RETHINKING REAL-WORLD THICK-SLICE CT SUPER-RESOLUTION VIA THICKNESS-AWARE DEGRADATION MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

Thick-slice CT super-resolution (SR) aims to recover high-resolution (HR) thin-slice volumes from low-resolution (LR) thick-slice scans. Existing methods primarily strengthen reconstruction priors, but commonly use geometric resampling (*e. g.*, nearest-neighbor or linear interpolation) to model the HR-to-LR degradation. As point-wise estimators, these geometric operators determine degradation weights only from slice-center locations while ignoring physical slice thickness, thereby failing to reproduce thickness-dependent partial-volume effects and limiting the reliability of SR models in real-world settings. We formulate a slice-thickness model (STM), a slice sensitivity profile (SSP)-parameterized linear operator that explicitly separates effective slice thickness from slice spacing. The full width at half maximum (FWHM) of the SSP specifies the effective thickness, whereas the sampling locations determine the LR slice centers. STM supports both analytical and data-fitted SSPs and can be used consistently for end-to-end supervised and unsupervised iterative SR. Experiments across real-paired CT datasets and multiple SR paradigms show that STM improves degradation fidelity, SR quality, and downstream anatomical segmentation. It further converts HR-only CT volumes into scalable synthetic supervision that complements limited real pairs. These results establish physically grounded degradation modeling as a key component of real-world thick-slice CT SR.

1 INTRODUCTION

Computed tomography (CT) is a central imaging modality for clinical diagnosis, treatment planning, and quantitative image analysis (Rubin, 2014; McCollough & Rajiah, 2023). Many routine CT volumes are anisotropic, with high in-plane resolution but substantially lower through-plane resolution. The resulting partial-volume effects and anatomical discontinuities in sagittal and coronal views (Abdulkareem et al., 2023; Yu et al., 2024) can obscure small structures and hinder quantitative analysis. Thick-slice CT super-resolution (SR) therefore aims to recover a high-resolution (HR) thin-slice volume from a low-resolution (LR) thick-slice observation, improving 3D visualization and supporting downstream tasks such as anatomical segmentation.

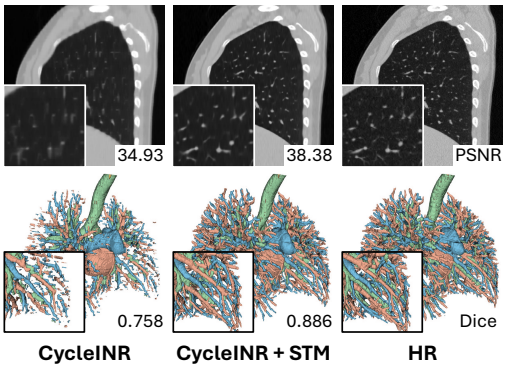


Figure 1: CycleINR (Fang et al., 2024), a leading supervised SR method, with and without STM for 5× through-plane SR on the *real-world* RPLHR-CT dataset (Yu et al., 2022). STM improves anatomical detail and segmentation.

Existing thick-slice CT SR methods primarily address this ill-posed inverse problem by strengthening the reconstruction prior (Wang et al., 2021b; Xiao et al., 2025). Classical optimization methods employ handcrafted regularizers, such as total variation (TV) (Rudin et al., 1992), to constrain the solution space. End-to-end supervised methods learn data-driven priors from paired LR-HR

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

volumes (Park et al., 2018; Zamir et al., 2022; Yu et al., 2022), while diffusion models introduce expressive generative priors for recovering plausible high-frequency anatomy (Moser et al., 2025; Chan & Rajapakse, 2023). However, prior learning alone cannot ensure reliable SR reconstructions. Equally important is degradation modeling, which characterizes the observation process underlying thick-slice CT.

The degradation model defines HR-to-LR formation, providing training pairs for supervised SR and data consistency for unsupervised iterative SR. Natural-image SR typically models blurring, spatial downsampling, and noise (Zhang et al., 2021; Liu et al., 2022), whereas thick-slice CT usually preserves in-plane resolution and mainly reduces through-plane sampling. This motivates geometric resampling, such as nearest-neighbor (NN) or cubic interpolation, which uses slice-center locations and treats LR slices as point samples (Peng et al., 2020; Fang et al., 2024; Song et al., 2024). However, a clinical thick slice integrates neighboring anatomy according to a slice sensitivity profile (SSP), whose full width at half maximum (FWHM) characterizes the effective slice thickness (Stierstorfer et al., 2004; Greene & Rong, 2014). Geometric resampling models spacing but ignores thickness, so it cannot distinguish acquisitions with identical spacing but different thicknesses or reproduce their distinct partial-volume effects. This mismatch limits real-world SR reliability.

In this work, we systematically investigate the impact of degradation modeling on thick-slice CT SR and formulate a slice-thickness model (STM) grounded in clinical CT formation. STM generates each LR thick slice through a normalized, SSP-weighted aggregation of neighboring HR thin slices. The SSP determines the distance-dependent contribution of each thin slice, its FWHM specifies the effective slice thickness, and the slice spacing determines the LR slice centers. By decoupling slice thickness from slice spacing, STM accommodates contiguous, gapped, and overlapping slice configurations while modeling thickness-dependent partial-volume effects. Experiments on three real-world CT datasets show that STM improves degradation fidelity, supervised and iterative SR, and downstream anatomical segmentation. STM also enables scalable supervision by synthesizing training pairs from HR-only CT volumes.

Our main contributions are summarized as follows:

- We identify a fundamental limitation of geometric resampling for thick-slice CT degradation: it encodes slice spacing but ignores slice thickness.
- We propose STM, a physics-grounded degradation model that explicitly decouples slice thickness from slice spacing through normalized SSP-weighted aggregation.
- Extensive experiments show that STM improves thick-slice synthesis, SR reconstruction, and downstream segmentation, while enabling scalable synthetic supervision from unpaired thin-slice CT volumes.

2 Related Work

Thick-Slice CT SR Thick-slice CT SR has progressed from handcrafted regularization to learned anatomical priors. Classical methods employ total variation (TV) (Rudin et al., 1992) or sparse representations (Jiang et al., 2018), whereas supervised methods learn direct LR-to-HR mappings from paired volumes (Park et al., 2018; You et al., 2019; Li et al., 2021). Subsequent studies incorporate spatial coordinates and inter-slice context (Peng et al., 2020; Song et al., 2024), while RPLHR-CT enables training and evaluation with real-paired thick- and thin-slice scans (Yu et al., 2022; 2024). Continuous implicit representations further support arbitrary-scale and zero-shot reconstruction (Wu et al., 2022; Chen et al., 2023; Fang et al., 2024; Li et al., 2023). To reduce reliance on paired supervision, self-supervised methods exploit in-plane details, cross-view consistency, or internal slice pairs (Xie et al., 2021; Fang et al., 2022; Peng et al., 2021), while generative methods introduce diffusion priors or known-operator restoration (Chan & Rajapakse, 2023; Wang et al., 2023). However, most of these methods still rely on geometric resampling to define the degradation process.

Degradation Modeling Natural-image SR commonly models degradation as blur, resampling, and noise. Blind SR studies have moved beyond fixed bicubic downsampling toward diverse kernels and practical synthesis pipelines (Cai et al., 2019; Zhang et al., 2021; Wang et al., 2021a; Liu et al., 2022), highlighting the importance of matching synthetic and real degradations. However, these pipelines do not characterize the physical formation of thick-slice CT.

In medical SR, LR volumes are commonly synthesized by through-plane NN or cubic resampling (Peng et al., 2020; Chen et al., 2023; Fang et al., 2024; Song et al., 2024; Wu et al., 2022), neighboring-slice averaging (Park et al., 2018; Xie et al., 2021), or their combination (Shi et al., 2023). These schemes are generally determined by sampling grids or fixed aggregation rules without explicitly parameterizing physical slice thickness. Tang et al. (2025) determine LR slice centers from physical coordinates and synthesize each thick slice using a normalized triangular weighting controlled by slice thickness, explicitly distinguishing thickness from slice increment. Building on this physical perspective, we develop a general SSP-based framework that decouples slice thickness from spacing and unifies analytical and data-fitted SSPs for supervised and iterative CT SR. We systematically analyze how SSP shape, slice thickness, and cross-paradigm integration affect degradation fidelity, SR quality, and downstream segmentation.

3 METHODOLOGY

3.1 PROBLEM STATEMENT

Let $\boldsymbol{I}_{\mathsf{HR}} \in \mathbb{R}^{H\times W\times D}$ denote a thin-slice HR CT volume, its thick-slice LR observation $\boldsymbol{I}_{\mathsf{LR}} \in \mathbb{R}^{H\times W\times D'}$ ($D' < D$) can be modeled as

$$\boldsymbol{I}_{\mathsf{LR}} = \boldsymbol{A}\boldsymbol{I}_{\mathsf{HR}} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{A} : \mathbb{R}^{H\times W\times D} \to \mathbb{R}^{H\times W\times D'}$ denotes the degradation model for formulating the physical acquisition of the LR volume $\boldsymbol{I}_{\mathsf{LR}}$ and $\boldsymbol{\epsilon}$ is the system noise.

The goal of thick-slice CT SR is to recover the unknown HR thin-slice volume $\boldsymbol{I}_{\mathsf{HR}}$ from an LR thick-slice scan $\boldsymbol{I}_{\mathsf{LR}}$. Due to information loss during data acquisition, this inverse problem is inherently ill-posed, with multiple suboptimal solutions. A typical optimization objective is:

$$\boldsymbol{I}^*_{\mathsf{HR}} = \arg\min_{\boldsymbol{I}_{\mathsf{HR}}} \underbrace{\frac{1}{2}\|\boldsymbol{A}\boldsymbol{I}_{\mathsf{HR}} - \boldsymbol{I}_{\mathsf{LR}}\|_2^2}_{\text{Data-fidelity term}} + \lambda \cdot \underbrace{\mathcal{R}(\boldsymbol{I}_{\mathsf{HR}})}_{\text{Prior term}}, \tag{2}$$

where the data-fidelity term ensures consistency with the LR observation $\boldsymbol{I}_{\mathsf{LR}}$, while the prior term enforces regularization for the desired HR volume $\boldsymbol{I}^*_{\mathsf{HR}}$. Both terms synergistically improve the SR reconstructions. Modern SR techniques (Zamir et al., 2022) mostly focus on the design of the prior term, such as data-driven priors learned by supervised networks and generative priors by diffusion models. However, the degradation model $\boldsymbol{A}$, which defines the LR-HR relationship, remains insufficiently investigated. This severely limits the reliability of SR images in anatomical structures.

To this end, this work studies a key question: ***What is the physics-grounded form of the degradation operator for the thick-slice CT SR problem?*** Technically, we study this question from a CT slice-formation perspective. Rather than treating $\boldsymbol{A}$ as a geometric resampling operator, we seek to construct a degradation operator that reflects the formation process of thick-slice CT. This may unify realistic LR synthesis for supervised learning and data-consistent iterative SR.

3.2 DEGRADATION MODELING FOR REAL-WORLD THICK-SLICE CT: FROM GEOMETRIC RESAMPLING TO SLICE-THICKNESS MODEL

Let $\{z_j\}_{j=1}^{D}$ and $\{\widetilde{z}_k\}_{k=1}^{D'}$ denote the center positions of HR thin slices and LR thick slices, respectively. Omitting noise, Eq. (1) becomes

$$\boldsymbol{I}_{\mathsf{LR}}(x, y, \widetilde{z}_k) = \sum_{j=1}^{D} \alpha_j^k \cdot \boldsymbol{I}_{\mathsf{HR}}(x, y, z_j), \tag{3}$$

where α_j^k denotes the contribution of the j-th HR slice to the k-th LR slice. To model real-world thick-slice CT, the degradation operator should satisfy three basic properties:

Property 1 (*Low-Pass Filtering*)**.** Through-plane degradation should act as a low-pass filter.

Property 2 (*HU Preservation*)**.** The weights should form a convex combination (*i. e.*, $\alpha_j^k \geq 0$ and $\sum_{j=1}^{D} \alpha_j^k = 1, \forall k$), so that homogeneous HU values are preserved.

Property 3 (*Spacing-Thickness Modeling*)**.** The weights should be jointly parameterized by slice spacing Δz and slice thickness T (*i. e.*, $\alpha_j^k : \mathbb{R}_+^2 \to [0, 1]$).

Physical Intuition Behind These Properties These properties capture the basic physical properties of thick-slice CT formation. *First*, a thick-slice scan is not obtained from a single anatomical plane, but from the aggregation of neighboring tissues within a finite through-plane range, which naturally induces low-pass filtering along the Z-axis. *Second*, this aggregation should be formulated as a normalized weighted average rather than an intensity accumulation, such that homogeneous HU values remain unchanged after thick-slice formation. *Third*, as shown in Fig. 2, slice thickness T and slice spacing Δz are two independent acquisition parameters. Their relationship gives rise to contiguous ($T = \Delta z$), gapped ($T < \Delta z$), and overlapping ($T > \Delta z$) acquisition protocols. A physics-grounded degradation model should therefore decouple thickness and spacing.

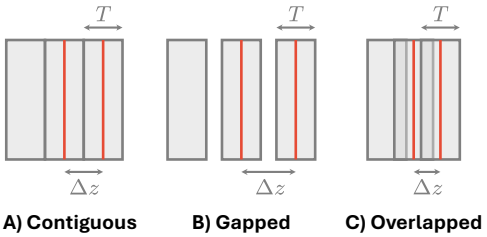


Figure 2: LR thick-slice formation under **A)** contiguous ($T = \Delta z$), **B)** gapped ($T < \Delta z$), and **C)** overlapping ($T > \Delta z$) acquisition. T and Δz denote slice thickness and spacing. Red lines indicate LR slice centers.

Geometric Resampling As illustrated in Fig. 3A, geometric resampling represents each LR slice by an image value estimated at its target center. Let Δz denote the LR slice spacing, such that $\widetilde{z}_k = \widetilde{z}_1 + (k-1)\Delta z$, and define the absolute offset $\Delta z_{j,k} = |z_j - \widetilde{z}_k|$. A generic interpolation kernel $\phi(\cdot)$ converts these offsets into normalized weights:

$$\alpha^k_{j,\text{geo}} = \frac{\phi(\Delta z_{j,k})}{\sum_{i=1}^{D} \phi(\Delta z_{i,k})}, \tag{4}$$

where a particular choice of ϕ specifies the resampling rule. NN (Peng et al., 2020; Fang et al., 2024; Chen et al., 2023; Song et al., 2024) sets $\alpha^k_{j,\text{geo}} = 1$ for the closest HR slice and zero for all others. Linear interpolation assigns nonzero weights to the two neighboring HR slices, with each weight decreasing linearly with its distance from $\widetilde{z}_k$.

This point-estimation formulation cannot jointly satisfy the three properties above. For Property 1, NN uses a one-hot kernel that copies a single HR slice. Its delta-like response does not attenuate high through-plane frequencies before subsampling, unlike the finite-width response of a clinical thick slice. For Property 2, the normalization in Eq. (4) enforces unit-sum weights. However, commonly used cubic interpolation kernels may assign negative weights, so the coefficients are not guaranteed to lie in $[0, 1]$ or form a convex combination. Most importantly, Property 3 is inherently violated because the weights depend only on slice-center offsets $\Delta z_{j,k}$ and do not include T. Consequently, scans with identical slice spacing but different slice thicknesses receive the same degradation, although their partial-volume effects should differ. Overall, geometric resampling can change the through-plane sampling grid or voxel spacing, but it does not satisfy all three properties above and therefore cannot fully characterize the physical formation of thick-slice CT.

Proposed Slice-Thickness Model To formulate real-world thick-slice CT, our STM makes two key contributions: 1) It introduces a thickness-parameterized SSP $h_T(\cdot)$ to model the finite through-plane response; 2) It calibrates the response width of h_T by setting its FWHM to the effective slice thickness T. Let $h_T(\tau) \geq 0$ denote a centered and symmetric SSP whose peak occurs at $\tau = 0$. We calibrate its response width to the effective slice thickness T by requiring

$$\left\{ \tau \in \mathbb{R} \;\middle|\; h_T(\tau) \geq \frac{h_T(0)}{2} \right\} = \left[-\frac{T}{2}, \frac{T}{2} \right]. \tag{5}$$

That is, $h_T(\Delta z_{j,k})$ remains above half of its peak value over an interval of width T, such that $\text{FWHM}(h_T) = T$. This definition calibrates different SSP shapes to the same effective slice thickness, allowing the through-plane response and the thickness-dependent partial-volume effects to be modeled consistently, rather than being determined solely by spacing-dependent kernel $\phi(\Delta z_{j,k})$ used in geometric resampling, as demonstrated in Fig. 3B.

For the k-th LR slice, STM weights each HR slice using its absolute offset $\Delta z_{j,k}$:

$$\alpha^k_{j,\text{STM}} = \frac{h_T(\Delta z_{j,k})}{\sum_{i=1}^{D} h_T(\Delta z_{i,k})}. \tag{6}$$

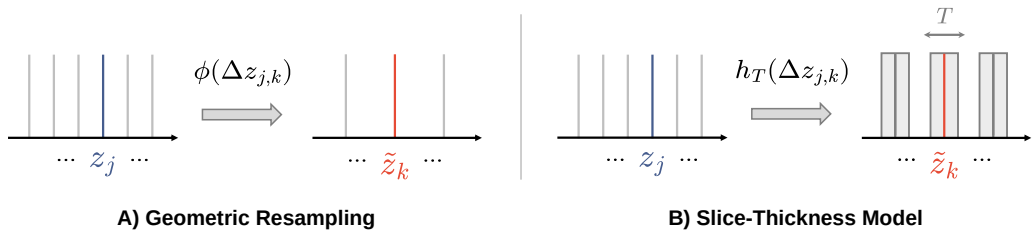


Figure 3: Geometric resampling versus our STM along the through-plane axis. **A)** Geometric resampling performs point-wise interpolation with kernel $\phi(\Delta z_{j,k})$ to estimate a point sample at $\widetilde{z}_k$, where $\Delta z_{j,k} = |z_j - \widetilde{z}_k|$. **B)** STM performs finite-thickness aggregation of neighboring HR slices using the SSP $h_T(\Delta z_{j,k})$, whose FWHM T specifies slice thickness.

As shown above, we explicitly decouple slice thickness T from slice spacing Δz, allowing STM to represent contiguous, gapped, and overlapping acquisition protocols. The normalization ensures that the weights form a convex combination, preserving homogeneous HU values. The finite-width response of h_T performs through-plane low-pass filtering, while its FWHM calibrates the response width to the effective slice thickness T. Overall, STM satisfies all three properties defined above and provides a physics-grounded degradation model for thick-slice CT.

SSP Implementation STM supports analytical SSPs with prescribed FWHM and data-fitted SSPs with learned width. For the analytical implementation, we use a triangular profile motivated by longitudinal weighting in helical CT reconstruction (Stierstorfer et al., 2004):

$$h_T(\Delta z_{j,k}) = \max\left(1 - \frac{|\Delta z_{j,k}|}{T}, 0\right). \tag{7}$$

This profile has support $[-T, T]$ and an FWHM of T. It is only one convenient implementation. Rectangular, Gaussian, or other profiles can be used without changing STM, provided that their FWHM is calibrated to the target slice thickness.

When paired LR-HR volumes are available, the continuous SSP h_T can instead be instantiated by a data-fitted discrete profile $\boldsymbol{h}$. Let $\boldsymbol{A}_{\boldsymbol{h},\Delta z}$ denote the corresponding STM operator induced by Eq. (6). We estimate $\boldsymbol{h}$ from N pairs by

$$\boldsymbol{h}^* = \arg\min_{\boldsymbol{h}\geq\boldsymbol{0},\, \sum_r h_r = 1,\; h_r = h_{-r}} \frac{1}{N}\sum_{n=1}^{N} \left\| \boldsymbol{A}_{\boldsymbol{h},\Delta z}\boldsymbol{I}_{\mathsf{HR}}^{(n)} - \boldsymbol{I}_{\mathsf{LR}}^{(n)} \right\|_2^2 + \beta \left\| \boldsymbol{D}_2 \boldsymbol{h} \right\|_2^2. \tag{8}$$

Here, $\boldsymbol{D}_2$ denotes the second-order difference operator and β controls SSP smoothness. The support size and β are selected on the validation set. Unlike the analytical profiles, $\mathrm{FWHM}(\boldsymbol{h}^*)$ is learned rather than fixed to T. The constraints enforce nonnegative, symmetric, normalized aggregation; fitting captures protocol-specific nonidealities.

3.3 Integrating STM into Supervised and Iterative SR

Given our STM operator $\boldsymbol{A}_{h_T,\Delta z}$, we next describe how to integrate it into two representative SR paradigms: end-to-end supervised learning and unsupervised iterative reconstruction. In both cases, STM provides a physics-grounded degradation model that changes only the LR-HR relationship while leaving the reconstruction prior unchanged.

End-to-End Supervised Learning. Let $\{\boldsymbol{I}_{\mathsf{HR}}^{(n)}\}_{n=1}^{N}$ be a collection of HR thin-slice volumes and $\mathcal{F}_{\boldsymbol{\theta}}$ an SR network with parameters $\boldsymbol{\theta}$. Applying STM produces the corresponding LR inputs as $\widetilde{\boldsymbol{I}}_{\mathsf{LR}}^{n} = \boldsymbol{A}_{h_T,\Delta z}\boldsymbol{I}_{\mathsf{HR}}^{n}$. Together with their HR targets, these observations form the training set $\{(\widetilde{\boldsymbol{I}}_{\mathsf{LR}}^{(n)}, \boldsymbol{I}_{\mathsf{HR}}^{(n)})\}_{n=1}^{N}$. The network is then optimized by

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{1}{N}\sum_{n=1}^{N} \mathcal{L}_{\mathrm{SR}}\left(\mathcal{F}_{\boldsymbol{\theta}}(\widetilde{\boldsymbol{I}}_{\mathsf{LR}}^{(n)}), \boldsymbol{I}_{\mathsf{HR}}^{(n)}\right), \tag{9}$$

where $\mathcal{L}_{\mathrm{SR}}$ is the task-specific reconstruction loss. At inference, $\widehat{\boldsymbol{I}}_{\mathsf{HR}} = \mathcal{F}_{\boldsymbol{\theta}^*}(\boldsymbol{I}_{\mathsf{LR}})$. Thus, STM changes supervision construction without altering the SR architecture or training objective.

Unsupervised Iterative Reconstruction. Unsupervised iterative reconstruction estimates an HR volume from each observed LR scan without paired LR-HR supervision. For traditional optimization-based methods such as ADMM-TV, STM is embedded in the data-fidelity term:

$$\boldsymbol{I}^*_{\mathsf{HR}} = \arg\min_{\boldsymbol{X}} \frac{1}{2} \left\| \boldsymbol{A}_{h_T,\Delta z}\boldsymbol{X} - \boldsymbol{I}_{\mathsf{LR}} \right\|_2^2 + \lambda \mathcal{R}(\boldsymbol{X}), \tag{10}$$

where $\mathcal{R}$ denotes a handcrafted prior and λ controls its strength. Generative iterative methods such as DDNM (Wang et al., 2023) instead use a diffusion prior to produce the current HR estimate $\boldsymbol{X}_t$. STM projects this estimate onto the measurement-consistent solution space through

$$\boldsymbol{X}_t^{\mathrm{dc}} = \boldsymbol{A}^{\dagger}_{h_T,\Delta z}\boldsymbol{I}_{\mathsf{LR}} + \left(\mathcal{I} - \boldsymbol{A}^{\dagger}_{h_T,\Delta z}\boldsymbol{A}_{h_T,\Delta z}\right)\boldsymbol{X}_t, \tag{11}$$

where $\boldsymbol{A}^{\dagger}_{h_T,\Delta z}$ and $\mathcal{I}$ are the Moore–Penrose pseudoinverse and identity operator, respectively. In the noise-free setting, $\boldsymbol{A}_{h_T,\Delta z}\boldsymbol{X}_t^{\mathrm{dc}} = \boldsymbol{I}_{\mathsf{LR}}$. The first term recovers the component determined by the LR observation, while the second preserves the null-space component supplied by the generative prior. Therefore, both traditional ADMM-TV and generative DDNM can use STM as a shared physical observation model while retaining their original priors and solvers.

4 Experiments

4.1 Experimental Settings

Datasets. We conduct experiments on three paired LR-HR datasets and one HR-only dataset. For all datasets, CT intensities are clipped to [-1024, 2048] HU and linearly normalized to [0,1]. Dataset-specific settings are described below:

1) RPLHR-CT (Yu et al., 2022) is a public real-paired benchmark for thick-slice CT SR. It contains 250 paired LR-HR cases, including 100 for training, 50 for validation, and 100 for testing. The thin-slice HR volumes have a slice thickness and spacing of 1 mm, while both values are 5 mm for the thick-slice LR volumes. We follow the official preprocessing and dataset splits.

2) TCIA (Clark et al., 2013) is a real-world paired dataset curated to evaluate cross-dataset generalization. After data cleaning and pair selection, we extract 20 paired LR-HR cases from the cancer image archive. The thin-slice HR volumes have a slice thickness and spacing of 1 mm, while both values are 5 mm for the thick-slice LR volumes. Their in-plane resolution is 0.6–0.8 mm.

3) AAPM-LDCT (McCollough et al., 2017) contains 10 paired LR-HR cases. The thin-slice HR volumes have a slice thickness of 1 mm and a spacing of 0.8 mm. The thick-slice LR volumes have a slice thickness of 3 mm and a spacing of 2 mm, resulting in a 1 mm overlap between adjacent slices. We use these cases to evaluate degradation fidelity under an overlapping acquisition setting.

4) CT-RATE (Hamamci et al., 2026) provides large-scale thin-slice CT volumes without paired thick-slice counterparts. We randomly select 2,000 HR volumes with 1 mm through-plane spacing and 0.6–0.8 mm in-plane resolution. We apply our STM to these volumes to synthesize LR-HR training pairs for the data scaling study.

Baselines and Metrics. We compare STM with three commonly used geometric baselines: NN, linear, and cubic interpolation (Song et al., 2024; Peng et al., 2020; Chen et al., 2023; Fang et al., 2024; Shi et al., 2023; Wu et al., 2022). Image quality is measured by PSNR and SSIM. Downstream anatomical utility is evaluated by Dice using the pre-trained TotalSegmentator (Wasserthal et al., 2023) on four fine lung structures. Dice measures segmentation consistency between predictions on SR volumes and their HR references. ***Further implementation details are provided in Appendix A.***

4.2 Experimental Results

Fidelity of Synthesized Thick-Slice CT. To verify degradation fidelity across acquisition protocols, we synthesize LR volumes from paired HR scans using NN, linear interpolation, cubic interpolation,

Table 1: Quantitative comparison of thick-slice LR volumes synthesized by geometric resampling (NN, linear, and cubic) and our STM on the real-world RPLHR-CT (Yu et al., 2022), TCIA (Clark et al., 2013), and AAPM-LDCT (McCollough et al., 2017) datasets.

Model	RPLHR-CT		TCIA		AAPM-LDCT	
	PSNR	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
NN	34.30±0.98	0.924±0.021	35.02±1.23	0.948±0.023	41.55±0.64	0.977±0.004
Linear	35.25±0.95	0.943±0.016	35.83±1.16	0.959±0.017	42.93±0.66	0.983±0.003
Cubic	34.61±0.98	0.930±0.019	35.29±1.21	0.951±0.021	42.38±0.66	0.980±0.003
STM (Ours)	**47.15**±2.40	**0.994**±0.005	**48.43**±3.68	**0.996**±0.005	**57.10**±0.93	**0.999**±0.000

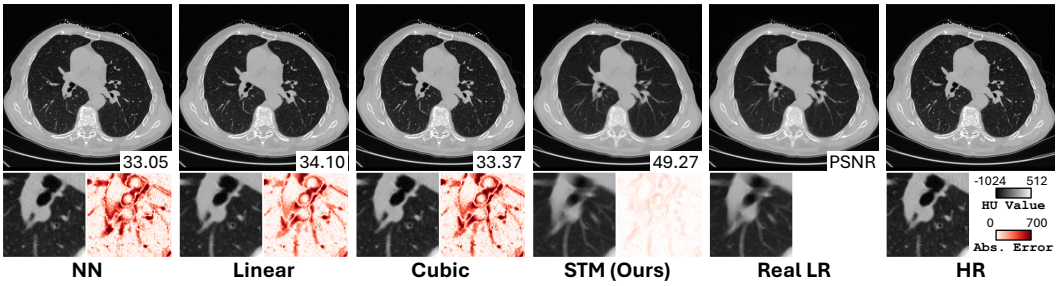


Figure 4: Qualitative comparison of thick-slice LR volumes synthesized by geometric resampling (NN, linear, and cubic) and our STM on a representative case from the real-world RPLHR-CT dataset (Yu et al., 2022).

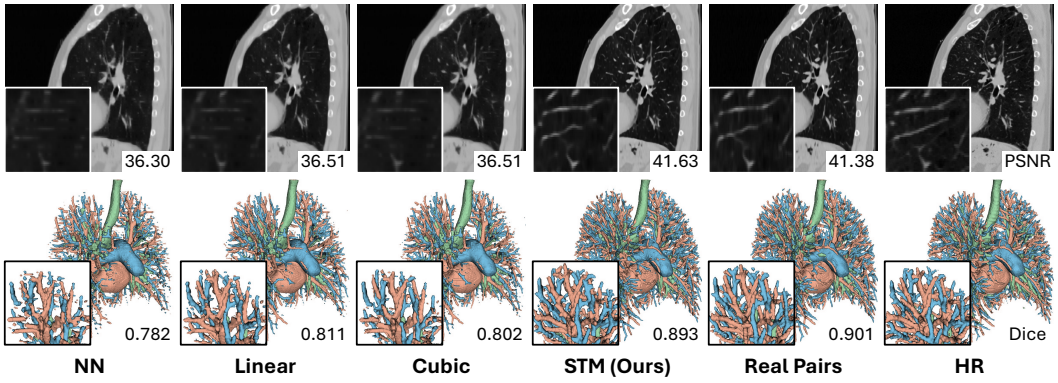


Figure 5: Qualitative comparison of supervised Restormer3D trained with synthetic LR-HR pairs constructed using geometric resampling (NN, linear, and cubic) or our STM, alongside a real-pair reference, on a representative case from the real-world RPLHR-CT dataset (Yu et al., 2022).

and STM. RPLHR-CT and TCIA provide contiguous LR acquisitions with $T = \Delta z = 5$ mm, whereas AAPM-LDCT provides overlapping LR acquisitions with $T = 3$ mm and $\Delta z = 2$ mm. Table 1 shows that STM improves PSNR over the strongest interpolation baseline by 11.90–14.17 dB and achieves the highest SSIM on all three datasets. Fig. 4 further shows smaller local errors and more faithful anatomical boundaries and intensity transitions. The consistent gains under both protocols validate the decoupled modeling of slice thickness and spacing, capturing the partial-volume effects induced by slice thickness.

Impact of Degradation Models across SR Paradigms. For controlled comparisons across SR paradigms, we replace only the degradation operator and keep other settings fixed; Real Pairs is a separate reference. For supervised SR, each operator constructs LR-HR pairs from the RPLHR-CT training and validation volumes for UNet3D and Restormer3D, with real-pair training as a reference. For iterative SR, each operator defines data consistency in ADMM-TV and DDNM. Table 2 shows that STM improves PSNR by 3.66–4.39 dB for the supervised backbones and by 3.90–5.59 dB for

Table 2: Quantitative comparison of degradation models across SR paradigms on the real-world RPLHR-CT (Yu et al., 2022) and TCIA (Clark et al., 2013) datasets. "Real Pairs" denotes supervised models trained on 100 real RPLHR-CT pairs, with 50 pairs reserved for validation.

Paradigm	Method	Model	RPLHR-CT			TCIA		
			PSNR	SSIM	Dice	PSNR	SSIM	Dice
Supervised	UNet3D	NN	33.68±0.93	0.889±0.030	0.788±0.021	34.33±1.19	0.924±0.032	0.805±0.051
		Linear	34.02±0.96	0.894±0.029	0.818±0.018	34.70±1.20	0.928±0.032	0.825±0.051
		Cubic	33.90±0.96	0.891±0.030	0.802±0.020	34.58±1.21	0.926±0.032	0.814±0.052
		STM (Ours)	37.68±1.31	0.927±0.024	**0.889±0.012**	**38.76±1.96**	0.956±0.027	**0.896±0.016**
		Real Pairs	**37.83±1.24**	**0.930±0.022**	0.886±0.013	38.58±1.77	**0.956±0.026**	0.893±0.019
	Restormer3D	NN	33.75±0.95	0.894±0.028	0.784±0.023	34.48±1.23	0.926±0.033	0.797±0.051
		Linear	34.03±0.96	0.898±0.028	0.816±0.019	34.75±1.22	0.929±0.032	0.823±0.051
		Cubic	34.00±0.97	0.896±0.028	0.811±0.019	34.70±1.24	0.928±0.032	0.816±0.051
		STM (Ours)	37.94±1.39	0.932±0.023	0.893±0.012	39.14±2.16	0.958±0.027	0.897±0.018
		Real Pairs	**38.39±1.38**	**0.936±0.021**	**0.899±0.013**	**39.18±2.02**	**0.958±0.026**	**0.903±0.018**
Iterative	ADMM-TV	NN	30.75±0.79	0.878±0.029	0.680±0.029	30.88±1.17	0.907±0.036	0.700±0.055
		Linear	31.38±0.81	0.882±0.029	0.695±0.029	31.62±1.22	0.912±0.036	0.705±0.051
		Cubic	31.38±0.81	0.883±0.029	0.401±0.069	31.58±1.25	0.912±0.035	0.552±0.065
		STM (Ours)	**35.28±1.10**	**0.914±0.025**	**0.786±0.027**	**36.01±1.49**	**0.943±0.029**	**0.807±0.042**
	DDNM	NN	30.65±0.85	0.840±0.027	0.666±0.029	30.85±1.02	0.858±0.044	0.660±0.093
		Linear	30.81±0.87	0.854±0.028	0.691±0.029	31.10±1.09	0.877±0.039	0.682±0.084
		Cubic	31.74±0.91	0.882±0.028	0.749±0.025	32.35±1.36	0.919±0.034	0.743±0.049
		STM (Ours)	**36.44±1.38**	**0.918±0.025**	**0.843±0.021**	**37.94±1.91**	**0.952±0.028**	**0.848±0.046**

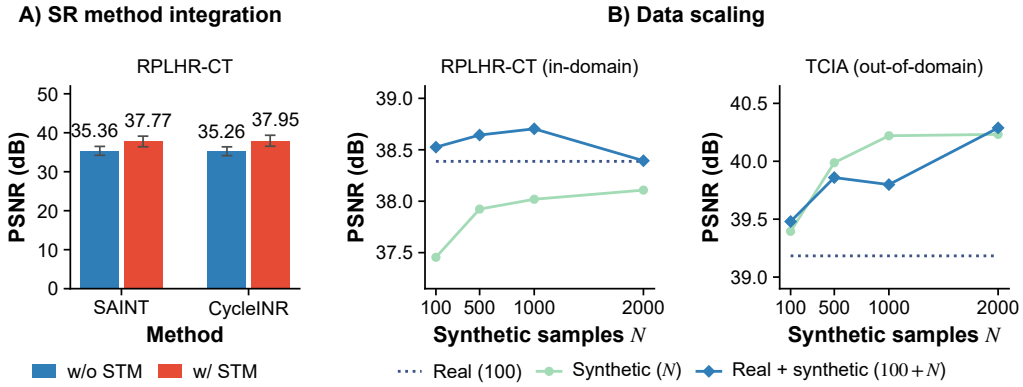


Figure 6: **A)** Integrating STM into SAINT and CycleINR on RPLHR-CT (Yu et al., 2022) (mean ± SD). **B)** Data scaling on RPLHR-CT and TCIA (Clark et al., 2013) using STM-synthesized CT-RATE pairs (Hamamci et al., 2026).

the iterative methods, with consistent SSIM gains and higher Dice in all evaluated segmentation settings. Models trained with STM-synthesized pairs approach their real-pair counterparts, while Fig. 5 shows more continuous pulmonary structures and segmentations closer to the HR reference. These results show that our STM benefits both reconstruction and downstream anatomical analysis across SR paradigms.

Integrating STM into Existing CT SR Methods. To assess compatibility with existing CT SR pipelines, we replace the degradation models in SAINT and CycleINR with STM while retaining their architectures. Fig. 6A shows PSNR gains of 2.41 and 2.69 dB; the full results in Table 4 of the Appendix also show corresponding SSIM gains and a Dice increase of 0.099 for both methods. Fig. 1 further shows finer anatomical structures and reduced partial-volume effects with STM. These consistent gains without architectural changes demonstrate that STM can strengthen established CT SR methods as a modular degradation model.

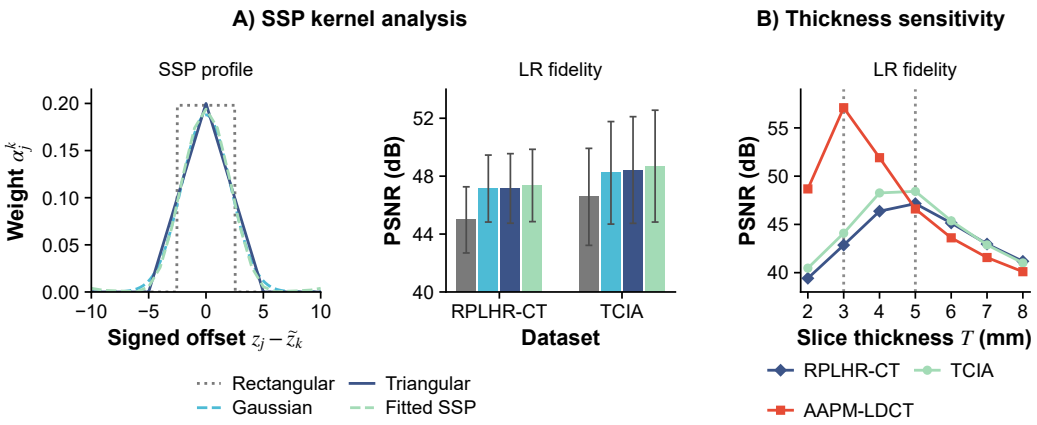


Figure 7: **A)** SSP kernel analysis on RPLHR-CT (Yu et al., 2022) and TCIA (Clark et al., 2013) (mean ± SD). Analytical FWHM: 5 mm; learned FWHM: 4.96 mm. **B)** Thickness sensitivity, additionally evaluated on AAPM-LDCT (McCollough et al., 2017). Dotted lines indicate the true thickness.

Data Scaling Study via STM. To assess whether STM can scale supervision beyond scarce real pairs, we use CT-RATE to synthesize N LR-HR pairs and train Restormer3D with N synthetic pairs, 100 real RPLHR-CT pairs, or their 100+N combination. Fig. 6B shows that performance generally improves with increasing N on the in-domain RPLHR-CT and out-of-domain TCIA test sets. On TCIA, 100 synthetic pairs already outperform real-only training, and the gain reaches about 1 dB at $N = 2000$. Combining real and synthetic data yields the best results on both test sets. The sustained out-of-domain gains show that STM converts HR-only volumes into effective supervision, complements limited paired acquisitions, and improves cross-dataset robustness.

Impact of SSP Kernel in STM. We compare rectangular, Gaussian, and triangular SSPs with FWHM fixed to $T = 5$ mm, alongside a data-fitted SSP with a learned FWHM of 4.96 mm. Fig. 7A shows that the rectangular profile performs worst because its uniform weights poorly reflect the longitudinal response, while the triangular profile is the strongest analytical choice consistent with helical CT weighting (Stierstorfer et al., 2004; Tang et al., 2025). The constrained SSP fitted on RPLHR-CT and applied unchanged to both test sets performs best by capturing protocol-specific nonidealities. All four variants outperform geometric interpolation, supporting thickness-aware aggregation with further gains from profile selection and fitting.

Impact of Slice Thickness T in STM. We vary T from 2 to 8 mm while fixing the triangular SSP and slice spacing. The true thickness is 5 mm for RPLHR-CT and TCIA and 3 mm for AAPM-LDCT. As shown in Fig. 7B, fidelity peaks at the corresponding thickness on all three datasets, reaching 47.15, 48.43, and 57.10 dB, respectively. Incorrect thickness under-aggregates anatomy or exaggerates partial-volume blurring, validating the physical meaning of T for both protocols.

5 Conclusion

This work rethinks degradation modeling for real-world thick-slice CT and introduces STM, which models clinical slice formation through normalized SSP-weighted aggregation while decoupling slice thickness from spacing. Experiments show that STM synthesizes more faithful thick slices, improves supervised and iterative SR and anatomical segmentation, integrates into established CT SR methods, and converts HR-only volumes into effective supervision. These findings establish accurate slice-formation modeling as essential for reliable real-world thick-slice CT SR.

Limitations STM is an effective image-domain approximation of through-plane slice formation, with HR inputs that themselves have finite slice thickness. It does not explicitly model the full scanner reconstruction process, spatially varying SSPs, or noise correlations; incorporating these effects remains a direction for future development and validation.

AI USE STATEMENT

We used generative AI tools solely for language polishing, table formatting, and figure editing. No generative AI tools were used to develop the research ideas or to design, implement, or conduct the experiments. We take responsibility for the final manuscript and all AI-assisted edits.

REFERENCES

Nashwan Karkhi Abdulkareem, Shereen Ismail Hajee, Fatiheea Fatihalla Hassan, Ilham Khalid Ibrahim, Ruaa Emad Hussein Al-Khalidi, and Noor Abubaker Abdulqader. Investigating the slice thickness effect on noise and diagnostic content of single-source multi-slice computerized axial tomography. *Journal of Medicine and Life*, 16(6):862, 2023.

Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3086–3095, 2019.

Trevor J Chan and Chamith S Rajapakse. A super-resolution diffusion model for recovering bone microstructure from ct images. *Radiology: Artificial Intelligence*, 5(6):e220251, 2023.

Zixuan Chen, Lingxiao Yang, Jian-Huang Lai, and Xiaohua Xie. 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.

Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al. The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of digital imaging*, 26(6): 1045–1057, 2013.

Chenhe Du, Qing Wu, Xuanyu Tian, Jingyi Yu, Hongjiang Wei, and Yuyao Zhang. Improving 2d diffusion models for 3d medical imaging with inter-slice consistent stochasticity. In *The Fourteenth International Conference on Learning Representations*, 2026.

Chaowei Fang, Liang Wang, Dingwen Zhang, Jun Xu, Yixuan Yuan, and Junwei Han. Incremental cross-view mutual distillation for self-supervised medical ct synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20677–20686, 2022.

Wei Fang, Yuxing Tang, Heng Guo, Mingze Yuan, Tony CW Mok, Ke Yan, Jiawen Yao, Xin Chen, Zaiyi Liu, Le Lu, 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.

Travis C Greene and X John Rong. Evaluation of techniques for slice sensitivity profile measurement and analysis. *Journal of applied clinical medical physics*, 15(2):281–294, 2014.

Ibrahim Ethem Hamamci, Sezgin Er, Chenyu Wang, Furkan Almas, Ayse Gulnihan Simsek, Sevval Nil Esirgun, Irem Dogan, Omer Faruk Durugol, Benjamin Hou, Suprosanna Shit, et al. Generalist foundation models from a multimodal dataset for 3d computed tomography. *Nature Biomedical Engineering*, pp. 1–19, 2026.

Changhui Jiang, Qiyang Zhang, Rui Fan, and Zhanli Hu. Super-resolution ct image reconstruction based on dictionary learning and sparse representation. *Scientific reports*, 8(1):8799, 2018.

Guangyuan Li, Lei Zhao, Jiakai Sun, Zehua Lan, Zhanjie Zhang, Jiafu Chen, Zhijie Lin, Huaizhong Lin, and Wei Xing. Rethinking multi-contrast mri super-resolution: Rectangle-window cross-attention transformer and arbitrary-scale upsampling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 21230–21240, October 2023.

Yinhao Li, Yutaro Iwamoto, Lanfen Lin, Rui Xu, Ruofeng Tong, and Yen-Wei Chen. Volumenet: A lightweight parallel network for super-resolution of mr and ct volumetric data. *IEEE Transactions on Image Processing*, 30:4840–4854, 2021. doi: 10.1109/TIP.2021.3076285.

Anran Liu, Yihao Liu, Jinjin Gu, Yu Qiao, and Chao Dong. Blind image super-resolution: A survey and beyond. *IEEE transactions on pattern analysis and machine intelligence*, 45(5):5461–5480, 2022.

Cynthia H McCollough and Prabhakar Shantha Rajiah. Milestones in ct: past, present, and future. *Radiology*, 309(1):e230803, 2023.

Cynthia H McCollough, Adam C Bartley, Rickey E Carter, Baiyu Chen, Tammy A Drees, Phillip Edwards, David R Holmes III, Alice E Huang, Farhana Khan, Shuai Leng, et al. Low-dose ct for the detection and classification of metastatic liver lesions: results of the 2016 low dose ct grand challenge. *Medical physics*, 44(10):e339–e352, 2017.

Brian B. Moser, Arundhati S. Shanbhag, Federico Raue, Stanislav Frolov, Sebastian Palacio, and Andreas Dengel. Diffusion models, image super-resolution, and everything: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 36(7):11793–11813, 2025. doi: 10.1109/TNNLS.2024.3476671.

Junyoung Park, Donghwi Hwang, Kyeong Yun Kim, Seung Kwan Kang, Yu Kyeong Kim, and Jae Sung Lee. Computed tomography super-resolution using deep convolutional neural network. *Physics in Medicine & Biology*, 63(14):145011, 2018.

Cheng Peng, Wei-An Lin, Haofu Liao, Rama Chellappa, and S Kevin Zhou. Saint: Spatially aware interpolation network for medical slice synthesis. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7747–7756. IEEE Computer Society, 2020.

Cheng Peng, S. Kevin Zhou, and Rama Chellappa. Da-vsr: Domain adaptable volumetric super-resolution for medical images. In Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pp. 75–85, Cham, 2021. Springer International Publishing. ISBN 978-3-030-87231-1.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi (eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4.

Geoffrey D Rubin. Computed tomography: revolutionizing the practice of medicine for 40 years. *Radiology*, 273(2S):S45–S74, 2014.

Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

Jiayang Shi, Daniël M Pelt, and K Joost Batenburg. Sr4zct: Self-supervised through-plane resolution enhancement for ct images with arbitrary resolution and overlap. In *International Workshop on Machine Learning in Medical Imaging*, pp. 52–61, 2023.

Haofei Song, Xintian Mao, Jing Yu, Qingli Li, and Yan Wang. I^3net: Inter-intra-slice interpolation network for medical slice synthesis. *IEEE Transactions on Medical Imaging*, 43(9):3306–3318, 2024.

Karl Stierstorfer, Annabella Rauscher, Jan Boese, Herbert Bruder, Stefan Schaller, and Thomas Flohr. Weighted fbp—a simple approximate 3d fbp algorithm for multislice spiral ct with good dose usage for arbitrary pitch. *Physics in Medicine & Biology*, 49(11):2209–2218, 2004.

Zeyu Tang, Xiaodan Xing, Gang Wang, and Guang Yang. Enhancing super-resolution network efficacy in ct imaging: Cost-effective simulation of training data. *IEEE Open Journal of Engineering in Medicine and Biology*, 2025.

Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1905–1914, 2021a.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *The Eleventh International Conference on Learning Representations*, 2023.

Zhihao Wang, Jian Chen, and Steven C. H. Hoi. Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3365–3387, 2021b. doi: 10.1109/TPAMI.2020.2982166.

Jakob Wasserthal, Hanns-Christian Breit, Manfred T Meyer, Maurice Pradella, Daniel Hinck, Alexander W Sauter, Tobias Heye, Daniel T Boll, Joshy Cyriac, Shan Yang, et al. Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. *Radiology: Artificial Intelligence*, 5 (5):e230024, 2023.

Qing Wu, Yuwei Li, Yawen Sun, Yan Zhou, Hongjiang Wei, Jingyi Yu, and Yuyao Zhang. 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.

Hanguang Xiao, Zhiying Yang, Tianqi Liu, Shihong Liu, Xiaoxuan Huang, and Jiahui Dai. Deep learning for medical imaging super-resolution: A comprehensive review. *Neurocomputing*, 630: 129667, 2025.

Huiqiao Xie, Yang Lei, Tonghe Wang, Zhen Tian, Justin Roper, Jeffrey D Bradley, Walter J Curran, Xiangyang Tang, Tian Liu, and Xiaofeng Yang. High through-plane resolution ct imaging with self-supervised deep learning. *Physics in medicine & biology*, 66(14):145013, 2021.

Chenyu You, Guang Li, Yi Zhang, Xiaoliu Zhang, Hongming Shan, Mengzhou Li, Shenghong Ju, Zhen Zhao, Zhuiyang Zhang, Wenxiang Cong, et al. Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle). *IEEE transactions on medical imaging*, 39(1):188–203, 2019.

Pengxin Yu, Haoyue Zhang, Han Kang, Wen Tang, Corey W Arnold, and Rongguo Zhang. Rplhr-ct dataset and transformer baseline for volumetric super-resolution from ct scans. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 344–353. Springer, 2022.

Pengxin Yu, Haoyue Zhang, Dawei Wang, Rongguo Zhang, Mei Deng, Haoyu Yang, Lijun Wu, Xiaoxu Liu, Andrea S Oh, Fereidoun G Abtin, et al. Spatial resolution enhancement using deep learning improves chest disease diagnosis based on thick slice ct. *npj Digital Medicine*, 7(1):335, 2024.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5728–5739, 2022.

Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4791–4800, 2021.

Minghui Zhang, Yangqian Wu, Hanxiao Zhang, Yulei Qin, Hao Zheng, Wen Tang, Corey Arnold, Chenhao Pei, Pengxin Yu, Yang Nan, et al. Multi-site, multi-domain airway tree modeling. *Medical image analysis*, 90:102957, 2023.

A Appendix

A.1 Details of Datasets

Table 3: Summary of the real-paired RPLHR-CT (Yu et al., 2022), TCIA (Clark et al., 2013), and AAPM-LDCT (McCollough et al., 2017) datasets and the HR-only CT-RATE dataset (Hamamci et al., 2026). All spatial resolutions are reported in millimeters.

Dataset	Type	Data Split	Data Role	HR Images		LR Images		Manufacturer
				T	Δz	T	Δz	
RPLHR-CT	Real paired	100/50/100	Internal Train/val/test	1.0 mm	1.0 mm	5.0 mm	5.0 mm	Philips
TCIA	Real paired	20	External Test	1.0 mm	1.0 mm	5.0 mm	5.0 mm	Philips/ Siemens/ GE
AAPM-LDCT	Real paired	10	External Test	1.0 mm	0.8 mm	3.0 mm	2.0 mm	Siemens
CT-RATE	HR only	2,000	Train for scaling study	1.0 mm	1.0 mm	5.0 mm	5.0 mm	Philips/ Siemens

We use three real-paired LR-HR datasets and one HR-only dataset. RPLHR-CT and TCIA represent contiguous acquisition settings, where LR slice thickness equals slice spacing. AAPM-LDCT provides an overlapping setting, where LR slice thickness is larger than slice spacing. CT-RATE provides large-scale thin-slice CT volumes for synthesizing additional supervision. Unless otherwise specified, CT intensities are clipped to [-1024, 2048] HU and linearly normalized to [0, 1]. Table 3 summarizes the dataset configurations and their roles in our experiments.

RPLHR-CT RPLHR-CT (Yu et al., 2022) is a public real-paired dataset for volumetric thick-slice CT SR. It contains 250 paired chest CT cases. For each case, the same acquisition is reconstructed into an HR thin-slice volume with a slice thickness and spacing of 1 mm and an LR thick-slice volume with a slice thickness and spacing of 5 mm. The paired volumes share the same in-plane grid, with resolutions ranging from 0.604 mm to 0.795 mm, and are spatially aligned according to their physical coordinates. The released volumes are stored in NIfTI format with an in-plane matrix of 512×512. HR and LR volumes contain 191–396 and 39–80 slices, respectively.

We follow the official split of 100 training, 50 validation, and 100 testing cases. For synthetic-pair SR models, the training set is used for training and the validation set for model selection and hyperparameter tuning. The test set remains held out for all quantitative and qualitative evaluations. For degradation-fidelity experiments, each degradation operator is applied to the HR volume of the test set, and the resulting synthetic LR volume is compared with the corresponding real LR volume. The Table 2 Real Pairs reference models use 100 real LR-HR pairs for training and 50 pairs for validation. Since the LR slice thickness and spacing are both 5 mm, RPLHR-CT represents a contiguous acquisition setting.

TCIA We curate an additional real-paired dataset from The Cancer Imaging Archive (TCIA) (Clark et al., 2013). After data cleaning and pair selection, 20 cases with paired thin- and thick-slice CT volumes are retained. The HR volumes have a slice thickness and spacing of 1 mm, while the corresponding LR volumes have a slice thickness and spacing of 5 mm. Their in-plane resolutions range from 0.6 mm to 0.8 mm. Similar to RPLHR-CT, the equal LR thickness and spacing define a contiguous slice configuration.

Note that no TCIA cases are used for network training or model selection. Instead, the entire curated dataset serves as an external test set. It is used to evaluate whether a degradation model trained on RPLHR-CT remains effective under cross-dataset variations in anatomy, scanner, and acquisition protocol. We report degradation fidelity by comparing synthetic LR volumes with their real counterparts. We further evaluate SR reconstruction and downstream anatomical segmentation using models trained only on RPLHR-CT or STM-synthesized CT-RATE data. This protocol provides an out-of-domain (OOD) assessment without adapting the SR models to the TCIA dataset.

AAPM-LDCT AAPM-LDCT (McCollough et al., 2017) originates from the 2016 NIH–AAPM–Mayo Clinic Low Dose CT Grand Challenge and contains contrast-enhanced abdominal CT data. From this resource, we use 10 cases that provide paired thin- and thick-slice image series. The HR volumes have a slice thickness of 1 mm and a slice spacing of 0.8 mm. The LR volumes have a slice thickness of 3 mm and a slice spacing of 2 mm. Therefore, adjacent LR slices overlap by 1 mm, providing a protocol in which slice thickness and spacing are explicitly different.

AAPM-LDCT is used only for evaluating degradation fidelity and the physical parameterization of STM; no cases are used to train the supervised SR networks. For each case, NN, linear, cubic, and STM degradation operators are applied to the HR volume using the LR physical coordinates. The synthesized volumes are then compared with the paired real LR volumes. This dataset complements the contiguous RPLHR-CT and TCIA settings by evaluating whether STM can separately model slice thickness and spacing under overlapping acquisition. It is also used in the slice-thickness sensitivity experiment, where the assumed FWHM parameter is varied while the LR slice spacing remains fixed.

CT-RATE CT-RATE (Hamamci et al., 2026) is a large-scale collection of non-contrast chest CT volumes accompanied by radiology reports. The full dataset contains multiple reconstructions from a large patient cohort and exhibits substantial variability in scanner manufacturer, spatial resolution, and reconstruction protocol. Because CT-RATE does not provide the real thick-slice counterparts required for paired SR evaluation, we use it exclusively as an HR-only dataset. We randomly select 2,000 thin-slice volumes with 1 mm through-plane spacing and in-plane resolutions between 0.6 mm and 0.8 mm. STM is applied to each selected HR volume to synthesize a corresponding LR volume, thereby forming synthetic LR-HR training pairs. For the data-scaling study, we construct subsets with $N \in 100, 500, 1000, 2000$ cases and train Restormer3D using: 1) N STM-synthesized pairs, 2) 100 real RPLHR-CT training pairs, or 3) their combination containing $100 + N$ pairs. All resulting models are evaluated on the held-out RPLHR-CT and TCIA test sets. CT-RATE itself is not used for evaluation, ensuring that the scaling experiment measures whether HR-only external data can improve real-world SR performance rather than performance on the source dataset.

Overall, the four datasets above cover complementary experimental settings. RPLHR-CT provides the primary real-paired benchmark and supervised reference; TCIA evaluates cross-dataset generalization; AAPM-LDCT evaluates overlapping slice acquisition; and CT-RATE measures the scalability of STM-generated supervision.

A.2 Details of Quantitative Metrics

Image Quality. SR performance is evaluated using peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). Both metrics are computed between the reconstructed SR volume and its corresponding HR reference. Before evaluation, CT intensities are clipped to [-1024, 2048] HU and normalized to [0, 1], consistent with the preprocessing used for network training. We report the mean and standard deviation of PSNR and SSIM across all test cases.

Downstream Anatomical Segmentation. We further assess whether the SR volumes preserve fine anatomical structures relevant to downstream analysis. We use the pre-trained *lung_vessels* model (`https://github.com/wasserth/TotalSegmentator`) of TotalSegmentator (Wasserthal et al., 2023) without additional fine-tuning. This model separately segments four pulmonary structures: pulmonary arteries, pulmonary veins, airways, and airway walls.

For each test case, the same segmentation model is independently applied to the reconstructed SR volume and its corresponding HR reference. The segmentation obtained from the HR volume serves as the reference, allowing us to evaluate anatomical consistency without requiring additional manual annotations. Dice scores are computed separately for the four structures between the SR-derived and HR-derived masks. We average the four class-wise scores to obtain a case-level Dice score and report the mean and standard deviation across the test set. All SR methods use the same TotalSegmentator model and inference configuration.

A.3 Details of Two SR Paradigms

The degradation operator plays different roles in the two investigated SR paradigms. In end-to-end supervised learning, it determines how synthetic LR-HR training pairs are constructed. In

unsupervised iterative reconstruction, it defines the data-consistency constraint for each observed LR volume. In controlled degradation comparisons, we modify only the operator while keeping the reconstruction prior and other method-specific settings unchanged; Real Pairs serves as a separate reference.

A.3.1 End-to-End Supervised Learning

For each HR training volume, we generate an LR volume on the native LR grid using NN, linear, cubic, or STM degradation. The resulting LR volume is then linearly interpolated back to the HR grid and used as the network input. For the Table 2 Real Pairs reference, the 100 acquired training pairs undergo the same LR interpolation procedure; 50 additional pairs are reserved for validation. Each degradation setting uses a separate model trained with an ℓ_1 loss.

UNet3D UNet3D (Ronneberger et al., 2015) contains five resolution levels with channel widths of $[64, 128, 256, 512, 1024]$. Each encoder level consists of two $3 \times 3 \times 3$ convolutions, followed by instance normalization and LeakyReLU with a negative slope of 0.1. Downsampling is performed using $2 \times 2 \times 2$ max pooling. The decoder uses trilinear upsampling, skip connections from the corresponding encoder levels, and two convolutional blocks at each resolution. A global residual connection adds the interpolated LR input to the predicted residual. The model contains approximately 94.13 million parameters.

Restormer3D Restormer3D (Zamir et al., 2022) is a three-dimensional adaptation of Restormer (Zamir et al., 2022). Its four encoder stages use channel widths of $[16, 32, 64, 128]$, with $[2, 2, 2, 2]$ Transformer blocks and $[1, 2, 4, 8]$ attention heads, respectively. The decoder mirrors the encoder and includes two refinement blocks at full spatial resolution. The feed-forward expansion factor is set to 2.66, and convolutional biases are disabled. Restormer3D also predicts a residual that is added to the interpolated LR input. The model contains 1,007,567 trainable parameters.

Hyper-parameters of Training and Inference Both networks are trained for 500 epochs using randomly sampled $64 \times 64 \times 64$ patches. The batch sizes are two for UNet3D and one for Restormer3D. We use AdamW with an initial learning rate of 10^{-3} and zero weight decay, and reduce the learning rate by a factor of 0.5 every 100 epochs. Automatic mixed precision is enabled during training. For synthetic-pair models, validation is performed every ten epochs using eight randomly sampled patches from each validation volume, and the checkpoint with the highest validation PSNR is retained. For the Real Pairs reference, we retain the final-epoch checkpoint trained on the 100 training pairs.

At inference, each complete CT volume is reconstructed using $128 \times 128 \times 128$ sliding windows with 50% overlap along each dimension. Predictions in overlapping regions are uniformly averaged, and the normalized output is clamped to $[0, 1]$. Optimization and inference settings are shared across degradation settings; the Real Pairs reference uses the final-epoch checkpoint specified above.

A.3.2 Unsupervised Iterative Reconstruction

Throughout this subsection, $\mathbf{A}$ denotes the degradation operator under evaluation, including nearest-neighbor, linear, cubic, or STM, and $\mathbf{y}$ denotes the observed thick-slice volume. All degradation operators act only along the through-plane axis while preserving the in-plane sampling grid.

ADMM-TV. We use a model-based reconstruction baseline that combines measurement fidelity with through-plane total variation:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda_{\mathrm{TV}} \|\mathbf{D}_z \mathbf{x}\|_1 , \tag{12}$$

where $\mathbf{D}_z$ is the first-order finite-difference operator along the z-axis. Since the observation process reduces resolution only in the through-plane direction, no in-plane TV regularization is applied.

With auxiliary variable $\mathbf{q} = \mathbf{D}_z\mathbf{x}$ and scaled dual variable $\mathbf{u}$, iteration k updates

$$\mathbf{x}^{k+1} = \mathrm{CG}\left(\mathbf{H}_\rho, \mathbf{b}^k\right), \tag{13}$$

$$\mathbf{q}^{k+1} = \mathrm{soft}\left(\mathbf{D}_z\mathbf{x}^{k+1} + \mathbf{u}^k, \frac{\lambda_{\mathrm{TV}}}{\rho}\right), \tag{14}$$

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \mathbf{D}_z\mathbf{x}^{k+1} - \mathbf{q}^{k+1}, \tag{15}$$

where $\mathbf{H}_\rho = \mathbf{A}^\mathsf{T}\mathbf{A} + \rho\mathbf{D}_z^\mathsf{T}\mathbf{D}_z + \varepsilon\mathbf{I}$ and $\mathbf{b}^k = \mathbf{A}^\mathsf{T}\mathbf{y} + \rho\mathbf{D}_z^\mathsf{T}(\mathbf{q}^k - \mathbf{u}^k)$. Here, $\mathrm{CG}(\mathbf{H}, \mathbf{b})$ denotes the approximate solution of the linear system $\mathbf{H}\mathbf{x} = \mathbf{b}$ obtained by conjugate gradients, and $\mathrm{soft}(\cdot, \cdot)$ denotes element-wise soft-thresholding.

We run ADMM for 8 iterations with $\lambda_{\mathrm{TV}} = 1.5 \times 10^{-3}$, $\rho = 0.5$, and $\varepsilon = 10^{-6}$. Each $\mathbf{x}$-update uses 20 CG iterations with a tolerance of 10^{-5}.

DDNM. DDNM (Wang et al., 2023) combines a pretrained diffusion model with measurement consistency through a range–null space decomposition. Given a clean-image estimate $\mathbf{x}_{0|s}$ at solver step s, the DDNM update rule can be expressed as:

$$\widetilde{\mathbf{x}}_{0|s} = \left(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A}\right)\mathbf{x}_{0|s} + \mathbf{A}^\dagger\mathbf{y}, \tag{16}$$

where the first term retains the component supplied by the diffusion prior, while the second term recovers the measurement-determined component.

Generative prior training. We train a pixel-space flow-matching model with a 2D U-Net backbone to reduce training time and computational cost compared with volumetric modeling. The prior is trained exclusively on thin-slice lung CT volumes from ATM22 (Zhang et al., 2023), with no overlap with any dataset used for SR training or evaluation in the main experiments. We retain volumes with a slice thickness no larger than 1 mm and at least 384 axial slices, from which sagittal slices are extracted. Each slice is resampled to 512×512, and slices with less than 15% body occupancy are discarded. The resulting training set contains 115,040 sagittal slices.

The single-channel 2D U-Net v_θ is trained to predict the velocity field using the flow-matching objective

$$\mathcal{L}_{\mathrm{FM}} = \mathbb{E}_{\mathbf{x}_0,t,\boldsymbol{\epsilon}}\left[\left\|v_\theta(\mathbf{x}_t, t) - (\mathbf{x}_0 - \boldsymbol{\epsilon})\right\|_2^2\right]. \tag{17}$$

We train the model for 17 epochs using AdamW with a learning rate of 10^{-4}, weight decay of 10^{-2}, and a global batch size of 40 (5 images per GPU on 8 GPUs).

DDNM reconstruction. At each solver step, one Euler update of the flow ODE produces the clean-image estimate

$$\mathbf{x}_{0|s} = \mathbf{x}_s + \alpha_s v_\theta(\mathbf{x}_s, t_s), \tag{18}$$

which is subsequently corrected using Eq. (16). Except at the final step, the corrected estimate is mapped to the next noise level according to

$$\mathbf{x}_{s'} = (1 - \alpha_{s'})\widetilde{\mathbf{x}}_{0|s} + \alpha_{s'}\eta\boldsymbol{\epsilon}', \tag{19}$$

where $\boldsymbol{\epsilon}'$ is a newly sampled Gaussian noise field. To improve coherence across the volume, noise fields in adjacent sagittal slices are correlated using ISCS (Du et al., 2026).

We use 20 DDNM solver steps with $\eta = 1.0$. The regularized pseudoinverse is solved using 10 CG iterations. All reported $5\times$ reconstructions use full volumes without spatial cropping.

A.4 Details of Two Leading SR Models

For SAINT and CycleINR, we compare the data simulation scheme adopted in the original method with our proposed STM-based simulation. Within each method, the network architecture, training data, optimization settings, and random seed are kept unchanged across the two simulation schemes. Both methods are trained using the 100 thin-slice volumes from the RPLHR-CT training set. The acquired 5-mm volumes are not used for training.

A.4.1 SAINT

Architecture SAINT (Peng et al., 2020) performs through-plane interpolation from sagittal and coronal views using an anisotropic meta-interpolation module. We reproduce the method based on the original paper and its officially released implementation, with minimal modernization in PyTorch to improve computational efficiency while preserving the original model design.

The interpolation network uses a 2D residual dense backbone to extract features from three adjacent marginal slices. A filter distance matrix encodes the relative positions between the observed slices and the target interpolation phases. The distance features are converted into spatially varying filters and applied to the image features to predict the five HR phases required for $5\times$ through-plane SR. The model contains approximately 2.79 million trainable parameters.

Training The model is trained for 150 epochs using an ℓ_1 loss and Adam with a learning rate of 10^{-4} and a batch size of 64. Both the original simulation and STM use this training configuration.

A.4.2 CycleINR

Architecture CycleINR (Fang et al., 2024) formulates volumetric SR as a continuous coordinate-based reconstruction problem. We implement the method following the architecture and training procedure described in the original paper and fix its arbitrary-scale formulation to $5\times$ SR.

The model consists of a 3D residual encoder, a local attention module, and a coordinate-based implicit decoder. The encoder maps the LR volume to a voxel-wise latent representation. For each continuous query coordinate, the local attention module aggregates features from its neighboring latent voxels, and the implicit decoder maps the aggregated feature and spatial coordinate to the corresponding CT intensity. The complete model contains approximately 2.63 million trainable parameters.

Training CycleINR is trained for 900 epochs. We use Adam with a batch size of 1 and an initial learning rate of 10^{-4}. The learning rate is multiplied by 0.8 every 200 epochs. All training settings are shared between the original simulation and STM.

A.5 Additional Results

Table 4 reports PSNR, SSIM, and Dice for SAINT and CycleINR with and without STM. Fig. 8 compares thick-slice LR synthesis, while Fig. 9 and Fig. 10 show additional SR reconstructions and downstream anatomical segmentations.

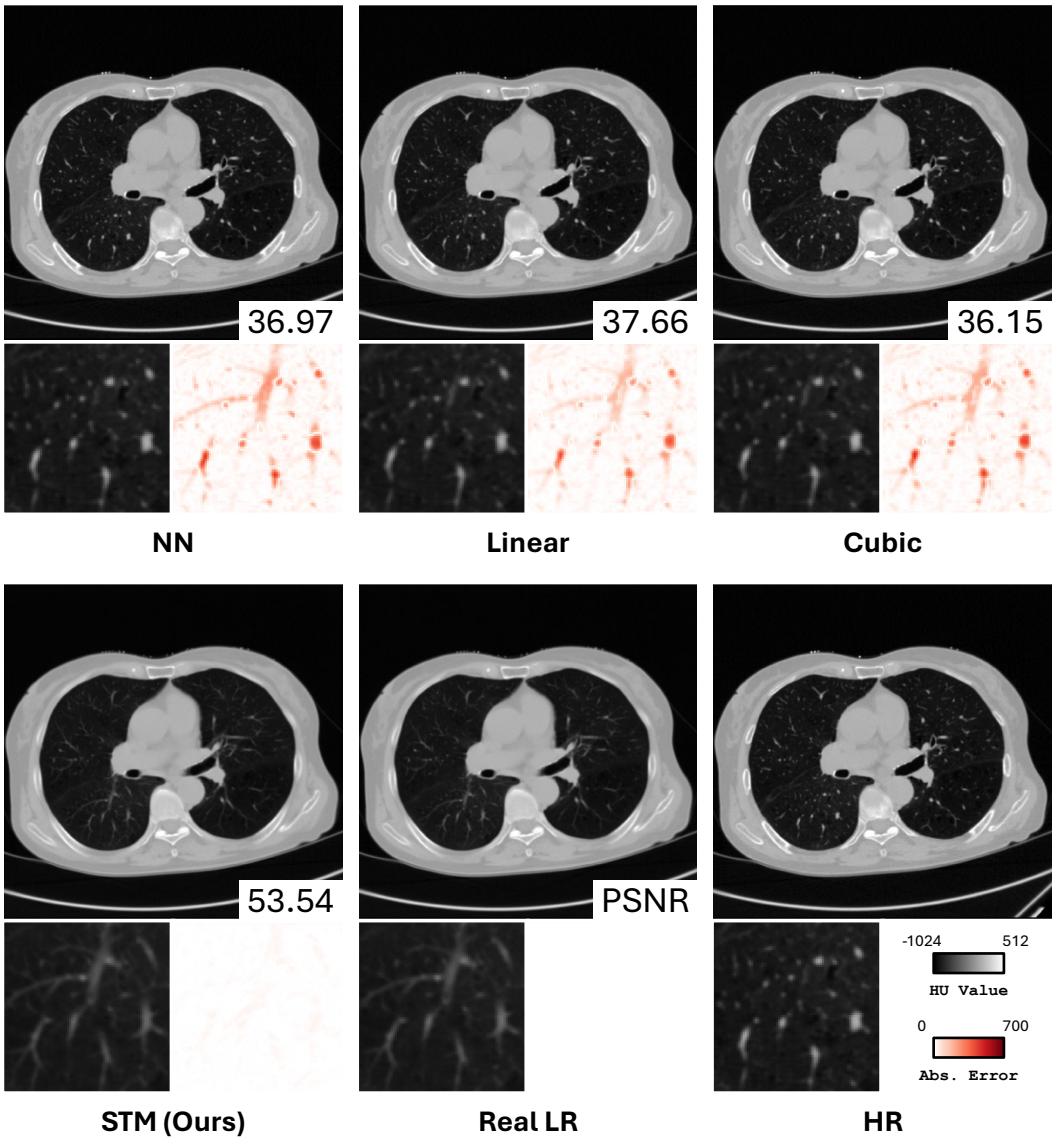


Figure 8: Qualitative comparison of thick-slice LR synthesis using geometric resampling (NN, linear, and cubic) and our STM on the real-world TCIA dataset (Clark et al., 2013).

Table 4: Quantitative comparison of SAINT (Peng et al., 2020) and CycleINR (Fang et al., 2024) with and without our STM on the real-world RPLHR-CT dataset (Yu et al., 2022).

Method	Setting	PSNR	SSIM	Dice
SAINT	w/o STM	35.36±1.14	0.906±0.028	0.791±0.025
	w/ STM	**37.77**±1.36	**0.928**±0.024	**0.890**±0.013
CycleINR	w/o STM	35.26±1.13	0.907±0.028	0.797±0.023
	w/ STM	**37.95**±1.41	**0.930**±0.024	**0.896**±0.012

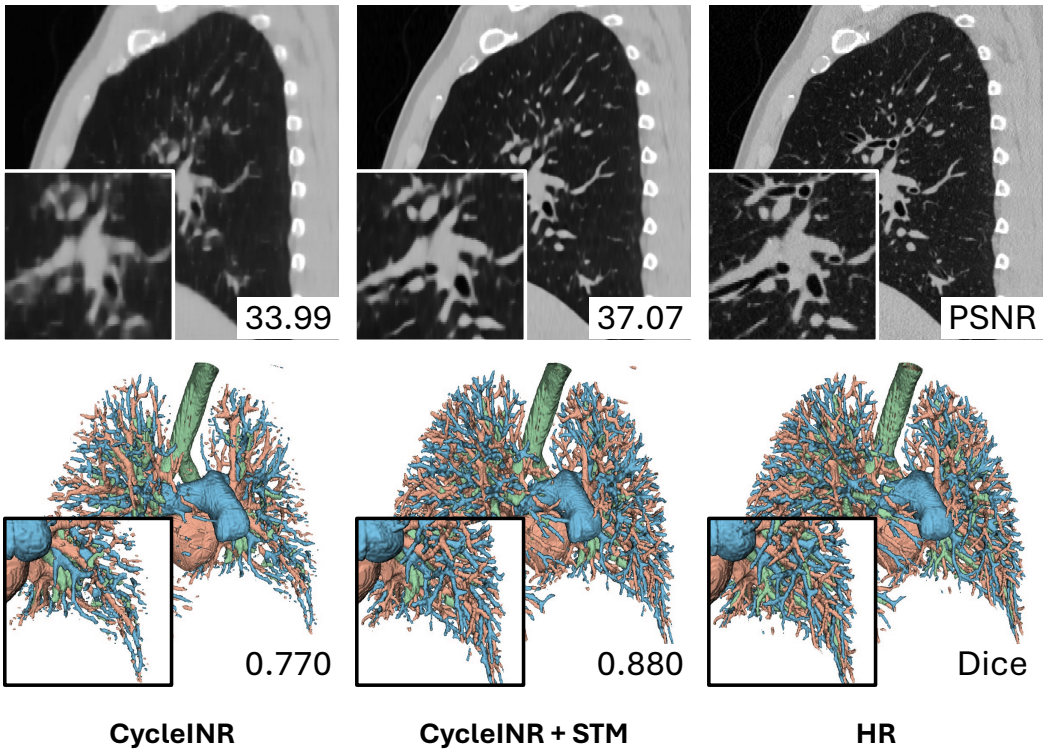


Figure 9: Qualitative comparison of CycleINR (Fang et al., 2024) with and without our STM for 5× through-plane SR on the real-world RPLHR-CT dataset (Yu et al., 2022).

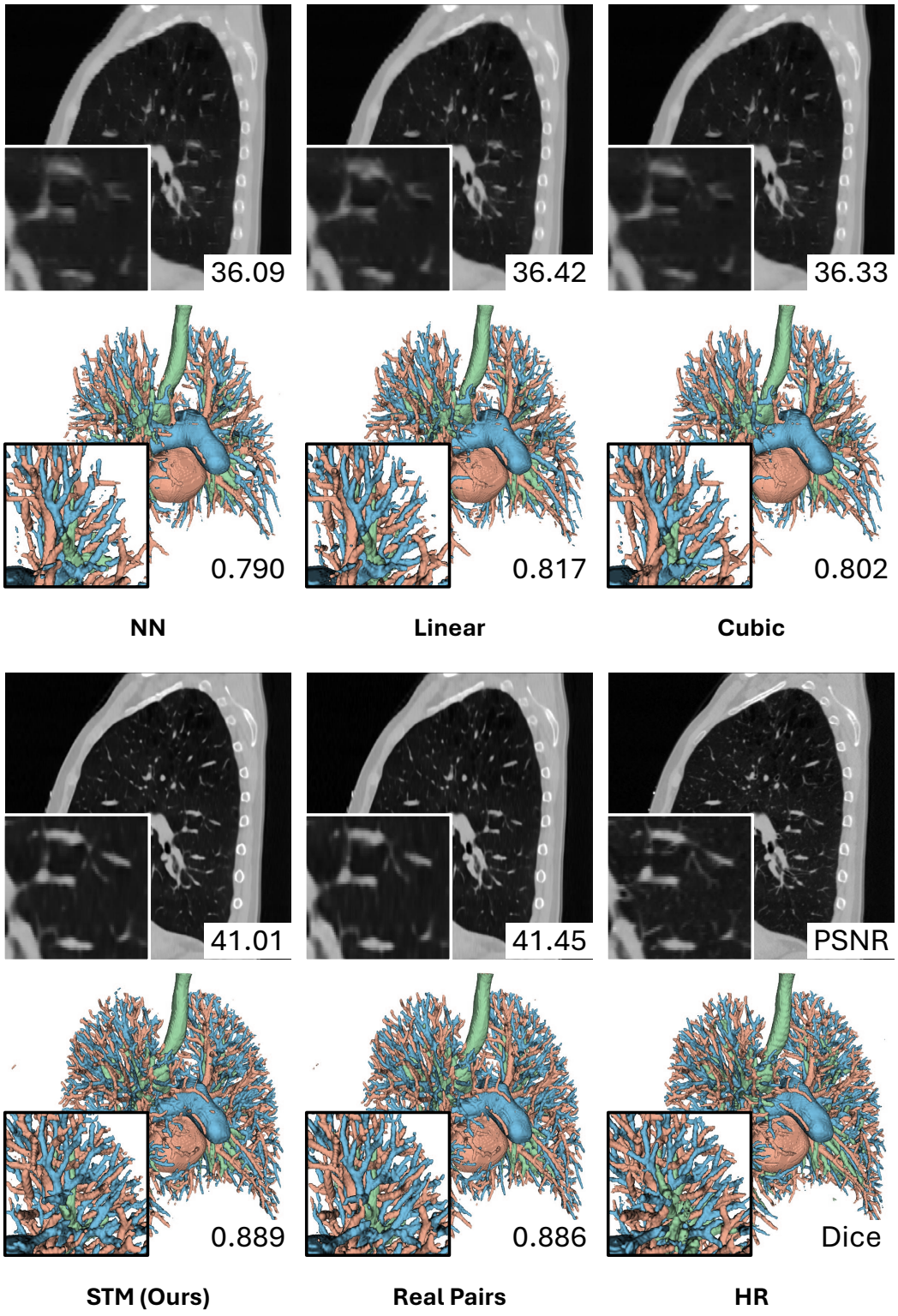


Figure 10: Qualitative comparison of supervised UNet3D trained with synthetic LR-HR pairs constructed using geometric resampling (NN, linear, and cubic) or our STM, alongside a real-pair reference, on the real-world TCIA dataset (Clark et al., 2013).