InsideOut: Integrated RGB-Radiative Gaussian Splatting for Comprehensive 3D Object Representation

Jungmin Lee[1],   Seonghyuk Hong[3],

Juyong Lee[2],   Jaeyoon Lee[2],   Jongwon Choi[1,2*]

[1]Dept. of Advanced Imaging, GSAIM, Chung-Ang University, Republic of Korea

[2]GS. of AI, Chung-Ang University, Republic of Korea

[3]Cultural Heritage Conservation Science Center, National Research Institute of Cultural Heritage, Republic of Korea

`jngmnlee@vilab.cau.ac.kr, h123kr@korea.kr,`

`{jylee, leejaeyoon}@vilab.cau.ac.kr, choijw@cau.ac.kr`

Abstract

We introduce InsideOut, an extension of 3D Gaussian splatting (3DGS) that bridges the gap between high-fidelity RGB surface details and subsurface X-ray structures. The fusion of RGB and X-ray imaging is invaluable in fields such as medical diagnostics, cultural heritage restoration, and manufacturing. We collect new paired RGB and X-ray data, perform hierarchical fitting to align RGB and X-ray radiative Gaussian splats, and propose an X-ray reference loss to ensure consistent internal structures. InsideOut effectively addresses the challenges posed by disparate data representations between the two modalities and limited paired datasets. This approach significantly extends the applicability of 3DGS, enhancing visualization, simulation, and non-destructive testing capabilities across various domains.

1. Introduction

RGB and X-ray imaging offer strong complementary advantages in various fields, such as medical imaging [45], cultural heritage preservation [12], and manufacturing [48]. RGB images provide a high-fidelity visual appearance on external surfaces, but lack depth in internal structures. In contrast, X-ray imaging reveals subsurface structure, but fails to capture surface texture. The fusion of these modalities enables a more comprehensive understanding of both external appearance and internal structure, facilitating applications such as interactive simulations and non-destructive testing.

Despite their potential, the modality gap between RGB and X-ray imaging remains a critical challenge for effective fusion. This challenge primarily arises from the fundamen-
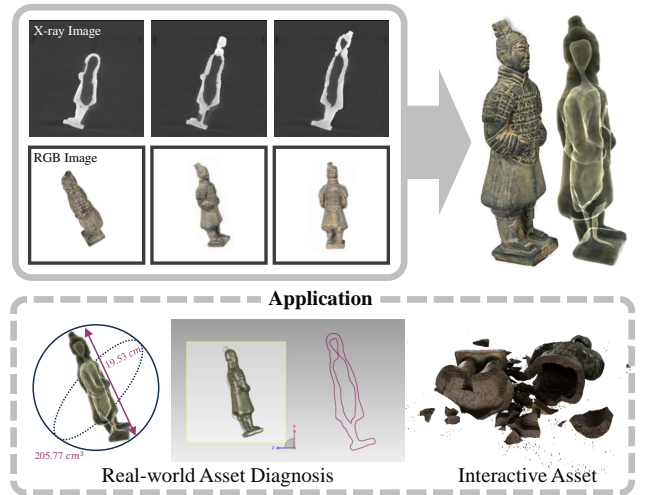

Figure 1. **Applications of InsideOut.** Our approach visualizes external appearance and internal structures, enabling analysis and physical simulation.

tal differences in how each modality captures visual information. X-rays penetrate objects and generate projection-based tomography with multi-layered structures, producing low-contrast grayscale images without external appearance details. In contrast, RGB imaging captures only the external appearance by reflecting light from object surfaces. These differences complicate feature alignment since X-ray images lack spatial correspondences with RGB features.

To address this challenge, we adopt 3D Gaussian splatting (3DGS) [17] as a novel approach for multi-modal data fusion. Optimizing both modalities within a unified 3D model allows for efficient spatial and visual alignment. The Gaussian splats in 3DGS encode position, opacity, and color parameters, enabling depth-aware modeling of multi-layered structures visible in X-ray images alongside RGB

surface details. As a result, as shown in Fig. 1, the unified 3D model can enhance analysis and visualization capabilities in various applications, such as real-world asset diagnosis and interactive assets.

However, an integrated 3DGS model that simultaneously utilizes both RGB and X-ray data presents several challenges. First, existing datasets lack both RGB image sets and corresponding X-ray images, necessitating new data collection procedures. Second, the geometric representations of RGB and X-ray data differ significantly, making geometric alignment difficult to achieve within a single 3D model. Third, the two modalities capture fundamentally different information, with RGB data providing detailed texture and color information, whereas X-ray data only captures structural information.

To address these challenges, we propose InsideOut, which is an extension of 3DGS that jointly represents external appearance from RGB images and internal structures from X-ray images. For evaluation, we first collect a new dataset containing paired RGB images and X-ray images. We then perform a hierarchical fitting procedure to align the scale and position of initial RGB and radiative Gaussian splats, ensuring geometric consistency. Guided by a novel X-ray reference loss, cross-sectional X-ray images direct the training of both RGB and radiative Gaussian splats, enabling them to be accurately aligned by internal structures. Finally, as a post-processing step, we selectively duplicate fine-scale RGB splats into the radiative splats, transferring color information from nearby RGB splats. Consequently, the radiative Gaussian splats form a unified 3D model that preserves both detailed surface features and internal structures. Our contributions are as follows:

- We propose a novel framework that integrates RGB and X-ray images into 3DGS, producing a unified 3D model with both external appearance and internal structures.
- We develop a hierarchical fitting approach that bridges the geometric domain gap between RGB and X-ray 3D models, ensuring consistent position and scale alignment.
- We introduce a new X-ray reference loss that reduces structural differences between RGB and X-ray 3D models based on cross-sectional X-ray guidance.
- We construct and release a new dataset of paired RGB and X-ray images and demonstrate the effectiveness of our framework with the new dataset.

2. Related works

2.1. Multi-modal Data Fusion

RGB Imaging-Based Fusion. RGB imaging preserves the texture of objects and finds wide utilization in various studies due to the natural similarity that RGB imaging shares with the human visual system [4, 16, 38]. Prior work has focused on visual interpretability by fusing RGB images with depth maps [31, 47, 51], thermal images [2, 43], infrared images [26, 30, 44], and multispectral images [40, 60]. For instance, combining RGB images with depth information enhances the realism of 3D reconstruction models [22], while recent work proposes implicit neural fusion of far-infrared images for improved 3D reconstruction [23]. However, RGB imaging captures only surface information and fails to provide insight into internal structures.

X-ray Imaging-Based Fusion. X-ray imaging can penetrate high-density materials, making it an essential modality for internal structure analysis [20]. Previous studies frequently combine X-ray imaging with other modalities such as MRI [53, 55, 61], structured light [35], and time-of-flight (ToF) cameras [8]. These methods integrate X-ray imaging with 3D structured light to improve the visualization of bone deformations [35] and with ToF cameras to enable cost-effective and accurate patient positioning [8]. However, these studies primarily focus on improving visual representation rather than generating a complete 3D model, since X-ray imaging produces density-based projections from absorption patterns, posing inherent challenges in matching with surface-capturing modalities.

While X-ray and RGB imaging reveal internal and external structures respectively, the integration challenges stem from fundamental physical differences. Previous studies have applied multi-resolution wavelet transformation to fuse these modalities [32] and calibrated geometry in X-ray images using RGB cameras [1]. Existing approaches focus on 2D image alignment, limited by geometric inconsistencies, camera parameter variations, and data interpretation differences. Full 3D reconstruction remains an unresolved problem, causing separate processing and independent analysis within the respective domains, even for the same object.

2.2. Gaussian Splatting

RGB Imaging-Based 3DGS. Previous 3DGS methods have achieved remarkable success in RGB image rendering tasks, particularly in 3D object generation [37, 54, 56] and surface reconstruction [10, 13, 50, 57]. In 3D object generation, Gaussian representations enable advanced object creation from sparse inputs and effective object-background separation. However, these approaches focus on surface rendering without integrating internal structures, filling with Gaussian splat noise. Recent work in surface reconstruction has evolved beyond 3D Gaussian object generation by incorporating depth information and normal vectors to enhance surface rendering quality. Although these studies have successfully converted Gaussian representations into external meshes, the methods still fail to address the limitations of internal structure representation.

X-ray Imaging-Based 3DGS. Several studies have extended 3DGS to X-ray imaging, primarily for 3D sparse-view reconstruction to reduce radiation exposure [5, 9, 33].
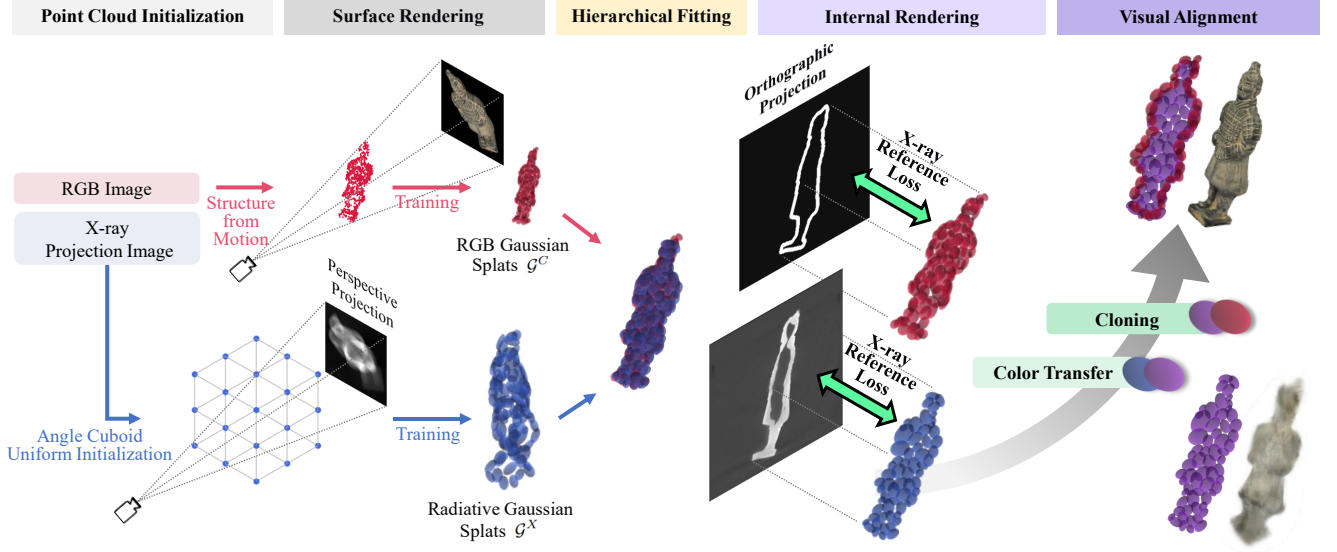
Figure 2. **Framework of InsideOut.** Our framework employs a five-stage pipeline to integrate RGB and X-ray images into a unified 3D representation. Starting with point cloud initialization from both modalities, the pipeline performs surface rendering to learn basic structures, hierarchical fitting to achieve geometric alignment, internal rendering with cross-sectional guidance for structural refinement, and visual alignment to transfer colors and details, ultimately producing a comprehensive 3D model capturing both external appearance and internal structures.

While these approaches enable X-ray scene rendering, the methods limit 3DGS to internal structure representation. Similarly, volumetric methods using Gaussian kernels [24, 59] focus on density reconstruction but remain confined to X-ray imaging. Existing work has developed 3DGS separately within RGB and X-ray domains. In contrast, our study bridges this gap by integrating both modalities to represent internal and external object structures.

3. Preliminaries

3.1. RGB Gaussian Splatting

The standard 3DGS [17] represents objects using a set of 3D Gaussian splats $\mathcal{G}^C$, where each individual Gaussian splat G_i^C encapsulates the essential geometric and appearance characteristics:

$$G_i^C = \{\mu_i^C, \Sigma_i^C, \alpha_i^C, \mathbf{c}_i^C\}, \tag{1}$$

where $\mu_i^C \in \mathbb{R}^3$ and $\Sigma_i^C \in \mathbb{R}^{3\times3}$ represent the center position and covariance matrix, respectively. $\alpha_i^C \in [0, 1]$ denotes the opacity, and $\mathbf{c}_i^C \in \mathbb{R}^3$ represents the RGB color of the i-th Gaussian splat. The complete representation is expressed as:

$$\mathcal{G}^C = \left\{ G_i^C \mid i = 1, 2, \ldots, N^C \right\}, \tag{2}$$

where N^C denotes the total number of Gaussian splats in the representation.

We first generate a 3D point cloud using Structure from Motion (SfM) [41] and define Gaussians centered at each point. During rendering, each Gaussian G_i^C is projected into 2D image space via the camera projection matrix $\mathbf{P}$. The projected 2D Gaussian is characterized by its mean position $\mathbf{p}_i = \mathbf{P}\mu_i$ and covariance matrix $\mathbf{J}_i\Sigma_i\mathbf{J}_i^T$, where $\mathbf{J}_i$ is the Jacobian of the camera projection at μ_i^C. The opacity α_i^C determines the transparency of the projected Gaussian in the final rendered image.

The final 2D image is generated by differentiable rasterization, where all projected Gaussian splats contribute to the final rendering. The photometric loss between the projected and target images is minimized using gradient-based optimization, refining the position, covariance, opacity, and color of each Gaussian.

3.2. Radiative Gaussian Splatting

We utilize X-Gaussian [5] to represent radiative Gaussian splats. Each individual radiative Gaussian splat G_j^X is formally defined as:

$$G_j^X = \{\mu_j^X, \Sigma_j^X, \alpha_j^X, \mathbf{f}_j^X\}, \tag{3}$$

where μ_j^X, Σ_j^X, and α_j^X denote the position, covariance, and opacity parameters, respectively, maintaining the same formulation as in standard RGB Gaussians. However, since X-ray imaging lacks color information, we use the feature vector $\mathbf{f}_j^X$, which represents the inherent radiative properties of the j-th Gaussian [5]. The complete set of radiative Gaussian splats is expressed as:

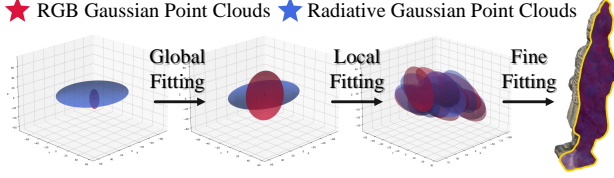$$\mathcal{G}^X = \{G_j^X \mid j = 1, 2, \ldots, N^X\}, \tag{4}$$

Figure 3. **Overview of hierarchical fitting.** The hierarchical fitting aligns RGB and radiative Gaussian point clouds through three progressive stages: global fitting for overall alignment, local fitting with clustering at multiple resolutions, and fine fitting for individual Gaussian alignment.

where N^X denotes the total number of radiative Gaussian splats in the representation.

We use an angle cuboid uniform initialization (ACUI) to calculate intrinsic and extrinsic matrices from X-ray scanner parameters. ACUI uniformly samples 3D points within a cuboid that encompasses the scanned object to initialize radiative Gaussian point cloud center positions. The initialized 3D point cloud undergoes differentiable radiative rasterization to generate rendered images from given view directions. This process maintains isotropic properties by ensuring consistent radiation intensity across viewing angles, producing the projected image. We optimize radiative Gaussian splats using photometric loss by minimizing differences between rendered and X-ray projection images.

4. Methodology

Fig. 2 illustrates the framework of InsideOut. Our framework integrates RGB and radiative Gaussian splats to create a comprehensive 3D object. First, we train RGB Gaussian splats using standard 3DGS [17] and radiative Gaussian splats using X-Gaussian [5] to learn the coarse internal and external structure. Second, we perform hierarchical fitting to align position, scale, and rotation between different modalities. Third, we use cross-sectional X-ray images to align surfaces and refine internal layers, clarifying boundaries and strengthening geometry. After geometric alignment, we achieve visual alignment by transferring colors to the X-ray 3D model and adding fine surface details from the RGB 3D model.

4.1. Hierarchical Fitting

Based on the coarse Gaussian splats, we perform hierarchical fitting, as shown in Fig. 3. This process ensures consistent spatial correspondence between the RGB Gaussian splat set $\mathcal{G}^C$ and the radiative Gaussian splat set $\mathcal{G}^X$. By considering the mean and covariance, hierarchical fitting achieves effective alignment even when shape differences exist between the two modalities. The hierarchical fitting process consists of three progressive stages, namely global fitting, local fitting, and fine fitting. The overall algorithm flow is given in Alg. 1.

Algorithm 1: Hierarchical Fitting

Input: $\mathcal{G}^X = \{G_j\}, \mathcal{G}^C = \{G_i\}$
    (each has N_X and N_C Gaussians)
Output: Aligned $\mathcal{G}^C$

2   Compute $(\bar{\mu}^X, \bar{\Sigma}^X) \leftarrow$ GlobalMean$(\mathcal{G}^X)$ and
    $(\bar{\mu}^C, \bar{\Sigma}^C) \leftarrow$ GlobalMean$(\mathcal{G}^C)$;

4   Construct global Gaussians $G^X(\mathbf{x})$, $G^C(\mathbf{x})$ using
    $(\bar{\mu}^X, \bar{\Sigma}^X)$ and $(\bar{\mu}^C, \bar{\Sigma}^C)$;

6   Align $G^C(\mathbf{x})$ to $G^X(\mathbf{x})$ using ICP [3];

8   **for** K *in* $\{N/10, \ N/2, \ N\}$ **do**

10   Cluster $\mathcal{G}^X$ and $\mathcal{G}^C$ into K clusters:

11      $C^X \leftarrow$ KMeansCluster$(\mathcal{G}^X, K)$,

12      $C^C \leftarrow$ KMeansCluster$(\mathcal{G}^C, K)$;

14   **for** *each cluster* C_k **do**

16   Compute cluster-level mean and
         covariance:

17          $(\bar{\mu}_k^X, \bar{\Sigma}_k^X) \leftarrow$ Mean(C_k^X),

18          $(\bar{\mu}_k^C, \bar{\Sigma}_k^C) \leftarrow$ Mean(C_k^C);

20   Align C_k^C to C_k^X using ICP;

22 Perform fine alignment on all Gaussians in $\mathcal{G}^C$;

In the global fitting stage, we approximate each Gaussian splat set, $\mathcal{G}^C$ and $\mathcal{G}^X$, as a single Gaussian G^C and G^X. We compute the mean position and covariance for each modality $s \in \{C, X\}$ by averaging across all Gaussian splats within the set:

$$\bar{\mu}^s = \frac{1}{N_s} \sum_{l=1}^{N_s} \mu_l^s, \quad \bar{\Sigma}^s = \frac{1}{N_s} \sum_{l=1}^{N_s} \Sigma_l^s, \tag{5}$$

where N_s denotes the number of Gaussian splats in the set $\mathcal{G}^s$, and μ_l^s and Σ_l^s represent the mean and covariance of the l-th Gaussian splat in the set, respectively.

After computing the global mean and covariance, we perform affine alignment to transform G^C to match G^X. SfM-generated RGB point clouds operate in relative coordinate systems, while X-ray point clouds use absolute physical coordinates from scanner parameters. To bridge this coordinate gap, we compute the affine transformation by aligning centroids, normalizing scales, and aligning orientations through PCA-derived rotation. The composite transformation is then applied to all individual Gaussian splats, updating their positions and covariances accordingly.

After global fitting, the alignment is progressively refined through hierarchical clustering. The Gaussian splats are first grouped into $\lfloor N/10 \rfloor$ clusters, then into $\lfloor N/2 \rfloor$ clusters, before finalizing alignment using all N Gaussian splats. We apply K-means clustering [11] to each modality separately, grouping spatially neighboring Gaussian splats. Each Gaussian splat $G_l \in \mathcal{G}^s$ is assigned to its nearest clus-

ter center, forming clusters C_k^s:

$$C_k^s = \left\{ G_l^s \ \middle| \ \|\mu_l^s - \bar{\mu}_k^s\|_2 = \min_{k' \in K} \|\mu_l^s - \bar{\mu}_{k'}^s\|_2 \right\}, \quad (6)$$

where C_k^s denotes the k-th cluster, μ_l^s represents the mean position of Gaussian splat G_l, $\bar{\mu}_{k'}$ denotes the centroid of cluster k', and K represents all possible cluster indices.

After clustering, the process refines the alignment by approximating each cluster using its mean and covariance as:

$$\tilde{\mu}_k^s = \frac{1}{|C_k^s|} \sum_{G_l \in C_k^s} \mu_l^s, \quad \tilde{\Sigma}_k^s = \frac{1}{|C_k^s|} \sum_{G_l \in C_k^s} \Sigma_l^s, \quad (7)$$

where $|C_k^s|$ denotes the number of Gaussian splats in cluster C_k^s. This strategy aligns both modalities spatially while preserving their geometric structures.

4.2. Internal Rendering

To sharpen the multi-layer boundaries of 3D models, we introduce cross-sectional images obtained through orthographic projection of X-ray volumetric data as pseudo-ground truth (GT). These images provide fine-grained supervision for internal alignment, reinforcing consistency between RGB Gaussian splats and radiative Gaussian splats throughout training. The pseudo-GT is generated by slicing the X-ray volume at 15mm intervals along the axial and coronal planes, refining the spatial distribution of radiative Gaussian splats.

For RGB Gaussian splats, we apply a Canny edge detector [6] to extract surface contours from the sliced images. When multiple edge layers are detected, we identify the outermost surface boundary by computing bounding boxes for each edge contour and selecting the one with the largest area. The extracted surface edges serve as pseudo-GT for training RGB Gaussian splats. This approach guides the splats to align along the object's surface while keeping the internal region empty. During optimization, we update only the geometric parameters μ_i^C, Σ_i^C, and α_i^C while excluding the color parameter $\mathbf{c}_i^C$, since the pseudo-GT provides only grayscale edge information.

To optimize the internal structure, we introduce an X-ray reference loss that combines multiple loss terms to achieve accurate layer separation and structural consistency. The total X-ray reference loss is:

$$\mathcal{L} = (1 - \lambda_s) \cdot \mathcal{L}_{\text{L1}} + \lambda_s \cdot \mathcal{L}_{\text{D-SSIM}} + \lambda_z \cdot \mathcal{L}_{\text{zero-one}}, \quad (8)$$

where $\mathcal{L}_{\text{L1}}$ and $\mathcal{L}_{\text{D-SSIM}}$ are photometric losses [49] that compare the rendered image with the X-ray image. The parameter λ_s controls the balance between L1 and D-SSIM losses for photometric reconstruction quality. Additionally, $\mathcal{L}_{\text{zero-one}}$ [10, 14, 28, 52] constrains the opacity values α_i of Gaussian splats to converge to either 0 or 1 during training,

with weight λ_z. This constraint enforces the formation of thin planar structures at specific layer locations, sharpening layer boundaries and accurately capturing radiation intensity. As a result, this approach removes internal noise from the RGB 3D model while promoting well-defined layer structures in the X-ray 3D model.

4.3. Visual Alignment

Once the RGB and X-ray 3D models are fully aligned, we transfer the surface color from the RGB Gaussian splats to the radiative Gaussian splats. Then, we duplicate the fine splats from the RGB object to the X-ray 3D model to create a richly detailed surface.

Color Transfer. To assign color to radiative Gaussian splats, we apply k-Nearest Neighbor (k-NN) matching by finding the closest RGB Gaussian splat for each radiative Gaussian splat:

$$\mathbf{c}_j^X = \mathbf{c}_{\text{argmin}_{i \in \{1,2,\dots,N_C\}} \|\mu_j^X - \mu_i^C\|_2}^C, \quad (9)$$

where the color attributes are transferred from the RGB Gaussian splat with minimum distance to each radiative Gaussian splat. Since radiative Gaussian splats are initially defined without color attributes, this process assigns color parameters $\mathbf{c}_j^X$ to ensure they reflect the surface appearance from RGB Gaussian splats, preserving the external visual appearance of the object.

Detail Cloning. To enhance surface detail while preserving structural consistency, we selectively duplicate high-frequency splats from the RGB 3D model into the X-ray 3D model. Gaussian splats are classified based on their covariance values, where smaller covariance indicates more detailed structures. We retain the bottom 95% of RGB Gaussian splats ranked by covariance magnitude, discarding the top 5% with large covariance values. First, we extract the largest eigenvalue from each RGB Gaussian splat's covariance matrix:

$$\Lambda^C = \{\lambda_{\max}(\Sigma_i^C) \mid i = 1, 2, \dots, N^C\}, \quad (10)$$

where $\lambda_{\max}(\Sigma_i^C)$ denotes the largest eigenvalue of the covariance matrix Σ_i^C for the i-th RGB Gaussian splat. The threshold is computed as the p-th percentile of the eigenvalue distribution:

$$T_p = \text{Quantile}(\Lambda^C, p/100), \quad (11)$$

where p represents the percentile value. The detailed Gaussian splats are selected based on the threshold:

$$G^{\text{detail}} = \{G_i^C \mid \lambda_{\max}(\Sigma_i^C) \leq T_p, i = 1, 2, \dots, N^C\}, \quad (12)$$

Empirical results show that setting $p = 95$ works well. These detailed Gaussian splats are incorporated into the radiative Gaussian splat set:

$$\hat{\mathcal{G}}^X = \mathcal{G}^X \cup \mathcal{G}^{\text{detail}}. \quad (13)$$

	Matryoshka	Pharaoh	Terracotta	Skull	Toy Gun
# RGB Images	555	644	210	90	416
Resolution (pixels)	4000×1848	4000×1848	4000×4032	3024×1848	5712×4284
# X-ray Projections	360	360	2160	360	360
# Detector Elements	2048	2048	2048	2048	2048
# Z Slices	2048	2048	2016	2048	2048
Voltage (kV)	230.00	330.00	185.00	230.00	230.00
Current (mA)	1.35	1.50	0.49	1.35	1.35

Table 1. **Dataset specifications.** RGB image collection details (top) and X-ray acquisition parameters (bottom) for five objects.

where $\hat{\mathcal{G}}^X$ denotes the enhanced radiative Gaussian splat set with detailed surface features. This integration enhances surface fidelity without disrupting the internal structure.

5. Experiments

We implement InsideOut using PyTorch [36] with CUDA [39]. The training pipeline consists of two optimization stages: surface appearance and internal structure optimization. For surface rendering, standard 3DGS [17] was trained for 20k iterations, while X-Gaussian [5] was trained for 5k iterations. The resulting Gaussian point cloud was then used as initial point cloud for internal rendering, which was further optimized for 20k iterations. We apply loss weights $\lambda_s = 0.2$ and $\lambda_z = 0.005$. We employ Adam optimizer [19] with an initial learning rate of 5e-4. All experiments were conducted on an NVIDIA A6000 GPU.

5.1. Dataset.

We constructed a novel RGB-X-ray paired dataset encompassing five objects across diverse application domains including medical imaging, cultural heritage preservation, and manufacturing quality control. Our dataset includes Matryoshka, Pharaoh, Terracotta, Skull, and Toy Gun, as detailed in Tab. 1. For RGB data collection, we captured 90-644 images per object using smartphones and DSLR cameras in natural lighting conditions with over 60% overlap between consecutive views to ensure comprehensive coverage. Image resolutions range from 3024×1848 to 5712×4284 pixels depending on the capture device. For X-ray data acquisition, we used a cone-beam CT scanner with 360° rotation. Acquisition parameters were optimized per object based on material properties, with voltage ranging from 185-330 kV and current from 0.49-1.50 mA, and 2048 detector elements with 2016-2048 Z-slices.

5.2. Internal Rendering

Fig. 4 presents a qualitative comparison of internal rendering, highlighting structural clarity and noise reduction improvements. Unlike 3DGS, which produces internal noise and scattered Gaussian splats, InsideOut eliminated these artifacts and concentrated the splats along the object's surface, creating a cleaner representation. Additionally, InsideOut enhanced the visibility of internal layers that X-Gaussian blurred, producing sharper and more well-defined
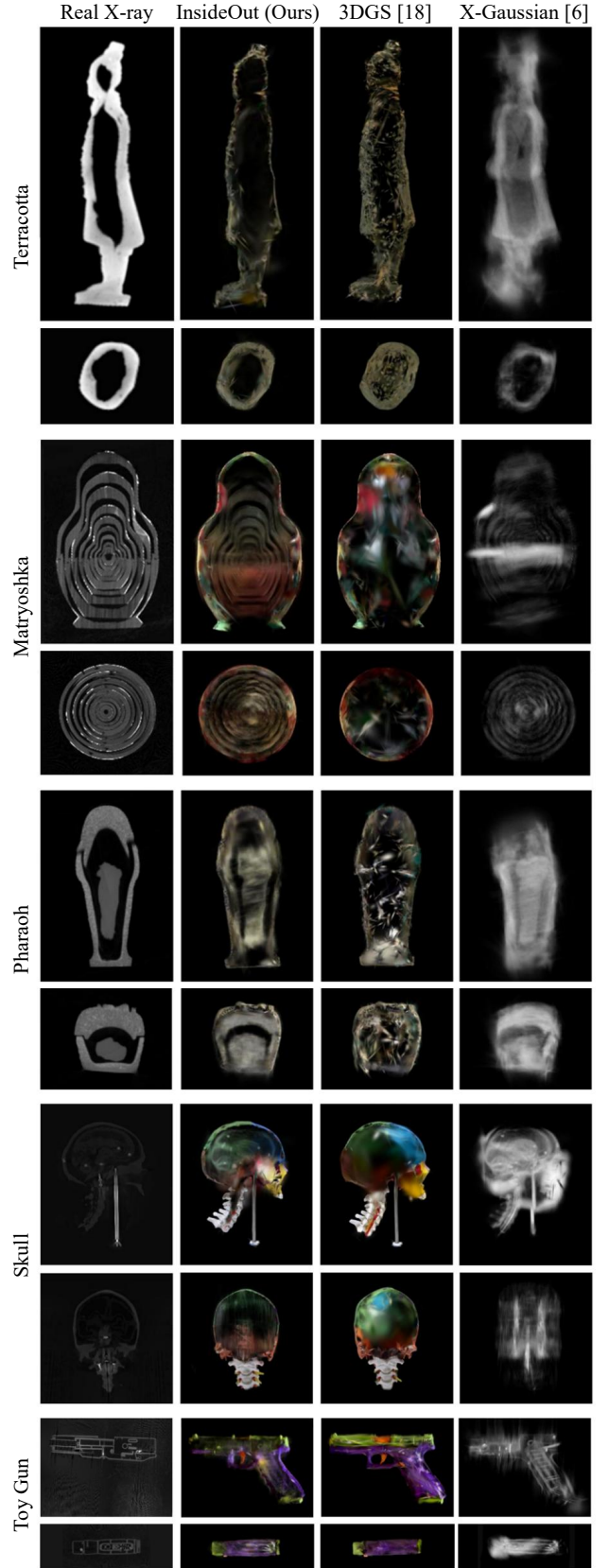

Figure 4. **Qualitative comparison of cross-sectional 3D models.** Coronal slices (upper) and axial slices (lower) are listed.

	Matryoshka			Pharaoh			Terracotta			Skull			Toy Gun		
	PIQE ↓	PSNR ↑	SSIM ↑	PIQE ↓	PSNR ↑	SSIM ↑	PIQE ↓	PSNR ↑	SSIM ↑	PIQE ↓	PSNR ↑	SSIM ↑	PIQE ↓	PSNR ↑	SSIM ↑
3DGS [17]	54.63	16.25	0.60	43.71	13.66	0.62	30.35	13.47	0.46	51.71	16.24	0.61	44.63	18.48	0.61
3DGS-MCMC [18]	50.84	16.12	0.59	41.63	13.82	0.64	30.36	13.98	0.46	50.35	16.81	0.61	43.22	18.68	0.61
DoF-Gaussian [42]	53.62	16.59	0.60	43.25	13.85	0.63	31.83	14.31	0.47	50.79	16.71	0.62	44.29	18.36	0.61
MaskGaussian [27]	54.34	16.43	0.60	42.15	13.41	0.63	30.02	13.82	0.46	56.66	16.35	0.61	43.58	18.26	0.61
Analytic-Splatting [25]	52.71	16.67	0.60	41.29	13.37	0.62	31.85	13.24	0.47	50.39	15.87	0.61	43.82	17.58	0.60
RS-NeRF [34]	52.96	16.88	0.60	44.08	13.64	0.62	32.75	13.57	0.47	52.32	15.41	0.61	44.67	17.59	0.61
NAF [58]	59.63	17.03	0.61	64.39	14.21	0.64	61.51	14.88	0.49	67.82	18.64	0.63	60.32	18.23	0.63
X-Gaussian [5]	29.82	17.46	**0.61**	33.33	14.78	**0.64**	**25.11**	15.20	0.49	23.71	19.42	0.64	21.27	19.53	0.64
InsideOut (Ours)	**25.28**	**18.86**	**0.61**	**30.35**	**15.14**	**0.64**	25.35	**19.56**	**0.55**	**20.01**	**20.33**	**0.65**	**20.74**	**20.87**	**0.66**

Table 2. **Comparison of quality for internal detail.** We present PIQE, PSNR, and SSIM scores for internal detail quality across five objects. All methods are evaluated on cross-sectional views for internal structure assessment. The best results are in bold, and our method is shown in blue.

(a) InsideOut (Ours)   (b) R^2-Gaussians [60]   (c) GS2Mesh [51]

Front View Surface Zoom Side View X-ray Shader
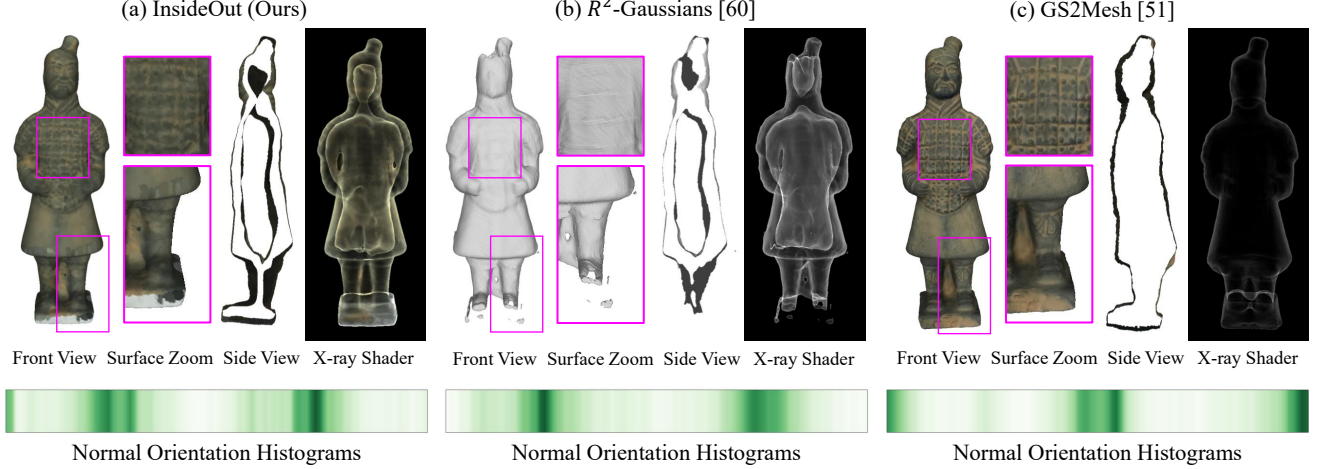
Normal Orientation Histograms

Figure 5. **Comparison of surface and internal structure in mesh reconstruction.** Mesh reconstruction quality is compared across multiple views: front view, surface zoom, side view, and X-ray shader rendering. Pink boxes indicate zoomed regions. Normal orientation histograms below each method illustrate the distribution of surface normals.
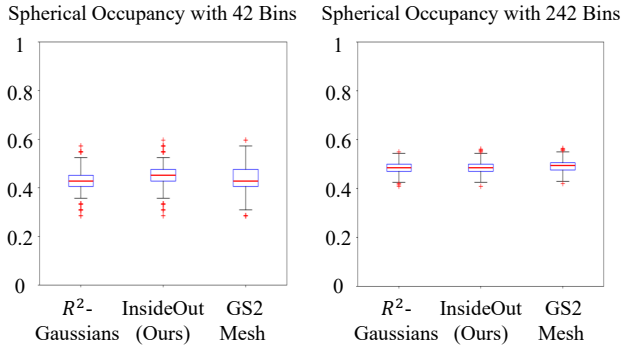

Figure 6. **Box plots for distribution of spherical occupancy.** We compare two different binning approaches: 42 bins and 242 bins.

we employ a no-reference image quality assessment. We slice the output into cross-sections and use PIQE metrics to evaluate the perceptual quality.

Tab. 2 presents the quantitative evaluation of InsideOut in cross-sectional views, where we compute PSNR and SSIM against reference X-ray images. On average, InsideOut improved PSNR by 1.67 dB over X-Gaussian and by 3.33 dB over 3DGS, demonstrating significant enhancement in reconstruction quality. Similarly, SSIM increased by 0.018 compared to X-Gaussian and by 0.042 compared to 3DGS, indicating improved structural consistency and perceptual quality.

5.3. Mesh Reconstruction

We construct 3D mesh models to assess the quality and applicability of InsideOut. We generate meshes from R^2-Gaussian [59], a radiative volume reconstruction method, by applying Marching Cubes [29]. We then compare the results with GS2Mesh [50], an RGB-based mesh reconstruction approach. For InsideOut, we apply the density voxelizer from R^2-Gaussian and then generate meshes using Marching Cubes.

structural contours. The results demonstrate that InsideOut achieved superior internal detail preservation while mitigating noise, thereby improving overall rendering fidelity.

According to Tab. 2, our method achieved the lowest perceptual image quality evaluator (PIQE) [46] score across all datasets, demonstrating superior internal detail preservation. Since no GT images exist for simultaneously evaluating both the internal and external structures of a 3D model,

★ RGB Gaussian Point Cloud   ★ Radiative Gaussian Point Cloud

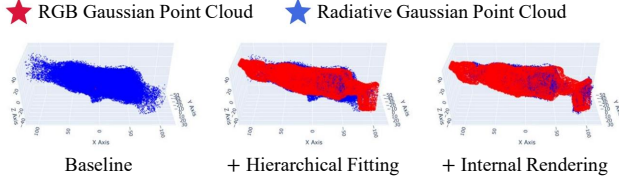Baseline   + Hierarchical Fitting   + Internal Rendering

Figure 7. **Qualitative results of ablation study.** The figure shows the progressive improvement achieved by each module: baseline (no module), hierarchical alignment only, and hierarchical alignment with internal rendering.

Fig. 5 qualitatively evaluates the three meshes, offering insights into structural fidelity. InsideOut captured finer geometric details than radiative meshes while preserving the overall shape of the object. Additionally, rendering with MeshLab's X-ray Shader [7] enabled comparison with RGB-based meshes, demonstrating that InsideOut reconstructed both the external surface and internal structure with high fidelity. As shown in Fig. 1, our method supports real-world asset diagnosis and physical simulation, enabling cross-sectional analysis and quantitative assessment of internal structures. However, since we employed a voxelization method for radiative Gaussian splats, the surface texture tends to be of lower quality compared to GS2Mesh.

To quantify fine-grained surface variations, we introduce the normal orientation histogram (NOH) and spherical occupancy [21]. Spherical occupancy measures the distribution of surface normals by partitioning the unit sphere into discrete regions n. We assign each normal to the nearest reference point, forming the NOH, which captures the directional distribution. We then compute spherical occupancy as the fraction of occupied regions in the NOH.

Fig. 5 presents the NOH for three mesh models, illustrating the distribution of surface normals across directional space. R^2-Gaussian exhibits minimal variation in normal directions, indicating a simple structure, whereas InsideOut shows a more diverse distribution of normals across different directions. The histograms provide insight into the geometric complexity of each 3D model by revealing variations in normal orientation density.

Fig. 6 presents spherical occupancy values for R^2-Gaussian, InsideOut, and GS2Mesh at bin resolutions of 42 and 242. At 42 bins, InsideOut achieved the highest occupancy (0.443 mean, 0.452 median), followed by GS2Mesh (0.437 mean, 0.429 median) and R²-Gaussian (0.431 mean, 0.429 median). At 242 bins, GS2Mesh led with 0.476 mean and 0.475 median, while InsideOut recorded 0.471 for both metrics, and R²-Gaussian remained lowest at 0.468 mean and 0.466 median. These findings indicate that InsideOut captures more diverse normal distributions than R²-Gaussian and represents internal layer normals that GS2Mesh cannot capture.

| Module | | Matryoshka | Pharaoh | Terracotta | Skull | Toy Gun |
Hierarchical Alignment	Internal Rendering					
		38.56	27.93	21.29	722.45	23.02
✓		8.41	12.52	11.44	9.53	4.59
✓	✓	**6.45**	**4.10**	**2.42**	**7.95**	**3.42**

Table 3. **Comparison of chamfer distance for geometrical alignment.** The table presents progressive improvement achieved by each module: baseline (no module), hierarchical alignment only, and hierarchical alignment with internal rendering.

5.4. Ablation Study

Geometrical Alignment. Fig. 7 shows that hierarchical fitting improved initial correspondence, while internal rendering refined the structure. The leftmost visualization shows the initial state where RGB Gaussian point clouds (red) appear much smaller than radiative Gaussian point clouds (blue) due to scale differences, making red points barely visible. With hierarchical fitting (middle), the two point clouds achieved substantial alignment. The rightmost shows the complete method, achieving optimal alignment. The progression demonstrates that hierarchical fitting provides primary alignment benefits, while internal rendering further refines the geometry.

Tab. 3 compares chamfer distance (CD) to evaluate the effects of hierarchical fitting and internal rendering. The results show that hierarchical fitting significantly reduced CD values across all 3D models, substantially enhancing initial alignment. Internal rendering further refined geometric accuracy, leading to the lowest CD values. To compute CD, we extracted surface points from both point clouds using Hidden Point Removal (HPR) [15]. This approach ensured that only visible surface points contributed to the alignment evaluation while excluding internal points from radiative Gaussian point clouds.

6. Conclusion

We introduce InsideOut, a novel 3DGS framework that integrates RGB and X-ray data to simultaneously capture an object's external surface and internal structure. By leveraging 3DGS, our approach effectively aligns these two modalities within a unified 3D representation. To achieve this, we first apply hierarchical fitting to structure Gaussian splats based on their position and scale, ensuring geometric consistency. Next, we refine the internal representation using cross-sectional X-ray guidance, allowing the model to delineate internal boundaries with higher precision. Finally, we transfer color and surface details from RGB Gaussians to radiative Gaussians, producing a comprehensive 3D model that preserves both external textures and internal structures. This advancement supports applications such as diagnostics, virtual simulations, and industrial inspection, making InsideOut a powerful tool for real-world use.

References

[1] Francisco Albiol, Alberto Corbi, and Alberto Albiol. Geometrical calibration of x-ray imaging with rgb cameras for 3d reconstruction. *IEEE transactions on medical imaging*, 35(8):1952–1961, 2016. 2

[2] Quincy G Alexander, Vedhus Hoskere, Yasutaka Narazaki, Andrew Maxwell, and Billie F Spencer Jr. Fusion of thermal and rgb images for automated deep learning based crack detection in civil infrastructure. *AI in Civil Engineering*, 1(1): 3, 2022. 2

[3] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992. 4

[4] Martin Brenner, Napoleon H Reyes, Teo Susnjak, and Andre LC Barczak. Rgb-d and thermal sensor fusion: a systematic literature review. *IEEE Access*, 2023. 2

[5] Yuanhao Cai, Yixun Liang, Jiahao Wang, Angtian Wang, Yulun Zhang, Xiaokang Yang, Zongwei Zhou, and Alan Yuille. Radiative gaussian splatting for efficient x-ray novel view synthesis. In *European Conference on Computer Vision*, pages 283–299. Springer, 2025. 2, 3, 4, 6, 7

[6] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986. 5

[7] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, and Guido Ranzuglia. MeshLab: an Open-Source Mesh Processing Tool. In *Eurographics Italian Chapter Conference*. The Eurographics Association, 2008. 8

[8] Omid Ehsani, M Pouladian, S Toosizadeh, and A Aledavood. Registration and fusion of 3d surface data from ct and tof camera for position verification in radiotherapy. *SN Applied Sciences*, 1(11):1347, 2019. 2

[9] Zhongpai Gao, Benjamin Planche, Meng Zheng, Xiao Chen, Terrence Chen, and Ziyan Wu. Ddgs-ct: Direction-disentangled gaussian splatting for realistic volume rendering. *Advances in Neural Information Processing Systems*, 37:39281–39302, 2024. 2

[10] Antoine Guédon and Vincent Lepetit. Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5354–5363, 2024. 2, 5

[11] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1):100–108, 1979. 4

[12] Michael Hess, Vid Petrovic, Dominique Meyer, Dominique Rissolo, and Falko Kuester. Fusion of multimodal three-dimensional data for comprehensive digital documentation of cultural heritage sites. In *2015 Digital Heritage*, pages 595–602. IEEE, 2015. 1

[13] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–11, 2024. 2

[14] Yingwenqi Jiang, Jiadong Tu, Yuan Liu, Xifeng Gao, Xiaoxiao Long, Wenping Wang, and Yuexin Ma. Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5322–5332, 2024. 5

[15] Sagi Katz, Ayellet Tal, and Ronen Basri. Direct visibility of point sets. In *ACM SIGGRAPH 2007 papers*, pages 24–es. 2007. 8

[16] Harpreet Kaur, Deepika Koundal, and Virender Kadyan. Image fusion techniques: a survey. *Archives of computational methods in Engineering*, 28(7):4425–4447, 2021. 2

[17] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42 (4):139–1, 2023. 1, 3, 4, 6, 7

[18] Shakiba Kheradmand, Daniel Rebain, Gopal Sharma, Weiwei Sun, Yang-Che Tseng, Hossam Isack, Abhishek Kar, Andrea Tagliasacchi, and Kwang Moo Yi. 3d gaussian splatting as markov chain monte carlo. *Advances in Neural Information Processing Systems*, 37:80965–80986, 2024. 7

[19] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[20] B Kozioziemski, B Bachmann, A Do, and R Tommasini. X-ray imaging methods for high-energy density physics applications. *Review of Scientific Instruments*, 94(4), 2023. 2

[21] Nikolas Lamb, Cameron Palmer, Benjamin Molloy, Sean Banerjee, and Natasha Kholgade Banerjee. Fantastic breaks: A dataset of paired 3d scans of real-world broken objects and their complete counterparts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4681–4691, 2023. 8

[22] Jianwei Li, Wei Gao, Yihong Wu, Yangdong Liu, and Yanfei Shen. High-quality indoor scene 3d reconstruction with rgb-d cameras: A brief review. *Computational Visual Media*, 8 (3):369–393, 2022. 2

[23] Xiangjie Li, Shuxiang Xie, Ken Sakurada, Ryusuke Sagawa, and Takeshi Oishi. Implicit neural fusion of rgb and far-infrared 3d imagery for invisible scenes. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 12501–12508. IEEE, 2024. 2

[24] Yingtai Li, Xueming Fu, Shang Zhao, Ruiyang Jin, and S Kevin Zhou. Sparse-view ct reconstruction with 3d gaussian volumetric representation. *arXiv preprint arXiv:2312.15676*, 2023. 3

[25] Zhihao Liang, Qi Zhang, Wenbo Hu, Lei Zhu, Ying Feng, and Kui Jia. Analytic-splatting: Anti-aliased 3d gaussian splatting via analytic integration. In *European conference on computer vision*, pages 281–297. Springer, 2024. 7

[26] Ke Liu, Hang Li, Dongdong Lu, and Huabing Huang. Rgb-infrared image fusion and classification based on tensor decomposition. In *Conference on Spectral Technology and Applications*, pages 928–936. SPIE, 2024. 2

[27] Yifei Liu, Zhihang Zhong, Yifan Zhan, Sheng Xu, and Xiao Sun. Maskgaussian: Adaptive 3d gaussian representation from probabilistic masks. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 681–690, 2025. 7

[28] Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. Neural volumes: Learning dynamic renderable volumes from images. *arXiv preprint arXiv:1906.07751*, 2019. 5

[29] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Seminal graphics: pioneering efforts that shaped the field*, pages 347–353. 1998. 7

[30] Yongyu Luo and Zhongqiang Luo. Infrared and visible image fusion: Methods, datasets, applications, and prospects. *Applied Sciences*, 13(19):10891, 2023. 2

[31] Benedict Marsh, Abdul Hamid Sadka, and Hamid Bahai. A critical review of deep learning-based multi-sensor fusion techniques. *Sensors*, 22(23):9364, 2022. 2

[32] Rekha R Nair and Tripty Singh. Multi-sensor, multi-modal medical image fusion for color images: A multi-resolution approach. In *2018 Tenth international conference on advanced computing (ICoAC)*, pages 249–254. IEEE, 2018. 2

[33] Emmanouil Nikolakakis, Utkarsh Gupta, Jonathan Vengosh, Justin Bui, and Razvan Marinescu. Gaspct: Gaussian splatting for novel ct projection view synthesis. *arXiv preprint arXiv:2404.03126*, 2024. 2

[34] Muyao Niu, Tong Chen, Yifan Zhan, Zhuoxiao Li, Xiang Ji, and Yinqiang Zheng. Rs-nerf: Neural radiance fields from rolling shutter images. In *European Conference on Computer Vision*, pages 163–180. Springer, 2024. 7

[35] Sławomir Paśko and Wojciech Glinkowski. Combining 3d structured light imaging and spine x-ray data improves visualization of the spinous lines in the scoliotic spine. *Applied Sciences*, 11(1):301, 2020. 2

[36] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 6

[37] Guocheng Qian, Jinjie Mai, Abdullah Hamdi, Jian Ren, Aliaksandr Siarohin, Bing Li, Hsin-Ying Lee, Ivan Skorokhodov, Peter Wonka, Sergey Tulyakov, and Bernard Ghanem. Magic123: One image to high-quality 3d object generation using both 2d and 3d diffusion priors. In *The Twelfth International Conference on Learning Representations*, 2024. 2

[38] Ayodeji Olalekan Salau, Shruti Jain, and Joy Nnenna Eneh. A review of various image fusion types and transform. *Indonesian Journal of Electrical Engineering and Computer Science*, 24(3):1515–1522, 2021. 2

[39] Jason Sanders. *CUDA by Example: An Introduction to General-Purpose GPU Programming*. Addison-Wesley Professional, 2010. 6

[40] Dioline Sara, Ajay Kumar Mandava, Arun Kumar, Shiny Duela, and Anitha Jude. Hyperspectral and multispectral image fusion techniques for high resolution applications: A review. *Earth Science Informatics*, 14(4):1685–1705, 2021. 2

[41] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4104–4113, 2016. 3

[42] Liao Shen, Tianqi Liu, Huiqiang Sun, Jiaqi Li, Zhiguo Cao, Wei Li, and Chen Change Loy. Dof-gaussian: Controllable depth-of-field for 3d gaussian splatting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 26462–26471, 2025. 7

[43] Ivana Shopovska, Ljubomir Jovanov, and Wilfried Philips. Deep visible and thermal image fusion for enhanced pedestrian visibility. *Sensors*, 19(17):3727, 2019. 2

[44] Changqi Sun, Cong Zhang, and Naixue Xiong. Infrared and visible image fusion techniques based on deep learning: A review. *Electronics*, 9(12):2162, 2020. 2

[45] Jing Ru Teoh, Jian Dong, Xiaowei Zuo, Khin Wee Lai, Khairunnisa Hasikin, and Xiang Wu. Advancing healthcare through multimodal data fusion: a comprehensive review of techniques and applications. *PeerJ Computer Science*, 10: e2298, 2024. 1

[46] Narasimhan Venkatanath, D Praneeth, Maruthi Chandrasekhar Bh, Sumohana S Channappayya, and Swarup S Medasani. Blind image quality evaluation using perception based features. In *2015 twenty first national conference on communications*, pages 1–6. IEEE, 2015. 7

[47] Haowen Wang, Mingyuan Wang, Zhengping Che, Zhiyuan Xu, Xiuquan Qiao, Mengshi Qi, Feifei Feng, and Jian Tang. Rgb-depth fusion gan for indoor depth completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6209–6218, 2022. 2

[48] Yue Wang, Jinlong Peng, Jiangning Zhang, Ran Yi, Yabiao Wang, and Chengjie Wang. Multimodal industrial anomaly detection via hybrid fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8032–8041, 2023. 1

[49] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 5

[50] Yaniv Wolf, Amit Bracha, and Ron Kimmel. GS2Mesh: Surface reconstruction from Gaussian splatting via novel stereo views. In *European Conference on Computer Vision*, 2024. 2, 7

[51] Lennard Wunsch, Christian Görner Tenorio, Katharina Anding, Andrei Golomoz, and Gunther Notni. Data fusion of rgb and depth data with image enhancement. *Journal of Imaging*, 10(3):73, 2024. 2

[52] Qiangeng Xu, Zexiang Xu, Julien Philip, Sai Bi, Zhixin Shu, Kalyan Sunkavalli, and Ulrich Neumann. Point-nerf: Point-based neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5438–5448, 2022. 5

[53] Yushen Xu, Xiaosong Li, Yuchan Jie, and Haishu Tan. Simultaneous tri-modal medical image fusion and super-resolution using conditional diffusion model. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 635–645. Springer, 2024. 2

[54] Chen Yang, Sikuang Li, Jiemin Fang, Ruofan Liang, Lingxi Xie, Xiaopeng Zhang, Wei Shen, and Qi Tian. Gaussianobject: High-quality 3d object reconstruction from four views with gaussian splatting. *ACM Transactions on Graphics*, 2024. 2

[55] Muhammad Yaqub, Feng Jinchao, Kaleem Arshid, Shahzad Ahmed, Wenqian Zhang, Muhammad Zubair Nawaz, and Tariq Mahmood. Deep learning-based image reconstruction for different medical imaging modalities. *Computational and Mathematical Methods in Medicine*, 2022(1):8750648, 2022. 2

[56] Taoran Yi, Jiemin Fang, Junjie Wang, Guanjun Wu, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Qi Tian, and Xinggang Wang. Gaussiandreamer: Fast generation from text to 3d gaussians by bridging 2d and 3d diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 2

[57] Zehao Yu, Torsten Sattler, and Andreas Geiger. Gaussian opacity fields: Efficient adaptive surface reconstruction in unbounded scenes. *ACM Transactions on Graphics*, 2024. 2

[58] Ruyi Zha, Yanhao Zhang, and Hongdong Li. Naf: neural attenuation fields for sparse-view cbct reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 442–452. Springer, 2022. 7

[59] Ruyi Zha, Tao Jun Lin, Yuanhao Cai, Jiwen Cao, Yanhao Zhang, and Hongdong Li. R^2-gaussian: Rectifying radiative gaussian splatting for tomographic reconstruction. In *Advances in Neural Information Processing Systems*, 2024. 3, 7

[60] Jiangsan Zhao, Ajay Kumar, Balaji Naik Banoth, Balram Marathi, Pachamuthu Rajalakshmi, Boris Rewald, Seishi Ninomiya, and Wei Guo. Deep-learning-based multispectral image reconstruction from single natural color rgb image—enhancing uav-based phenotyping. *Remote Sensing*, 14(5):1272, 2022. 2

[61] Meng Zhou, Yuxuan Zhang, Xiaolan Xu, Jiayi Wang, and Farzad Khalvati. Edge-enhanced dilated residual attention network for multimodal medical image fusion. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 4108–4111. IEEE, 2024. 2

InsideOut: Integrated RGB-Radiative Gaussian Splatting for Comprehensive 3D Object Representation

Supplementary Material

A. Effect of Hyperparameters.

Loss Balancing. Tab. A shows that the values of the optimal parameters are $\lambda_s = 0.2$ and $\lambda_z = 0.005$. The value of $\lambda_s = 0.2$ is commonly used as a standard in the field, and our experimental results confirm its effectiveness. For λ_z, values that are too low introduce excessive noise, while values that are too high cause distortion in RGB Gaussian splats and fading in radiative Gaussian splats. Our experiments reveal that $\lambda_z = 0.005$ provides the ideal balance, effectively flattening noise while clearly distinguishing between layers.

	RGB Gaussian Splatting		Radiative Gaussian Splatting	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
$\lambda_s = 0.0, \lambda_z = 0.005$	34.66	0.92	27.42	0.97
$\lambda_s = 0.1, \lambda_z = 0.005$	37.71	0.94	27.43	0.97
$\mathbf{\lambda_s = 0.2, \lambda_z = 0.005}$	37.74	0.94	28.14	0.97
$\lambda_s = 0.3, \lambda_z = 0.005$	37.02	0.94	28.07	0.97
$\lambda_s = 0.2, \lambda_z = 0.000$	35.54	0.92	27.41	0.96
$\lambda_s = 0.2, \lambda_z = 0.001$	37.26	0.94	27.45	0.97
$\mathbf{\lambda_s = 0.2, \lambda_z = 0.005}$	37.74	0.94	28.14	0.97
$\lambda_s = 0.2, \lambda_z = 0.010$	36.87	0.93	27.61	0.96

Table A. **Analysis of weighting parameters.** The optimal values are highlighted in bold.

Fine Gaussian Selection. As shown in Fig. A, setting $p = 95$ retains sufficient fine Gaussian splats to cover the full surface. Lower p values result in incomplete surface reconstruction, whereas higher values introduce redundant elements without significant improvement. Thus, $p = 95$ provides an optimal balance between surface fidelity and structural efficiency.

B. Computational Efficiency.

Our proposed modules—hierarchical fitting and visual alignment—are highly efficient, requiring only 9.83 minutes (46,129 MB) and 2 seconds (697 MB), respectively, compared to 32.66 minutes (12,332 MB) for rendering on an NVIDIA A6000 GPU. The training of our algorithm takes approximately 45 minutes—slightly longer than rendering X-ray and RGB images separately—yet it maintains the inference speed of baseline methods.

C. Real-world Applicability.

Our method enables multi-domain applications using only RGB and X-ray data without calibration. The elimination of complex calibration procedures significantly reduces setup
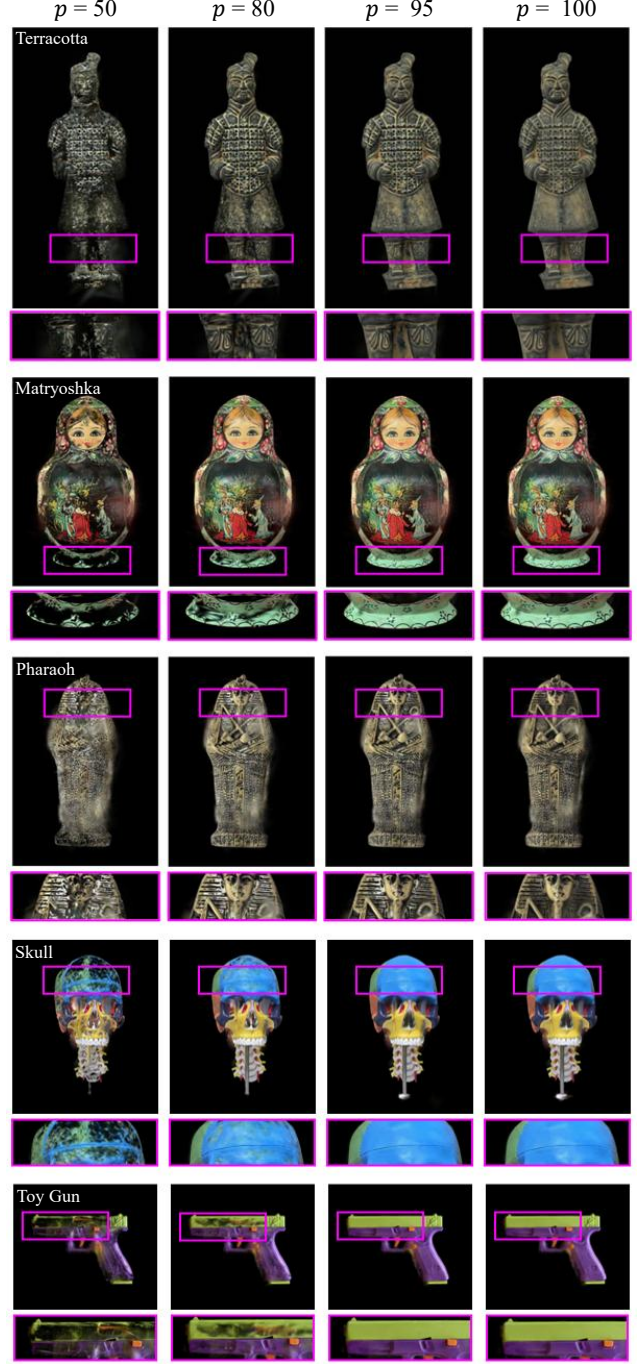

Figure A. **Optimal selection threshold p for efficient surface representation.** Pink boxes indicate zoomed regions.

time and makes the technology more accessible across different environments. In surgical planning, we provide visualization of underlying fractures beneath bruising, helping surgeons optimize incisions and employ real-time navigation. For cultural heritage preservation, we expose internal cracks connecting to the surface and provide museum visitors with augmented reality experiences to observe artifacts' interior and exterior simultaneously. In industrial defect detection, we examine both external appearance and internal structure to identify hidden defects, improving correlation analysis and root cause identification in production.