensed under a Creative Commons Attribution 4.0 International License. *MM '26, Rio de Janeiro, Brazil.*

© 2026 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-2213-4/2026/11  
<https://doi.org/10.1145/3767308.3835699>



**Figure 1: Motivation of Classical Painting-to-3D (CP3D) and the domain gap between classical paintings and natural images. (a) CP3D aims to recover 3D structure from a classical painting while preserving appearance fidelity. (b) Paintings provide sparse, stylized, and inconsistent geometric cues, violating the assumptions of existing reconstruction methods. (c) Natural images provide physically consistent cues that support reliable 3D reconstruction.**

This difference makes reliable geometry recovery in CP3D particularly challenging, since the available spatial cues are often weak, stylized, and ambiguous. As shown in Fig. 1 (b), classical paintings frequently exhibit flattened shading, symbolic layout, and misleading depth cues, making perspective, illumination, and spatial structure difficult to interpret in physically consistent terms. By contrast, natural images usually offer more coherent perspective, richer illumination variation, and stronger depth cues, as illustrated in Fig. 1 (c). Existing monocular reconstruction models [19, 22, 26, 33] therefore struggle in the CP3D setting because their underlying assumptions, including photometric consistency, physically grounded image formation, and natural-image statistics, are frequently violated in classical paintings.

To address this tension, we propose **GeAR**, a training-free framework for **G**eometry **G**rounding and **A**pppearance **R**estitution. The key insight of GeAR is that the representation best suited for stable geometry recovery is not necessarily the one that best preserves the painterly appearance of the source artwork. In classical paintings, the visual patterns that obstruct physically grounded reconstruction often remain essential to brushwork, texture, and stylistic identity. GeAR therefore does not force a single representation to simultaneously maximize structural recoverability and painterly fidelity. Instead, it assigns these two requirements to different stages of the reconstruction pipeline. Geometry Grounding first transforms the input painting into a geometry-oriented representation with more coherent shading and illumination cues, reducing ambiguity introduced by stylized depiction and enabling a pretrained reconstruction model to infer a more stable 3D representation. Because this process may attenuate fine textures and painterly details, GeAR then performs Appearance Restitution on the grounded 3D scene under spatial constraints and multi-view consistency, restoring the

visual characteristics weakened during grounding while preserving the recovered structure. This design improves geometric stability without collapsing the reconstructed result toward a natural-image appearance, while retaining the visual identity of the original painting throughout the reconstruction process.

To support systematic study of CP3D, we construct **HeriArch**, a large-scale benchmark for classical painting-to-3D reconstruction. Existing cultural-heritage datasets [13, 30, 32] mainly focus on 2D restoration and preservation, and are therefore not suitable for evaluating 3D reconstruction from classical planar artworks. By contrast, HeriArch contains 10,160 high-resolution images spanning murals, ukiyo-e prints, and religious paintings, covering diverse artistic styles, compositional conventions, and degrees of geometric ambiguity. This diversity makes HeriArch a suitable testbed for systematic evaluation of CP3D across varied artistic conventions and levels of geometric ambiguity.

Our contributions can be summarized as follows:

- We introduce **Classical Painting-to-3D (CP3D)**, a new problem setting for inferring a plausible 3D representation from a single classical painting while jointly enforcing geometric plausibility and fidelity to the source artwork.
- We propose **GeAR**, a training-free framework that improves geometric recoverability by grounding classical paintings into a geometry-oriented representation, and restores painterly fidelity by recovering artwork-faithful appearance on the grounded 3D scene.
- We construct **HeriArch**, a 10,160-image benchmark for systematic evaluation of CP3D, and show through extensive qualitative, quantitative, and human evaluations that GeAR achieves more stable geometry and more faithful appearance than strong baselines.

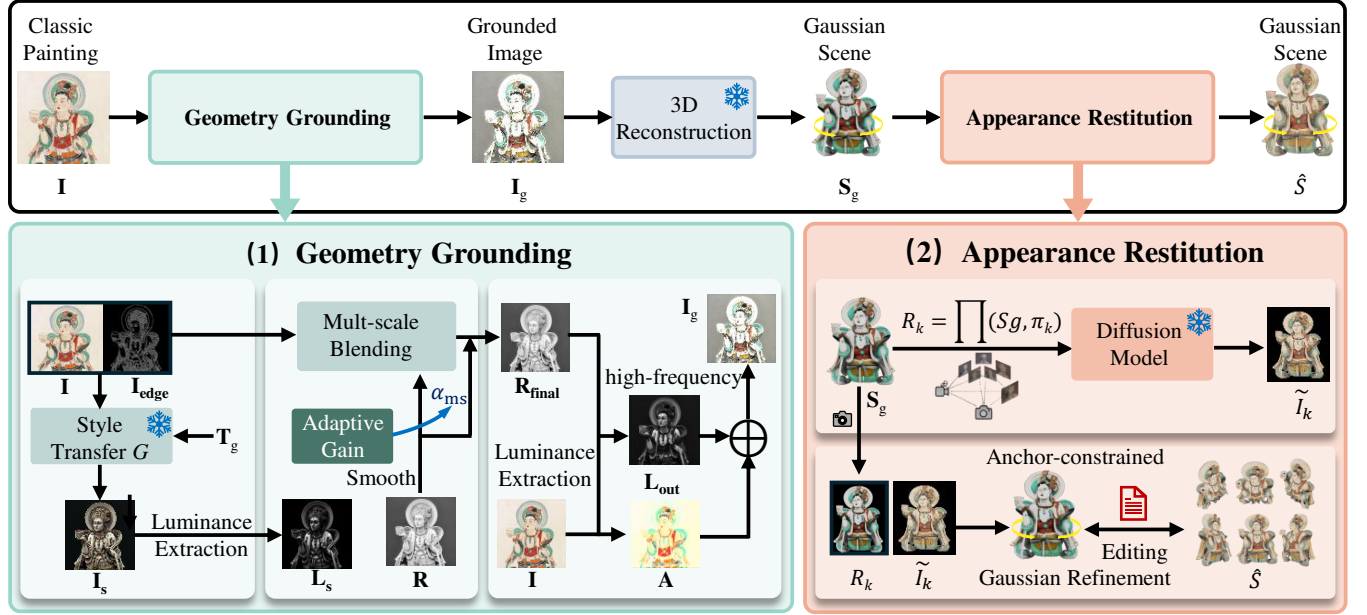


Figure 2: Overview of GeAR. We first perform Geometry Grounding to transform a classical painting into a geometry-oriented representation for stable 3D reconstruction, yielding a grounded Gaussian scene  $S_g$ . We then apply Appearance Restitution, where multi-view targets guided by a source-derived prompt are used to refine the scene via anchor-constrained optimization. The final result  $S^*$  is both geometrically plausible and faithful to the original artwork.

## 2 Related Work

**Neural Single-Image 3D Reconstruction.** Reconstructing 3D scenes from a single image is a fundamental yet ill-posed task in computer vision, owing to missing depth cues and occluded geometry [4, 9]. Existing methods often adopt either neural implicit representations [15, 17, 22] or point-based explicit approaches [3, 7, 10]. NeRF [15] models continuous 3D scenes via MLPs but struggles under sparse-view inputs. Recent extensions introduce structural priors for regularization, such as depth smoothing [17], visibility priors [22], and frequency constraints [33]. In parallel, monocular 3D Gaussian Splatting (3DGS) is gaining traction. Splatter Image [27] maps pixels to Gaussians in a feed-forward manner. Flash3D [26] predicts layered Gaussians for occlusion modeling. Triplane-based approaches [20] and transformer-based pipelines [19] further improve efficiency. However, these methods are largely developed for natural images and typically assume physically consistent scene geometry, perspective, and appearance cues, assumptions that may not hold for classical or heritage paintings.

**Generative 3D via Diffusion and Latent Priors.** Diffusion models and latent priors have recently been exploited for 3D reconstruction. SiTH produces unseen-view textures and reconstructs textured meshes [8]. VideoScene distills video diffusion into a one-shot 3D scene generator [29], while DRiVE creates riggable 3D Gaussians from a single image [25]. Other notable methods include RenderDiffusion, which integrates rendering into the denoiser to enforce multi-view consistency [1]; LN3Diff, which performs diffusion in a compact latent 3D space for efficient monocular reconstruction and generation [11]; and DiffRF, which directly generates volumetric

radiance fields under rendering guidance [16]. These approaches are also primarily developed for natural-image domains, and their assumptions may not transfer directly to heritage imagery characterized by non-physical perspective and artist-mediated structure. **Artistic and Heritage Image Datasets.** Existing cultural heritage datasets, including MuralDH [32], DunHuang-Mural [30], and ARMCD [13], mainly focus on 2D restoration problems and therefore provide annotations such as synthetic degradations, restoration targets, or segmentation masks rather than explicit 3D structural supervision. This design makes them valuable for preservation-oriented image restoration, but less suitable for studying monocular 3D reconstruction. In addition, their samples often contain repeated content, local crops, or region-specific annotations, which weakens the global compositional and spatial cues needed for geometric reasoning. 3DSynBrush introduces the CME dataset for single-image 3D reconstruction of Dunhuang mural elements [18], but remains specific to one heritage domain. HeriArch fills this gap as a large-scale benchmark for monocular 3D reconstruction from diverse heritage imagery. To our knowledge, it is the first large-scale benchmark spanning a broad range of non-photorealistic heritage artworks, with non-redundant samples and diverse styles.

## 3 Method

We first formalize the Classical Painting-to-3D (CP3D) task and then present GeAR, a training-free framework that decomposes reconstruction into geometry grounding and appearance restitution. The key idea is to avoid forcing a single representation to simultaneously maximize geometric recoverability and painterly fidelity. Instead,

GeAR first converts the input painting into a geometry-oriented representation that is more suitable for stable 3D reconstruction, and then restores artwork-faithful appearance on the recovered 3D scene under structural constraints.

### 3.1 Problem Formulation

Given a single classical painting  $\mathbf{I} \in \mathbb{R}^{H \times W \times 3}$ , CP3D aims to recover a 3D scene representation  $\hat{\mathbf{S}}$  that supports plausible novel-view rendering while remaining faithful to the source artwork:

$$\hat{\mathbf{S}} = \{g_i\}_{i=1}^N, \quad (1)$$

where  $N$  denotes the number of Gaussian primitives, and each primitive  $g_i$  is parameterized by its center, scale, rotation, opacity, and appearance attributes. Let  $\Pi(\mathbf{S}, \pi)$  denote the differentiable rendering of a Gaussian scene  $\mathbf{S}$  under camera pose  $\pi$ .

Unlike conventional single-image 3D reconstruction, CP3D takes as input an artist-mediated planar depiction rather than a physical observation of a geometrically consistent 3D scene. Consequently, perspective, shading, and depth cues in the painting may be stylized, ambiguous, or internally inconsistent, making reliable geometry inference difficult. At the same time, the reconstructed scene should remain faithful to the painterly appearance of the source artwork. CP3D therefore requires jointly satisfying two objectives: recovering geometrically plausible scene structure from an under-constrained painting and preserving fidelity to the source artwork in the rendered appearance.

### 3.2 GeAR

GeAR comprises two stages: **Geometry Grounding** and **Appearance Restitution**, as illustrated in Figure 2. Geometry Grounding first improves geometric recoverability from classical paintings, whose structural, shading, and illumination cues are often weak, stylized, or unreliable for direct 3D reconstruction. Specifically, the input painting is mapped to a geometry-oriented representation that provides more coherent cues for pretrained monocular reconstruction, thereby producing an initial 3D scene with stabilized geometry. Appearance Restitution then restores the painterly appearance and artwork-specific details that may be weakened during geometry grounding. Specifically, the grounded 3D scene is refined with view-aligned appearance targets and constrained Gaussian refinement, thereby recovering fidelity to the source artwork while preserving the grounded geometry.

**3.2.1 Geometry Grounding.** The goal of Geometry Grounding is to stabilize 3D reconstruction from an underconstrained painting by constructing a geometry-oriented intermediate representation with more coherent structural and illumination cues. This stage is not intended to preserve the final artwork appearance. Instead, it focuses on suppressing depiction-specific ambiguities that would otherwise be absorbed into the recovered geometry.

**Structure-preserving geometry translation.** A straightforward translation from paintings to a more natural-image-like domain may alter composition or distort object boundaries, which is undesirable because subsequent 3D recovery depends strongly on spatial layout. To preserve scene layout during translation, we first extract a contour-based structural prior:

$$\mathbf{I}_{\text{edge}} = f_{\text{edge}}(\mathbf{I}), \quad (2)$$

where  $f_{\text{edge}}(\cdot)$  denotes an edge detector applied to the input painting. The resulting contour map preserves object boundaries and coarse scene organization, which serve as structural guidance during translation. Conditioned on this structural prior, we then generate a geometry-oriented image:

$$\mathbf{I}_s = G(\mathbf{I}_{\text{edge}}, \mathbf{I}, T_g), \quad (3)$$

where  $G(\cdot)$  denotes a text-guided image translation model, and  $T_g$  is a prompt that encourages physically coherent shape, shading, and illumination cues. This translation is not intended to beautify the painting, but to suppress depiction-specific ambiguity while retaining the original compositional structure. The translated image therefore provides a more suitable input for downstream 3D reconstruction.

**Illumination grounding.** Building on the structure-preserving translated image, we further improve geometric recoverability by grounding its illumination cues. Because monocular 3D reconstruction depends strongly on shading and luminance structure, this step regularizes the translated image toward more spatially coherent illumination, yielding a more reliable input for downstream reconstruction.

Let  $L_s$  denote the luminance of the translated image  $\mathbf{I}_s$  after linearization to the radiance domain. We first compute a log-ratio illumination field:

$$R = \log(L_s + \varepsilon) - \log(\text{gmean}(L_s + \varepsilon)), \quad (4)$$

where  $\varepsilon > 0$  avoids numerical instability, and  $\text{gmean}(\cdot)$  denotes the geometric mean over all pixels. The resulting ratio field  $R$  captures relative illumination variation while reducing sensitivity to the global intensity scale. We then apply local mean filtering to obtain a smoothed ratio field that enforces spatial consistency in illumination variation. For each pixel  $p$ , the smoothed ratio field is computed as:

$$R_{\text{smooth}}(p) = \frac{1}{|\Omega(p)|} \sum_{q \in \Omega(p)} R(q), \quad (5)$$

where  $p$  indexes a pixel location,  $\Omega(p)$  denotes the local filter window centered at  $p$ ,  $q$  indexes pixels inside that window, and  $|\Omega(p)|$  is the number of pixels in  $\Omega(p)$ . To balance global illumination consistency against local structural detail, we further refine  $R_{\text{smooth}}$  in a multi-scale, boundary-aware manner. Specifically, we construct an illumination pyramid to capture global trends at coarse scales and local shading variations at fine scales, while deriving inverse boundary weights from the edge map to suppress halo artifacts near structural boundaries. The resulting multi-scale estimate is denoted by  $R_{\text{ms}}$ . A fixed fusion weight cannot accommodate the large variation in illumination contrast across paintings: it tends to over-regularize some cases while under-regularizing others. We therefore adapt the fusion strength on a per-image basis.

**Adaptive Gain Control.** To this end, we introduce an adaptive gain control mechanism that regulates the fusion strength according to the global contrast of the multi-scale illumination estimate. Specifically, we measure the global contrast of the multi-scale ratio field using its standard deviation:

$$\sigma_R = \text{std}(R_{\text{ms}}), \quad (6)$$

which quantifies the overall contrast of the multi-scale illumination estimate. We then define an adaptive fusion weight as:

$$\tilde{\alpha}_{ms} = \alpha_{base} \cdot \frac{\alpha_{target}}{\sigma_R + \varepsilon}, \quad (7)$$

where  $\alpha_{base}$  is the base fusion strength,  $\alpha_{target}$  denotes the target contrast level, and  $\varepsilon > 0$  is a small constant for numerical stability. To ensure stable adaptation, we constrain the fusion weight to lie within a bounded range:

$$\alpha_{ms} = \text{clip}(\tilde{\alpha}_{ms}, \alpha_{min}, \alpha_{max}), \quad (8)$$

where  $\text{clip}(\cdot, \cdot, \cdot)$  restricts the fusion weight to  $[\alpha_{min}, \alpha_{max}]$  for numerical stability and perceptual coherence. This form inversely scales the fusion weight with the contrast of  $R_{ms}$ , so that images with stronger illumination variation receive weaker smoothing, while flatter cases receive stronger regularization. The final illumination ratio is then obtained as

$$R_{final} = \alpha_{ms} R_{ms} + (1 - \alpha_{ms}) R_{smooth}. \quad (9)$$

Notably,  $\alpha_{ms}$  is not confined to  $[0, 1]$ . When  $\alpha_{ms} > 1$ , Eq. (9) performs controlled extrapolation instead of convex blending, which empirically amplifies informative illumination in flat paintings. **Illumination-consistent reconstruction.** We then reconstruct a geometry-grounded image in the linear radiance domain by combining the grounded illumination with reflectance estimated from the original painting. Let  $L$  denote the luminance of the original painting after linearization. We recover the grounded illumination map by injecting the refined illumination ratio in the log domain:

$$L_{out} = \exp(\log L + \beta R_{final}), \quad (10)$$

where  $\beta$  controls the strength of illumination grounding. We estimate the reflectance map  $A$  from the original painting by dividing its linearized RGB values by the original luminance:

$$A = \frac{\text{Linearize}(\mathbf{I})}{L + \varepsilon}, \quad (11)$$

where  $\varepsilon > 0$  ensures numerical stability. The grounded linear image is then reconstructed by combining the preserved reflectance with the grounded illumination:

$$\mathbf{I}_{g,linear} = A \odot L_{out}. \quad (12)$$

To recover fine brushwork and local texture, we extract high-frequency components from the original artwork and add them back:

$$\mathbf{I}_g = \mathbf{I}_{g,linear} + \text{HF}(\text{Linearize}(\mathbf{I})), \quad (13)$$

where  $\text{HF}(\cdot)$  denotes a high-frequency extraction operator. The resulting grounded image  $\mathbf{I}_g$  is finally fed into a pretrained monocular 3D reconstruction network [31] to obtain a grounded Gaussian scene  $\mathbf{S}_g$ . In practice, this grounded input yields more stable geometry than reconstructing directly from the original painting.

**3.2.2 Appearance Restitution.** Although Geometry Grounding improves structural stability, it does not fully preserve the appearance of the original artwork. In particular, the grounding process tends to suppress high-frequency visual patterns that may be geometrically unreliable but remain essential to painterly appearance. Naively restoring each view independently in 2D is insufficient, because the resulting edits often introduce severe cross-view inconsistency

when used as supervision for 3D refinement. To resolve this issue, we restore appearance directly on the grounded 3D scene through multi-view refinement under geometric constraints.

**Multi-view appearance target generation.** We first render the grounded scene  $\mathbf{S}_g$  from a set of sampled viewpoints  $\{\pi_k\}_{k=1}^n$ :

$$\mathbf{R}_k = \Pi(\mathbf{S}_g, \pi_k), \quad k = 1, \dots, n. \quad (14)$$

These rendered views provide geometrically consistent observations of the grounded scene under known camera poses. We then edit each rendered view with an appearance-oriented diffusion model [2] to recover artwork-faithful texture and style:

$$\tilde{\mathbf{I}}_k = D(\mathbf{R}_k, T_a), \quad (15)$$

where  $D(\cdot)$  denotes the diffusion-based appearance editing model, and  $T_a$  is an appearance prompt derived from the source painting to encourage preservation of its artwork-specific texture, style, and semantic content. Because each target  $\tilde{\mathbf{I}}_k$  is generated from a rendered view  $\mathbf{R}_k$  under a known camera pose, it preserves the coarse layout and viewpoint of the current 3D representation. These restored views therefore serve as approximately view-aligned appearance targets for subsequent 3D refinement.

**Geometry-preserving appearance refinement.** We then leverage GaussianEditor [5] to refine the Gaussian representation so that its rendered views match the appearance-restored targets while preserving the grounded geometry. Formally, starting from the grounded initialization  $\mathbf{S}_g$ , we obtain the final representation  $\hat{\mathbf{S}}$  by optimizing:

$$\hat{\mathbf{S}} = \arg \min_{\mathbf{S}} \sum_{k=1}^n \mathcal{L}_{\text{edit}}(\Pi(\mathbf{S}, \pi_k), \tilde{\mathbf{I}}_k) + \lambda_{\text{anc}} \mathcal{L}_{\text{anchor}}(\mathbf{S}, \mathbf{S}_g), \quad (16)$$

where  $\mathcal{L}_{\text{edit}}$  encourages the rendered appearance of  $\mathbf{S}$  to match the restored view targets, while  $\mathcal{L}_{\text{anchor}}$  regularizes the optimization toward the grounded initialization  $\mathbf{S}_g$  to suppress geometric drift. In practice, the anchor term suppresses large geometric drift and keeps the refinement focused on texture, color, and painterly detail rather than structural change.

## 4 The HeriArch Dataset

The HeriArch dataset is a comprehensive collection of 10,160 unique images of classical hand-painted artworks, spanning from the 3rd to the 20th century and encompassing over 17 centuries of artistic evolution. This dataset encompasses a broad spectrum of artistic styles and traditions, offering an invaluable resource for research on style consistency and geometric structure recovery in cultural heritage artworks. The dataset is organized into several key subsets, each representing a distinct artistic tradition: (1) **Mural**: 2,361 images, including 1,591 from the DhMural dataset [14] and additional murals from Dunhuang, China. (2) **Court Ladies Paintings**: 775 depictions of historical court life from the Tang to Qing dynasties. (3) **Tibetan Thangka**: 1,783 images spanning 13 centuries, illustrating the evolution of Tibetan Buddhist art. (4) **Indian Art**: 709 artworks from the 18th–20th centuries, sourced from the Smithsonian Institution, showcasing three centuries of Indian artistic development. (5) **Ukiyo-e**: 4,241 Japanese woodblock prints from the Edo period, documenting the transition from monochrome *sumi-e* to multicolor *nishiki-e*. (6) **Persian Miniature**: 291 classical Persian manuscript and illuminated paintings from the Safavid and Timurid eras.

| Method             | VEND $\uparrow$ | DDI <sub>15°</sub> $\uparrow$ | DDI <sub>30°</sub> $\uparrow$ | DDI <sub>45°</sub> $\uparrow$ | DDI <sub>60°</sub> $\uparrow$ | DDI <sub>75°</sub> $\uparrow$ | DDI <sub>90°</sub> $\uparrow$ | DDI <sub>avg</sub> $\uparrow$ |
|--------------------|-----------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| SplatterImage [27] | 66.65           | 4.11                          | 4.03                          | 3.96                          | 3.82                          | 3.90                          | 3.18                          | 3.83                          |
| LGM [28]           | 73.84           | 5.67                          | 4.98                          | 4.16                          | 4.23                          | 4.47                          | 4.56                          | 4.68                          |
| TRELLIS [31]       | 77.64           | 6.35                          | 6.30                          | 6.58                          | 6.96                          | 8.12                          | 9.82                          | 7.37                          |
| GeAR (Ours)        | <b>85.29</b>    | <b>10.66</b>                  | <b>10.56</b>                  | <b>10.62</b>                  | <b>10.87</b>                  | <b>11.29</b>                  | <b>12.29</b>                  | <b>11.05</b>                  |

**Table 1: Quantitative comparison on final reconstructions. We report geometry-oriented metrics on the final outputs to evaluate whether the recovered geometry remains plausible after the full two-stage pipeline.**

## 5 Experiments

### 5.1 Experimental Setup

**Evaluation Protocol and Metrics.** Since ground-truth novel views are unavailable for classical paintings, CP3D cannot be adequately evaluated with a single family of metrics. We therefore evaluate the final reconstructions from two complementary perspectives: *geometric plausibility* and *artwork fidelity*. For geometric plausibility, we report two geometry-oriented metrics, *Volumetric Extent and Normal Directional Diversity* (VEND) and *Depth Detail Integrity* (DDI), computed on the final outputs to assess whether the recovered geometry remains plausible after the full two-stage pipeline. For artwork fidelity, we conduct a texture-focused MLLM evaluation and a human study on the final reconstructions to assess how well they preserve the painterly appearance of the source artwork and their overall perceptual preference. (1) VEND measures both the volumetric extent of the reconstructed shape and the directional diversity of surface normals, thereby reflecting global structural plausibility and local geometric variation. Higher VEND scores indicate reconstructions with less flattening and richer surface-orientation diversity. (2) DDI evaluates the integration of local surface detail and global volumetric structure by combining high-frequency geometric variation with shape anisotropy. Higher DDI values indicate that the reconstruction preserves finer depth variations while maintaining stable three-dimensional structure across viewpoints. (3) Inspired by [12], we conduct a texture-focused MLLM evaluation on the final reconstructions to assess how well they preserve the texture and painterly appearance of the source painting. (4) We additionally conduct a human study on the final reconstructions to assess overall perceptual quality and human preference. The study involves over 300 participants, and each sample is evaluated at least 30 times in total under randomized presentation.

**Implementation details.** Unless otherwise specified, we use the same settings across all experiments. For 3D reconstruction, we use TRELLIS as the pretrained monocular reconstruction backbone. For metric computation, each reconstructed scene is rendered from six offset viewpoints at 15°, 30°, 45°, 60°, 75°, and 90° relative to the input view. In Geometry Grounding, we set  $\alpha_{\text{base}} = 0.8$ ,  $\alpha_{\text{target}} = 1.0$ ,  $\alpha_{\text{min}} = 0.4$ , and  $\alpha_{\text{max}} = 1.5$ . In Appearance Restitution, we render  $n = 7$  viewpoints from the grounded Gaussian scene to construct multi-view appearance targets. The rendered views of each painting are edited using the same source-derived appearance prompt  $T_a$ , and the edited views are used to supervise Gaussian refinement with GaussianEditor.

**Comparison with SOTA methods.** Since CP3D is a new setting and no prior method is designed specifically for monocular 3D reconstruction of classical hand-painted artworks, we compare

GeAR with three strong monocular 3D reconstruction baselines based on 3D Gaussian representations: SplatterImage [27], LGM [28], and TRELLIS [31]. All baselines are evaluated using their official or publicly released inference pipelines on the same test set and input resolution for fair comparison.

| Method             | TF $\uparrow$ | GA $\uparrow$ | VC $\uparrow$ | AC $\uparrow$ |
|--------------------|---------------|---------------|---------------|---------------|
| SplatterImage [27] | 6.12          | 3.32          | 3.92          | 4.36          |
| LGM [28]           | 4.35          | 2.79          | 4.25          | 4.11          |
| TRELLIS [31]       | 5.92          | 6.54          | 6.55          | 5.52          |
| GeAR w/o AR        | 7.97          | 7.43          | 7.14          | 7.29          |
| GeAR w/ AR         | <b>8.71</b>   | <b>8.03</b>   | <b>7.95</b>   | <b>7.97</b>   |

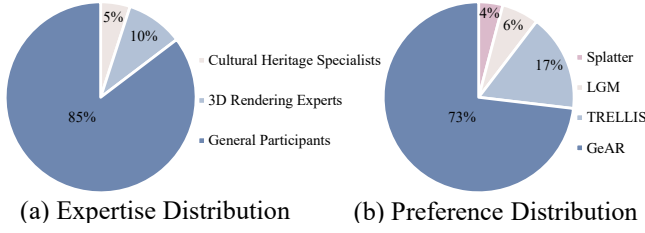
**Table 2: The results of different methods evaluated by MLLMs on four appearance dimensions. AR denotes Appearance Restitution.**

### 5.2 Quantitative Evaluation of Geometric Plausibility

GeAR achieves the best geometry-oriented performance on the final reconstructions, as shown in Table 1. Compared with TRELLIS, GeAR improves VEND from 77.64 to 85.29 and increases the average DDI from 7.37 to 11.05, with consistent gains across all viewpoint offsets. These results indicate that GeAR produces final reconstructions with stronger volumetric plausibility and richer depth variation under novel views. The improvements also suggest that the geometric gains introduced by Geometry Grounding are preserved after Appearance Restitution, rather than being degraded by the final refinement stage. Overall, the results show that GeAR achieves more plausible final geometry than strong monocular 3D reconstruction baselines in the CP3D setting.

### 5.3 MLLM-Based Evaluation of Appearance Fidelity

Geometric metrics alone are insufficient to assess the perceptual quality of the final reconstruction. Table 2 therefore reports an additional evaluation of the final outputs using a blind MLLM judge following the LLM-as-a-judge paradigm [12], with randomized method order and four evaluation dimensions: Texture Fidelity (TF), Geometric Accuracy (GA), View Consistency (VC), and Aesthetic Consistency (AC). GeAR with Appearance Restitution achieves the best performance on all four dimensions. The gain in TF indicates that Appearance Restitution restores artwork-specific textures and painterly details weakened during Geometry Grounding, while the improvements in GA and VC suggest that the restored appearance

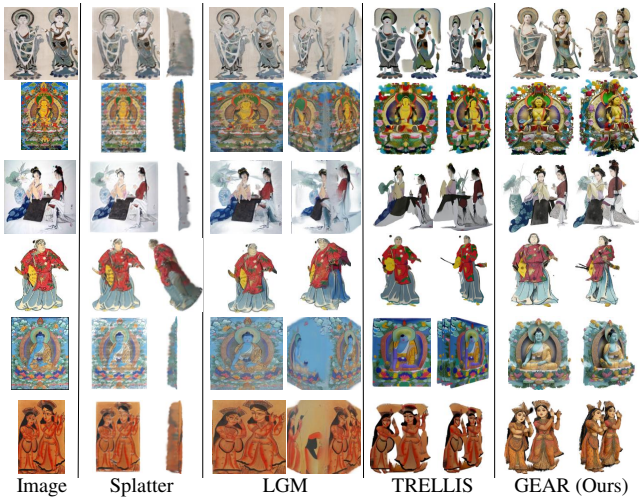


**Figure 3: (a) Distribution of participants across three expertise groups; (b) Distribution of participant preferences among the evaluated methods.**

remains compatible with the grounded geometry. The higher AC further indicates stronger overall perceptual coherence. Together, these results show that GeAR achieves a better overall balance between geometric plausibility and artwork fidelity in the final reconstruction.

#### 5.4 User Study

We conducted a comprehensive user study involving over 300 participants, including 29 experts in 3D rendering and 15 specialists in cultural heritage. The goal was to compare different monocular 3D reconstruction methods for hand-painted artworks based on user preference. Specifically, we randomly sampled 300 classic hand-painted artworks from the HeriArch Dataset and generated corresponding monocular 3D reconstructions using various methods. The reconstructed results were randomly shuffled and presented to participants, who were asked to evaluate them from three perspectives: perceived depth realism, aesthetic quality, and overall preference based on personal intuition. As shown in Figure 3, GeAR received overwhelmingly positive feedback, being strongly preferred by participants for its significant improvement in reconstruction quality.



**Figure 4: Comparisons of generated 3D Gaussians for image-to-3D. Our method generates Gaussian splatting with better visual quality on various challenging hand-painted artworks.**

#### 5.5 Qualitative Results

We compare GeAR with recent monocular image-to-3D Gaussian reconstruction methods. Figure 4 presents rendered results from the reconstructed 3D models under different viewpoints. GeAR achieves better preservation of artistic details such as figure posture, fabric folds, and object structure. SplatterImage and TRELIS lose substantial detail, particularly in complex regions such as faces and clothing, while LGM recovers some structure but remains blurry and distorted in high-frequency areas. In contrast, GeAR produces more accurate geometry, preserves local details more effectively, and maintains better global shape consistency. It also recovers lighting variation and material appearance more faithfully, especially in shadow transitions, fabric reflections, and object glossiness. Overall, GeAR produces more stable and visually faithful reconstructions than competing methods.

#### 5.6 Ablation Study

In this section, we perform ablation studies to examine the contribution of key components in GeAR. Unless otherwise specified, all results are reported on the final reconstructions. For analyses related to Geometry Grounding, we additionally report NIQE on the grounded intermediate as a supplementary diagnostic of natural-image compatibility rather than a holistic metric of final CP3D quality, since Appearance Restitution intentionally restores painting-specific textures and stylistic patterns that may deviate from natural-image statistics.

| Method      | VEND $\uparrow$ | DDI <sub>avg</sub> $\uparrow$ | NIQE <sub>avg</sub> $\downarrow$ |
|-------------|-----------------|-------------------------------|----------------------------------|
| GeAR w/o AR | 85.84           | 6.68                          | 4.75                             |
| GeAR w/ AR  | <b>87.61</b>    | <b>10.24</b>                  | 6.12                             |

**Table 3: Stage-wise ablation of GeAR on a 1,000-sample subset of the main evaluation set. We report VEND, DDI, and NIQE for the output w/ or w/o Appearance Restitution.**

**Stage-wise Effect of Appearance Restitution.** We conduct a stage-wise ablation on 1000 samples to examine the role of the two stages in GeAR. Specifically, we compare the grounded output without Appearance Restitution and the final reconstruction with Appearance Restitution, using both quantitative metrics in Table 3 and qualitative comparisons in Figure 5. As shown in Table 3, adding Appearance Restitution improves both VEND and DDI, indicating that the second stage preserves and further enhances geometric plausibility after Geometry Grounding. This behavior is also visually consistent with Figure 5: the grounded result exhibits more stable structure but loses part of the original painterly texture, while Appearance Restitution recovers texture, decorative patterns, and stylistic details on top of the grounded reconstruction. As expected, NIQE increases after Appearance Restitution, since this stage intentionally restores artwork-specific textures and stylistic patterns that depart from natural-image statistics.

**Effect of Different Prompt Styles.** We examine how the *Sculpture* prompt influences reconstruction performance compared to other styles. As shown in Table 4, the sculpture-style prompt significantly improves both VEND and DDI scores, indicating better 3D structure



**Figure 5: Qualitative Comparison of Reconstruction Results W/ and W/o Appearance Restitution**

| Prompt Style         | VEND $\uparrow$ | DDI <sub>avg</sub> $\uparrow$ | NIQE <sub>avg</sub> $\downarrow$ |
|----------------------|-----------------|-------------------------------|----------------------------------|
| <i>Sculpture</i>     | <b>85.58</b>    | <b>11.05</b>                  | <b>4.79</b>                      |
| <i>Pencil sketch</i> | 83.86           | 2.46                          | 7.25                             |
| <i>Woodcut</i>       | 80.46           | 2.95                          | 6.79                             |

**Table 4: Results of different prompt styles on geometry-oriented metrics, with NIQE reported as a supplementary diagnostic of natural-image compatibility.**

recovery and more accurate surface details. The physically consistent lighting with clear highlight and shadow transitions provides strong depth cues for the model. In contrast, prompts like *Pencil sketch* and *Woodcut* lack continuous tonal transitions, leading to incomplete shading and reduced geometric fidelity. Furthermore, the sculpture prompt achieves the lowest NIQE, indicating that it better aligns the grounded intermediate with the natural-image statistics assumed by pretrained monocular 3D reconstruction models.

| Range of $\alpha$ | VEND $\uparrow$ | DDI <sub>avg</sub> $\uparrow$ | NIQE <sub>avg</sub> $\downarrow$ |
|-------------------|-----------------|-------------------------------|----------------------------------|
| Fixed             | <b>86.01</b>    | 7.45                          | 5.09                             |
| 0.5-1.2           | 85.35           | 10.35                         | <b>4.79</b>                      |
| 0.4-1.5           | 85.29           | <b>11.05</b>                  | <b>4.79</b>                      |
| 0.3-1.8           | 85.27           | 11.04                         | <b>4.79</b>                      |

**Table 5: Performance Comparison of Fixed vs. Adaptive  $\alpha_{ms}$  Fusion Coefficient.**

**$\alpha_{ms}$  V.S.  $\alpha_{adaptive}$ .** We compare a fixed  $\alpha_{ms}$  with an adaptive  $\alpha_{ms}$  scheme. The fixed setting uses  $\alpha_{ms} = 0.8$ , while the adaptive scheme is evaluated with  $[\alpha_{min}, \alpha_{max}] = [0.5, 1.2]$ ,  $[0.4, 1.5]$ , and  $[0.3, 1.8]$ . As shown in Table 5, the fixed  $\alpha_{ms} = 0.8$  achieves a slightly higher VEND, likely because stronger relighting contrast provides more pronounced shading cues for volumetric recovery. However, this more aggressive enhancement also leads to overexposure, which

weakens depth-detail preservation and reduces compatibility with the natural-image statistics assumed by pretrained reconstruction models, as reflected by the lower DDI and higher NIQE. In contrast, the adaptive scheme yields a better trade-off between structural extent and detail preservation. Among the tested ranges,  $[\alpha_{min}, \alpha_{max}] = [0.4, 1.5]$  gives the best overall balance, achieving the highest DDI and the lowest NIQE while maintaining competitive VEND. Overall, these results suggest that adaptive  $\alpha_{ms}$  provides a more robust setting for Geometry Grounding across diverse paintings.

| Method            | VEND $\uparrow$ | DDI <sub>avg</sub> $\uparrow$ | NIQE <sub>avg</sub> $\downarrow$ |
|-------------------|-----------------|-------------------------------|----------------------------------|
| TRELLIS           | 77.64           | 7.37                          | 5.53                             |
| TRELLIS + HE      | 75.86           | 3.89                          | 5.07                             |
| TRELLIS + Retinex | 75.40           | 3.24                          | 5.39                             |
| TRELLIS + Ge      | <b>85.29</b>    | <b>11.05</b>                  | <b>4.79</b>                      |

**Table 6: Comparison of geometry grounding with generic image enhancement baselines. Ge: Geometry Grounding.**

**Geometry Grounding vs. Generic Image Enhancement.** We further compare Geometry Grounding with two generic image enhancement baselines, Histogram Equalization (HE) and Retinex, both integrated into the same TRELLIS reconstruction pipeline (Table 6). Although HE and Retinex slightly reduce NIQE, they substantially degrade VEND and DDI, indicating that generic enhancement alone does not improve geometric recoverability from classical paintings. By contrast, Geometry Grounding consistently improves all three metrics, showing that the gains in Table 1 do not come from arbitrary preprocessing, but from constructing a more geometry-compatible intermediate representation for monocular 3D reconstruction.

## 6 Conclusion

We introduce **CP3D**, a new task that aims to reconstruct 3D scenes from classical paintings while preserving fidelity to the source artwork. To address the geometric ambiguity and appearance loss inherent in this setting, we propose **GeAR**, a training-free two-stage framework that combines Geometry Grounding with Appearance Restitution. We further present **HeriArch**, a benchmark for systematic evaluation of classical painting-to-3D reconstruction. Through extensive experiments, we show that GeAR produces more stable geometry and more faithful final reconstructions than strong monocular 3D baselines. These results highlight the importance of separating geometric recoverability from appearance restoration in CP3D.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 62571051 and 62271393), the National Key R&D Program of China (Grant No. 2024YFF0907605), the Technology Innovation Leading Project of Shaanxi (Grant No. 2024QY-SZX-11), and the General Projects of the Shaanxi Provincial Department of Science and Technology (Grant No. 2025JC-YBQN-801).

## References

- [1] Titas Anciukevičius, Zexiang Xu, Matthew Fisher, Paul Henderson, Hakan Bilen, Niloy J Mitra, and Paul Guerrero. 2023. Renderdiffusion: Image diffusion for 3d reconstruction, inpainting and generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 12608–12618.
- [2] Black Forest Labs, Stephen Batifol, Andreas Blattmann, Frederic Boesel, Saksham Consul, Cyril Diagne, Tim Dockhorn, Jack English, Zion English, Patrick Esser, Sumith Kulal, Kyle Lacey, Yam Levi, Cheng Li, Dominik Lorenz, Jonas Müller, Dustin Podell, Robin Rombach, Harry Saini, Axel Sauer, and Luke Smith. 2025. Flux. 1 kontext: Flow matching for in-context image generation and editing in latent space. arXiv:2506.15742 <https://arxiv.org/abs/2506.15742>
- [3] David Charatan, Sizhe Lester Li, Andrea Tagliasacchi, and Vincent Sitzmann. 2024. pixelsplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 19457–19467.
- [4] Ruiyang Chen, Mohan Yin, Jiawei Shen, and Wei Ma. 2024. Recon3D: High Quality 3D Reconstruction from a Single Image Using Generated Back-View Explicit Priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2802–2811.
- [5] Yiwen Chen, Zilong Chen, Chi Zhang, Feng Wang, Xiaofeng Yang, Yikai Wang, Zhongang Cai, Lei Yang, Huaping Liu, and Guosheng Lin. 2024. Gaussianeditor: Swift and controllable 3d editing with gaussian splatting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 21476–21485.
- [6] Andreas Dahn, Leszek Plichta, Simon Spielmann, Eduard Schäfer, and Justus Blönnigen. 2024. Fate of the Minotaur: A scalable location based VR experience. In *ACM SIGGRAPH 2024 Immersive Pavilion*. 1–2.
- [7] Zhiwen Fan, Kairun Wen, Wenyang Cong, Kevin Wang, Jian Zhang, Xinghao Ding, Danfei Xu, Boris Ivanovic, Marco Pavone, Georgios Pavlakos, Zhangyang Wang, and Yue Wang. 2024. InstantSplat: Sparse-view SfM-free Gaussian Splatting in Seconds. arXiv:2403.20309 [cs.CV] <https://arxiv.org/abs/2403.20309>
- [8] I Ho, Jie Song, Otmar Hilliges, et al. 2024. Sith: Single-view textured human reconstruction with image-conditioned diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 538–549.
- [9] Zixuan Huang, Mark Boss, Aaryaman Vasishtha, James M Rehg, and Varun Jampani. 2025. Spar3d: Stable point-aware reconstruction of 3d objects from single images. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 16860–16870.
- [10] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics* 42, 4, Article 139 (2023), 14 pages. doi:10.1145/3592433
- [11] Yushi Lan, Fangzhou Hong, Shuai Yang, Shangchen Zhou, Xuyi Meng, Bo Dai, Xingang Pan, and Chen Change Loy. 2024. Ln3diff: Scalable latent neural fields diffusion for speedy 3d generation. In *European Conference on Computer Vision*. Springer, 112–130.
- [12] Dawei Li, Bohan Jiang, Liangjie Huang, Alimohammad Beigi, Chengshuai Zhao, Zhen Tan, Amrita Bhattacharjee, Yuxuan Jiang, Canyu Chen, Tianhao Wu, et al. 2025. From generation to judgment: Opportunities and challenges of llm-as-a-judge. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*. 2757–2791.
- [13] Haolong Li, Chenghao Du, Ziheng Jiang, Yifan Zhang, Jiawei Ma, and Chen Ye. 2024. Towards automated Chinese ancient character restoration: a diffusion-based method with a new dataset. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 3073–3081.
- [14] Luxi Li, Qin Zou, Fan Zhang, Hongkai Yu, Long Chen, Chengfang Song, Xianfeng Huang, and Xiaoguang Wang. 2022. Line Drawing Guided Progressive Inpainting of Mural Damages. arXiv:2211.06649 <https://arxiv.org/abs/2211.06649>
- [15] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (2021), 99–106.
- [16] Norman Müller, Yawar Siddiqui, Lorenzo Porzi, Samuel Rota Buló, Peter Kontschieder, and Matthias Nießner. 2023. Diffirf: Rendering-guided 3d radiance field diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4328–4338.
- [17] Michael Niemeyer, Jonathan T Barron, Ben Mildenhall, Mehdi SM Sajjadi, Andreas Geiger, and Noha Radwan. 2022. Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5480–5490.
- [18] Xianlin Peng, Jingyu Wang, Qiyao Hu, Nuo Xu, Jun Wang, Shuyi Qu, and Jinye Peng. 2026. 3DSynBrush: A High Quality 3D Reconstruction Framework for Single Dunhuang Murals. *npj Heritage Science* 14 (2026), 154.
- [19] Qiuhong Shen, Zike Wu, Xuanyu Yi, Pan Zhou, Hanwang Zhang, Shuicheng Yan, and Xinchao Wang. 2025. Gamba: Marry Gaussian Splatting With Mamba for Single-View 3D Reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2025), 1–14.
- [20] J Ryan Shue, Eric Ryan Chan, Ryan Po, Zachary Ankner, Jiajun Wu, and Gordon Wetzstein. 2023. 3d neural field generation using triplane diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20875–20886.
- [21] Jocelyn Penny Small. 2019. Circling round Vitruvius, linear perspective, and the design of Roman wall painting. *Arts* 8, 3 (2019), 118. doi:10.3390/arts8030118
- [22] Nagabhushan Somraj and Rajiv Soundararajan. 2023. Vip-nerf: Visibility prior for sparse input neural radiance fields. In *ACM SIGGRAPH 2023 conference proceedings*. 1–11.
- [23] Huang SongSheng, Azian Tahir, and Issarezal Ismail. 2025. Re-examining the Meaning of Flatness Based on the Theory Evolution of Medium Specificity in Painting. *Asian Journal of Research in Education and Social Sciences* 7, 1 (2025), 93–104.
- [24] Matthias Springstein, Stefanie Schneider, Javad Rahnama, Julian Stalter, Maximilian Kristen, Eric Müller-Budack, and Ralph Ewerth. 2024. Visual narratives: Large-scale hierarchical classification of art-historical images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 7220–7230.
- [25] Mingze Sun, Junhao Chen, Junting Dong, Yurun Chen, Xinyu Jiang, Shiwei Mao, Puhua Jiang, Jingbo Wang, Bo Dai, and Ruqi Huang. 2025. Drive: Diffusion-based rigging empowers generation of versatile and expressive characters. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 21170–21180.
- [26] Stanislaw Szymanowicz, Eldar Insafutdinov, Chuanxia Zheng, Dylan Campbell, Joao F Henriques, Christian Rupprecht, and Andrea Vedaldi. 2025. Flash3d: Feed-forward generalisable 3d scene reconstruction from a single image. In *2025 International Conference on 3D Vision (3DV)*. IEEE, 670–681.
- [27] Stanislaw Szymanowicz, Christian Rupprecht, and Andrea Vedaldi. 2024. Splatler image: Ultra-fast single-view 3d reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10208–10217.
- [28] Jiaxiang Tang, Zhaoxi Chen, Xiaokang Chen, Tengfei Wang, Gang Zeng, and Ziwei Liu. 2024. Lgm: Large multi-view gaussian model for high-resolution 3d content creation. In *European Conference on Computer Vision*. Springer, 1–18.
- [29] Hanyang Wang, Fangfu Liu, Jiawei Chi, and Yueqi Duan. 2025. VideoScene: Distilling video diffusion model to generate 3D scenes in one step. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16475–16485.
- [30] Han-Lei Wang, Ping-Hsuan Han, Yu-Mu Chen, Kuan-Wen Chen, XinYi Lin, Ming-Sui Lee, and Yi-Ping Hung. 2018. Dunhuang mural restoration using deep learning. In *SIGGRAPH Asia 2018 technical briefs*. 1–4.
- [31] Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jialong Yang. 2025. Structured 3d latents for scalable and versatile 3d generation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 21469–21480.
- [32] Zishan Xu, Yuqing Yang, Qianzhen Fang, Wei Chen, Tingting Xu, Jueting Liu, and Zehua Wang. 2024. A comprehensive dataset for digital restoration of Dunhuang murals. *Scientific Data* 11, 1 (2024), 955.
- [33] Jiawei Yang, Marco Pavone, and Yue Wang. 2023. Freenerf: Improving few-shot neural rendering with free frequency regularization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8254–8263.
- [34] Shuo Zhao, Yifei Huang, Xiaoyang He, Xin Tong, Xin Li, and Dan Wu. 2025. Reviving Mural Art through Generative AI: A Comparative Study of AI-Generated and Hand-Crafted Recreations. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.