PointGT: Simultaneous Geometry and Texture Editing for Point-Based Representations

Yanshu Zhang[1], George Shramko[1], Pratul P. Srinivasan[2], and Ke Li[1]

[1] Simon Fraser University
[2] Google DeepMind

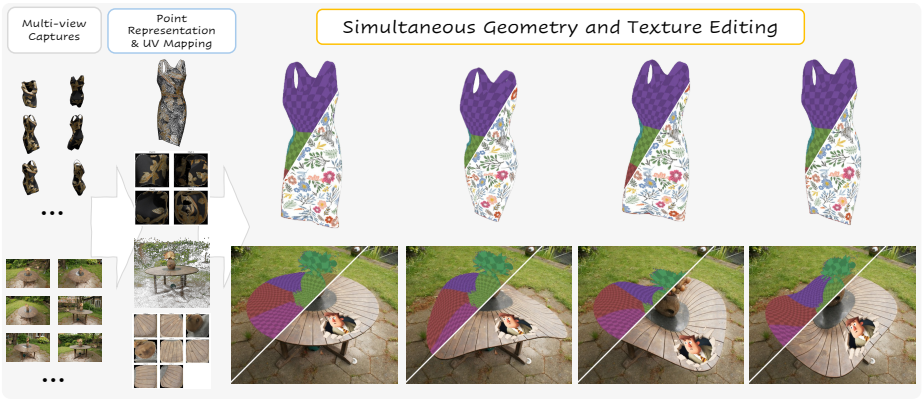


Fig. 1: Overview Given multi-view captures (left), PointGT reconstructs a point-based 3D representation and learns a global UV mapping onto 2D charts (middle left). By maintaining a deformation-aware correspondence between canonical and deformed space, PointGT ensures persistent texture edits under non-rigid geometry changes. Users can simultaneously deform and retexture a reconstructed dress (top) or a real-world table (bottom), with geometry and appearance fully decoupled and editable.

Abstract. We present PointGT, a point-based 3D representation that enables simultaneous editing of object geometry and appearance. Existing reconstruction and view synthesis techniques produce volumetric 3D representations that are high-quality and photorealistic, but are difficult to edit. In particular, recent efforts to enable texture editing for 3D Gaussian Splatting representations are not compatible with geometry edits and deformations. Our method combines a point-based representation that is well-suited for geometry deformations with a learned UV mapping technique that enables high-resolution texture editing. We demonstrate that PointGT enables fine-grained editing of both geometry and appearance in point-based neural representations while maintaining high rendering quality.

Keywords: Point neural representation · Geometry and texture editing

1 Introduction

Point-based radiance field representations, most notably 3D Gaussian Splatting (3DGS) [7], have emerged as a powerful approach to novel view synthesis and 3D

reconstruction. These methods represent scenes as point clouds, where each point is associated with a volumetric primitive (*e.g.* a 3D Gaussian distribution) that has an opacity and a single view-dependent color. Rasterizing these volumetric primitives is efficient, so techniques such as 3DGS can quickly optimize large collections of primitives to represent detailed scenes.

However, since each Gaussian is assigned a single (view-dependent) color, 3DGS can only represent detailed appearance and textures by using many (often millions) of Gaussian primitives. This coupling of geometry and appearance resolution prevents scenes reconstructed by 3DGS from being easily edited. Editing appearance is difficult because the textures of reconstructed scenes cannot be modified at resolutions finer than the sizes of the reconstructed Gaussians, and editing geometry is difficult because it is not clear how to non-rigidly transform Gaussian primitives based on user manipulations.

Recent efforts to decouple appearance and geometry in Gaussian splatting have extended 3DGS to include either global [22] or per-primitive [2,11,13,17,18, 25] UV maps. However, these techniques are still unable to enable both geometry deformations and appearance editing. Learned global UV maps from 3D surface points to 2D UV coordinates are no longer valid after geometry deformations, and per-primitive UV maps have difficulties handling deformations that are than the size of the Gaussian primitives.

We focus on another point-based representation that uses attention to render a point cloud [1, 9, 26], which effectively avoids those limitations by modeling scenes as a collection of infinitesimal points with learned features. These methods render rays with a cross-attention mechanism between the ray and the K points closest to the ray. This defines the aggregated feature of a ray as the interpolation of the features from those K nearest points, which enables editing scene geometry since these interpolation weights are still valid after deformations of the scene points. While attention-based point representation is better suited for geometry editing than 3DGS [26], appearance editing is still challenging as it is not possible to locally edit textures by modifying the optimized point features.

In this paper, we propose PointGT, a representation that extends the attention-based point renderer PAPR [26] to enable simultaneous editing of geometry and appearance. Our key contributions are:

- We extend PAPR to include a learned UV mapping from 3D points on objects to a learned 2D texture map. (Sec. 4.2)
- We introduce two novel geometry regularizers that significantly improve the fidelity of PAPR's geometry representation and demonstrate that these are crucial for effective UV mapping. (Sec. 4.1)
- We propose a deformation-aware correspondence mechanism that explicitly maps deformed-space ray intersections back to canonical space via displacement fusion, preserving texture attachment under non-rigid edits. (Sec. 4.3)

2 Related Work

Editing the geometry and appearance of 3D assets is a core part of the computer graphics workflow for digital content creation. While there are many established techniques and tools for editing widely-used mesh-based representations, editing the 3D representations produced by recent view synthesis and 3D reconstruction methods is much more challenging.

The geometry of neural scene representations such as Neural Radiance Fields (NeRFs) [8] cannot be directly edited, so many works [5, 10, 19, 21, 23, 24, 27, 28] have focused on associating NeRFs with proxy geometry which can be edited, and then converting this edited proxy geometry back into the NeRF representation. However, converting NeRFs to proxy geometry and back can introduce approximation errors, and the choice of proxy geometry can limit the types of edits that are possible.

On the other hand, the appearance of reconstructed NeRFs can be edited by optimizing UV mapping functions that map from 3D coordinates on the object to 2D texture coordinates [15, 20]. In this work, we use ideas from Nuvo [15] to develop a UV mapping technique for point-based reconstruction techniques.

2.1 3D Shape Deformation and Correspondence

Classical geometry processing has a rich history of editing 3D shapes, leveraging methods like As-Rigid-As-Possible (ARAP) deformation [14], Moving Least Squares (MLS), and embedded deformation graphs [16] to manipulate surface meshes or cages flexibly. While these approaches effectively manipulate geometry and preserve local structural rigidity, applying appearance edits to the deformed surfaces is relatively straightforward for meshes because texture coordinates (UVs) are explicitly attached to vertices and interpolate naturally across triangles.

However, projecting and interpolating high-resolution 2D textures onto unstructured, dynamically-weighted point clouds is fundamentally difficult. As point-based neural representations (like PAPR or 3DGS) lack explicit connectivity, the set of neighbors and attention weights used to define a surface point can shift non-rigidly under deformation. PointGT directly targets this novel problem of establishing canonical correspondence within dynamic attention fields to enable robust texture persistence.

More recent view synthesis techniques are based on 3D Gaussian Splatting (3DGS), which is a point-based representation that optimizes parameters of a 3D Gaussian along with a view-dependent color (parameterized using spherical harmonics) for each point. 3DGS is more suitable for direct geometry editing, since it is straightforward to move the position, orientation, and scale of any particle in the representation.

While Gaussian splatting has proved to be a powerful representation for numerous tasks, this paradigm has several limitations: (1) it often requires millions of Gaussian primitives to enable high rendering quality and accurate scene reconstructions, (2) it's challenging to make appearance edits that are finer than the

size of an individual Gaussian, and (3) free-form geometry deformation remains inconvenient without introducing geometry proxies.

Recent works tackle the first two issues by decoupling the appearance and geometry of Gaussian splatting. Texture-GS [22] learns a global UV mapping for all the ray-Gaussian intersections and stores features of these intersections on a 2D texture map. Instead of learning a global UV mapping with surface parameterization constraints (e.g., simple topologies), another line of works [2,11,13,17,18,25] learn a per-primitive texture map for each Gaussian, where the Gaussians are often restricted to be a 2D surfels [4] so texture for each point on the surfel can easily be queried from the texture map with their 2D UV coordinates. Since the frequency of surface textures is often much higher than that of geometry, decoupling the resolutions of geometry and texture representations allows these methods to render scenes photo-realistically with only 5K points [2,11,17,22,25] and enables manipulation of the textures with fine-grained details [11,18,22].

However, none of these methods can perform simultaneous geometry deformation and texture editing. This limitation is rooted in the fundamental representation of Gaussian splatting, where each point is represented as a volume of space. In the case where per-primitive texture maps are learned, preserving textures after deformation requires not only the correct transformation (rotation, scaling) of each Gaussian but also accurate splitting or cloning of primitives to represent deformations that are finer than a Gaussian (as Gaussians are rigid and cannot be bent). Such a process is often intractable in practice. In the case where a global UV mapping is learned, in addition to the difficulties described above, the UV mapping must be correctly found for the new intersection points after deformation. The combination of these factors makes the simultaneous editing of texture and geometry for Gaussian splatting extremely difficult.

3 Preliminaries

We briefly introduce the relevant background on attention-based point representations and neural UV mappings, and establish the notation used throughout.

3.1 Attention-Based Point Renderers

Attention-based point renderers (like PAPR [26] and Pointersect [1]) represent scenes as unstructured point clouds coupled with an attention interpolation mechanism. For instance, PAPR represents a scene as N learnable points $\mathcal{P} = \{(\mathbf{p}_j, \mathbf{f}_j)\}_{j=1}^{N}$, where $\mathbf{p}_j \in \mathbb{R}^3$ is the spatial position and $\mathbf{f}_j \in \mathbb{R}^h$ is a feature vector. To render ray $\mathbf{r}_i$, PAPR selects the K points nearest to it by perpendicular distance as the local neighborhood. A learned cross-attention mechanism then assigns a softmax weight a_{ij} to each neighbor point j, and the ray's feature $\mathbf{f}_i^{\text{ray}}$ is computed as the weighted sum of their features:

$$\mathbf{f}_i^{\text{ray}} = \sum_{j=1}^{K} a_{ij}\, \mathbf{f}_{ij}, \tag{1}$$

where $\mathbf{f}_{ij}$ denotes the feature associated with the j-th selected neighbor point $\mathbf{p}_{ij}$ for ray i. The aggregated feature map is then decoded by a UNet [12] to

predict the RGB colors. Given the same attention weights, these methods typically predict the intersection point $\mathbf{x}_i \in \mathbb{R}^3$ where the ray hits the surface by interpolating the point positions:

$$\mathbf{x}_i = \sum_{j=1}^{K} a_{ij}\, \mathbf{p}_{ij}. \tag{2}$$

Note that $\mathbf{x}_i$ is an *attention-weighted average* of point positions, not a true geometric intersection: its quality depends on how concentrated the weights are and how well-placed the supporting points are. This sensitivity motivates the geometry regularizers introduced in Sec. 4.1.

Canonical and deformed space. We call the point positions after training the **canonical positions** $\mathbf{p}_j^{\text{can}}$. The corresponding **canonical intersection**

$$\mathbf{x}_i^{\text{can}} = \sum_{j=1}^{K} a_{ij}^{\text{can}}\, \mathbf{p}_{ij}^{\text{can}} \tag{3}$$

is what the UV mapping is trained on. At edit time, a non-rigid deformation displaces each point to $\mathbf{p}_j^{\text{def}}$, producing a **deformed intersection**

$$\mathbf{x}_i^{\text{def}} = \sum_{j=1}^{K} a_{ij}^{\text{def}}\, \mathbf{p}_{ij}^{\text{def}} \tag{4}$$

in deformed space, where direct UV lookup is no longer valid. Bridging this gap is the central problem addressed in Sec. 4.3.

3.2 Neural UV Mapping

Neural UV mapping methods (like Nuvo [15] and NeuTex [20]) parameterize surface appearances by learning continuous mappings from 3D space to 2D texture atlases. For example, Nuvo learns a multi-chart UV atlas and texture map from a set of surface points $\mathcal{G}$ (in our setting, $\mathcal{G} = \{\mathbf{x}_i^{\text{can}}\}$). It uses cycle-consistency (bijectivity) losses and additional regularizers to learn the UV mappings; we refer readers to [15] for full details.

To enable texture editing and to encourage an even allocation of texture resolution over the scene, Nuvo also penalizes mapping distortion using conformal and stretch regularizers that operate on the differential of each chart's 3D$\rightarrow$2D UV mapping [15]. These regularizers require the surface normal at each surface point to construct tangent-space directions, which are not explicitly available for raw attention-interpolated intersection points $\mathbf{x}_i^{\text{can}}$. We therefore use a normal-free distortion loss in Sec. 4.2 that removes the need for surface normals.

4 Method

The goal of PointGT is to enable persistent texture edits under non-rigid deformation for point-based neural representations. While applying appearance edits to deformed surfaces is relatively straightforward for meshes—where texture coordinates (UVs) are explicitly attached to vertices and interpolate naturally across triangles—point clouds lack this explicit connectivity. Consequently,

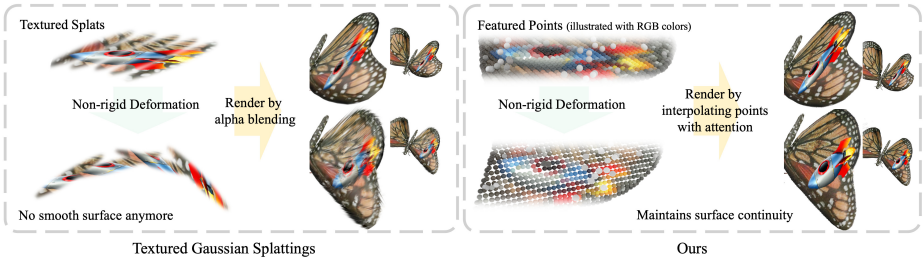


Fig. 2: Comparison of geometry deformations of textured Gaussian splatting methods and PointGT. **Left:** Since each Gaussian is a linear primitive, these methods struggle to preserve surface smoothness under large non-rigid deformations. **Right:** PointGT maintains surface continuity after deformation by predicting ray–surface points via attention-based interpolation on a point cloud.

maintaining a consistent mapping between the deformed geometry and a 2D texture map becomes fundamentally difficult, often causing texture edits to drift or break when the underlying points move non-rigidly.

To tackle this challenge, the choice of the underlying point-based representation is critical. As discussed, Gaussian Splatting-based methods struggle to preserve surface continuity under non-rigid deformations (Fig. 2, left), and may break continuous texture mapping. In contrast, attention-based renderers like PAPR [26] maintain surface continuity via smooth point interpolation (Fig. 2, right), providing the stable surface points required for consistent UV mapping.

Building upon this insight, PointGT introduces a framework for simultaneous geometry and appearance editing. To ensure the attention-based surface points are suitable for learning a multi-chart UV atlas and texture map [15], we first propose **geometry regularization** losses that condition the ray–surface intersections to accurately reflect the true geometry (Sec. 4.1). Then, to preserve texture attachment during edits, we introduce a **deformation-aware canonical correspondence** mechanism (Sec. 4.3). This explicitly maps the deformed-space intersections reliably back to the original canonical space for consistent UV lookup, ensuring that appearance edits persist seamlessly regardless of the applied geometric deformation.

4.1 Learning PAPR with Geometry Regularization

To enable reliable UV learning and stable editing, we require accurate and well-conditioned ray–surface intersection points. In PAPR, the predicted surface point for ray i is an attention-weighted average of K ray-nearest neighbors (Eq. 2), which can be sensitive to sparse attention and imperfect point geometry. Without additional constraints, predicted intersection points may *collapse* toward discrete support points and/or be supported by noisy off-surface neighborhoods (Fig. 7 and 8a), which harms the sampled surface point set used to train the UV atlas. We therefore add two geometry regularizers during PAPR pre-training.

On-ray regularizer (close-to-ray). As shown in Fig. 7, sparse attention weights can cause predicted intersection points to cluster around discrete point supports.

This degrades the effective geometric resolution of $\mathbf{x}_i$ to the resolution of the point cloud and yields unstable surface samples for UV supervision. Since a valid ray–surface intersection point should lie on its ray by definition, we minimize the distance between each predicted point $\mathbf{x}_i$ and its orthogonal projection $\mathbf{x}'_i$ onto the corresponding ray $\mathbf{r}_i$:

$$\mathcal{L}_{\text{close2ray}} = \frac{1}{M} \sum_{i=1}^{M} \|\mathbf{x}_i - \mathbf{x}'_i\|_2, \tag{5}$$

where M is the number of rays in a batch.

On-surface neighborhood regularizer (close-to-surface). Even when intersections can be predicted, the underlying point cloud may contain off-surface points. This is problematic when starting from sparse points and densifying: new points may gather around an off-surface location, and a ray may then select a neighborhood whose convex hull does not cover the true surface. To regularize neighborhood geometry, we encourage each selected neighbor point $\mathbf{p}_{ij}$ to move toward the predicted surface point $\mathbf{x}_i$:

$$\mathcal{L}_{\text{close2surface}} = \frac{1}{MK} \sum_{i=1}^{M} \sum_{j=1}^{K} \|\mathbf{p}_{ij} - \text{sg}(\mathbf{x}_i)\|_2 \tag{6}$$

that encourages all the nearby points to move closer to the predicted intersection point. Note that we stop the gradients backpropagating to $\mathbf{x}_i$ from this regularizer for training stability. To prevent early geometry collapse, we apply $\mathcal{L}_{\text{close2surface}}$ in the middle of training after a coarse geometry is learned.

Objective. We combine the regularizers with PAPR's rendering loss for pretraining:

$$\mathcal{L}_{\text{rendering}} + \gamma\ \mathcal{L}_{\text{close2ray}} + \eta\ \mathcal{L}_{\text{close2surface}}, \tag{7}$$

where $\mathcal{L}_{\text{rendering}}$ is PAPR's rendering loss, γ and η are the coefficients of the regularizers.

4.2 Learning UV Mapping

Building upon the regularized point cloud geometry, we extract ray–surface intersections to learn a multi-chart UV atlas and a 2D texture map. We sample foreground rays from the training views, compute their canonical-space intersection points $\mathbf{x}_i^{\text{can}}$, and optimize a continuous neural UV parameterization over these surface samples.

Distortion loss without normals (Jacobian regularization). To enable consistent texture editing and encourage an even allocation of texture resolution across the scene, it is crucial to minimize the distortion of the learned UV mappings. While existing surface parameterization methods [15] typically rely on ground-truth geometry normals to compute conformal and stretch penalties, point-based representations natively lack this explicit geometric connectivity and normal information. To overcome this, we define a distortion loss directly over the Jacobians of the texture coordinate mappings.

Specifically, for each chart k in our multi-chart texture maps, we compute the Jacobian matrix $\mathbf{J}_{ik} \in \mathbb{R}^{2\times 3}$ of the UV mapping by taking the gradients of the texture coordinates $\mathbf{u}_{ik} \in \mathbb{R}^2$ w.r.t. the 3D point coordinates $\mathbf{x}_i$. The Jacobian's singular values σ_{ik}^1 and σ_{ik}^2 characterize the local scaling behavior of the mapping. To minimize anisotropic scaling distortion, we minimize the squared difference of the singular values across all charts:

$$\mathcal{L}_{\text{scaling}} = \frac{1}{Gn} \sum_{i=1}^{G} \sum_{k=1}^{n} \left(\sigma_{ik}^1 - \sigma_{ik}^2\right)^2, \tag{8}$$

where G is the total number of 3D points. This encourages the texture coordinate network to learn mappings with isotropic scaling (where $\sigma^1 = \sigma^2$). While $\mathcal{L}_{\text{scaling}}$ minimizes anisotropy, it does not explicitly prevent area shrinkage or blow-up. To address this, we also define an area distortion penalty, $\mathcal{L}_{\text{area}} = \frac{1}{Gn} \sum_{i=1}^{G} \sum_{k=1}^{n} (\log(\sigma_{ik}^1 \sigma_{ik}^2))^2$, which regularizes the local scale to remain close to 1 and heavily penalizes near-zero areas, preventing degenerate mappings. We define the distortion loss as the sum of these two terms:

$$\mathcal{L}_{\text{distortion}} = \mathcal{L}_{\text{scaling}} + \mathcal{L}_{\text{area}}. \tag{9}$$

In practice, we optimize the UV atlas by minimizing a weighted sum of Nuvo [15]'s losses (excluding the normal-based distortion terms) and the proposed distortion loss $\mathcal{L}_{\text{distortion}}$. We jointly learn a 2D RGB texture map by supervising rays with PAPR-predicted colors using bilinear sampling on the atlas; we use the learned texture map for appearance editing.

4.3 Geometry and Appearance Editing

Appearance editing in texture space. PointGT supports appearance editing by directly modifying the 2D texture atlas. Given an edited texture map $\mathbf{T}'$, we generate a mask map $\mathbf{M} \in [0,1]^{n\times h\times w}$ by comparing the original atlas $\mathbf{T}$ and the edited atlas $\mathbf{T}'$, where the pixel values are marked as 1 in edited regions and 0 otherwise. During rendering, for each ray we compute the UV coordinate $\mathbf{u}_i$ of its intersection and replace the sampled color with the edited texture when $m_i = \mathbf{M}(\mathbf{u}_i) > 0.95$ (both using bilinear interpolation on the atlas).

Geometry editing and the correspondence problem. Geometry editing applies a non-rigid deformation to the learned point cloud, producing deformed point positions $\mathbf{p}_j^{\text{def}}$ from canonical positions $\mathbf{p}_j^{\text{can}}$ (Sec. 3.1). Rendering after deformation produces a deformed-space ray–surface point by interpolating deformed neighbor positions with deformed-space attention weights:

$$\hat{\mathbf{x}}_i^{\text{def}} = \sum_{j=1}^{K} a_{ij}^{\text{def}}\, \mathbf{p}_{ij}^{\text{def}}, \tag{10}$$

where the top-K neighbors $\mathbf{p}_{ij}^{\text{def}}$ are selected by ray proximity in the deformed point cloud, and the weights a_{ij}^{def} are *re-inferred* in deformed space (i.e., both the neighbor set and weights are recomputed after deformation). However, the UV atlas is trained in canonical space, so we must transfer $\hat{\mathbf{x}}_i^{\text{def}}$ back to a canonical coordinate before UV lookup.

Why a naïve transfer fails. A natural approach is to reuse the deformed attention weights to directly interpolate *canonical* neighbor positions:

$$\mathbf{x}^{\text{can}}_{i,\text{naive}} = \sum_{j=1}^{K} a^{\text{def}}_{ij}\, \mathbf{p}^{\text{can}}_{ij}. \tag{11}$$

In practice, attention weights in PAPR are often sparse; Eq. 11 therefore collapses many rays' canonical lookup points onto a small subset of support points, causing discontinuous UV jumps across neighboring rays and visible texture popping/drift under deformation. While the on-ray regularizer (Sec. 4.1) improves training stability, it does not guarantee that intersections remain exactly on rays under extreme non-rigid, non-training deformations.

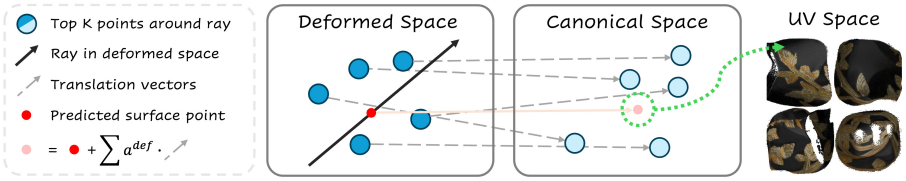


Fig. 3: Illustration of the deformation-aware canonical correspondence. In deformed space (left), a ray intersects the deformed point cloud to predict a surface point (red). Each deformed neighbor (dark blue) has a stored canonical counterpart (light blue); the per-neighbor displacements (dash gray arrow) are fused with the deformed-space attention weights to transfer the intersection back to canonical space (middle), where UV lookup retrieves the texture (right).

Deformation-aware canonical correspondence. To address this, we propose an edit-time transfer rule that maps each deformed-space ray–surface point back to canonical space while keeping samples well-distributed. **Step 1 (point-to-ray projection).** We strictly enforce the on-ray constraint by orthogonally projecting $\hat{\mathbf{x}}^{\text{def}}_i$ onto its ray $\mathbf{r}_i$:

$$\mathbf{x}^{\text{def}}_i = \Pi_{\mathbf{r}_i}\big(\hat{\mathbf{x}}^{\text{def}}_i\big), \tag{12}$$

Step 2 (attention-weighted displacement fusion). Since point identity is preserved under deformation, each deformed neighbor $\mathbf{p}^{\text{def}}_{ij}$ has a corresponding canonical position $\mathbf{p}^{\text{can}}_{ij}$. We compute per-neighbor displacements and fuse them using the *same* deformed-space attention weights, then transfer the ray-projected deformed intersection back to canonical space:

$$\mathbf{t}_i = \sum_{j=1}^{K} a^{\text{def}}_{ij}\, \big(\mathbf{p}^{\text{can}}_{ij} - \mathbf{p}^{\text{def}}_{ij}\big), \tag{13}$$

$$\mathbf{x}^{\text{can}}_i = \mathbf{x}^{\text{def}}_i + \mathbf{t}_i. \tag{14}$$

We then query the UV atlas at $\mathbf{x}^{\text{can}}_i$ and sample the (edited) texture map to render appearance. Intuitively, even if attention weights are sparse, ray projection keeps $\mathbf{x}^{\text{def}}_i$ distributed along rays, and the fused displacement provides a stable canonical transfer for UV lookup.

Edit-order interchangeability. PointGT supports flexible and simultaneous geometry and texture editing; the order of applying geometry and texture edits is

interchangeable, which is important for applications such as 3D character design where edits are applied iteratively in varying order.

5 Experiments

We evaluate PointGT along two axes: (i) **simultaneous geometry and appearance editing** under *non-rigid deformation*, where texture edits should persist without drifting (Sec. 4.3), and (ii) **novel view synthesis** quality compared to recent textured Gaussian splatting baselines.

5.1 Implementation Details

When pre-training PAPR with geometry regularization (Sec. 4.1), we use $\gamma = 0.002$ and $\eta = 0.01$ for the close-to-ray and close-to-surface losses. Since geometry is unreliable early in training, we start applying the regularizers at step 25,000 (i.e., after 1/10 of total steps). When learning the UV atlas and 2D texture map (Sec. 4.2), we follow Nuvo's optimization settings and use weight 0.4 for the proposed $\mathcal{L}_{\text{distortion}}$ and 0.04 for $\mathcal{L}_{\text{texture}}$. We use $n = 4$ or 8 charts depending on the complexity of the geometry, and fix memory usage by using a texture resolution of $256\sqrt{2/n} \times 256\sqrt{2/n}$ for each chart. For editing evaluation, we use Objaverse assets with artist-created motion sequences to evaluate persistent editing under realistic deformation scenarios (Sec. 4.3).

5.2 Simultaneous Geometry and Appearance Editing

PointGT supports appearance editing by modifying the 2D texture atlas, and geometry editing by applying non-rigid deformations to the point cloud while preserving canonical UV lookup via our edit-time canonical correspondence (Sec. 4.3). We evaluate this capability under non-rigid deformations and both **local** edits (e.g., paste an icon into a localized region) and **global** edits (e.g., recolor or material-style changes).

Non-rigid deformation with local and global edits. Figure 4 shows qualitative results on Objaverse assets with artist-created motion. PointGT preserves surface continuity under deformation and maintains edit attachment: the edited appearance remains aligned with the deforming surface without noticeable drifting or popping. In particular, we encourage readers to view the supplementary video, which visually demonstrates this robust texture-attachment (e.g., tracking a checkerboard pattern) under continuous deformation sequences.

Comparison to GSTex. We choose GSTex [11] as the baseline for simultaneous geometry and appearance editing. While Texture-GS [22] learns a texture map, the quality of its learned texture map is insufficient for reliable texture-space editing in practice; we provide evidence in the supplementary. The remaining NVS baselines (e.g., Textured Gaussians, NeST-Splatting) do not provide an appearance editing pipeline in their released codebases. We therefore compare against GSTex in the main paper for editing, and include additional discussion/analysis in the supplementary.

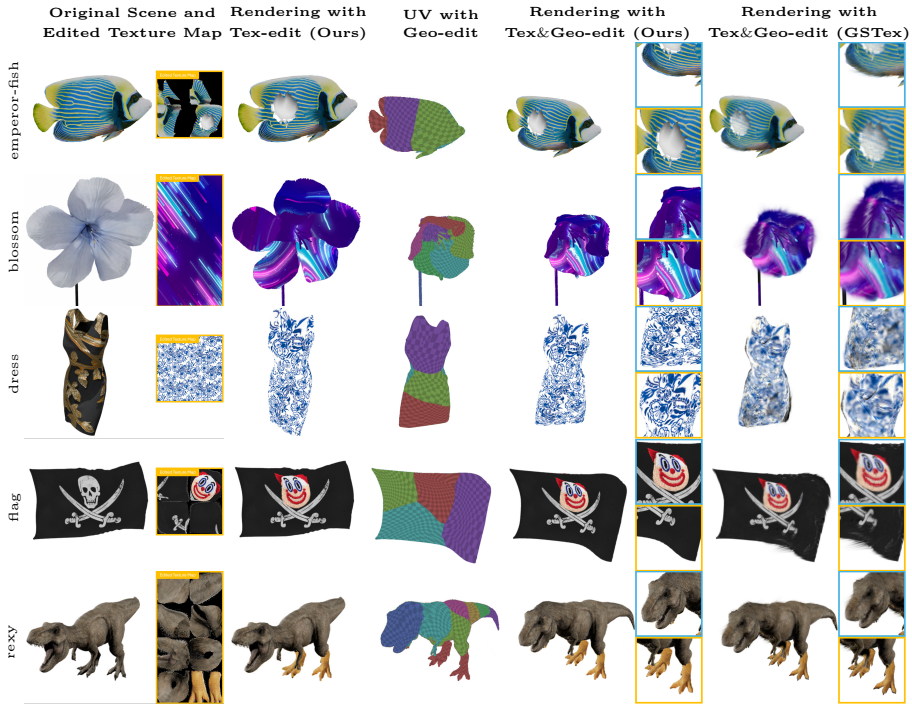


Fig. 4: Simultaneous geometry and appearance editing under non-rigid deformation with both local and global texture edits.

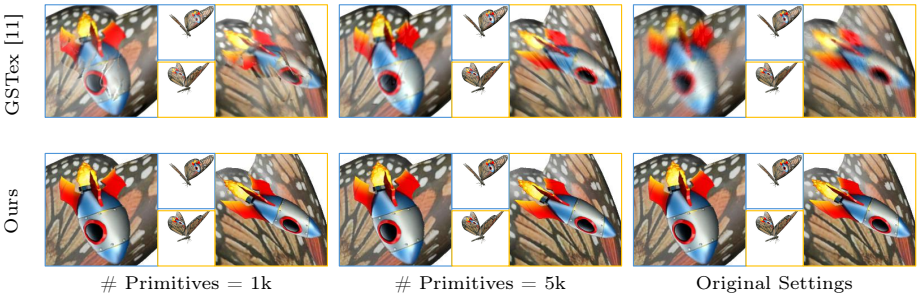


Fig. 5: Comparison of appearance editing with GSTex [11] with different number of primitives. In their original settings, GSTex uses 63k Gaussians, and we use 30k points for this scene.

In Figure 5 we add a "rocket" icon to a butterfly scene from Objaverse [3]. We edit PointGT by pasting the icon into the desired region on the 2D texture map. GSTex performs appearance editing by editing a rendered view and projecting the edited pixels back to Gaussians. To apply the same local edit to the same region for a fair comparison, we use the edited pixels from our method's rendered view as the editing input for GSTex. As shown, our method maintains sharp edit details across different point budgets (1,000 to 30,000), while GSTex renders blurred appearance, especially with 1,000 Gaussians and with its original (63,000

Table 1: Quantitative comparison of Novel-view Synthesis.

	DTU [6]				**Nerf Synthetic** [8]			
Method	PSNR ↑	SSIM ↑	LPIPS ↓	# Pts	PSNR ↑	SSIM ↑	LPIPS ↓	# Pts
Authors' recommended primitive counts								
Texture-GS [22]	30.53	0.920	0.083	90k	28.97	0.938	0.055	90k
GSTex [11]	32.87	0.956	0.038	186k	33.25	0.969	0.024	100k
Textured-Gaussian [2]	33.61	0.970	0.056	240k	33.24	0.967	0.043	190k
NEST Splatting [25]	**33.65**	0.964	0.042	80k	33.37	0.967	0.032	73k
PAPR [26]	29.34	0.952	0.073	30k	32.07	0.971	0.038	30k
Ours	33.48	**0.973**	**0.023**	30k	**33.57**	**0.982**	**0.021**	30k
Fixed primitive count (5k)								
Texture-GS [22]	26.81	0.833	0.206	5k	19.29	0.795	0.213	5k
GSTex [11]	28.92	0.932	0.072	5k	30.20	0.897	0.149	5k
Textured-Gaussian [2]	29.58	0.912	0.080	5k	26.21	0.919	0.086	5k
NEST Splatting [25]	32.68	0.966	0.056	5k	30.48	0.958	0.057	5k
PAPR [26]	24.87	0.846	0.040	5k	30.38	0.963	0.048	5k
Ours	**33.25**	**0.970**	**0.024**	5k	**31.01**	**0.976**	**0.039**	5k

Gaussians) setting. Moreover, as shown in Figure 4 our method maintains surface continuity after non-rigid deformation, while GSTex exhibits extruded Gaussians and surface artifacts under deformations.

5.3 Novel View Synthesis

We compare PointGT for novel view synthesis to PAPR [26], Texture-GS [22], GSTex [11], NeST-Splatting [25], and Textured Gaussians [2] on the Blender [8] dataset and the DTU dataset [4, 6]. Table 1 shows quantitative results for novel view synthesis on the Blender and DTU datasets. We compare against baselines in two settings: (1) the original settings reported in prior work (no cap on the number of primitives), and (2) capping all methods at 5,000 primitives. As shown, our method outperforms baselines while using far fewer primitives (30,000 points versus 73,000–240,000 points for baselines). Furthermore, PointGT significantly outperforms all baselines under the 5,000 primitives cap. Figure 6 shows a qualitative comparison with 5,000 primitives. As shown, PointGT better models fine details (e.g., chair patterns, labels on Scan 97) with the same number of primitives.

5.4 Ablation Study

Evaluation of the Geometry Regularizers We ablate the two regularizers described in Sec. 4.1.

On-ray regularizer. Figure 7 shows the ablation of the on-ray loss (close-to-ray). We visualize the predicted surface points from a camera view, the same points in UV space (across 4 atlas maps), and crops of the final renderings using varying loss weights. Without the on-ray loss, the predicted surface points collapse and cluster around discrete support points. Consequently, querying the texture map with these points fails to smoothly interpolate the texture, degrading the effective

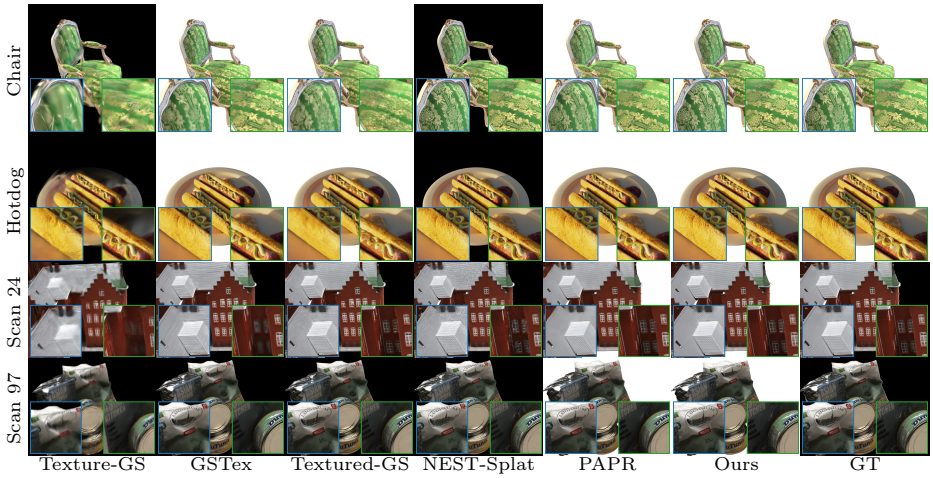


Fig. 6: Qualitative comparison of novel view synthesis with 5,000 Gaussians/points.

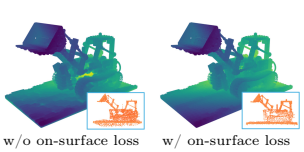
Fig. 7: Ablation of the on-ray regularizer with different weights. For each weight, we show the predicted surface points in 3D (left), the corresponding UV coordinates in the 2D atlas (middle), and the final rendered texture from texture maps (right). As shown, the on-ray loss ensures continuous interpolation of surface points, which preserves texture resolution and yields clearer high-frequency details (e.g., text).

texture resolution. With the on-ray loss, the surface points are continuous and smoothly interpolate the surface, yielding much clearer details in the renderings.

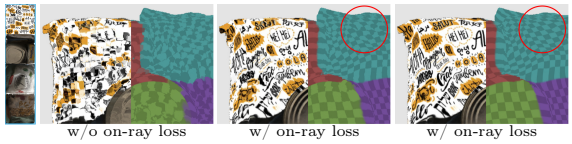
On-surface regularizer. Figure 8a shows the ablation of the on-surface loss (close-to-surface). We visualize a depth map from a selected view alongside a side-view of the learned point cloud. Without the on-surface loss, the point cloud contains many noisy, off-surface points. Adding the on-surface loss significantly reduces these noisy points, resulting in a much smoother depth map with fewer holes.

Evaluation of Distortion Loss Figure 9 shows how the learned texture map depends on the weight of the proposed distortion loss in Sec. 4.2. Increasing the distortion-loss weight reduces UV distortion, which is crucial for preserving the shape of edited appearance.

Correspondence Stability Under Deformation We ablate the edit-time canonical correspondence used for UV lookup after deformation (Sec. 4.3). Specifically, we compare our deformation-aware canonical transfer (Eq. 14) against the naïve alternative that directly interpolates canonical neighbor positions using deformed-space attention weights (Eq. 11). Figure 8b shows that the naïve transfer leads to discontinuous UV jumps and visible texture drift/popping un-



(a) The on-surface regularizer reduces noisy points and produces a smoother surface geometry.



(b) Ablation of canonical correspondence. Naïve transfer (Eq. 11) exhibits texture drift; our deformation-aware correspondence (Eq. 14) improves stability.

Fig. 8: Ablation studies on geometry regularization and canonical correspondence.

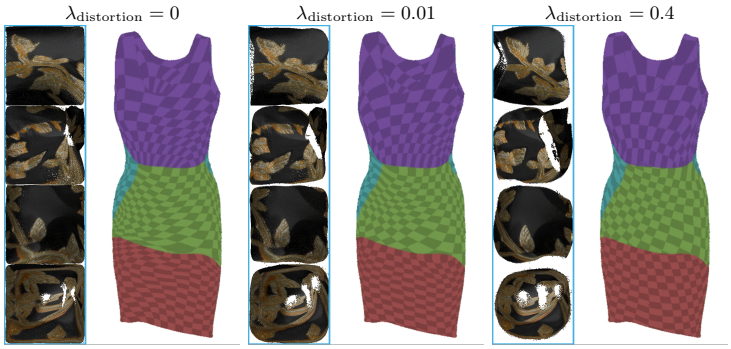


Fig. 9: Visualization of learned texture maps under different weights of the proposed distortion loss.

der non-rigid deformations, while our transfer better preserves correspondence and keeps texture edits attached.

6 Conclusion and Limitation

We presented PointGT, a representation based on PAPR that enables simultaneous editing of geometry and appearance for point-based representations. PointGT successfully achieves this by learning a UV mapping which maps 3D points to a learned 2D texture map. To improve geometry accuracy, we introduced two novel geometry regularizers, which have been demonstrated to improve the geometry significantly. We also introduced a novel correspondence mechanism to maintain UV mapping under deformation, which has been demonstrated to be highly effective. Comparison with baselines showed that PointGT preserves appearance edits under large deformations while the baseline exhibits artifacts and surface discontinuities. Furthermore, we showed that PointGT outperforms competitive baselines on novel view synthesis with a much smaller number of primitives. While PointGT achieves a highly compact representation (20MB vs 100MB for GS-based methods) and competitive synthesis, the reliance on an attention-based renderer caps interactive editing speeds at $\sim$2 FPS. Improving the computational efficiency of attention-based point rendering remains important for future work. Additionally, extreme topological tearing can introduce erroneous spatial neighbours, degrading the localized mapping and causing artifacts.

References

1. Chang, J.H.R., Chen, W.Y., Ranjan, A., Yi, K.M., Tuzel, O.: Pointersect: Neural rendering with cloud-ray intersection. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 8359–8369 (2023)
2. Chao, B., Tseng, H.Y., Porzi, L., Gao, C., Li, T., Li, Q., Saraf, A., Huang, J.B., Kopf, J., Wetzstein, G., Kim, C.: Textured gaussians for enhanced 3d scene appearance modeling. 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 8964–8974 (2024), https://api.semanticscholar.org/CorpusID:274306094
3. Deitke, M., Schwenk, D., Salvador, J., Weihs, L., Michel, O., VanderBilt, E., Schmidt, L., Ehsani, K., Kembhavi, A., Farhadi, A.: Objaverse: A universe of annotated 3d objects. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 13142–13153 (2022), https://api.semanticscholar.org/CorpusID:254685588
4. Huang, B., Yu, Z., Chen, A., Geiger, A., Gao, S.: 2d gaussian splatting for geometrically accurate radiance fields. ACM SIGGRAPH 2024 Conference Papers (2024), https://api.semanticscholar.org/CorpusID:268692234
5. Jambon, C., Kerbl, B., Kopanas, G., Diolatzis, S., Leimkühler, T., Drettakis, G.: Nerfshop. Proceedings of the ACM on Computer Graphics and Interactive Techniques **6**, 1 – 21 (2023), https://api.semanticscholar.org/CorpusID:258718613
6. Jensen, R.R., Dahl, A., Vogiatzis, G., Tola, E., Aanæs, H.: Large scale multi-view stereopsis evaluation. 2014 IEEE Conference on Computer Vision and Pattern Recognition pp. 406–413 (2014), https://api.semanticscholar.org/CorpusID:18412989
7. Kerbl, B., Kopanas, G., Leimkuehler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics (TOG) **42**, 1 – 14 (2023), https://api.semanticscholar.org/CorpusID:259267917
8. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: Representing scenes as neural radiance fields for view synthesis. Commun. ACM **65**, 99–106 (2020)
9. Ost, J., Laradji, I., Newell, A., Bahat, Y., Heide, F.: Neural point light fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18419–18429 (2022)
10. Peng, Y., Yan, Y., Liu, S., Cheng, Y., Guan, S., Pan, B., Zhai, G., Yang, X.: Cagenerf: Cage-based neural radiance field for generalized 3d deformation and animation. In: Neural Information Processing Systems (2022), https://api.semanticscholar.org/CorpusID:258509626
11. Rong, V., Chen, J., Bahmani, S., Kutulakos, K.N., Lindell, D.B.: Gstex: Per-primitive texturing of 2d gaussian splatting for decoupled appearance and geometry modeling. 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) pp. 3508–3518 (2024), https://api.semanticscholar.org/CorpusID:272753260
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. ArXiv **abs/1505.04597** (2015)
13. Song, Y., Lin, H., Lei, J., Liu, L., Daniilidis, K.: Hdgs: Textured 2d gaussian splatting for enhanced scene rendering. ArXiv **abs/2412.01823** (2024), https://api.semanticscholar.org/CorpusID:274436413
14. Sorkine, O., Alexa, M.: As-rigid-as-possible surface modeling. In: Symposium on Geometry processing. vol. 4, pp. 109–116. Citeseer (2007)

15. Srinivasan, P.P., Garbin, S.J., Verbin, D., Barron, J.T., Mildenhall, B.: Nuvo: Neural uv mapping for unruly 3d representations. ArXiv **abs/2312.05283** (2023), https://api.semanticscholar.org/CorpusID:266162619
16. Sumner, R.W., Schmid, J., Pauly, M.: Embedded deformation for shape manipulation. ACM SIGGRAPH 2007 papers (2007), https://api.semanticscholar.org/CorpusID:35682459
17. Svitov, D., Morerio, P., de Agapito, L., Bue, A.D.: Billboard splatting (bbsplat): Learnable textured primitives for novel view synthesis. ArXiv **abs/2411.08508** (2024), https://api.semanticscholar.org/CorpusID:273993982
18. Tang, K., Ai, K., Han, J., Wang, C.: Texgs-volvis: Expressive scene editing for volume visualization via textured gaussian splatting. ArXiv **abs/2507.13586** (2025), https://api.semanticscholar.org/CorpusID:280054473
19. Wang, C., He, M., Chai, M., Chen, D., Liao, J.: Mesh-guided neural implicit field editing. ArXiv **abs/2312.02157** (2023), https://api.semanticscholar.org/CorpusID:265609951
20. Xiang, F., Xu, Z., Havsan, M., Hold-Geoffroy, Y., Sunkavalli, K., Su, H.: Neutex: Neural texture mapping for volumetric neural rendering. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 7115–7124 (2021), https://api.semanticscholar.org/CorpusID:232075944
21. Xu, T., Harada, T.: Deforming radiance fields with cages. ArXiv **abs/2207.12298** (2022)
22. Xu, T., Hu, W., Lai, Y.K., Shan, Y., Zhang, S.: Texture-gs: Disentangling the geometry and texture for 3d gaussian splatting editing. In: European Conference on Computer Vision (2024), https://api.semanticscholar.org/CorpusID:268510017
23. Yang, B., Bao, C., Zeng, J., Bao, H., Zhang, Y., Cui, Z., Zhang, G.: Neumesh: Learning disentangled neural mesh-based implicit field for geometry and texture editing. ArXiv **abs/2207.11911** (2022), https://api.semanticscholar.org/CorpusID:251040986
24. Yuan, Y.J., Sun, Y.T., Lai, Y.K., Ma, Y., Jia, R., Gao, L.: Nerf-editing: Geometry editing of neural radiance fields. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 18332–18343 (2022)
25. Zhang, X., Chen, A., Xiong, J., Dai, P., Shen, Y., Xu, W.: Neural shell texture splatting: More details and fewer primitives. ArXiv **abs/2507.20200** (2025), https://api.semanticscholar.org/CorpusID:280322742
26. Zhang, Y., Peng, S., Moazenipourasil, S.A., Li, K.: PAPR: Proximity attention point rendering. In: Thirty-seventh Conference on Neural Information Processing Systems (2023)
27. Zheng, C., chang Lin, W., Xu, F.: Editablenerf: Editing topologically varying neural radiance fields by key points. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 8317–8327 (2022), https://api.semanticscholar.org/CorpusID:254408978
28. Zhou, K., Hong, L., Xie, E., Yang, Y., Li, Z., Zhang, W.: Serf: Fine-grained interactive 3d segmentation and editing with radiance fields. ArXiv **abs/2312.15856** (2023), https://api.semanticscholar.org/CorpusID:266551558

A Additional Novel View Synthesis Results

We show additional novel view synthesis comparison using the originally suggested number of primitives for each method in Figure 10. As shown, our method achieves rendering quality comparable to or better than the baselines while using only 30,000 primitives.

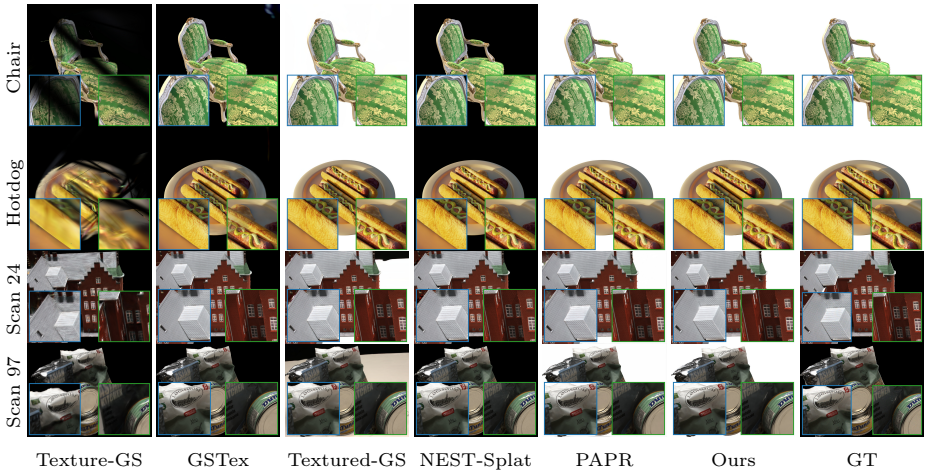


Fig. 10: Qualitative comparison of novel view synthesis with the original number of Gaussians/points proposed by the authors.

B Discussion of Other Gaussian Splatting Baselines

As noted in the main paper, we choose GSTex [11] as the primary baseline for simultaneous geometry and appearance editing. Here we provide additional discussion and evidence for why the other baselines are limited in their ability to perform simultaneous geometry and appearance editing.

Texture-GS [22] supports appearance editing through a single-chart global texture map. However, this representation often struggles with complex geometries, resulting in texture maps that are insufficient for reliable texture-space editing in practice (see Figure 11). For a fair comparison, Figure 12 shows the comparison of simultaneous geometry and appearance editing results on the 'Butterfly' scene from Objaverse, where Texture-GS successfully generates a valid map. As shown, unlike Texture-GS, which suffers from broken surface continuity and loss of resolution after deformation, our method ensures that appearance edits remain sharp and continuous.

Textured Gaussians [2] adopts the same core idea as GSTex [11]: both methods attach a learnable per-Gaussian texture map to each Gaussian primitive and query it via ray–Gaussian intersection with UV mapping defined by the Gaussian's local coordinate axes. Both also decompose appearance into a spatially

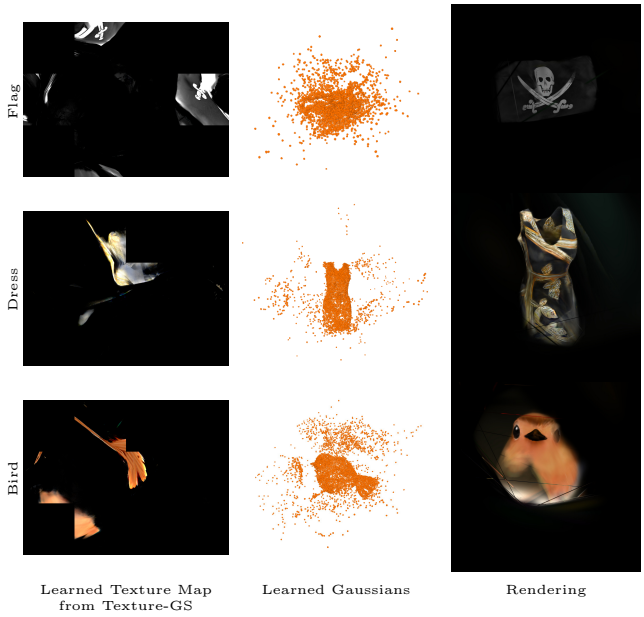


Fig. 11: Texture-GS [22] may fail to learn a reasonable texture map that is valid for appearance editing.

varying texture and spatially constant spherical harmonics, and both follow a two-stage optimization that first trains a standard Gaussian model before adding textures. The main differences are that Textured Gaussians builds on 3D Gaussians rather than 2D, uses a fixed texture resolution across all primitives (instead of adapting the resolution to each Gaussian's scale), and optionally includes an alpha channel in the texture map. Given the strong architectural similarity and the fact that appearance editing is not the main focus of Textured Gaussians (their released codebase does not include an editing pipeline), we compare against GSTex as a representative method from this family.

Fig. 12: Simultaneous geometry and appearance editing comparisons with Texture-GS [22] on the Butterfly from Objaverse [3]. We use the original settings of Texture-GS without capping the number of Gaussians.

NeST-Splatting [25] decouples geometry and appearance by replacing per-Gaussian spherical harmonics with a global shell texture, parameterized as a multi-resolution hash grid encoding. During rendering, each ray-Gaussian intersection is queried against the hash grid in *world space* to obtain appearance features. While this design yields a compact representation that achieves high-fidelity rendering with fewer primitives, the hash grid remains anchored in world space rather than attached to individual Gaussians. Consequently, displacing or deforming the Gaussian geometry causes the ray-Gaussian intersections to sample the hash grid at the wrong locations. NeST-Splatting also does not provide an editable appearance pipeline in their released codebase. As a result, NeST-Splatting is unable to perform geometry editing without the use of additional proxies, such as cages [21] or explicit meshes [24], to establish a correspondence between the deformed and canonical space.