

MedGSSR: Generalizable Medical Image Super-Resolution 3D Reconstruction via Hierarchical Feed-forward Gaussian Splatting

Chengkai Wang^{1*}, Luoyu Hong^{2*}, Yiting Zhao², Jiamin Wang³, Xiang Feng³,
Feiwei Qin², Zhenzhong Kuang², Xuefei Yin⁴
Ali Bashashati¹ , and Yanming Zhu⁴ 

¹ University of British Columbia, Vancouver, Canada

² Hangzhou Dianzi University, Hangzhou, China

³ ShanghaiTech University, Shanghai, China

⁴ Griffith University, Gold Coast, Australia

Abstract. High-resolution volumetric medical imaging is critical for clinical diagnosis, yet acquisition is often limited by scanner hardware, scan time, and for CT, radiation dose. Medical 3D Super-Resolution (Med3DSR) offers a computational alternative, but existing methods commonly rely on per-subject optimization, pretrained priors, or coordinate-based implicit representations, which compromise anatomical fidelity and limit efficiency. To address these limitations, we present **MedGSSR**, a fully end-to-end feed-forward framework that represents volumes as an explicit 3D Gaussian field for Med3DSR. Unlike coordinate-based implicit functions, our explicit 3D Gaussian representation naturally enhances signal continuity and local high-frequency fidelity. Specifically, MedGSSR explicitly decouples the reconstruction process into coarse-grained structural preservation and fine-grained textural refinement through the proposed Pyramid Anatomical Encoder and a Hierarchical Gaussian Projector. To support arbitrary-scale super-resolution, we introduce sub-voxel Gaussian decomposition and a Differentiable Gaussian Voxelizer that directly queries the continuous 3D intensity field, reducing discretization artifacts. Extensive experiments on MRI and CT benchmarks demonstrate that MedGSSR significantly outperforms state-of-the-art methods. Notably, our framework exhibits robust generalizability across unseen datasets without requiring per-subject optimization, enabling fast inference and high-fidelity volumetric super-resolution in practical clinical settings. Our project webpage, including code, is at <https://william2ai.github.io/medgssr>

Keywords: 3D Medical Image Super-Resolution · Feed-forward Gaussian Splatting · Arbitrary-Scale Reconstruction

1 Introduction

High-resolution (HR) volumetric medical imaging, such as MRI and CT, is fundamental to accurate clinical diagnosis and quantitative analysis. However, ob-

* Equal contribution.  Corresponding authors.

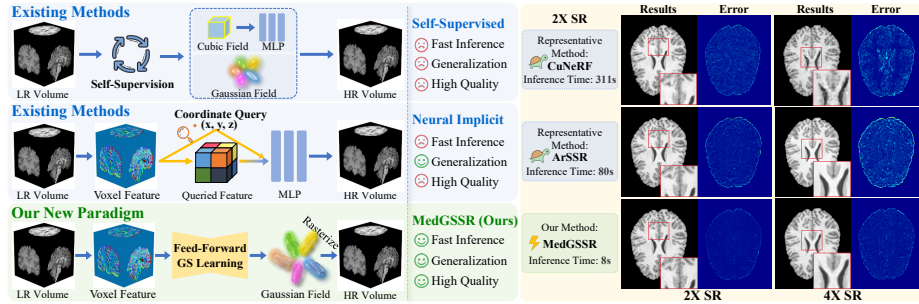


Fig. 1: MedGSSR introduces a paradigm shift for Medical 3D Super-Resolution. Left: Existing self-supervised methods (relying on isolated optimization) and neural implicit networks (relying on coordinate querying) fail to simultaneously achieve fast inference, robust generalization, and high-quality reconstruction. MedGSSR introduces a generalized feed-forward 3DGS paradigm that explicitly satisfies all three clinical requirements. Right: Cross-dataset Visual comparisons (trained on MSD, tested on HCP) at 2× and 4× upsampling scales. Compared to baselines that produce geometric artifacts or over-smoothed textures, MedGSSR accurately restores high-frequency anatomical structures with the lowest error (dark blue in difference maps) while accelerating inference by an order of magnitude.

taining isotropic HR volumes is often constrained by physical and practical limitations. High-field MRI scanners remain scarce, and prolonged scanning times required for high spatial resolution increase the risk of motion artifacts [12, 24]. Similarly, in CT imaging, increasing resolution inevitably requires higher radiation doses, heightening safety risks [15]. To bridge the gap between the observed low-resolution (LR) data and the clinically desired HR volumes, Med3DSR has emerged as a critical research focus, aiming to computationally recover missing spatial information without modifying the imaging hardware.

Despite recent progress, clinical-grade Med3DSR remains challenged by two persistent bottlenecks, as shown in Fig. 1. First, continuous zooming does not necessarily translate to high-frequency fidelity. Implicit Neural Representations (INRs) enable efficient, arbitrary-scale querying by learning a coordinate-to-intensity mapping [25]. However, coordinate-based implicit functions face intrinsic representational limitations. Because a single set of shared network weights must represent the entire volumetric signal, coordinate-based INRs are prone to spectral bias [20], often fitting low-frequency components and producing over-smoothed reconstructions that miss fine anatomical structures and pathological boundaries. Second, limited isotropic 3D data hinders learning 3D-consistent priors and generalization. Due to the scarcity of high-quality isotropic 3D medical data, explicit reconstruction pipelines retreat to either isolated per-subject optimization [5] or 2D proxy supervision [18]. Per-subject optimization treats each scan as an independent fitting problem, which is computationally prohibitive and prevents transferable anatomical priors from being learned. Lacking holistic 3D spatial context, these slice-wise 2D proxies often introduce inter-slice incon-

sistencies [13] and texture artifacts, making the attainable 3D reconstruction fidelity strongly dependent on the quality of the 2D pseudo-labels [10, 11].

To address these limitations, we propose MedGSSR, a novel paradigm that reformulates Med3DSR as a generalized, purely volumetric feed-forward prediction of an explicit 3D Gaussian Splatting (3DGS) field. In contrast to coordinate-based INRs that encode the entire volume in shared global weights, MedGSSR represents anatomy as a set of local-support Gaussian primitives, which reduces global coupling across distant regions and facilitates localized modeling of high-frequency structures. Furthermore, since medical modalities directly measure continuous physical density (e.g., Hounsfield units in CT or proton density in MRI), modeling the anatomy as an explicit superposition of continuous Gaussian distributions ensures strict physical consistency. Moreover, unlike optimization-based methods that isolate each scene, MedGSSR learns a direct feed-forward mapping function from sparse LR voxel grids to continuous HR 3DGS parameters. By training on large-scale multi-subject datasets, our model autonomously learns 3D-specific anatomical priors, reconstructing accurate high-frequency details via continuous Gaussian representations across unseen subjects without test-time optimization.

To effectively adapt this paradigm for highly heterogeneous medical data, we design a specialized architecture. We introduce a Pyramid Anatomical Encoder that extracts multi-scale volumetric features and organizes them into two complementary representations capturing global anatomy and local appearance. Subsequently, a Hierarchical Gaussian Projector explicitly factorizes reconstruction into two branches: a coarse branch dedicated to preserving macroscopic organ geometry, and a fine branch focused on refining local tissue textures. To bridge the gap between discrete feature grids and continuous biological signals, we employ a sub-voxel decomposition strategy, where each voxel is represented by a set of adaptive Gaussian primitives, effectively eliminating discretization artifacts. Finally, a Differentiable Gaussian Voxelizer aggregates these primitives into a continuous 3D volumetric field. Consequently, this formulation allows us to render HR volumes at any desired resolution by simply querying the intensity field at corresponding spatial coordinates, entirely within the 3D domain.

The core contributions of our study are as follows:

- We reformulate Med3DSR as a generalized, purely volumetric feed-forward mapping from sparse LR voxel grids to an explicit continuous 3DGS field, shifting the paradigm from INRs and per-subject optimization to a generalizable physics-consistent explicit representation that enables fast inference, high-fidelity reconstruction, and cross-subject generalization.
- We propose a Hierarchical Gaussian Projector with a sub-voxel decomposition mechanism, which explicitly decouples structural preservation from textural refinement to handle the strong heterogeneity of medical images.
- Extensive experiments on MRI and CT benchmarks demonstrate that the proposed MedGSSR outperforms SOTA approaches, achieving superior high-frequency reconstruction fidelity and robust cross-dataset generalization.

2 Related Work

2.1 3D Reconstruction

The evolution of recent neural 3D scene reconstruction has been heavily driven by Implicit Neural Representations, notably NeRF [16], which models continuous volumetric fields using MLPs. Subsequent variants improved rendering quality and optimization speed by integrating multi-scale representations [1], hash-grid encodings [17], and tensorial decomposition [3]. To overcome the computational bottleneck of volume rendering, 3DGS [14] introduced an explicit point-based formulation for efficient rasterization, with methods like Mip-Splatting [30] further addressing aliasing artifacts.

In the domain of 3D super-resolution (3DSR), early implicit methods [23, 29] and recent 3DGS-based models [9, 26] rely heavily on pretrained 2D single-image super-resolution (SISR) networks to generate pseudo-HR labels. While improving visual quality, these methods require time-consuming per-subject optimization. Furthermore, relying on 2D models to process individual viewpoints independently fundamentally lacks 3D spatial awareness. This mechanism inevitably introduces cross-view inconsistencies and caps reconstruction fidelity at the upper bound of the 2D pseudo-labels.

To bypass per-subject optimization, recent vision models [2, 22, 27, 28] have explored the direct feed-forward prediction of 3DGS parameters from sparse images. However, these architectures target sparse natural scenes and are structurally ill-equipped for the dense, continuous topologies of volumetric medical data. In contrast to these approaches, our method explicitly adapts the feed-forward 3DGS paradigm for medical volumes to establish generalized, 3D-consistent structural priors.

2.2 Medical Image Super-Resolution

Early data-driven medical image super-resolution methods predominantly relied on 3D CNNs, including SRCNN3D [19], mDCSRN [4], and ResCNN [7]. While these architectures effectively restore local textures compared to traditional interpolation, their heavy computational demands and restriction to fixed integer upsampling scales limit their clinical flexibility.

To achieve arbitrary-scale super-resolution, the focus shifted toward INRs [5, 8, 25], which map continuous spatial coordinates to intensity values. Additionally, self-supervised learning frameworks [21] train exclusively on the input volume to avoid the need for external HR data. However, coordinate-based MLPs inherently suffer from spectral bias, prioritizing low-frequency structures and producing over-smoothed textures. Concurrently, self-supervised methods fail to leverage large-scale anatomical priors due to their isolated training mechanisms.

Recently, researchers have attempted to apply explicit 3D representations to medical super-resolution. Due to the scarcity of paired 3D HR data, methods such as MedNeRF [6] and diffusion-guided NAB-GS [18] project volumes into 2D X-ray domains to exploit 2D generative priors. This 3D-to-2D-to-3D pipeline

inevitably introduces modality gaps and geometric ambiguity during projection. Unlike these projection-based or implicit approaches, MedGSSR operates purely in the continuous 3D domain, avoiding spectral bias and modality gaps entirely.

3 Preliminaries

In this section, we briefly review the standard formulation of 3DGS and discuss the paradigm shift toward feed-forward prediction, which forms the theoretical foundation of our MedGSSR framework.

3.1 3D Gaussian Splatting (3DGS)

3DGS models a volumetric scene using a collection of anisotropic 3D Gaussian primitives $\mathcal{G} = \{G_1, \dots, G_N\}$. In contrast to INRs that implicitly encode scene properties within neural network weights, 3DGS explicitly parameterizes geometry and appearance. Each primitive G_i is characterized by a center position $\boldsymbol{\mu}_i \in \mathbb{R}^3$, a covariance matrix $\boldsymbol{\Sigma}_i \in \mathbb{R}^{3 \times 3}$, an opacity scalar $\xi_i \in [0, 1]$, and appearance features $\mathbf{c}_i$. The spatial contribution of the i -th Gaussian at a query point $\mathbf{x} \in \mathbb{R}^3$ is given by:

$$G_i(\mathbf{x}) = \xi_i \cdot \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right). \quad (1)$$

These Gaussians are projected into screen space and rasterized via alpha blending to render novel views. To ensure $\boldsymbol{\Sigma}_i$ is symmetric positive definite (and thus invertible), it is typically parameterized as a scaling vector $\mathbf{s}_i$ and a rotation quaternion $\mathbf{q}_i$. Consequently, the full learnable parameter set becomes $\Theta = \{\boldsymbol{\mu}_i, \mathbf{q}_i, \mathbf{s}_i, \xi_i, \mathbf{c}_i\}$, which is optimized via gradient descent to minimize the reconstruction error between rendered projections and ground-truth images.

3.2 Feed-Forward 3D Gaussian Prediction

In contrast to standard 3DGS which entails computationally intensive per-subject optimization and lacks generalization across subjects, recent advances in computer vision have recast scene reconstruction as a direct regression task. While existing feed-forward models primarily map sparse 2D images to 3D scenes, we reconceptualize this paradigm for continuous volumetric medical data.

Specifically, we formulate the Med3DSR task as a direct voxel-to-Gaussian mapping. Given a low-resolution input observation $\mathbf{V}_{LR}$ (i.e., a 3D voxel grid), our objective is to learn a mapping function Ψ_θ , parameterized by a neural network, that directly predicts the Gaussian attributes for any given spatial location $\mathbf{x}$:

$$\{\boldsymbol{\mu}, \mathbf{q}, \mathbf{s}, \xi, \mathbf{c}\}_{\mathbf{x}} = \Psi_\theta(E(\mathbf{V}_{LR}))_{\mathbf{x}}, \quad (2)$$

where $E(\cdot)$ denotes a volumetric feature extraction encoder. This explicit formulation allows the network to predict a dense, continuous field of Gaussian primitives directly from 3D medical scans.

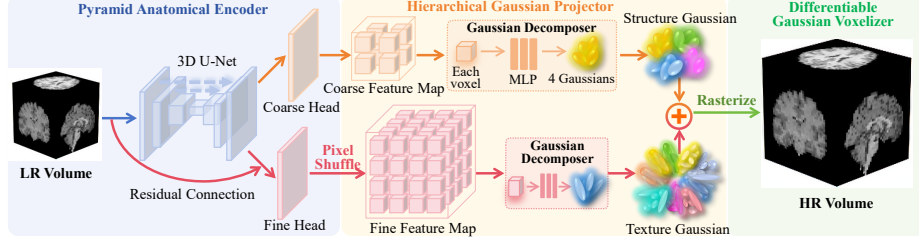


Fig. 2: Overview of MedGSSR. Given an LR input volume, the Pyramid Anatomical Encoder (PAE) built on a 3D U-Net extracts a coarse feature map and a $2\times$ upsampled fine feature map via input-residual connection and PixelShuffle. The Hierarchical Gaussian Projector (HGP) performs sub-voxel Gaussian decomposition by predicting multiple Gaussian primitives per voxel with lightweight MLP heads, producing structural Gaussians from the coarse stream and textural Gaussians from the fine stream. These primitives are then aggregated by the Differentiable Gaussian Voxelizer (DGV) through differentiable rasterization to form a continuous volumetric intensity field, from which the HR volume is rendered.

4 Methodology

4.1 Overview

Given a low-resolution (LR) medical volumetric input $\mathbf{V}_{LR} \in \mathbb{R}^{H \times W \times D}$, our goal is to reconstruct a high-fidelity super-resolved volume $\mathbf{V}_{HR}$ at an arbitrary upsampling scale κ . MedGSSR reformulates this task as a direct prediction problem, learning a generalizable mapping function $\Psi : \mathbf{V}_{LR} \rightarrow \mathcal{G}$, where $\mathcal{G}$ represents a continuous explicit field of 3D Gaussian primitives, as illustrated in Fig. 2.

4.2 Pyramid Anatomical Encoder

To effectively capture the heterogeneous nature of medical scans, where macroscopic organ geometry and microscopic tissue textures coexist, we design the PAE as a dual-stream feature extractor built upon a 3D U-Net architecture. We deliberately eschew Batch Normalization layers throughout the network to preserve absolute intensity information (e.g., Hounsfield Units in CT), which is clinically critical for accurate diagnosis. Specifically, the 3D U-Net processes $\mathbf{V}_{LR}$ and outputs the final decoder feature map $\mathbf{H}_{dec}$. The coarse stream then maps $\mathbf{H}_{dec}$ to a coarse feature map $\mathbf{F}_{coarse} \in \mathbb{R}^{H \times W \times D \times C}$ via a 3D convolution layer. Defined on the original $1\times$ voxel grid, $\mathbf{F}_{coarse}$ preserves spatial alignment while capturing organ-scale context. Compared with $\mathbf{F}_{fine}$ defined below, each voxel in $\mathbf{F}_{coarse}$ corresponds to a $2\times$ larger spacing along each spatial axis on the same underlying volume, and thus covers an $8\times$ larger physical volume. This coarse representation therefore provides a large receptive field for capturing global anatomical structures.

For the fine-grained stream, to enhance high-frequency details while retaining structural cues, we introduce an *input-residual connection* through concatenation. Specifically, we concatenate the LR input $\mathbf{V}_{LR}$ as a residual signal with the decoder feature map $\mathbf{H}_{dec}$ along the channel dimension. The concatenated tensor is then passed through a 3D convolution layer to expand channels, followed by a 3D PixelShuffle operator with an upsampling factor of 2:

$$\mathbf{F}_{fine} = \text{PixelShuffle}(\text{Conv}_{3D}([\mathbf{H}_{dec}, \mathbf{V}_{LR}])) \in \mathbb{R}^{2H \times 2W \times 2D \times C'}, \quad (3)$$

where $[\cdot, \cdot]$ denotes channel-wise concatenation. This dual-resolution design provides complementary coarse-to-fine features for subsequent Gaussian projection.

4.3 Hierarchical Gaussian Projector (HGP)

Structure-Texture Decoupling. The HGP comprises two parallel projection heads implemented as dedicated MLPs to directly regress Gaussian attributes. The structure-texture decoupling is fundamentally enabled by our dual-resolution feature design. The coarse head operates on the $1\times$ resolution feature map $\mathbf{F}_{coarse}$, where each voxel corresponds to a larger physical neighborhood on the underlying volume and thus emphasizes organ-scale, low-frequency anatomical context. Conversely, the fine head operates on the $2\times$ upsampled feature map $\mathbf{F}_{fine}$, which provides denser spatial sampling and richer local cues, and is tailored to refine microscopic, high-frequency details and boundaries.

Sub-voxel Gaussian Decomposition. A core challenge in volumetric SR is the discretization artifact arising from sparse voxel grids, which limits the ability to represent sub-voxel intensity variations. To address this, we propose a Sub-voxel Gaussian Decomposition strategy within the HGP. Instead of assigning a single Gaussian primitive to each voxel location, we predict a set of m Gaussian primitives with learnable sub-voxel offsets and scales in the local neighborhood of the voxel center. This increases local representational capacity by enabling multiple continuous primitives to describe fine structural transitions within one voxel, which is beneficial around tissue boundaries and highly textured regions.

For the coarse branch, given a feature vector $\mathbf{f}_{v_c}^c \in \mathbf{F}_{coarse}$ located at the 3D grid coordinate $\mathbf{v}_c$, the MLP predicts m ‘‘Structural Gaussian’’ $\mathcal{G}_{v_c}^c = \{G_j^c\}_{j=1}^m$. Similarly, for the fine branch, given a feature vector $\mathbf{f}_{v_f}^f \in \mathbf{F}_{fine}$ at the upsampled grid coordinate $\mathbf{v}_f$, the fine MLP predicts m ‘‘Textural Gaussians’’ $\mathcal{G}_{v_f}^f = \{G_j^f\}_{j=1}^m$. Specifically, for the j -th sub-Gaussian in either branch, the projector regresses the 11-dimensional primitive attributes relative to its anchor voxel center $\mathbf{v} \in \{\mathbf{v}_c, \mathbf{v}_f\}$:

$$\boldsymbol{\mu}_j = \mathbf{v} + \delta \cdot \tanh(\Delta\boldsymbol{\mu}_j), \quad \mathbf{s}_j = \text{softplus}(\Delta\mathbf{s}_j - \mathbf{s}_{base}), \quad (4)$$

$$\mathbf{q}_j = \frac{\mathbf{q}_{raw,j}}{\|\mathbf{q}_{raw,j}\|}, \quad \alpha_j = \text{sigmoid}(\rho_j - \alpha_{base}), \quad (5)$$

where δ restricts the positional offset to enforce local spatial connectivity, preventing primitive collapse, and α_j denotes the intensity amplitude of the Gaussian. Crucially, we apply distinct adaptive initialization strategies to enforce

the role of each branch. Structural Gaussians are activated with a larger s_j and higher α_j to cover uniform regions and form the macroscopic anatomical backbone. Textural Gaussians utilize a smaller s_j and lower α_j , specializing in describing microscopic variations.

To form the complete volumetric representation, we concatenate the output of both branches along the point dimension. Let $N_{coarse} = m \times (HWD)$ and $N_{fine} = m \times (8HWD)$. The final explicit representation is a unified set of $K = N_{coarse} + N_{fine}$ Gaussian primitives:

$$\mathcal{G} = \bigcup_{\mathbf{v}_c} \mathcal{G}_{\mathbf{v}_c}^c \cup \bigcup_{\mathbf{v}_f} \mathcal{G}_{\mathbf{v}_f}^f = \{G_i\}_{i=1}^K. \quad (6)$$

4.4 Differentiable Gaussian Voxelizer

Unlike standard view synthesis where Gaussians are projected onto a 2D plane via alpha blending along camera rays, our Med3DSR task requires reconstructing a purely 3D scalar field. Since medical volumes represent continuous-valued signals, we formulate the reconstructed continuous intensity function $I(\mathbf{x}) : \mathbb{R}^3 \rightarrow \mathbb{R}$ at any arbitrary spatial query coordinate $\mathbf{x}$ as the direct spatial superposition of the Gaussian primitives.

The reconstructed intensity is modeled as the aggregation of contributions from all K Gaussians in the set $\mathcal{G}$. Formally, the voxelizer evaluates the intensity by aggregating contributions from all primitives:

$$I(\mathbf{x}) = \sum_{i=1}^K \alpha_i \cdot \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right), \quad (7)$$

where $\boldsymbol{\mu}_i$ and α_i denote the center and intensity amplitude of the i -th Gaussian, respectively. We use the unnormalized Gaussian kernel as a smooth, distance decaying weighting function. The covariance matrix is parameterized as $\boldsymbol{\Sigma}_i = \mathbf{R}(\mathbf{q}_i) \text{diag}(\mathbf{s}_i^2) \mathbf{R}(\mathbf{q}_i)^\top$, with scaling $\mathbf{s}_i \in \mathbb{R}_+^3$ and a unit quaternion $\mathbf{q}_i$ defining $\mathbf{R}(\mathbf{q}_i)$. For efficiency, we approximate the summation in Eq. (7) by evaluating only primitives within a fixed truncation radius in Mahalanobis space, namely those satisfying $d_M(\mathbf{x}, \boldsymbol{\mu}_i) < 3$ (a commonly used 3σ like confidence interval), where $d_M(\mathbf{x}, \boldsymbol{\mu}_i) = \sqrt{(\mathbf{x} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)}$. Primitives outside this neighborhood contribute negligibly and are omitted to reduce computation, while Eq. (7) defines a continuous and differentiable intensity field.

Arbitrary-Scale Inference. The continuous mathematical formulation of Eq. (7) fundamentally decouples the target resolution from the network architecture. During inference, to reconstruct a volume at an arbitrary upsampling scale $\boldsymbol{\kappa} = (\kappa_h, \kappa_w, \kappa_d)$, we simply instantiate a dense spatial coordinate grid of size $(\lceil \kappa_h H \rceil \times \lceil \kappa_w W \rceil \times \lceil \kappa_d D \rceil)$. By querying the continuous intensity function $I(\mathbf{x})$ at each corresponding grid coordinate $\mathbf{x}$, MedGSSR renders the HR volume at any desired resolution using a single trained model.

4.5 Optimization Objective

Since the DGV module enables back-propagation through volumetric rendering, MedGSSR can be trained end-to-end. We supervise the reconstructed volume $\mathbf{V}_{pred}$ (sampled at the target HR grid resolution) against the ground-truth high-resolution volume $\mathbf{V}_{GT}$ using a voxel-wise L_1 reconstruction loss:

$$\mathcal{L}_{rec} = \|\mathbf{V}_{GT} - \mathbf{V}_{pred}\|_1. \quad (8)$$

This objective effectively drives the network to recover accurate anatomical structures and textures through the explicit 3D Gaussian representation.

5 Experiments

5.1 Experimental Setup

Datasets. We conduct comprehensive evaluations on two widely adopted medical imaging modalities: MRI and CT. For MRI, we use the Medical Segmentation Decathlon (MSD) dataset, which contains high-quality T1-weighted brain volumes for intra-domain training and evaluation. To assess generalization, we select a testing subset from the Human Connectome Project (HCP) dataset, which features different scanner characteristics and population demographics from the MSD dataset. For CT, we train on the MELA dataset, containing diverse chest CT scans, and test generalization on an independent subset of the Ultra-High-Resolution CT (UHRCT) dataset, which captures fine-grained anatomical lung structures. Following standard protocols, low-resolution inputs are generated by downsampling the ground-truth volumes using cubic interpolation.

Baselines and Metrics. MedGSSR is compared against state-of-the-art (SOTA) methods grouped into four paradigms: (1) Traditional: Trilinear and Cubic interpolation; (2) Self-supervised: CuNeRF [5] and 2D-diffusion-guided NAB-GS [18]; (3) Neural Implicit: INR-based ArSSR [25]; and (4) the proposed Feed-Forward 3DGS paradigm (FF-3DGS): MedGSSR. Reconstruction fidelity is quantified using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS).

Implementation Details. We implement MedGSSR in PyTorch. The PAE is built upon a 3-layer 3D U-Net architecture without batch normalization to preserve absolute intensity values. Given an input LR patch (e.g., $1 \times 32 \times 32 \times 32$), the PAE extracts a coarse feature grid of size 128×32^3 . Concurrently, a fine-grained feature grid of size 128×64^3 is generated via a 3D PixelShuffle upsampling operation. The HGP consists of parallel MLPs for the coarse and fine branches. Each MLP contains three fully connected layers with GELU activations, mapping the 128-dimensional features to 11-dimensional Gaussian attributes $(\alpha, \mathbf{s}, \boldsymbol{\mu}, \mathbf{q})$ for $m = 4$ sub-voxels. To explicitly enforce the structure-texture decoupling, we apply distinct initializations for the two branches: Structural Gaussians use a smaller s_{base} 4.0 and amplitude bias α_{base} 2.0 to model dense macroscopic structures, whereas Textural Gaussians are initialized with a larger s_{base} 6.0

Table 1: Comparison of intra-domain 3D super-resolution performance on the MSD (MRI) dataset. Models are both trained and evaluated on the MSD dataset.

Paradigm	Methods	2×			3×			4×		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Traditional	Trilinear	33.31	0.9670	0.1134	30.77	0.9377	0.2026	29.08	0.9100	0.2558
	Cubic	33.99	0.9725	0.0981	31.13	0.9422	0.2083	29.27	0.9118	0.2711
Self-supervised	CuNeRF [5]	32.60	0.9715	0.0513	30.07	0.9524	0.0878	28.26	0.9309	0.1271
Neural Implicit	ArSSR [25]	32.87	0.9742	0.0581	29.56	0.9442	0.0730	28.51	0.9315	0.0853
FF-3DGS	MedGSSR	35.91	0.9821	0.0337	33.97	0.9700	0.0706	32.10	0.9541	0.1246

Table 2: Comparison of intra-domain 3D super-resolution performance on the MELA (CT) dataset. Models are both trained and evaluated on the MELA dataset.

Paradigms	Methods	2×			3×			4×		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Traditional	Trilinear	37.51	0.9530	0.1561	34.83	0.9166	0.2302	33.84	0.8843	0.3291
	Cubic	37.64	0.9636	0.1179	35.14	0.9237	0.2325	34.09	0.8888	0.3263
Self-supervised	CuNeRF [5]	37.12	0.9654	0.1201	35.42	0.9235	0.1863	33.94	0.8957	0.2456
	NAB-GS [18]	-	-	-	-	-	-	34.13	0.9518	-
Neural Implicit	ArSSR [25]	38.53	0.9644	0.0914	36.02	0.9413	0.1576	34.63	0.9244	0.2261
FF-3DGS	MedGSSR	42.08	0.9738	0.0705	39.95	0.9558	0.1193	37.31	0.9362	0.1472

and amplitude bias α_{base} 4.0 to capture sparse high-frequency edges. The entire framework is trained end-to-end for 1,000,000 iterations using the Adam optimizer. The learning rate is initialized at 1×10^{-4} and decays logarithmically to 1×10^{-5} . All experiments are conducted on four NVIDIA RTX 4090 GPUs.

5.2 Main Results

Quantitative Analysis. We first assess the reconstruction fidelity of MedGSSR within the primary MRI and CT domains. As reported in Table 1, MedGSSR establishes a new state-of-the-art across spatial scales on the MSD dataset. At the highly challenging 4× upsampling scale, our method achieves a PSNR of 32.10 dB, delivering a remarkable improvement (+3.59 dB) over the neural implicit method ArSSR. These results quantitatively prove that representing anatomy with explicit, localized Gaussian primitives successfully bypasses the spectral bias of coordinate-based MLPs, enabling the network to reconstruct high-frequency structural variations without the typical over-smoothing penalty. Table 2 details the corresponding results for the CT modality on the MELA dataset. Consequently, at the challenging 4× scale, our method achieves the highest geometric accuracy (37.31 dB PSNR), outperforming ArSSR by +2.68 dB. Additionally, for the diffusion-guided method NAB-GS, due to the unavailability of its open-source code, we directly report the 4× performance quoted from their original paper. Notably, MedGSSR still surpasses this reported per-

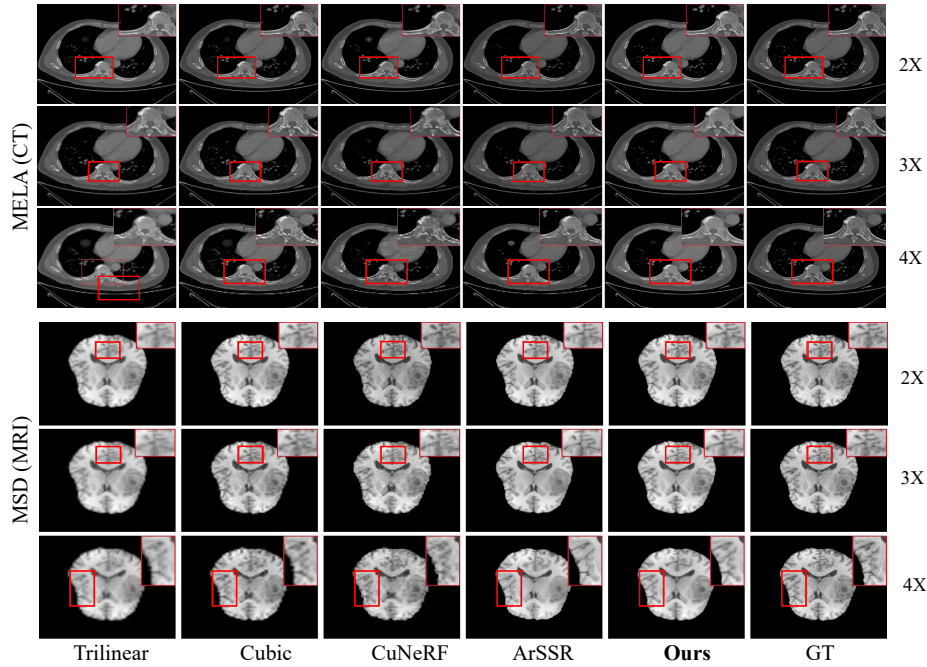


Fig. 3: Qualitative comparison of intra-domain 3DSR on MELA (CT) and MSD (MRI). Axial views are shown at $2\times$, $3\times$ and $4\times$ scales. MedGSSR better recovers fine-grained skeletal and vascular structures, avoiding the smoothing artifacts of ArSSR and the noise commonly observed in other methods.

formance by a substantial +3.18 dB margin. These results provide strong evidence that learning a 3D-native feed-forward mapping directly from voxel grids is fundamentally superior to paradigms constrained by isolated optimization, coordinate-based mappings, or 2D proxy supervision.

Qualitative Visualizations. Fig. 3 shows qualitative comparisons on MELA (CT) and MSD (MRI) in the axial plane at $2\times$, $3\times$, and $4\times$ upsampling factors. The self-supervised CuNeRF and the neural implicit ArSSR produce noticeable blurring, and ArSSR yields over-smoothed tissue boundaries. In contrast, MedGSSR reconstructs sharper anatomical structures, such as clearer skeletal edges and finer vessels, supporting the effectiveness of our dual-resolution structure-texture decoupling. Additional visualizations are provided in the **Supplementary Material**.

5.3 Generalization Results

A fundamental limitation of isolated optimization methods is their inability to generalize to unseen datasets. To assess the cross-dataset generalizability of MedGSSR, we evaluate models trained on the source datasets (MSD/MELA)

Table 3: Comparison of cross-domain 3D super-resolution performance on the HCP (MRI). Models are trained on the MSD dataset and evaluated on the HCP dataset.

Paradigms	Methods	2×			3×			4×		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Traditional	Trilinear	36.45	0.9425	0.1048	33.75	0.9221	0.1634	30.76	0.9026	0.2675
	Cubic	36.47	0.9588	0.0885	34.01	0.9289	0.1454	31.86	0.9125	0.2483
Self-supervised	CuNeRF [5]	36.91	0.9716	0.0724	34.48	0.9547	0.1362	33.12	0.9382	0.1587
Neural Implicit	ArSSR [25]	36.07	0.9655	0.0615	33.56	0.9429	0.0892	31.87	0.9220	0.1141
FF-3DGS	MedGSSR	40.81	0.9849	0.0674	37.94	0.9702	0.0945	35.84	0.9533	0.1253

Table 4: Comparison of cross-domain 3D super-resolution performance on the UHRCT (CT). Models are trained on the MELA dataset and evaluated on the UHRCT dataset.

Paradigms	Methods	4×			6×			8×		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Traditional	Trilinear	31.96	0.9404	0.1051	26.73	0.9027	0.1706	23.96	0.8436	0.2605
	Cubic	32.88	0.9407	0.0898	28.51	0.9189	0.1654	26.07	0.8579	0.2517
Self-supervised	CuNeRF [5]	38.03	0.9628	0.1252	33.76	0.9164	0.2465	30.50	0.8690	0.3510
Neural Implicit	ArSSR [25]	35.92	0.9471	0.0918	32.88	0.9332	0.1191	29.38	0.8994	0.2145
FF-3DGS	MedGSSR	41.20	0.9818	0.0468	37.74	0.9698	0.0765	35.24	0.9233	0.1253

directly on held-out target datasets (HCP for MRI and UHRCT for CT), without any fine-tuning. Tables 3 and 4 summarize these generalization results. As observed, the method ArSSR experiences notable performance degradation when encountering unseen datasets. Conversely, MedGSSR maintains exceptional reconstruction fidelity, securing a PSNR improvement of 3.97 dB over ArSSR at the 4× scale on the HCP dataset. More impressively, on the UHRCT dataset (Table 4), MedGSSR is evaluated at extreme upsampling scales (4×, 6×, 8×). Even at an 8× resolution gap, our method yields a PSNR surge (+5.86 dB) over ArSSR and a gain (+4.74 dB) over CuNeRF. As visualized in Fig. 4, baseline methods tend to produce blurred or noisy reconstructions under domain shift. In contrast, MedGSSR better preserves structurally consistent cortical folds and tissue boundaries. These results support that our model learns transferable priors from the source data, enabling robust feed-forward recovery of high-frequency details on unseen domains. Additional visualizations are provided in the **Supplementary Material**.

Clinical Validation. To verify that the visual improvements of MedGSSR translate into actual clinical utility, we perform an automated brain tissue segmentation task using a pre-trained nnUNet model. The segmentation is conducted on the 4× upsampled MRI volumes from the HCP dataset. Table 5 reports the Dice Similarity Coefficients across five brain regions (CSF, GM, WM, dGM, BS+CB). MedGSSR achieves the highest average Dice score (0.8178), outperforming CuNeRF (which completely collapses to 0.3618 due to severe artifacts) and ArSSR (+3.4% relative increase). As visualized in Fig. 5, the

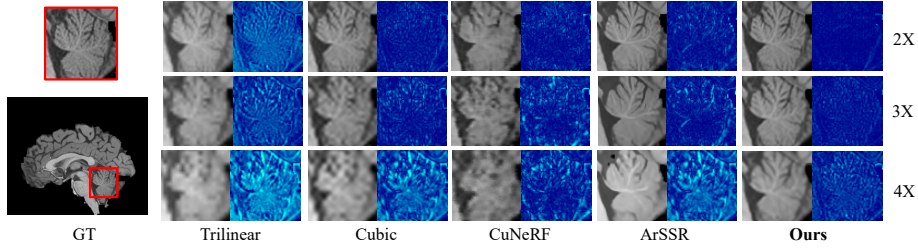


Fig. 4: Qualitative comparison of cross-domain 3DSR on the HCP (MRI). All models are trained on the MSD and evaluated on the unseen HCP. The error maps (dark blue indicates lower error) highlight that MedGSSR produces the sharpest anatomical details with minimal reconstruction deviation compared to the baselines.

Table 5: Comparison of Dice scores. The segmentation is performed using a pre-trained nnUNet on $4\times$ super-resolved MRI volumes (HCP). Best results are highlighted in bold.

Methods	CSF	GM	WM	dGM	BS+CB	Average
Trilinear	0.4937	0.7265	0.8617	0.8098	0.6701	0.7124
Cubic	0.5410	0.7616	0.8860	0.8616	0.8518	0.7804
CuNeRF [5]	0.3668	0.5330	0.7347	0.1285	0.0458	0.3618
ArSSR [25]	0.6782	0.8318	0.9187	0.9140	0.5612	0.7908
Ours (MedGSSR)	0.5779	0.8406	0.9295	0.9407	0.8006	0.8178

volumes reconstructed by ArSSR exhibit blurry boundaries, leading to under-segmentation and bleeding at tissue margins. Thanks to our hierarchical structure-texture decoupling, MedGSSR restores sharp anatomical edges, yielding segmentation masks that align with the original ground truth. Additional visualizations are provided in the **Supplementary Material**.

5.4 Ablation Study

Unless otherwise specified, all ablation experiments are conducted under the intra-domain setting on the HCP MRI dataset for $4\times$ 3D SR.

Structure-Texture Decoupling. To validate the design of MedGSSR, we study the effect of dual-resolution structure-texture decoupling. Fig. 6 reports a qualitative and quantitative comparison between a coarse-only variant and the full model. The coarse branch captures global anatomy (35.24 dB PSNR, 0.9802 SSIM), but fails to recover fine-scale boundaries, resulting in blurred cortical folds. Adding the fine branch explicitly models high-frequency residuals, improving reconstruction quality to 36.60 dB PSNR and 0.9870 SSIM, while reducing LPIPS from 0.0789 to 0.0363. These results indicate that the decoupled hierarchical prediction is important for preserving fine anatomical details.

Impact of Sub-voxel Count. The hyperparameter m determines the number of continuous Gaussian spheres split from each discrete latent feature location. We examine its effect on reconstruction quality in Table 6. With $m = 1$, the

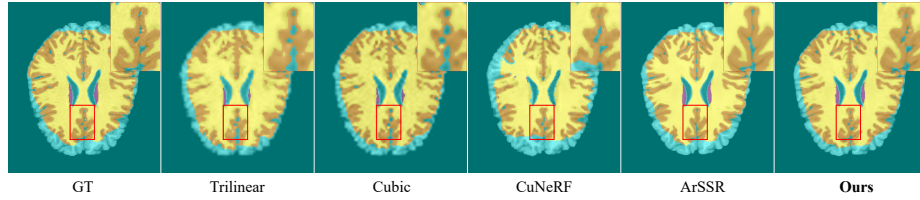


Fig. 5: Downstream brain tissue segmentation on the HCP dataset ($4\times$ SR). The structural fidelity provided by MedGSSR translates to precise automated boundary delineation, avoiding the tissue-bleeding artifacts observed in the others.

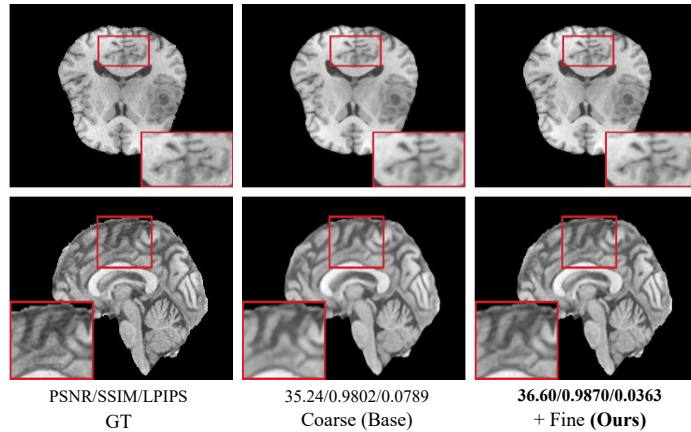


Fig. 6: Ablation on Structure-Texture Decoupling on the HCP dataset ($4\times$ SR). The coarse branch captures global anatomy, and adding the fine branch restores high-frequency textures, improving sharpness and quantitative metrics.

limited geometric capacity is insufficient to model complex anatomical boundaries, resulting in inferior performance (35.98 dB PSNR). Increasing m to 2 and 4 improves both fidelity and perceptual quality, reaching the best results at $m = 4$ (36.60 dB PSNR). Further increasing to $m = 8$ brings marginal saturation and slightly worse metrics (36.52 dB PSNR). We therefore set $m = 4$ by default to balance reconstruction quality and computational cost.

Effect of Gaussian Support, Frequency Supervision, and Data Scale.

We further investigate the effects of the Gaussian support radius, an auxiliary frequency-domain objective, and training data scale on the HCP dataset under the cross-domain setting for $4\times$ 3D SR. As shown in Table 7, adding the FFT loss improves LPIPS from 0.1253 to 0.1078, suggesting better perceptual high-frequency recovery, but slightly decreases PSNR and SSIM. This indicates a trade-off between perceptual sharpness and voxel-wise fidelity. In contrast, reducing the truncation radius to 1σ causes a large degradation across all metrics, confirming that sufficient Gaussian support is necessary for aggregating neighboring primitives and maintaining a continuous volumetric field. We also examine

Table 6: Ablation on sub-voxel count m under intra-domain setting on the HCP dataset ($4\times$ SR). Setting $m = 4$ delivers the best balance between accuracy and computational cost. Best results are highlighted in bold.

Settings	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
$m = 1$	35.98	0.9760	0.0423
$m = 2$	36.06	0.9771	0.0388
$m = 4$ (Default)	36.60	0.9870	0.0363
$m = 8$	36.52	0.9862	0.0365

Table 7: Ablation study on effects of frequency-domain supervision, Gaussian truncation radius, and training data under the cross-domain setting on the HCP dataset ($4\times$ SR). Best results are highlighted in bold.

Variants	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
Ours	35.84	0.9533	0.1253
Ours + FFT Loss	35.72	0.9472	0.1078
Ours w/ 1σ truncation	31.05	0.9142	0.1608
Ours w/ 50% training data	34.26	0.9405	0.1452
Ours w/ 75% training data	35.49	0.9512	0.1283

the robustness of MedGSSR under reduced training data. The results show that MedGSSR remains robust when using 75% of the training data. MedGSSR remains competitive with 75% training data, achieving 35.49 dB PSNR and 0.9512 SSIM, while performance drops more noticeably with 50% data. Additional ablations are provided in the **Supplementary Material**.

6 Conclusion

In this paper, we present MedGSSR, a feed-forward framework that formulates medical 3D super-resolution as a generalized volumetric reconstruction problem. By leveraging explicit 3D Gaussian Splatting as continuous basis functions, MedGSSR removes the need for isolated per-volume optimization in self-supervised pipelines and alleviates over-smoothing effects commonly observed in implicit neural representations. To address anatomical heterogeneity, we introduce a Hierarchical Gaussian Projector with sub-voxel decomposition, which decouples global structural preservation from fine-scale detail refinement. Moreover, the proposed Differentiable Gaussian Voxelizer enables end-to-end supervision directly in the 3D domain, avoiding reliance on 2D proxy projections and reducing geometric ambiguity. Extensive experiments across MRI and CT benchmarks demonstrate strong medical super-resolution with high fidelity, and robust cross-dataset generalization under domain shift. Finally, the improved structural accuracy of MedGSSR consistently benefits downstream clinical tasks, highlighting its potential as an efficient and reliable foundation for volumetric medical image enhancement.

References

1. Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P.: Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5855–5864 (2021)
2. Charatan, D., Li, S.L., Tagliasacchi, A., Sitzmann, V.: pixelsplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 19457–19467 (2024)
3. Chen, A., Xu, Z., Geiger, A., Yu, J., Su, H.: Tensorf: Tensorial radiance fields. In: European conference on computer vision. pp. 333–350. Springer (2022)
4. Chen, Y., Shi, F., Christodoulou, A.G., Xie, Y., Zhou, Z., Li, D.: Efficient and accurate mri super-resolution using a generative adversarial network and 3d multi-level densely connected network. In: International conference on medical image computing and computer-assisted intervention. pp. 91–99. Springer (2018)
5. Chen, Z., Yang, L., Lai, J.H., Xie, X.: Cunerf: Cube-based neural radiance field for zero-shot medical image arbitrary-scale super resolution. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 21185–21195 (2023)
6. Corona-Figueroa, A., Frawley, J., Bond-Taylor, S., Bethapudi, S., Shum, H.P., Willcocks, C.G.: Mednerf: Medical neural radiance fields for reconstructing 3d-aware ct-projections from a single x-ray. In: 2022 44th annual international conference of the IEEE engineering in medicine & Biology society (EMBC). pp. 3843–3848. IEEE (2022)
7. Du, J., He, Z., Wang, L., Gholipour, A., Zhou, Z., Chen, D., Jia, Y.: Super-resolution reconstruction of single anisotropic 3d mr images using residual convolutional neural network. *Neurocomputing* **392**, 209–220 (2020)
8. Fang, W., Tang, Y., Guo, H., Yuan, M., Mok, T.C., Yan, K., Yao, J., Chen, X., Liu, Z., Lu, L., et al.: Cycleinr: cycle implicit neural representation for arbitrary-scale volumetric super-resolution of medical data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11631–11641 (2024)
9. Feng, X., He, Y., Chen, L., Yang, Y., Wang, C., Chen, Y., Zhong, Y., Kuang, Z., Yin, X., Zhu, Y., et al.: Srgs: Super-resolution 3d gaussian splatting. *arXiv preprint arXiv:2404.10318* (2024)
10. Feng, X., He, Y., Wang, Y., Wang, C., Kuang, Z., Ding, J., Qin, F., Yu, J., Fan, J.: Zs-srt: An efficient zero-shot super-resolution training method for neural radiance fields. *Neurocomputing* **590**, 127714 (2024)
11. Feng, X., Zhong, T., Chang, S., Wang, W., Wang, C., Chen, Y., Hu, T., Wang, Y., Kuang, Z., Yin, X., et al.: Ie-srgs: An internal-external knowledge fusion framework for high-fidelity 3d gaussian splatting super-resolution. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 3957–3965 (2026)
12. Forigua, C., Escobar, M., Arbelaez, P.: Superformer: Volumetric transformer architectures for mri super-resolution. In: International workshop on simulation and synthesis in medical imaging. pp. 132–141. Springer (2022)
13. Høeg, A.L., Bardenfleth, S.W., Kjer, H.M., Dyrby, T.B., Dahl, V.A., Dahl, A.: Mtvnet: Mapping using transformers for volumes–network for super-resolution with long-range interactions. *arXiv preprint arXiv:2412.03379* (2024)
14. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., et al.: 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* **42**(4), 139–1 (2023)

15. Krug, R., Burghardt, A.J., Majumdar, S., Link, T.M.: High-resolution imaging techniques for the assessment of osteoporosis. *Radiologic Clinics of North America* **48**(3), 601 (2010)
16. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* **65**(1), 99–106 (2021)
17. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)* **41**(4), 1–15 (2022)
18. Noh, J., Oh, H.J., Chae, B., Jeong, W.K.: Zero-shot volumetric ct super-resolution using 3d gaussian splatting with upsampled 2d x-ray projection priors. *arXiv preprint arXiv:2508.15151* (2025)
19. Pham, C.H., Ducournau, A., Fablet, R., Rousseau, F.: Brain mri super-resolution using deep 3d convolutional networks. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). pp. 197–200 (2017). <https://doi.org/10.1109/ISBI.2017.7950500>
20. Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y., Courville, A.: On the spectral bias of neural networks. In: International conference on machine learning. pp. 5301–5310. PMLR (2019)
21. Shocher, A., Cohen, N., Irani, M.: “zero-shot” super-resolution using deep internal learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3118–3126 (2018)
22. Szymanowicz, S., Rupperecht, C., Vedaldi, A.: Splatter image: Ultra-fast single-view 3d reconstruction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10208–10217 (2024)
23. Wang, C., Wu, X., Guo, Y.C., Zhang, S.H., Tai, Y.W., Hu, S.M.: Nerf-sr: High quality neural radiance fields using supersampling. In: Proceedings of the 30th ACM International Conference on Multimedia. pp. 6445–6454 (2022)
24. Wang, C., Zhang, Y., Wu, C., Liu, J., Huang, X., Wu, L., Wang, Y., Feng, X., Lu, Y., Wang, Y.: Mmdental-a multimodal dataset of tooth cbct images with expert medical records. *Scientific Data* **12**(1), 1172 (2025)
25. Wu, Q., Li, Y., Sun, Y., Zhou, Y., Wei, H., Yu, J., Zhang, Y.: An arbitrary scale super-resolution approach for 3d mr images via implicit neural representation. *IEEE Journal of Biomedical and Health Informatics* **27**(2), 1004–1015 (2022)
26. Xie, S., Wang, Z., Zhu, Y., Pan, C.: Supergs: Super-resolution 3d gaussian splatting via latent feature field and gradient-guided splitting. *arXiv preprint arXiv:2410.02571* **1** (2024)
27. Xu, H., Peng, S., Wang, F., Blum, H., Barath, D., Geiger, A., Pollefeys, M.: Depth-splat: Connecting gaussian splatting and depth. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 16453–16463 (2025)
28. Ye, B., Liu, S., Xu, H., Li, X., Pollefeys, M., Yang, M.H., Peng, S.: No pose, no problem: Surprisingly simple 3d gaussian splats from sparse unposed images. *arXiv preprint arXiv:2410.24207* (2024)
29. Yoon, Y., Yoon, K.J.: Cross-guided optimization of radiance fields with multi-view image super-resolution for high-resolution novel view synthesis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12428–12438 (2023)
30. Yu, Z., Chen, A., Huang, B., Sattler, T., Geiger, A.: Mip-splatting: Alias-free 3d gaussian splatting. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 19447–19456 (2024)