PRISM: PROBABILISTIC 3D RESTORATION VIA INTER-BAND SPECTRAL MODELING

Anonymous authors
Paper under double-blind review

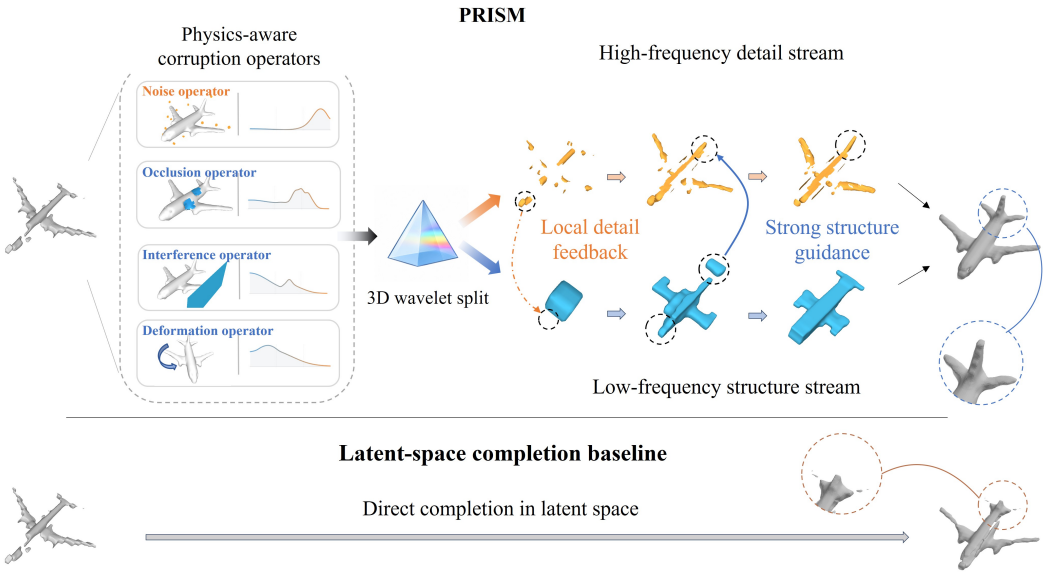

Figure 1: Overview of PRISM. A fixed one-level 3D wavelet split reveals band-wise corruption responses. PRISM models low-frequency structure and high-frequency details with asymmetric cross-band interaction, combining strong structural guidance with local-detail feedback, while the baseline completes directly in latent space without explicit spectral decomposition.

ABSTRACT

Real-world 3D point clouds are often affected by mixed corruptions, including noise, occlusion, and background interference. Existing methods generally lack explicit physical modeling of corruption, making it difficult to distinguish stochastic disturbances from true geometry or use object-level geometric context to constrain local completion. We therefore reformulate restoration as a "physics-aware spectral generation" problem. Using a fixed 3D discrete wavelet transform, we analyze the frequency-band responses of four corruption operators, express observation changes as operator-induced coefficient deviations, and characterize band-wise differences between stochastic disturbances and true geometry, as well as cross-band relations between the low-frequency skeleton and high-frequency details induced by structural degradation. Building on this analysis, we propose PRISM (Probabilistic 3D Restoration via Inter-band Spectral Modeling), a physics-aware wavelet-based generative framework. Following the Knothe-Rosenblatt (KR) rearrangement theorem, we design an Inter-band Dual-Stream Probabilistic Bridge that decomposes complex joint-distribution inference into cascaded inference of the marginal distribution of the low-frequency skeleton and the conditional distribution of high-frequency details. Experiments show that PRISM provides a physically interpretable inference path and significantly outperforms existing methods on challenging mixed-corruption benchmarks. Code will be released soon.

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.

1 INTRODUCTION

With the rapid advancement of 3D perception, point clouds acquired from LiDAR or depth cameras have become a core data source for autonomous driving, robotic navigation, augmented reality, and related applications. However, due to sensor noise, object occlusion, and background interference, raw observations often exhibit multiple forms of degradation Tesema et al. (2025); Pathak et al. (2025), which we collectively refer to as "corruption." Recovering clean and complete geometry from highly non-ideal observations remains a pressing challenge. Current research has mainly evolved along two mainstream directions. The first is point cloud completion based on spatial-domain geometric features Yu et al. (2021). These methods extract deterministic geometric features to handle complex degradations. For example, recent work such as DWCNet Tesema et al. (2025), grounded in extensive observations and statistics of real-world data, was among the first to focus on and define mixed degradation settings involving non-target object interference, background clutter, and sensor jitter. However, most of these methods still operate in the original spatial domain, where observation disturbances and target geometry jointly enter feature extraction and can become mixed in feature representations, leading to remaining noise or structural blurring. The second is generative 3D completion, which leverages the strong distribution-modeling capability of generative models to infer missing geometry Chu et al. (2023); Kong et al. (2025). In real-world scenarios, however, models may mistake mixed noise and background interference for target geometry, and such distribution-level confusion can generate erroneous geometry that deviates from the true object topology.

The common root of these two limitations is that, although existing methods focus on geometric feature extraction and data distribution modeling, respectively, they generally lack explicit physical modeling of the corruption process. On the one hand, spatial feature extraction does not explicitly distinguish stochastic disturbances in local observations from true geometric details (e.g., sharp edges), so their local differences can be obscured after feature aggregation. On the other hand, local occlusion and structural interference can disrupt inference of the target object's holistic geometric skeleton, thereby weakening the constraint of holistic geometric context on completing locally missing regions. To gain deeper insight into the physical properties of corruption, grounded in the above comprehensive empirical observations of real-world scenarios, we further formalize these mixed degradations beyond phenomenological descriptions as four composable basic physical operators: geometric deformation, interference, occlusion, and noise. To address the highly nonlinear coupling among these operators, we adopt a fixed 3D DWT as a unified analysis basis, map observations into a coefficient space with explicit frequency-band structure, and analyze the response behavior of the four corruption operators in this space, thereby establishing the core "spectral decoupling mechanism." This mechanism shows that the observation deviations induced by different physical corruptions can be expressed in the wavelet domain as operator-induced coefficient deviations with distinct band-wise characteristics, providing an inductive bias for explicitly analyzing corruption physics. This theoretical reformulation not only makes stochastic observation disturbances and geometric structures easier to distinguish in spectral statistics, but also characterizes cross-band relations between the low-frequency skeleton and high-frequency details, providing a theoretical basis for using holistic geometric context to constrain local detail generation. Figure 1 summarizes these spectral responses.

Building on the physical inductive bias provided by the spectral decoupling mechanism, our core insight is to reformulate the physical corruption inverse process, which is difficult to solve directly in Euclidean space, as a spectral generation problem in coefficient space. Generative completion must not only model the distribution of complete geometry but also explicitly exploit structural relationships across frequency bands, enabling the low-frequency skeleton to constrain high-frequency detail generation. Accordingly, we construct PRISM (Probabilistic 3D Restoration via Inter-band Spectral Modeling), reformulating completion as generative Bayesian inference in the wavelet representation space: band-wise statistical differences among degradations are used to distinguish local true geometry from stochastic observation disturbances, while cross-band relations allow holistic geometric context to constrain local detail generation. To avoid an empirically partitioned engineering design and construct a probabilistic modeling framework, we follow the Knothe-Rosenblatt (KR) rearrangement theorem and decompose complex joint-distribution generation into cascaded inference of the marginal distribution of the low-frequency skeleton and the conditional distribution of high-frequency details. This leads to our Inter-band Dual-Stream Probabilistic Bridge, which learns conditional probability transport from incomplete observations to complete geometry in the wavelet latent space.

The main contributions of this paper are:

- We establish a physical operator model of 3D corruption and formally analyze its frequency-band responses in the wavelet domain, characterizing band-wise differences among corruption components and the cross-band responses of structural degradation, thereby providing a physical basis for understanding mixed corruption mechanisms.
- We propose a generative framework that combines fixed one-level wavelet representations with an inter-band dual-stream probabilistic bridge, reformulating 3D completion as conditional probabilistic inference in the wavelet representation space and providing a physically interpretable inference path for generative 3D completion.
- Experiments show that PRISM achieves significant performance improvements over existing methods under mixed-corruption scenarios involving diverse noise, complex occlusion, and background interference, validating the effectiveness of the framework for robust 3D completion.

2 Related Work

2.1 Geometric Completion and Robustness

SnowflakeNet progressively refines local geometry Xiang et al. (2021), while PoinTr and AdaPoinTr use geometry-aware attention to capture structural relationships Yu et al. (2021; 2023). IF-Nets and Convolutional Occupancy Networks combine local features with continuous implicit surfaces Chibane et al. (2020); Peng et al. (2020). For unseen categories, PatchComplete learns reusable multi-resolution patch priors Rao et al. (2022). Robustness requires additional attention to observation quality: score-based denoising estimates clean-surface directions Luo & Hu (2021), CS-Net couples completion with outlier segmentation Ma et al. (2022), and DWCNet jointly addresses denoising and completion on mixed-corruption inputs Tesema et al. (2025). RealPC further documents the gap between conventional benchmarks and real-world acquisition Pathak et al. (2025). These studies motivate explicit treatment of corruption alongside shape priors.

2.2 Generative Completion and Diffusion Bridges

Generative shape priors accommodate ambiguity in unobserved regions; WaLa uses wavelet-based latent diffusion for 3D generation Sanghi et al. (2024). Representative approaches include SDF-StyleGAN Zheng et al. (2022), continuous generative cellular automata Zhang et al. (2022), and the discrete shape representations of AutoSDF and ShapeFormer Mittal et al. (2022); Yan et al. (2022). Building on denoising diffusion Ho et al. (2020), PDR combines conditional point generation with refinement Lyu et al. (2021), DiffComplete injects spatially aligned, hierarchical observation features Chu et al. (2023), and SC-Diff performs multimodally conditioned completion in a compressed latent space Schaefer et al. (2024). Diffusion Schrödinger bridges model transport between prescribed endpoint distributions De Bortoli et al. (2021); I^2SB applies paired bridges to restoration Liu et al. (2023), P2P-Bridge to point-cloud denoising Vogel et al. (2024), and BridgeShape to latent shape completion Kong et al. (2025). PRISM follows this paired-endpoint setting while retaining explicit low/high-frequency streams, asymmetric interaction, and persistent observation conditioning.

3 Method

PRISM combines operator-based spectral analysis, a Wavelet VQ-VAE, and an interacting dual-stream probabilistic bridge.

3.1 Problem Formulation & Physics-Informed Representation

Let $S \in \mathcal{F} = L^2(\Omega)$ represent a shape signal on a bounded reconstruction domain Ω. We model the observation as

$$S_{in} = \mathcal{G}(S_{GT}, \Theta) = (\mathcal{N}_{sig} \circ \mathcal{K}_{occ} \circ \mathcal{K}_{ext} \circ \mathcal{K}_{geom})(S_{GT}), \tag{1}$$

where the operators denote sensor noise, occlusion, external interference, and geometric deformation (Appendix A). Because occlusion loses information, restoration is an ill-posed Bayesian inverse

problem with MAP objective

$$\widehat{S}_{GT} = \underset{S \in \mathcal{F}}{\operatorname{argmax}} P(S_{in} \mid S) P(S). \tag{2}$$

Physical objects consist of smooth surfaces and sharp boundaries (singularities). The coupled likelihood and geometric prior motivate a representation that distinguishes coarse structure from local detail. A fixed one-level 3D DWT $\mathcal{W}$ provides an explicit representation $C = \mathcal{W}S = \{C_{low}, C_{high}\}$: one approximation subband describes coarse structure and seven detail subbands describe local geometry. This Low/High split is cross-band, not a decomposition across multiple wavelet levels.

3.2 Spectral Analysis of Corruption Operators

The four operators abstract the corruption families in Sec. 3.1; Appendix A gives their signal-domain definitions. For each operator $\mathcal{O}$, define the operator-induced coefficient-space deviation

$$\Delta C_{\mathcal{O}}(S) = \mathcal{W}(\mathcal{O}(S)) - \mathcal{W}(S).$$

This identity assumes neither physical linearity nor independence among mechanisms. The following properties distinguish exact identities, assumption-dependent bounds, and modeling motivations; supporting analysis is given in Appendix B.

Proposition 1 (Bounded Response to Geometric Deformation). Define $\Delta C_{geom} = \mathcal{W}\mathcal{K}_{geom}S - \mathcal{W}S$. Since the fixed DWT is bounded and linear, $\|\Delta C_{geom}\| \leq \|\mathcal{W}\| \, \|\mathcal{K}_{geom}S - S\|$. Exactly by definition,

$$\mathcal{W}\mathcal{K}_{geom}S = \mathcal{W}S + \Delta C_{geom} = \{C_{low} + \Delta C^{low}_{geom}, C_{high} + \Delta C^{high}_{geom}\}. \tag{3}$$

Under small deformation and sufficient local signal regularity, the deviation admits a controlled first-order characterization (Appendix B); the identity itself requires no small-deformation approximation.

Proposition 2 (Energy Distribution of White Noise). For the signal-noise operator $\mathcal{N}_{sig}S = S + \epsilon_{noise}$, linearity of the wavelet transform strictly preserves additivity in coefficient space: $\mathcal{W}\mathcal{N}_{sig}S = \mathcal{W}S + \mathcal{W}\epsilon_{noise}$. For sampled white noise, an orthonormal DWT preserves coefficient variance and decorrelates distinct coefficients before branch rescaling. Writing the Low/High projections of $\mathcal{W}\epsilon_{noise}$ as ΔC^{low}_{noise} and ΔC^{high}_{noise} gives exactly

$$\mathcal{W}\mathcal{N}_{sig}S = \mathcal{W}S + \{\Delta C^{low}_{noise}, \Delta C^{high}_{noise}\} = \{C_{low} + \Delta C^{low}_{noise}, C_{high} + \Delta C^{high}_{noise}\}. \tag{4}$$

For white noise under an orthonormal one-level 3D DWT, the seven High subbands carry seven times the expected total squared noise energy of the single Low subband before branch rescaling. This is an aggregate statistical property, not a per-sample inequality or exact separation of noise from geometry.

Proposition 3 (Cross-band Response of Structural Corruption). For an input signal S, define the structural deviation $\Delta C_{struc} = \mathcal{W}(\mathcal{K}_{occ} \circ \mathcal{K}_{ext})(S) - \mathcal{W}S$. Occlusion and interference can introduce discontinuities or creases with broadband wavelet responses. The same structural event can therefore affect both Low and High projections:

$$\mathcal{W}(\mathcal{K}_{occ} \circ \mathcal{K}_{ext})(S) = \mathcal{W}S + \Delta C_{struc} = \{C_{low} + \Delta C^{low}_{struc}, C_{high} + \Delta C^{high}_{struc}\}. \tag{5}$$

These projections provide complementary cross-band evidence; a broadband response alone does not establish statistical dependence.

Proposition 4 (Complementary Cross-band Evidence for Structural Restoration). Low and High responses to the same structural disturbance preserve complementary geometric information. High-frequency edges and discontinuity locations can thus provide spatial evidence for recovering corrupted low-frequency structure. This motivates retaining both bands in the conditional model,

$$P(C^{low}_{in} \mid C_{low}, C_{high}), \tag{6}$$

rather than restricting it to $P(C^{low}_{in} \mid C_{low})$. This is a modeling choice motivated by the spectral analysis, not a proved probability inequality.

For the composite process, let $S^{(0)} = S_{GT}$ and let $S^{(1)}, \ldots, S^{(4)} = S_{in}$ denote its successive states. Define $\Delta C_k = \mathcal{W}S^{(k)} - \mathcal{W}S^{(k-1)}$. With $\Delta C_{geom} = \Delta C_1$, $\Delta C_{struc} = \Delta C_2 + \Delta C_3$, and $\Delta C_{noise} = \Delta C_4$, telescoping gives the coefficient-space deviation decomposition:

$$C^{low}_{in} = C^{low}_{GT} + \Delta C^{low}_{noise} + \Delta C^{low}_{struc} + \Delta C^{low}_{geom}, \tag{7}$$

$$C^{high}_{in} = C^{high}_{GT} + \Delta C^{high}_{noise} + \Delta C^{high}_{struc} + \Delta C^{high}_{geom}. \tag{8}$$

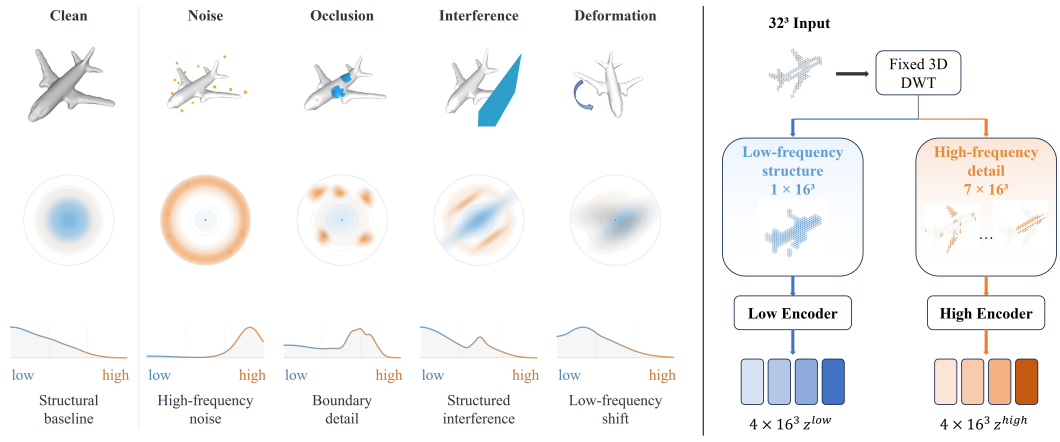

Figure 2: (a) Schematic wavelet-domain corruption signatures. (b) Stage 1: wavelet latent encoding.

The plus signs account for sequential coefficient-space deviations; they imply neither linearity of the composite physical process nor independence among corruptions. Propositions 1–3 characterize these terms: bounded geometric deviation under suitable regularity, exact additive noise under the linear DWT, and broadband structural responses. This is the scope of the "spectral decoupling mechanism" and motivates the subsequent inter-band interaction. Figure 2(a) schematically illustrates these band-wise responses to aid interpretation of the analysis.

3.3 PRISM Framework: Inter-band Dual-Stream Probabilistic Bridge

PRISM learns a complete-shape representation and a paired conditional bridge, using cross-band evidence to couple structure and detail recovery.

3.3.1 Conditional Decomposition

The expanded MAP formulation and Terms A/B are given in Appendix B.6. The posterior admits the exact chain-rule factorization

$$P(C_{low}, C_{high} \mid C_{in}) = P(C_{low} \mid C_{in}) P(C_{high} \mid C_{low}, C_{in}). \tag{9}$$

This ordering motivates low-frequency structure modeling followed by conditional detail modeling. It organizes the conditional model rather than prescribing independent band-wise optimization or strictly sequential network execution. The Knothe–Rosenblatt construction provides a related triangular transport perspective under absolute continuity and a fixed variable ordering (Appendix B). Our interacting network is not constrained to realize the unique KR map: Proposition 4 motivates retaining High-to-Low feedback alongside Low-to-High guidance.

3.3.2 Stage 1: Wavelet VQ-VAE Representation

A fixed one-level 3D DWT splits the complete shape into $(C_{GT}^{low}, C_{GT}^{high}) = \mathcal{W}S_{GT}$. Separate convolutional encoders and vector quantization learn compact latent representations Rombach et al. (2022), as shown in Figure 2(b).

The decoder reconstructs complete geometry through the inverse wavelet transform. This stage performs only self-reconstruction of complete shapes, learning a compact representation of complete geometry without directly performing completion. Subsequent generation operates on the continuous pre-quantization latents to preserve a continuous transport space. Thus, Stage 1 learns the complete-shape coordinate system, while Stage 2 learns completion within that representation.

3.3.3 Stage 2: Dual-Stream Probabilistic Bridge

Figure 3 shows the dual-stream bridge and its persistent observation conditioning.

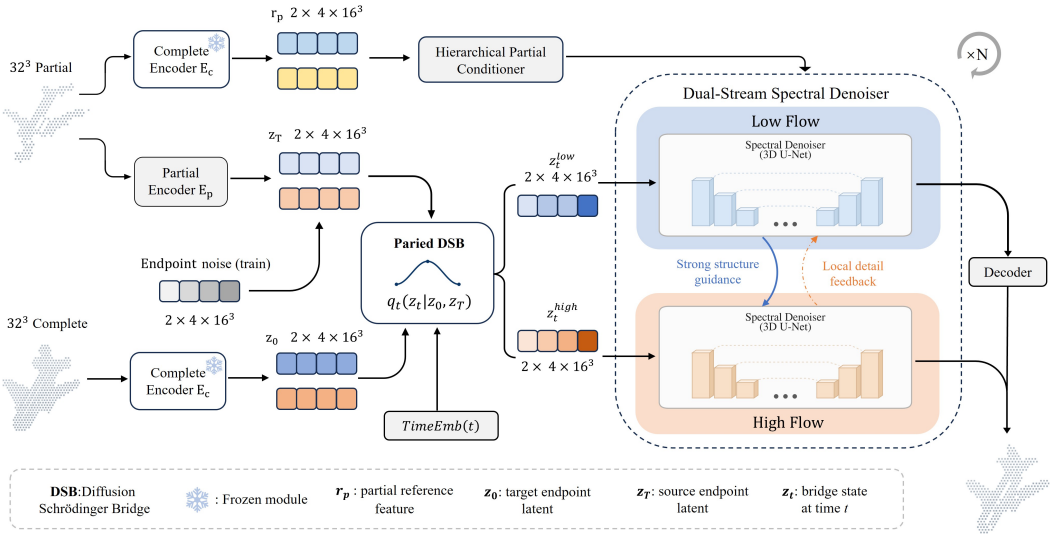

Figure 3: PRISM's dual-stream bridge combines asymmetric spectral interaction with persistent partial-reference conditioning.

A frozen Complete-VAE encoder defines the target $z_0 = (z_0^{low}, z_0^{high})$, while a trainable Partial Encoder defines the source $z_T = (z_T^{low}, z_T^{high})$. Following paired restoration bridges Liu et al. (2023); Kong et al. (2025), the Partial Encoder is initialized from the Complete-VAE encoder and jointly optimized with the bridge; Appendix D.2 tests source and initialization alternatives.

Given paired endpoints z_0 and z_T, we construct an intermediate state under the paired DSB schedule:

$$z_t^b = \mu_0(t) z_0^b + \mu_T(t) \tilde{z}_T^b, \qquad b \in \{low, high\}, \tag{10}$$

where $\tilde{z}_T^b = z_T^b + \eta_b$ is the source latent after training-time endpoint perturbation, with $\eta_b \sim \mathcal{N}(0, \sigma_{end,b}^2 I)$. The weights satisfy $\mu_0(t) = 1 - \mu_T(t)$, with $\mu_T(t)$ determined by the cumulative forward and backward variances:

$$\mu_T(t) = \frac{\sigma_{fwd}^2(t)}{\sigma_{fwd}^2(t) + \sigma_{bwd}^2(t)}. \tag{11}$$

The model takes z_t^b as input and predicts the normalized bridge residual relative to the target endpoint, supervised by

$$\epsilon_t^{b,*} = \frac{z_t^b - z_0^b}{\sigma_{fwd}(t)}, \qquad b \in \{low, high\}, \tag{12}$$

and reconstructs the complete endpoint as

$$z_0^b = z_t^b - \sigma_{fwd}(t) \epsilon_\theta^b. \tag{13}$$

We model the conditional bridge process separately in the low- and high-frequency streams and exchange cross-band information through inter-band interaction. To match the probabilistic decomposition in Sec. 3.3.1, where low-frequency structure is primary and high-frequency detail is conditionally refined, we use asymmetric interaction: low-to-high exchange receives a larger weight, while high-to-low exchange provides only weaker local structural feedback.

In addition, to prevent incomplete-observation information from being compressed only into the source endpoint z_T, we extract reference features from the incomplete input using the frozen complete-shape encoder,

$$r_P = \mathrm{sg}(E_C(S_{in})), \tag{14}$$

where E_C denotes the frozen Complete-VAE pre-quantization encoder and sg denotes stop-gradient. After multi-resolution pooling and projection, r_P is persistently injected into both streams as zero-initialized additive residuals Zhang et al. (2023).

Stream 1: Structure Stream (Term A). This stream predicts the low-frequency bridge residual ϵ_θ^{low} from the low-frequency bridge state z_t^{low} and uses it to recover the target low-frequency latent z_0^{low}, thereby reconstructing the complete geometric skeleton. Motivated by the complementary cross-band evidence in Proposition 4, low-frequency structure recovery uses local geometric information from the high-frequency stream. The Structure Stream therefore receives local structural information from the high-frequency branch at each resolution and uses it as an auxiliary condition when updating its own low-frequency features:

$$z_t^{low} \to \mathrm{ResBlocks}_t^{low} \to \alpha_{H\to L}\,\mathrm{CrossBand}_{H\to L} \to \mathrm{PartialCond}_{low} \to \epsilon_\theta^{low}. \tag{15}$$

Here, $\mathrm{CrossBand}_{H\to L}$ transfers High-to-Low features with weight $\alpha_{H\to L} = 0.25$, while $\mathrm{PartialCond}_{low}$ injects the corresponding-resolution observation reference. This limits the feedback strength so that it mainly provides auxiliary structural correction.

Stream 2: Detail Stream (Term B). This stream predicts the high-frequency bridge residual ϵ_θ^{high} from z_t^{high}, recovers the target high-frequency latent z_0^{high}, and generates local boundaries and details under the guidance of low-frequency geometry. According to the conditional decomposition of Term B, high-frequency details depend not only on their own band state but also on low-frequency geometry and incomplete observations. The Detail Stream therefore receives structural information from the low-frequency branch at each resolution while fusing the corresponding high-frequency reference features to keep generated details consistent with observed geometry:

$$z_t^{high} \to \mathrm{ResBlocks}_t^{high} \to \alpha_{L\to H}\,\mathrm{CrossBand}_{L\to H} \to \mathrm{PartialCond}_{high} \to \epsilon_\theta^{high}. \tag{16}$$

Here, $\alpha_{L\to H} = 1$ makes Low-to-High guidance dominant, and $\mathrm{PartialCond}_{high}$ injects multi-resolution observation references. Thus, high-frequency generation is constrained by the complete geometric skeleton while retaining direct awareness of observed details, corresponding to conditional detail recovery in Term B. Joint spectral attention at the compact bottleneck exchanges global information between the two streams.

4 Experiments

4.1 Datasets and Evaluation Metrics

Datasets. We evaluate on two 3D completion benchmarks covering synthetic shapes and real-world scans. **3D-EPN** Dai et al. (2017b), as a classic shape completion benchmark, contains 8 known categories from ShapeNet Chang et al. (2015). This dataset provides virtual scan inputs with varying degrees of incompleteness for each instance. This benchmark tests reconstruction fidelity and preservation of core geometric structures. **PatchComplete** Rao et al. (2022) focuses on evaluating the model's zero-shot generalization capability on completely unseen categories, and is divided into two parts: synthetic ShapeNet data and real-world ScanNet data Dai et al. (2017a). The ScanNet data, in particular, provides scanned objects extracted from real-world bounding boxes Rao et al. (2022); Avetisyan et al. (2019), which are often accompanied by severe sensor noise, background clutter, and extreme sparsity. It serves as our core validation platform for assessing the model's robust noise resistance and restoration capabilities under highly coupled and adverse observation environments.

Evaluation Metrics. To align with current state-of-the-art generative methods, we adopt standard metrics to quantitatively evaluate the quality of completion and restoration. For the 3D-EPN dataset, we calculate the average l_1-error between the predicted truncated unsigned distance field (TUDF) voxels and the ground truth. For the PatchComplete evaluation Rao et al. (2022), we employ the l_1 Chamfer Distance (CD) and Intersection over Union (IoU) to measure the geometric similarity and spatial occupancy accuracy of the predicted shapes.

4.2 Experimental Results

3D-EPN. Table 1 reports known-category completion results. PRISM achieves an average l_1-error of 0.035, compared with 0.039 for BridgeShape, with the lowest or tied-lowest error in every category.

PatchComplete-ShapeNet. Table 2 evaluates generalization to eight unseen ShapeNet categories. PRISM achieves the best average CD and IoU (3.60 and 73.0), improving on BridgeShape (4.06 and 70.1). Category-wise results show that IF-Nets retains the lowest Basket CD.

Table 1: Completion on 3D-EPN (l_1-error↓). Best in bold; second best underlined.

l_1-err. (↓)	Chair	Table	Sofa	Lamp	Plane	Car	Dresser	Watercraft	Avg.
3D-EPN(2017) Dai et al. (2017b)	0.418	0.377	0.392	0.388	0.421	0.259	0.381	0.356	0.374
ConvONet(2020) Peng et al. (2020)	0.210	0.247	0.254	0.234	0.185	0.195	0.250	0.184	0.220
SDF-StyleGAN(2022) Zheng et al. (2022)	0.321	0.256	0.289	0.280	0.295	0.224	0.273	0.282	0.278
cGCA(2022) Zhang et al. (2022)	0.174	0.212	0.179	0.239	0.170	0.161	0.204	0.143	0.185
AutoSDF(2022) Mittal et al. (2022)	0.201	0.258	0.226	0.275	0.184	0.187	0.248	0.157	0.217
ShapeFormer(2022) Yan et al. (2022)	0.104	0.175	0.133	0.176	0.136	0.127	0.157	0.119	0.141
PatchComplete(2022) Rao et al. (2022)	0.134	0.095	0.084	0.087	0.061	0.053	0.134	0.058	0.088
DiffComplete(2023) Chu et al. (2023)	0.070	0.073	0.061	0.059	0.015	0.025	0.086	0.031	0.053
BridgeShape(2025) Kong et al. (2025)	0.055	0.059	0.047	0.038	**0.012**	**0.023**	0.055	0.022	0.039
PRISM (Ours)	**0.054**	**0.050**	**0.038**	**0.034**	**0.012**	**0.023**	**0.050**	**0.018**	**0.035**

Table 2: Unseen-category completion on ShapeNet (CD↓, IoU↑; both $\times 100$). Best in bold; second best underlined.

Method	Bag		Lamp		Bathtub		Bed		Basket		Printer		Laptop		Bench		Avg.	
	CD↓	IoU↑	CD	IoU	CD	IoU	CD	IoU	CD	IoU	CD	IoU	CD	IoU	CD	IoU	CD	IoU
3D-EPN Dai et al. (2017b)	5.01	73.8	8.07	47.2	4.21	57.9	5.84	58.4	7.90	54.0	5.15	73.6	3.90	62.0	4.54	48.3	5.58	59.4
IF-Nets Chibane et al. (2020)	4.77	69.8	5.70	50.8	4.72	55.0	5.34	60.7	**4.44**	50.2	5.83	70.5	6.47	58.3	5.03	49.7	5.29	58.1
Auto-SDF Mittal et al. (2022)	5.81	56.3	6.57	39.1	5.17	41.0	6.01	44.6	6.70	39.8	7.52	49.9	4.81	51.1	4.31	39.5	5.86	45.2
ConvONet Peng et al. (2020)	5.10	70.8	5.42	52.6	4.96	60.4	5.42	63.2	6.16	54.6	5.56	72.1	4.78	57.3	4.69	49.6	5.26	60.1
PatchComplete Rao et al. (2022)	3.94	77.6	4.68	56.4	3.78	66.3	4.49	66.8	5.15	61.0	4.63	77.6	3.77	63.8	3.70	53.9	4.27	65.4
DiffComplete Chu et al. (2023)	3.86	78.3	4.80	57.9	3.52	68.9	4.16	67.1	4.94	65.5	4.40	76.8	3.52	67.4	3.56	58.2	4.10	67.5
BridgeShape Kong et al. (2025)	3.70	80.0	4.94	62.7	3.41	69.8	4.32	69.6	5.50	65.1	4.09	81.4	3.13	72.7	3.38	59.1	4.06	70.1
PRISM (Ours)	**3.38**	**82.4**	**4.30**	**63.6**	**3.31**	**71.6**	**3.67**	**75.0**	4.56	**69.9**	**3.68**	**83.9**	**2.82**	**76.9**	**3.08**	**60.4**	**3.60**	**73.0**

PatchComplete-ScanNet. Table 3 evaluates unseen real-world ScanNet objects. PRISM achieves the best average CD and IoU (6.95 and 53.9), compared with 6.99 and 53.5 for BridgeShape. Performance varies by category, with other methods retaining advantages on some metrics.

Table 3: Unseen-category completion on ScanNet (CD↓, IoU↑; both $\times 100$). Best in bold; second best underlined.

Method	Bag		Lamp		Bathtub		Bed		Basket		Printer		Avg.	
	CD↓	IoU↑	CD	IoU	CD	IoU	CD	IoU	CD	IoU	CD	IoU	CD	IoU
3D-EPN Dai et al. (2017b)	8.83	53.7	14.3	20.7	7.56	41.0	7.76	47.8	7.74	36.5	8.36	63.0	9.09	44.0
IF-Nets Chibane et al. (2020)	8.96	44.2	10.2	24.9	7.19	39.5	8.24	44.9	6.74	42.7	8.28	60.7	8.26	42.6
Auto-SDF Mittal et al. (2022)	9.30	48.7	11.2	24.4	7.84	36.6	7.91	38.0	7.54	36.1	9.66	49.9	8.90	38.9
ConvONet Peng et al. (2020)	9.12	52.5	9.83	20.3	7.93	41.2	8.14	41.6	7.39	37.0	7.62	64.9	8.34	42.9
PatchComplete Rao et al. (2022)	8.23	58.3	9.42	28.4	6.77	48.0	7.24	48.4	6.60	45.5	6.84	70.5	7.52	49.8
DiffComplete Chu et al. (2023)	**7.05**	48.5	**6.84**	30.5	8.22	48.5	7.20	46.6	7.42	**59.2**	6.36	**74.5**	7.18	51.3
BridgeShape Kong et al. (2025)	7.67	**61.2**	8.07	36.3	6.28	50.9	**6.87**	51.3	6.20	50.2	6.83	71.0	6.99	53.5
PRISM (Ours)	8.15	58.3	8.24	**38.2**	**6.03**	**52.2**	6.89	**51.7**	**6.07**	49.5	**6.31**	73.2	**6.95**	**53.9**

4.3 Ablation Analysis

We conduct a series of ablation experiments to verify the contributions of the core architectural and training components. Unless otherwise specified, all variants are evaluated at the common 100k ShapeNet test point with EMA weights, three reverse steps, and zero evaluation endpoint noise. Table 4 organizes the controls directly by the corresponding design claim rather than by internal experiment names.

The first group of controls examines the spectral architecture. Early fusion concatenates the low- and high-frequency latents into a single stream, while the no-interaction variant preserves two streams but removes cross-stream communication. We further isolate the compact-bottleneck joint spectral attention while leaving the remaining cross-stream interaction unchanged. The second group evaluates the two training and conditioning components directly by replacing loss-aware bridge-time allocation with uniform sampling, removing persistent partial-reference conditioning, or removing both. Together, these controls separate the contributions of explicit dual-stream modeling, cross-stream interaction, global joint spectral interaction, adaptive bridge-time allocation, and persistent observation evidence. Table 7 further examines the High→Low interaction strength. Source-endpoint controls, including the Gaussian-source and Partial-Encoder alternatives, are also reported separately in Appendix D.2.

Table 4: Architectural and training ablations on ShapeNet. Best in bold; second best underlined.

Variant	Dual Stream	Cross-Stream Interaction	Joint Attention	Loss-Aware Sampling	Persistent Conditioning	CD×100 ↓	IoU×100 ↑
Early-Fusion	–	–	–	✓	✓	3.786	71.874
w/o Cross-Stream Interaction	✓	–	–	✓	✓	3.727	70.842
w/o Joint Attention	✓	✓	–	✓	✓	<u>3.631</u>	72.584
w/o Loss-Aware Sampling	✓	✓	✓	–	✓	3.669	72.001
w/o Persistent Conditioning	✓	✓	✓	✓	–	3.712	<u>72.600</u>
w/o Both	✓	✓	✓	–	–	3.718	72.142
PRISM (Ours)	✓	✓	✓	✓	✓	**3.599**	**72.952**

✓/–: enabled/disabled; "w/o Both" removes loss-aware sampling and persistent conditioning.

4.4 Visualization

Figure 4 compares PRISM with 3D-EPN and PatchComplete. The examples show structural completion and local-detail recovery, with remaining differences in fine surface geometry. Input* filtering affects display only; model inputs, predictions, and metrics are unchanged. Appendix H provides category-wise examples.

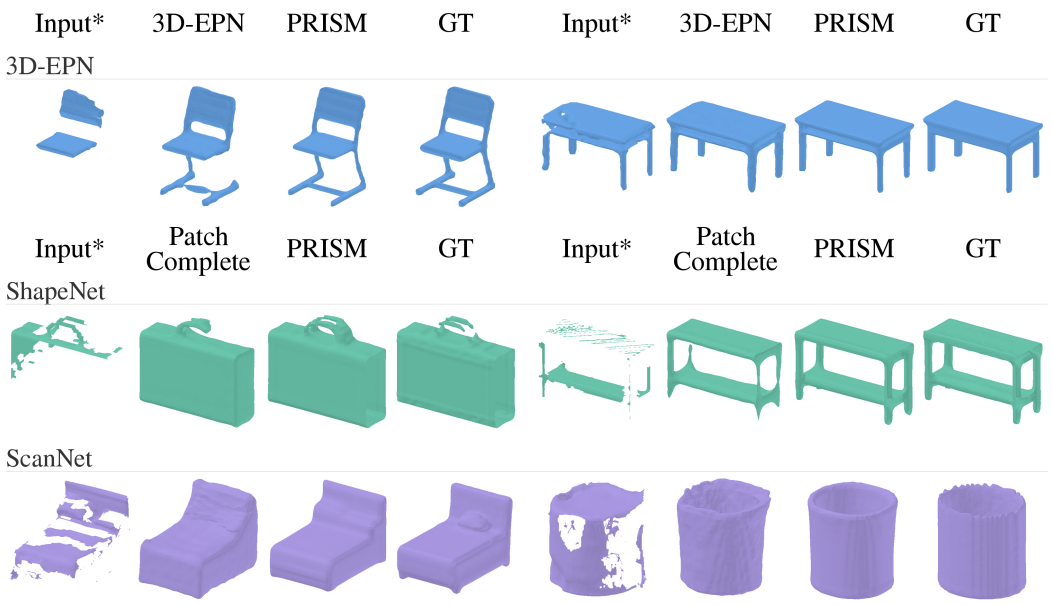

Figure 4: Qualitative completion on 3D-EPN, ShapeNet, and ScanNet. Baselines: 3D-EPN (top) and PatchComplete (middle and bottom).

5 Conclusion

In this paper, we propose PRISM, a physics-aware 3D restoration framework based on spectral analysis and a paired probabilistic bridge. A fixed one-level 3D DWT characterizes nonlinear corruptions through operator-induced coefficient deviations, while a Wavelet VQ-VAE provides compact complete-shape representations. The dual Low/High streams combine inter-band interaction with persistent observation conditioning. Experiments demonstrate the effectiveness of this physical inductive bias for restoration and provide an interpretable modeling perspective for generative inverse problems.

AI USE STATEMENT

In this work, we used generative AI tools to polish manuscript language, search for related work, and assist in generating portions of the research code. All AI-assisted code was reviewed, tested, and validated by the authors to ensure its correctness. We did not use generative AI tools to generate the core scientific conclusions or experimental results. We reviewed all AI-assisted content and manually verified the factual claims, mathematical statements, and experimental results presented in the paper. The authors take full responsibility for the final content of this work and its correctness.

REPRODUCIBILITY STATEMENT

Sections 3–4 provide a complete theoretical and architectural description of PRISM, including the corruption model, the dual-stream paired probabilistic bridge, and the Wavelet VQ-VAE with a fixed one-level 3D DWT. All datasets used in this work (3D-EPN, ShapeNet, and PatchComplete-ScanNet) are publicly available. Appendices E and F detail the training configurations, implementation details, computational resources, hyperparameters, and inference settings; Section 4 provides the full evaluation protocol and benchmark settings. To enable independent reproduction of the main experimental results, we will publicly release the complete source code, training and inference configurations, pretrained model checkpoints, and evaluation scripts.

REFERENCES

Armen Avetisyan, Manuel Dahnert, Angela Dai, Manolis Savva, Angel X. Chang, and Matthias Nießner. Scan2CAD: Learning CAD model alignment in RGB-D scans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. URL https://arxiv.org/abs/1811.11187.

Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. ShapeNet: An information-rich 3D model repository. *arXiv preprint arXiv:1512.03012*, 2015. URL https://arxiv.org/abs/1512.03012.

Julian Chibane, Thiemo Alldieck, and Gerard Pons-Moll. Implicit functions in feature space for 3D shape reconstruction and completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. URL https://arxiv.org/abs/2003.01456.

Ruihang Chu, Enze Xie, Shentong Mo, Zhenguo Li, Matthias Nießner, Chi-Wing Fu, and Jiaya Jia. DiffComplete: Diffusion-based generative 3D shape completion. *Advances in neural information processing systems*, 36:75951–75966, 2023. URL https://arxiv.org/abs/2306.16329.

Matthew S. Crouse, Robert D. Nowak, and Richard G. Baraniuk. Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Signal Processing*, 46(4):886–902, 1998. doi: 10.1109/78.668544. URL https://repository.rice.edu/items/5b84d16a-255a-4f1c-8870-0840f5ace39c.

Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. ScanNet: Richly-annotated 3D reconstructions of indoor scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017a. URL https://arxiv.org/abs/1702.04405.

Angela Dai, Charles Ruizhongtai Qi, and Matthias Nießner. Shape completion using 3D-encoder-predictor CNNs and shape synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5868–5877, 2017b. URL https://openaccess.thecvf.com/content_cvpr_2017/papers/Dai_Shape_Completion_Using_CVPR_2017_paper.pdf.

Ingrid Daubechies. *Ten Lectures on Wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, 1992. doi: 10.1137/1.9781611970104. URL https://doi.org/10.1137/1.9781611970104.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems*, volume 34, pp. 17695–17709, 2021. URL https://arxiv.org/abs/2106.01357.

David L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41 (3):613–627, 1995. doi: 10.1109/18.382009. URL https://www.fceia.unr.edu.ar/~jcgomez/wavelets/Donoho_1995.pdf.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://arxiv.org/abs/2006.11239.

Jingyu Hu, Ka-Hei Hui, Zhengzhe Liu, Ruihui Li, and Chi-Wing Fu. Neural wavelet-domain diffusion for 3D shape generation, inversion, and manipulation. *ACM Transactions on Graphics*, 43(2), 2024. doi: 10.1145/3635304. URL https://doi.org/10.1145/3635304.

Ka-Hei Hui, Aditya Sanghi, Arianna Rampini, Kamal Rahimi Malekshan, Zhengzhe Liu, Hooman Shayani, and Chi-Wing Fu. Make-A-Shape: A ten-million-scale 3D shape model. In *Forty-first International Conference on Machine Learning*, 2024. URL https://proceedings.mlr.press/v235/hui24a.html.

Dequan Kong, Honghua Chen, Zhe Zhu, and Mingqiang Wei. BridgeShape: Latent diffusion schrödinger bridge for 3D shape completion. *arXiv preprint arXiv:2506.23205*, 2025. URL https://arxiv.org/abs/2506.23205. Conference version: Proceedings of the AAAI Conference on Artificial Intelligence, 40(7):5726–5734, 2026, doi:10.1609/aaai.v40i7.37493.

Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos Theodorou, Weili Nie, and Anima Anandkumar. I^2SB: Image-to-image schrödinger bridge. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 22042–22062, 2023. URL https://proceedings.mlr.press/v202/liu23ai.html.

Shitong Luo and Wei Hu. Score-based point cloud denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4583–4592, 2021. URL https://openaccess.thecvf.com/content/ICCV2021/html/Luo_Score-Based_Point_Cloud_Denoising_ICCV_2021_paper.html.

Zhaoyang Lyu, Zhifeng Kong, Xudong Xu, Liang Pan, and Dahua Lin. A conditional point diffusion-refinement paradigm for 3d point cloud completion. *arXiv preprint arXiv:2112.03530*, 2021. URL https://arxiv.org/abs/2112.03530.

Changfeng Ma, Yang Yang, Jie Guo, Chongjun Wang, and Yanwen Guo. Completing partial point clouds with outliers by collaborative completion and segmentation. *arXiv preprint arXiv:2203.09772*, 2022. URL https://arxiv.org/abs/2203.09772.

Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012. URL https://www.di.ens.fr/~mallat/College/CPAM-Mallat-Scat.pdf.

Stéphane Mallat and Wen Liang Hwang. Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, 38(2):617–643, 1992. doi: 10.1109/18.119727. URL https://www.di.ens.fr/~mallat/papiers/MallatSingDet92.pdf.

Stéphane G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989. URL https://www.di.ens.fr/~mallat/papiers/MallatTheory89.pdf.

Youssef Marzouk, Tarek Moselhy, Matthew Parno, and Alessio Spantini. Sampling via measure transport: An introduction. In *Handbook of Uncertainty Quantification*. Springer, 2017. doi: 10.1007/978-3-319-12385-1_23. URL https://doi.org/10.1007/978-3-319-12385-1_23.

Paritosh Mittal, Yen-Chi Cheng, Maneesh Singh, and Shubham Tulsiani. AutoSDF: Shape priors for 3D completion, reconstruction and generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 306–315, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Mittal_AutoSDF_Shape_Priors_for_3D_Completion_Reconstruction_and_Generation_CVPR_2022_paper.html.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 8162–8171, 2021. URL https://proceedings.mlr.press/v139/nichol21a.html.

Stuti Pathak, Prashant Kumar, Dheeraj Baiju, Nicholus Mboga, Gunther Steenackers, and Rudi Penne. Revisiting point cloud completion: Are we ready for the real-world? In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025. URL https://openaccess.thecvf.com/content/ICCV2025/papers/Pathak_Revisiting_Point_Cloud_Completion_Are_We_Ready_For_The_Real-World_ICCV_2025_paper.pdf.

Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *European Conference on Computer Vision*, 2020. URL https://arxiv.org/abs/2003.04618.

Hao Phung, Quan Dao, and Anh Tran. Wavelet diffusion models are fast and scalable image generators. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/papers/Phung_Wavelet_Diffusion_Models_Are_Fast_and_Scalable_Image_Generators_CVPR_2023_paper.pdf.

Yuchen Rao, Yinyu Nie, and Angela Dai. PatchComplete: Learning multi-resolution patch priors for 3D shape completion on unseen categories. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://arxiv.org/abs/2206.04916.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. URL https://arxiv.org/abs/2112.10752.

Murray Rosenblatt. Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472, 1952. doi: 10.1214/aoms/1177729394. URL https://doi.org/10.1214/aoms/1177729394.

Aditya Sanghi, Aliasghar Khani, Pradyumna Reddy, Arianna Rampini, Derek Cheung, Kamal Rahimi Malekshan, Kanika Madan, and Hooman Shayani. Wavelet latent diffusion (WaLa): Billion-parameter 3D generative model with compact wavelet encodings. *arXiv preprint arXiv:2411.08017*, 2024. URL https://arxiv.org/abs/2411.08017.

Simon Schaefer, Juan D Galvis, Xingxing Zuo, and Stefan Leutenegger. SC-Diff: 3D shape completion with latent diffusion models. *arXiv preprint arXiv:2403.12470*, 2024. URL https://arxiv.org/abs/2403.12470.

Eero P Simoncelli. Modeling the joint statistics of images in the wavelet domain. In *Wavelet Applications in Signal and Image Processing VII*, volume 3813, pp. 188–195. SPIE, 1999. URL https://www.cns.nyu.edu/~eero/ABSTRACTS/simoncelli99b-abstract.html.

Andrew M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 2010. doi: 10.1017/S0962492910000061. URL https://doi.org/10.1017/S0962492910000061.

Keneni W. Tesema, Lyndon Hill, Mark W. Jones, and Gary K. L. Tam. Denoising-While-Completing Network (DWCNet): Robust point cloud completion under corruption. *Computers & Graphics*, 132:104401, 2025. doi: 10.1016/j.cag.2025.104401. URL https://doi.org/10.1016/j.cag.2025.104401.

Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://arxiv.org/abs/1711.00937.

Mathias Vogel, Keisuke Tateno, Marc Pollefeys, Federico Tombari, Marie-Julie Rakotosaona, and Francis Engelmann. P2P-Bridge: Diffusion bridges for 3D point cloud denoising. In *European Conference on Computer Vision*, pp. 184–201. Springer, 2024. URL https://arxiv.org/abs/2408.16325.

Peng Xiang, Xin Wen, Yu-Shen Liu, Yan-Pei Cao, Pengfei Wan, Wen Zheng, and Zhizhong Han. SnowflakeNet: Point cloud completion by snowflake point deconvolution with skip-transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5499–5509, 2021. URL https://openaccess.thecvf.com/content/ICCV2021/html/Xiang_SnowflakeNet_Point_Cloud_Completion_by_Snowflake_Point_Deconvolution_With_Skip-Transformer_ICCV_2021_paper.html.

Xingguang Yan, Liqiang Lin, Niloy J Mitra, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. ShapeFormer: Transformer-based shape completion via sparse representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6239–6249, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Yan_ShapeFormer_Transformer-Based_Shape_Completion_via_Sparse_Representation_CVPR_2022_paper.html.

Xumin Yu, Yongming Rao, Ziyi Wang, Zuyan Liu, Jiwen Lu, and Jie Zhou. PoinTr: Diverse point cloud completion with geometry-aware transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 12498–12507, 2021. URL https://openaccess.thecvf.com/content/ICCV2021/papers/Yu_PoinTr_Diverse_Point_Cloud_Completion_With_Geometry-Aware_Transformers_ICCV_2021_paper.pdf.

Xumin Yu, Yongming Rao, Ziyi Wang, Jiwen Lu, and Jie Zhou. AdaPoinTr: Diverse point cloud completion with adaptive geometry-aware transformers. *arXiv preprint arXiv:2301.04545*, 2023. URL https://arxiv.org/abs/2301.04545.

Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. PCN: Point completion network. In *2018 international conference on 3D vision (3DV)*, pp. 728–737. IEEE, 2018. URL https://arxiv.org/abs/1808.00671.

Dongsu Zhang, Changwoon Choi, Inbum Park, and Young Min Kim. Probabilistic implicit scene completion. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=BnQhMqDfcKG.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3836–3847, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Zhang_Adding_Conditional_Control_to_Text-to-Image_Diffusion_Models_ICCV_2023_paper.html.

Xin-Yang Zheng, Yang Liu, Peng-Shuai Wang, and Xin Tong. SDF-StyleGAN: Implicit SDF-based StyleGAN for 3D shape generation. *Computer Graphics Forum*, 41(5):51–63, 2022. doi: 10.1111/cgf.14602. URL https://doi.org/10.1111/cgf.14602.

A Formalization of Physics-based Corruption Operators

To rigorously define and invert physical-world degradation processes, we move beyond the limitations of the original spatial domain and formulate the problem in a continuous function space. We model the observation process as a mapping from the ground-truth shape signal $S_{GT} \in \mathcal{F}$ to the observed signal $S_{in} \in \mathcal{F}$, where $\mathcal{F}$ is a Hilbert space such as $L^2(\Omega)$ on a bounded reconstruction domain $\Omega \subset \mathbb{R}^3$, containing the TUDF signals. This process is unified as a nonlinear forward operator $\mathcal{G}(\cdot, \Theta)$ composed of four basic physical operators, covering the corruption modes in the CPCCD dataset introduced by DWCNet Tesema et al. (2025).

1. Geometric Warp Operator ($\mathcal{K}_{geom}$). At the initial stage of physical imaging, an object may undergo a rigid transformation or scale change relative to the sensor coordinate system. We model this as an affine transformation operator $\mathcal{K}_{geom}$ acting on the spatial domain. Formally, for an affine transform $T : x \mapsto Ax + b$, the observed signal is the pull-back of the original signal under the coordinate transform:

$$S_{warped}(x) = \mathcal{K}_{geom}(S_{GT}, T)(x) = S_{GT}(T^{-1}x). \tag{17}$$

This operator directly corresponds to *Triaxial Rotation* (T_R) and *Isometric Scaling* (I_S) in CPCCD. In the wavelet representation, the signal deviation induced by a small geometric deformation appears as a bounded coefficient perturbation, providing a basis for controlled modeling in the spectral domain.

2. Structural Interference Operator ($\mathcal{K}_{ext}$). Real-world scans inevitably contain non-target objects. We model such external interference with the structural interference operator $\mathcal{K}_{ext}$. Let $S_{clutter}$ denote the TUDF representation of the interfering environment, such as walls, floors, or other objects. Since a TUDF is defined by the distance to the nearest surface, the union of geometric sets is equivalent to a pointwise minimum in function space:

$$S_{interfered}(x) = \mathcal{K}_{ext}(S_{GT}, S_{clutter})(x) = \min\big(S_{GT}(x), S_{clutter}(x)\big). \tag{18}$$

This operator describes *External Object Interference* (E_{OI}) and *Background Interference* in CPCCD, including *Walls* (BI_W) and *Floors* (BI_F). Mathematically, the $\min(\cdot, \cdot)$ operation introduces non-differentiable "creases," which constitute a major source of high-frequency artifacts.

3. Structural Occlusion Operator ($\mathcal{K}_{occ}$). Viewport limitations or mutual occlusion between objects cause partial loss of geometric information. We define the occlusion operator $\mathcal{K}_{occ}$ using a binary observation mask field $M(x) \in \{0, 1\}$. In the TUDF setting, regions where the mask is zero (the occluded region Ω_{occ}) are treated as invalid or infinitely distant observations, which in a truncated distance field is typically represented by the truncation threshold τ. For operator-algebra analysis, we formalize this as

$$S_{occluded}(x) = \mathcal{K}_{occ}(S_{interfered}, M)(x) = M(x)S_{interfered}(x) + (1 - M(x))\tau. \tag{19}$$

This operator corresponds to *Occlusion By Other Objects* (O_{BOO}) and sensor self-occlusion in CPCCD. In the wavelet domain, such spatial discontinuities produce broadband responses across frequency bands.

4. Signal Noise Operator ($\mathcal{N}_{sig}$). Finally, inherent uncertainty in sensor sampling introduces stochastic noise. We model it with the signal noise operator $\mathcal{N}_{sig}$. Physical *Dynamic Jitter* (D_{JT}) perturbs point coordinates; its induced signal-value deviation need not be white or Gaussian. For the noise analysis, we adopt an idealized additive Gaussian white-noise model $\epsilon_{noise} \sim \mathcal{N}(0, \sigma^2 I)$ on the sampled signal:

$$S_{in}(x) = \mathcal{N}_{sig}(S_{occluded})(x) = S_{occluded}(x) + \epsilon_{noise}(x). \tag{20}$$

This operator covers sensor measurement errors and trajectory noise induced by D_{JT}.

In summary, the observed corrupted signal S_{in} results from composing the above operators in the chosen modeling order:

$$S_{in} = \mathcal{G}(S_{GT}) = (\mathcal{N}_{sig} \circ \mathcal{K}_{occ} \circ \mathcal{K}_{ext} \circ \mathcal{K}_{geom})(S_{GT}). \tag{21}$$

This formalization transforms the complex and discrete *Random Combined Corruptions* (R_{CC}) described by DWCNet into a structured operator-algebra problem, providing a mathematical basis for the subsequent spectral analysis of degradation in the wavelet domain.

B THEORETICAL ANALYSIS OF SPECTRAL PROPERTIES

Classical background. Early completion methods used volumetric prediction and direct point-set decoding Dai et al. (2017b); Yuan et al. (2018). The theoretical background comprises Bayesian inverse problems Stuart (2010), multiresolution wavelets Mallat (1989); Daubechies (1992), singularity analysis and denoising Mallat & Hwang (1992); Donoho (1995), cross-scale statistical dependencies Crouse et al. (1998); Simoncelli (1999), and ordered conditional transport via the Knothe–Rosenblatt construction Rosenblatt (1952); Marzouk et al. (2017). Vector quantization provides the representation-learning background van den Oord et al. (2017). Scattering transforms add nonlinearities and cascades Mallat (2012); PRISM retains a fixed linear DWT.

Wavelet-based generation. Multiresolution wavelets separate coarse structure from local detail and support cross-scale modeling. PRISM uses a fixed linear DWT; Appendix B reviews the classical foundations and their distinction from scattering transforms. Wavelet-based generation includes WaveDiff for images Phung et al. (2023), coarse/detail wavelet-domain diffusion for 3D shapes Hu et al. (2024), and the compact representations of Make-A-Shape and WaLa Hui et al. (2024); Sanghi et al. (2024). Thus, wavelet representations and coarse-to-fine generation are established tools. PRISM combines them with operator-induced deviation analysis and interacting latent bridges for completion. The Knothe–Rosenblatt construction provides background for ordered conditional transport; it does not by itself prove that an interacting neural architecture realizes the unique monotone triangular map.

This section provides supporting mathematical analysis for the wavelet-domain corruption model. The fixed DWT defines operator-induced coefficient deviations exactly. We distinguish the assumptions used to characterize geometric deviations and white noise from the structural-response arguments motivating inter-band interaction. Classical scattering results do not directly establish deformation stability for the fixed DWT used here.

B.1 ANALYSIS OF PROPOSITION 1: BOUNDED RESPONSE TO GEOMETRIC DEFORMATION

Proposition B.1. *For a bounded linear DWT, the norm of the geometric coefficient-space deviation is bounded by the corresponding signal-space deviation.*

Proof. Let $L_u S(x) = S(x - u(x))$. Define the deviation exactly by

$$\Delta C_{geom} = \mathcal{W}(L_u S) - \mathcal{W}S, \qquad \mathcal{W}(L_u S) = \mathcal{W}S + \Delta C_{geom}. \tag{22}$$

Bounded linearity Daubechies (1992) gives

$$\|\Delta C_{geom}\|_2 \le \|\mathcal{W}\| \, \|L_u S - S\|_2. \tag{23}$$

For a Lipschitz signal with constant L on a neighborhood of the bounded domain Ω, $\|L_u S - S\|_2 \le L|\Omega|^{1/2}\|u\|_\infty$. On smooth regions with bounded second derivatives, a local first-order characterization is

$$\Delta C_{geom} = -\mathcal{W}(\nabla S \cdot u) + \mathcal{W}R_u, \qquad \|R_u\|_2 = O(\|u\|_\infty^2). \tag{24}$$

The first-order characterization requires this additional regularity; it is not asserted across TUDF singularities. The exact definition and operator-norm bound do not require a Taylor expansion. Neither statement makes the physical warp an additive physical process. □

B.2 ANALYSIS OF PROPOSITION 2: ENERGY DISTRIBUTION OF WHITE NOISE

Proposition B.2. *For sampled Gaussian white noise and an orthonormal DWT, noise remains additive and isotropic before branch rescaling. The High branch carries more expected total noise energy because it contains seven subbands.*

Proof. Let $\epsilon_{noise} \sim \mathcal{N}(0, \sigma^2 I)$ and let W be the orthonormal discrete analysis matrix, including its boundary convention. Linearity yields

$$W(S + \epsilon_{noise}) = WS + W\epsilon_{noise}. \tag{25}$$

Orthogonality preserves the covariance, and Gaussianity makes distinct orthogonal coefficients independent Donoho (1995); Daubechies (1992):

$$\mathrm{Cov}(W\epsilon_{noise}) = \sigma^2 WW^\top = \sigma^2 I, \qquad \mathrm{Cov}(\Delta C^{low}_{noise}, \Delta C^{high}_{noise}) = 0. \tag{26}$$

If each of the eight subbands contains n coefficients, then

$$\mathbb{E}\|\Delta C^{low}_{noise}\|_2^2 = n\sigma^2, \qquad \mathbb{E}\|\Delta C^{high}_{noise}\|_2^2 = 7n\sigma^2.$$

These statements concern the unscaled orthonormal transform. PRISM's subsequent fixed High-branch rescaling changes relative variances and energies. The energy comparison is in expectation and does not establish exact separation of noise from geometric detail. □

B.3 Analysis of Proposition 3: Cross-band Response of Structural Corruption

Proposition B.3. *Structural discontinuities or creases can produce broadband wavelet responses affecting both Low and High branches.*

Supporting analysis. For a signal S entering the structural operators, define the spatial deviation $\delta_{struc} = (\mathcal{K}_{occ} \circ \mathcal{K}_{ext})(S) - S$. Under the masked-fill model, occlusion of a signal S' contributes $(1-M)(\tau - S')$; interference contributes $\min(S, S_{clutter}) - S$. These deviations can contain discontinuities or derivative discontinuities. Localized wavelets respond to such structural changes Mallat & Hwang (1992), while local averages can also change. Thus,

$$\Delta C_{struc} = \mathcal{W}\delta_{struc} = \{P_L \mathcal{W}\delta_{struc},\, P_H \mathcal{W}\delta_{struc}\}. \tag{27}$$

Here P_L and P_H select the Low and High coefficients. Both projections can respond to the same event, but neither must be nonzero for every geometry or alignment. This broadband-response argument motivates joint band modeling; it does not prove a statistical dependency or a universal wavelet-scale law.

B.4 Analysis of Proposition 4: Complementary Cross-band Evidence for Structural Restoration

Proposition B.4. *Local edge and discontinuity responses in the High branch can provide complementary spatial evidence for recovering corrupted low-frequency structure.*

Modeling motivation. Using the spatial structural deviation from Proposition 3, its coefficient projections are

$$\Delta C^{low}_{struc} = P_L \mathcal{W}\delta_{struc}, \qquad \Delta C^{high}_{struc} = P_H \mathcal{W}\delta_{struc}. \tag{28}$$

Low-frequency coefficients retain local averages, whereas localized high-frequency coefficients respond to edges and creases Mallat & Hwang (1992). For example, a truncated edge can alter both the local average and the detail response near its break point. This complementary information motivates conditioning low-frequency recovery on both bands, rather than enforcing independent band-wise recovery. It does not establish a universal probability inequality or guarantee that High information improves every instance. The restricted High-to-Low exchange is a modeling choice assessed by the ablations.

B.5 Synthesis: Coefficient-Space Deviation Decomposition

Telescoping Decomposition of Operator-Induced Deviations. Define the sequential states of the chosen composite operator:

$$\begin{aligned} S^{(0)} &= S_{GT}, & S^{(1)} &= \mathcal{K}_{geom}(S^{(0)}), \\ S^{(2)} &= \mathcal{K}_{ext}(S^{(1)}), & S^{(3)} &= \mathcal{K}_{occ}(S^{(2)}), \\ S^{(4)} &= \mathcal{N}_{sig}(S^{(3)}) = S_{in}. \end{aligned}$$

With $\Delta C_k = \mathcal{W}S^{(k)} - \mathcal{W}S^{(k-1)}$, cancellation of intermediate states yields

$$C_{in} = C_{GT} + \sum_{k=1}^{4} \Delta C_k = C_{GT} + \Delta C_{geom} + \Delta C_{struc} + \Delta C_{noise}, \tag{29}$$

where $\Delta C_{geom} = \Delta C_1$, $\Delta C_{struc} = \Delta C_2 + \Delta C_3$, and $\Delta C_{noise} = \Delta C_4$. The identity is exact by telescoping and does not imply physical linearity or statistical independence. Each deviation is evaluated on the preceding corrupted state, so interactions among mechanisms are retained.

Propositions 1–3 characterize the individual components: deformation gives a bounded deviation when its signal-space change is controlled; additive noise is preserved exactly by the linear DWT; structural discontinuities can affect both Low and High projections. Proposition 4 motivates using their complementary information in inter-band interaction. The decomposition accounts for deviations in coefficient space without asserting independent physical corruptions.

B.6 Expanded MAP and Conditional-Transport Formulation

Based on the frequency-band responses and cross-band relations analyzed in Sec. 3.2, to solve the Bayesian inverse problem defined in Sec. 3.1, we first reformulate the conditional structure of its MAP inference and implement PRISM as a two-stage generative model. Stage 1 learns a compact geometric representation with a fixed Wavelet VQ-VAE, motivated by the differing band-wise characteristics of signal and noise. Stage 2 introduces an Inter-band Dual-Stream Probabilistic Bridge that explicitly models dependencies across frequency bands and transports incomplete observations toward the distribution of complete geometry.

Specifically, using the probability chain rule, we first decompose the observation likelihood $P(C_{in} \mid C)$:

$$P(C_{in} \mid C) = P(C_{in}^{high} \mid C_{low}, C_{high}, C_{in}^{low}) \cdot P(C_{in}^{low} \mid C_{low}, C_{high}). \tag{30}$$

Combining this with the cross-band prior factorization $P(C) = P(C_{high} \mid C_{low}) \cdot P(C_{low})$, in which high-frequency details depend on low-frequency structure, the MAP objective becomes a product of two conditionally dependent probability terms:

$$\begin{aligned} \widehat{C}_{GT} = \underset{C_{low}, C_{high}}{\operatorname{argmax}} & \left[P(C_{in}^{low} \mid C_{low}, C_{high}) P(C_{low}) \right] \\ & \cdot \left[P(C_{in}^{high} \mid C_{high}, C_{low}, C_{in}^{low}) P(C_{high} \mid C_{low}) \right]. \end{aligned} \tag{31}$$

This factorization does not imply physical independence. Proposition 4 motivates retaining both clean bands in the first likelihood term; Proposition 3 motivates allowing the second term to retain both bands and the low-frequency observation. These conditional arguments are modeling choices compatible with the chain rule, rather than necessary dependencies proved by the propositions:

$$\begin{aligned} \widehat{C}_{GT} = \underset{C_{low}, C_{high}}{\operatorname{argmax}} & \underbrace{P(C_{in}^{low} \mid C_{low}, C_{high}) P(C_{low})}_{\text{Term A}} \\ & \cdot \underbrace{P(C_{in}^{high} \mid C_{high}, C_{low}, C_{in}^{low}) P(C_{high} \mid C_{low})}_{\text{Term B}}. \end{aligned} \tag{32}$$

Motivated by these cross-band relations, we organize the model into two coupled tasks: structure recovery and conditional detail recovery. Their posterior factorization is

$$P(C_{low}, C_{high} \mid C_{in}) = P(C_{low} \mid C_{in})\, P(C_{high} \mid C_{low}, C_{in}). \tag{33}$$

To solve this high-dimensional posterior, rather than seeking only a single optimum, we generalize the problem to finding a probability transport map between paired observation and complete-shape endpoint distributions, conditioned on the observation. Because the wavelet coefficients C have an cross-band conditional structure—the low-frequency geometric skeleton determines the spatial support to which high-frequency texture attaches—an ordered conditional transport provides a useful modeling perspective. According to the Knothe-Rosenblatt (KR) rearrangement theorem, for absolutely continuous source and target distributions under a given variable ordering, the KR rearrangement yields an almost-everywhere unique monotone triangular transport map. This result motivates the following schematic ordering of the conditional transport in coefficient space:

$$\mathcal{T}_C = [\mathcal{T}_{low}(C_{low} \mid C_{in}),\ \mathcal{T}_{high}(C_{high} \mid C_{low}, C_{in})]\,. \tag{34}$$

This ordering motivates the asymmetric generative structure of "low-frequency structure modeling–high-frequency conditional detail modeling," without asserting that the learned interacting bridge

realizes the unique KR map. The actual network retains limited cross-band information exchange to model the cross-band relations revealed in Sec. 3.2.

Based on this decomposition, we construct a two-stage generative framework that separately learns a compact spectral representation and conditional probability transport.

C Loss-Aware Bridge-Time Sampling

C.1 Definition

Different bridge-time positions have different learning difficulty. We allocate timestep sampling probabilities according to the second moment of the historical primary loss Nichol & Dhariwal (2021):

$$p_t \propto \sqrt{\mathbb{E}[L_t^2]}, \tag{35}$$

where L_t denotes the primary bridge-regression loss across the two frequency branches at timestep t:

$$L_t = \sum_{b\in\{low,high\}} \lambda_b \left\| \epsilon_\theta^b - \epsilon_t^{b,*} \right\|_2^2, \tag{36}$$

with λ_b denoting the loss weight of branch b. We further apply the importance weight

$$w_t = \frac{1}{T p_t}, \tag{37}$$

thereby prioritizing difficult bridge-time positions while preserving the original uniform-time objective in expectation Nichol & Dhariwal (2021). In implementation, each timestep maintains a loss history of length 10; the sampler is activated after every timestep has been observed, with a uniform sampling floor of 0.001. The history is updated only from the unweighted primary DSB loss and is restored together with the checkpoint state.

C.2 Effect of Loss-Aware Sampling

To evaluate this sampling strategy, we compare uniform timestep sampling with the proposed loss-aware sampling while keeping all other settings identical. All variants use the same training budget, model architecture, and evaluation protocol.

Loss-aware sampling improves all three metrics over uniform sampling (Table 5).

Table 5: Loss-aware bridge-time sampling. Best in bold.

Sampling Strategy	CD×100 ↓	IoU×100 ↑	F1×100 ↑
Uniform	3.669	72.001	82.179
Loss-aware (Ours)	**3.599**	**72.952**	**82.856**

D Additional Ablation and Design Analysis

D.1 Spectral Representation Design

D.1.1 3D DWT Representation

We use a fixed one-level 3D DWT to obtain a compact, local, multiband, linear, and exactly invertible decomposition directly in the volumetric domain Mallat (1989); Daubechies (1992). Unlike axis-wise 1D or slice-wise 2D decomposition, the 3D transform jointly characterizes geometry along all three spatial axes; meanwhile, compared with redundant wavelet or scattering representations, it preserves a non-redundant linear coefficient space without introducing an additional learned spectral transform. A one-level 3D DWT naturally produces one low-frequency approximation subband and seven high-frequency detail subbands, which provides a direct interface for the subsequent Low/High dual-stream probabilistic bridge. We therefore use 3D DWT as the fixed spectral representation throughout PRISM.

D.1.2 High-Frequency Scale Calibration and Sensitivity

The high-frequency scale is fixed from training-set coefficient statistics rather than selected according to validation or test completion performance. Specifically, for each complete shape in the ShapeNet training split, we apply the same one-level 3D DWT used by PRISM and measure the root-mean-square (RMS) magnitude of the low-frequency coefficients and the pooled seven high-frequency subbands. We estimate a dataset-level reference ratio

$$s_{\text{stat}} = \frac{\mathbb{E}_{x\sim\mathcal{D}_{\text{train}}}\left[\text{RMS}\left(x^L\right)\right]}{\mathbb{E}_{x\sim\mathcal{D}_{\text{train}}}\left[\text{RMS}(x^H)\right]}, \tag{38}$$

where x^H pools the seven high-frequency subbands. This data-only calibration provides the reference scale before Completion training and is independent of downstream test metrics; we use the resulting fixed reference scale of 4.5 in the main model. We then evaluate a coarse range around this reference value to test whether performance is sensitive to the exact scaling choice. Because the scale changes the VAE representation and hence the subsequent latent space, each setting uses its corresponding VAE and Completion training pipeline.

The reference scale of 4.5 yields the best metrics among the tested scales (Table 6).

Table 6: High-frequency scale sensitivity (default: 4.5). Best in bold.

High-frequency scale	CD×100 ↓	IoU×100 ↑	F1×100 ↑
1.0	3.647	72.506	82.532
4.5	**3.599**	**72.952**	**82.856**
8.0	3.604	72.831	82.753

D.1.3 Interaction-Strength Sensitivity

We further examine sensitivity to the specific weight of the restricted High→Low information transfer. The Low→High weight is fixed to 1.0, while only the High→Low weight is varied. The final model uses 0.25, and all settings follow the same final training and evaluation protocol.

A High→Low weight of 0.25 yields the best metrics among the tested settings (Table 7).

Table 7: High→Low interaction strength (default: 0.25). Best in bold.

High→Low weight	CD×100 ↓	IoU×100 ↑	F1×100 ↑
0.15	3.633	71.953	82.211
0.25	**3.599**	**72.952**	**82.856**
0.50	3.605	72.674	82.665
0.75	3.620	72.751	82.691

D.2 Bridge Source and Partial-Encoder Controls

The main model uses a sample-specific partial source endpoint produced by an additional Partial Encoder that is initialized from the Complete-VAE encoder and jointly optimized with the probabilistic bridge. Following the paired-endpoint formulation used by BridgeShape Kong et al. (2025), this keeps incomplete and complete shapes connected through paired latent endpoints rather than generating the complete shape from an unrelated random source. We provide two controls to separate the role of the source distribution from the role of Partial-Encoder adaptation.

D.2.1 Moment-Matched Gaussian Source

A standard isotropic Gaussian source can introduce an additional confound when the learned latent coordinates are not normalized to unit variance. We therefore construct a moment-matched Gaussian source using only the ShapeNet training split. Let z_T^{low} and z_T^{high} denote the partial-source latents

produced by the final Partial Encoder on the training data. For each branch $b \in \{low, high\}$, we estimate its first- and second-order statistics,

$$\mu_b = \mathbb{E}_{x\sim\mathcal{D}_{\text{train}}}[z_T^b(x)], \qquad \sigma_b^2 = \text{Var}_{x\sim\mathcal{D}_{\text{train}}}[z_T^b(x)], \tag{39}$$

and replace the sample-specific source endpoint with

$$z_{T,\text{Gauss}}^b = \mu_b + \sigma_b \odot \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I). \tag{40}$$

This control matches the marginal scale of the Low/High source distributions while removing their sample-specific geometric correspondence. All remaining architecture, conditioning, optimization, and evaluation settings are kept unchanged.

The learned partial source outperforms the moment-matched Gaussian source (Table 8).

Table 8: Learned partial versus moment-matched Gaussian source. Best in bold.

Source Endpoint	CD×100 ↓	IoU×100 ↑	F1×100 ↑
Moment-Matched Gaussian	4.277	68.467	79.428
Learned Partial Source (Ours)	**3.599**	**72.952**	**82.856**

D.2.2 Partial-Encoder Initialization and Adaptation

The final Partial Encoder is initialized as an exact copy of the corresponding Complete-VAE encoder and is then optimized jointly with the bridge. We use two complementary controls to separate initialization from adaptation. *Frozen Complete-copy* keeps the same Complete-VAE-copy initialization but freezes the source encoder throughout Completion training, isolating the effect of joint adaptation. *Random Init.* keeps the same Partial-Encoder architecture and trainability as the final model but replaces the Complete-VAE-copy initialization with the module's native random initialization, isolating the benefit of starting in the Complete latent coordinate system. We do not introduce a separately pretrained Partial-VAE initialization here because it would also change the pretraining objective and latent coordinate system rather than only the initialization.

Joint adaptation from the Complete-VAE initialization yields the best metrics (Table 9).

Table 9: Partial-encoder initialization and adaptation. Best in bold.

Partial Encoder	CD×100 ↓	IoU×100 ↑	F1×100 ↑
Frozen Complete-copy	3.996	70.622	81.099
Random Init.	3.607	72.637	82.640
Trainable Complete-copy (Ours)	**3.599**	**72.952**	**82.856**

D.3 Training Protocol

Table 10: ShapeNet augmentation ablation and PatchComplete reference. PRISM variants use 100k steps; PatchComplete uses its original training setting. Best in bold.

Protocol	Spatial transforms	Other aug.	CD×100 ↓	IoU×100 ↑	F1×100 ↑
No Aug.	None	None	4.117	67.944	79.078
Yaw-4	Yaw-4	None	3.616	71.982	82.145
PC-Aug (Ours)	Yaw-4, flips	Walls, ±2-voxel jitter	**3.599**	**72.952**	**82.856**
PatchComplete Rao et al. (2022)	Flips, x/z swap	±2-voxel jitter	4.27	65.4	–

D.3.1 Data Augmentation

We evaluate the augmentation protocol under the final matched ShapeNet Completion setting at 100k steps. All architecture, bridge, sampling, conditioning, optimizer, dropout, and evaluation settings are

kept identical; only the augmentation protocol is changed. *No Aug.* disables Yaw-4, wall corruption, reflection, and paired voxel jitter. *Yaw-4* retains only the joint yaw rotation. *PC-Aug* denotes the final protocol combining Yaw-4 with PatchComplete-style wall corruption, reflection, and paired $\pm$2-voxel jitter Rao et al. (2022).

The full PC-Aug protocol improves all three metrics over the reduced augmentation settings (Table 10).

D.3.2 DROPOUT

We further report matched controls for ordinary model dropout. All three variants use the final ShapeNet Completion setting and are compared at 100k on the test split, with only the ordinary U-Net dropout value changed. This experiment concerns the ordinary U-Net dropout value, not the persistent-conditioning path.

Dropout 0.25 gives the lowest CD, while the default 0.1 gives the highest IoU and F1 (Table 11).

Table 11: U-Net dropout on ShapeNet (100k steps). Best in bold.

Dropout	CD$\times$100 $\downarrow$	IoU$\times$100 $\uparrow$	F1$\times$100 $\uparrow$
0	3.612	72.920	82.836
0.1	3.599	**72.952**	**82.856**
0.25	**3.575**	72.670	82.656

E TRAINING RESOURCES

The reference environment uses one NVIDIA GeForce RTX 5090 (32 GB), PyTorch 2.8.0, and CUDA 12.8. The host has Intel Xeon Gold 6459C CPUs; the current training container is allocated a quota equivalent to 16 logical CPUs and 92 GiB of RAM. The saved ShapeNet Completion configuration specifies 100k optimization steps, batch size 24, BF16 mixed precision, gradient checkpointing, eight data-loader workers, and training seed 3407. Stage 1 VAE training uses seed 42. The bridge and Partial Encoder are jointly optimized with AdamW at learning rate 2×10^{-4}, weight decay 10^{-5}, and $\beta = (0.9, 0.999)$; EMA decay is 0.9999. Inputs have resolution 32^3, with two 4×16^3 latent streams. Inference profiling below uses FP32 and batch size 1.

F MODEL COMPLEXITY AND INFERENCE EFFICIENCY

F.1 MODEL COMPLEXITY

PRISM has 44.073M active inference parameters across the Partial Encoder, frozen reference encoder, bridge with conditioning, and decoder; unused VQ codebooks are excluded. Table 12 combines published baseline costs with PRISM measurements. Baseline accuracy is taken from Table 2; costs follow PatchComplete (Appendix E, Table 11), DiffComplete (Appendix D), and BridgeShape (Table 8) Rao et al. (2022); Chu et al. (2023); Kong et al. (2025). Different hardware and counting conventions preclude a controlled efficiency ranking.

Profiling protocol. We profile the 100k-step EMA checkpoint on the RTX 5090 with batch size 1, FP32, autograd disabled, and one CPU thread. cuDNN TF32 is enabled; matrix-multiplication TF32 and cuDNN benchmarking are disabled. Synchronized latency averages 100 runs after 20 warm-ups on a fixed GPU-resident input. The measured path includes DWT, both encoders, all reverse updates and conditioning, decoding, and inverse DWT; data transfer, loading, meshing, and metrics are excluded. Sampling uses the 1000-position symmetric schedule with deterministic OT-ODE updates, zero endpoint noise, and no latent normalization or VQ lookup.

We count convolutions, transposed convolutions, linear layers, and attention matrix products with two FLOPs per multiply–accumulate, excluding bias, normalization, activations, softmax, pooling, and elementwise operations. An independent tensor-shape count matches the profiler and verifies N

reverse-network evaluations:

$$G(N) = 91.278017 + 40.314601N \quad \text{GFLOPs}. \tag{41}$$

The fixed cost comprises two encoders (29.218 GFLOPs each), the decoder (32.842), and wavelet transforms (< 0.001). Thus, one-step completion still requires 131.593 GFLOPs; reducing reverse steps does not remove dense 3D encoding and decoding.

F.2 Accuracy–Compute Trade-off across Reverse Steps

Table 12: ShapeNet accuracy and inference cost: published baselines (top) and PRISM reverse-step settings (bottom). Best accuracy in bold.

Method	Params (M)	Steps	GFLOPs / step	Total GFLOPs	Time (ms)	CD×100 ↓	IoU×100 ↑
PatchComplete Rao et al. (2022)	182.85	–	–	–	63	4.27	65.4
DiffComplete Chu et al. (2023)	43.09	100	159.5	∼15,950	∼3000–4000	4.10	67.5
BridgeShape Kong et al. (2025)	–	3	33.6	∼100.8	40	4.06	70.1
PRISM	44.073	1	40.315	131.593	21.69	3.602	72.395
		2		171.907	34.93	3.601	72.681
		3*		212.222	45.79	**3.599**	72.952
		5		292.851	74.01	3.607	**73.021**

∗: default; –: unavailable or not applicable. PRISM counts active inference parameters and includes encoding/decoding in total FLOPs. Hardware and counting conventions differ.

Table 12 shows the reverse-step trade-off; three steps give the lowest CD, while five give the highest IoU. The main experiments use three reverse bridge steps. We evaluate the same checkpoint at 1, 2, 3, and 5 steps on the complete ShapeNet test split (5,300 partial/complete pairs from 1,325 objects in eight categories), using EMA weights, evaluation seed 42, batch size 4, and zero inference endpoint noise. All settings use the original PatchComplete-compatible metric implementation, 10,240 sampled surface points, and category-macro aggregation. Runtime and FLOPs are measured separately at batch size 1 using the protocol above; parameter count is independent of N.

G Limitation

Limitations and future work. This work focuses on establishing and validating PRISM's spectral modeling and dual-stream restoration mechanisms, with experiments primarily conducted in controlled object-level 3D completion settings. A natural next step is to exploit the framework's low-/high-frequency spectral structure to extend it to higher-resolution, hierarchical, or sparse 3D representations, supporting geometric restoration at larger scales. PRISM's formulation is not restricted to object-level completion, motivating further investigation of its applicability to scene-level restoration, multi-object modeling, and other 3D representations.

H Additional Visualizations

Figures 5–6, 7–8, and 9–10 show eight 3D-EPN, six ScanNet, and eight ShapeNet categories, respectively, with two examples per category. Each example uses a shared viewpoint, scale, and color across methods.

The 3D-EPN baseline uses the released EPN-unet checkpoint, ported to PyTorch, without retrieval-based shape synthesis; ShapeNet and ScanNet use released PatchComplete checkpoints. Input* denotes display-only filtering of unsupported truncation-boundary surfaces on ShapeNet and ScanNet; 3D-EPN inputs are unchanged.

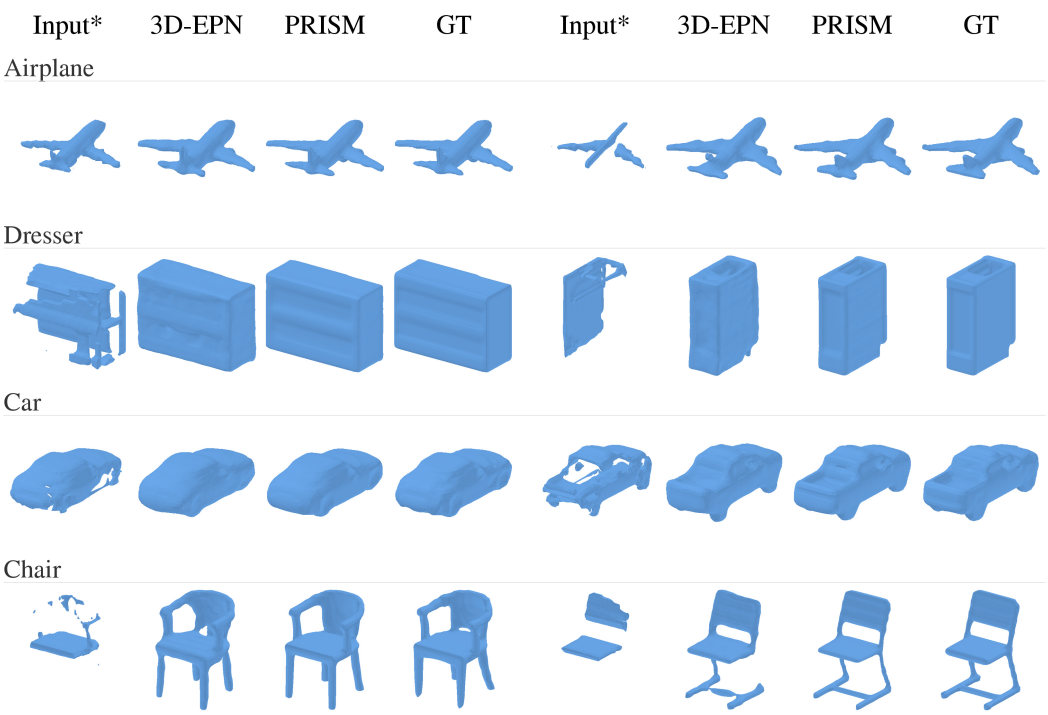

Figure 5: 3D-EPN: airplane, dresser, car, and chair.

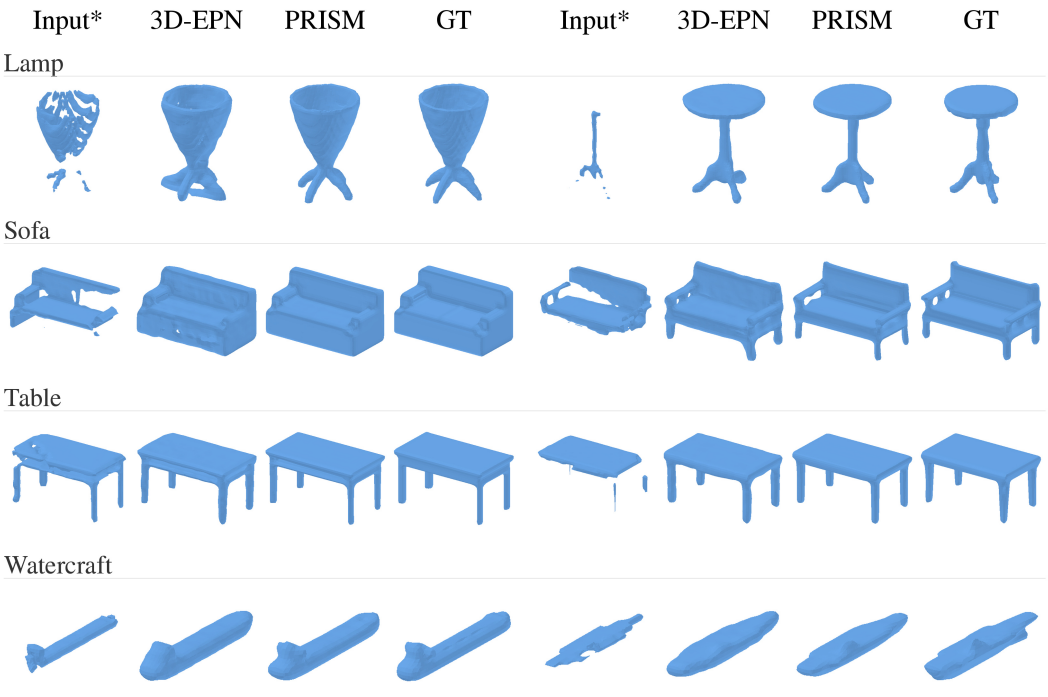

Figure 6: 3D-EPN: lamp, sofa, table, and watercraft.

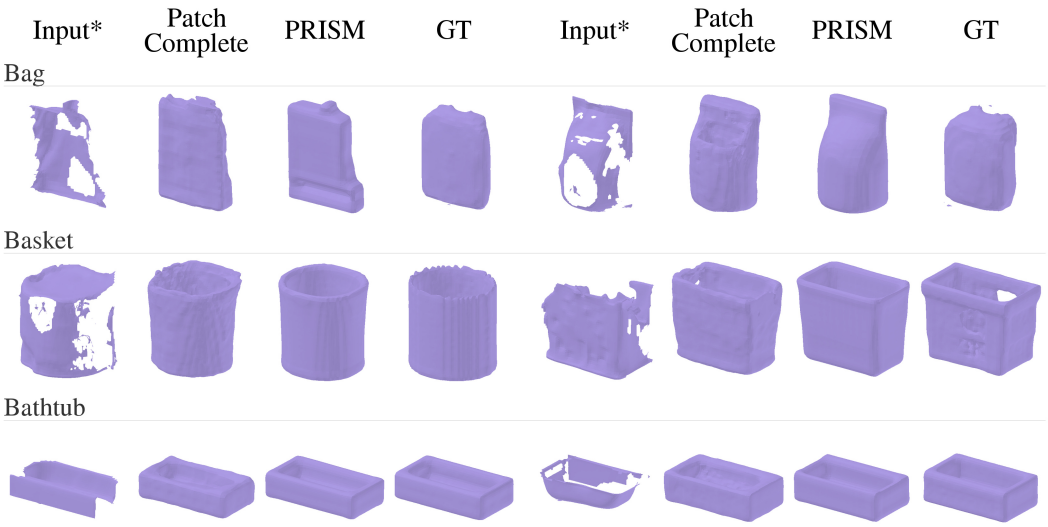

Figure 7: ScanNet: bag, basket, and bathtub.

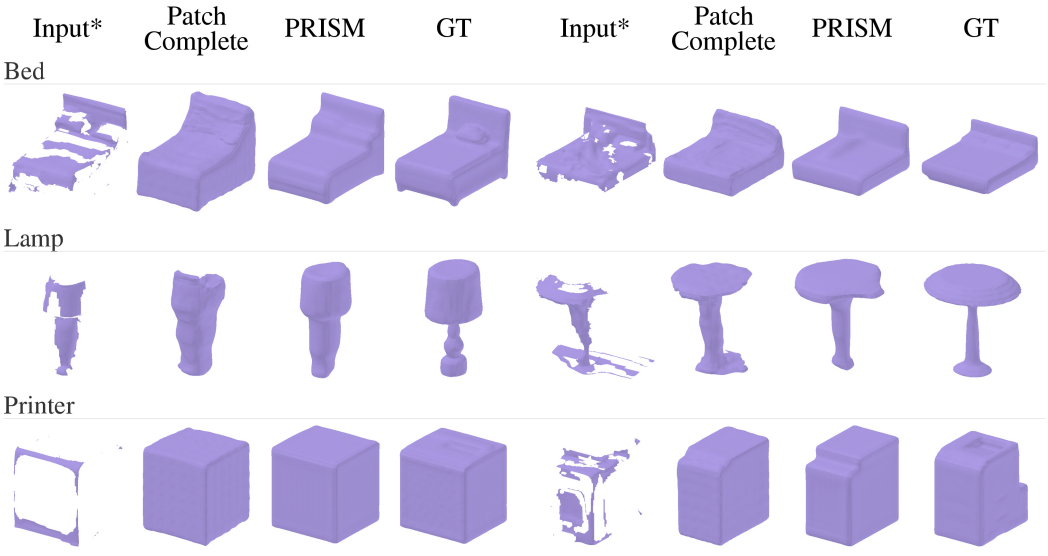

Figure 8: ScanNet: bed, lamp, and printer.

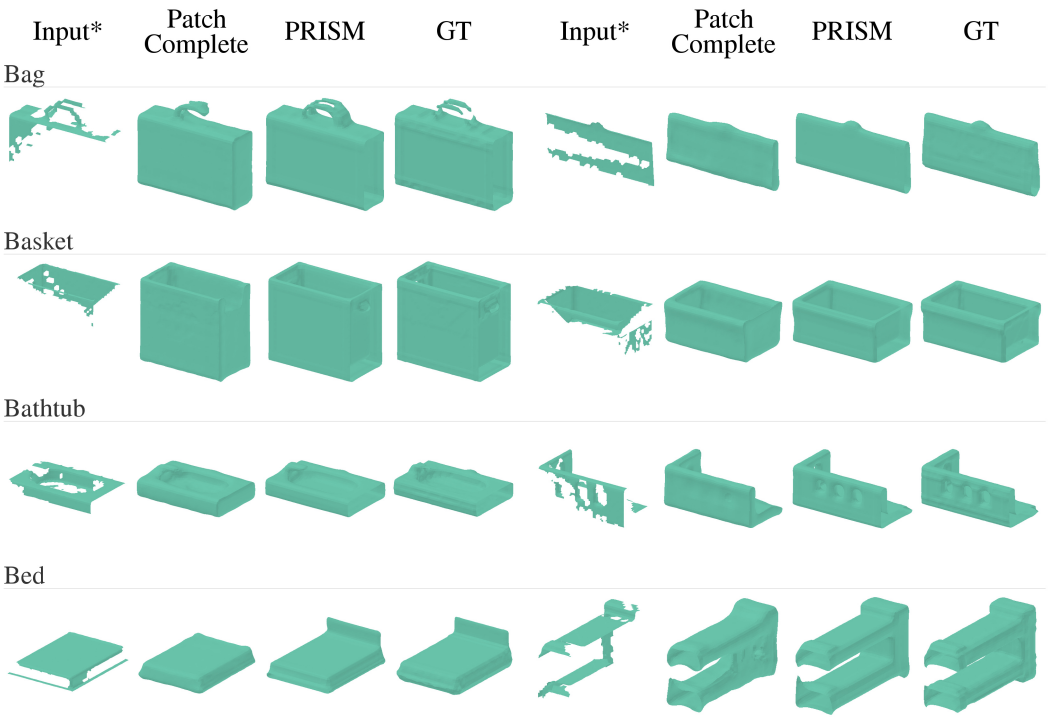

Figure 9: ShapeNet: bag, basket, bathtub, and bed.

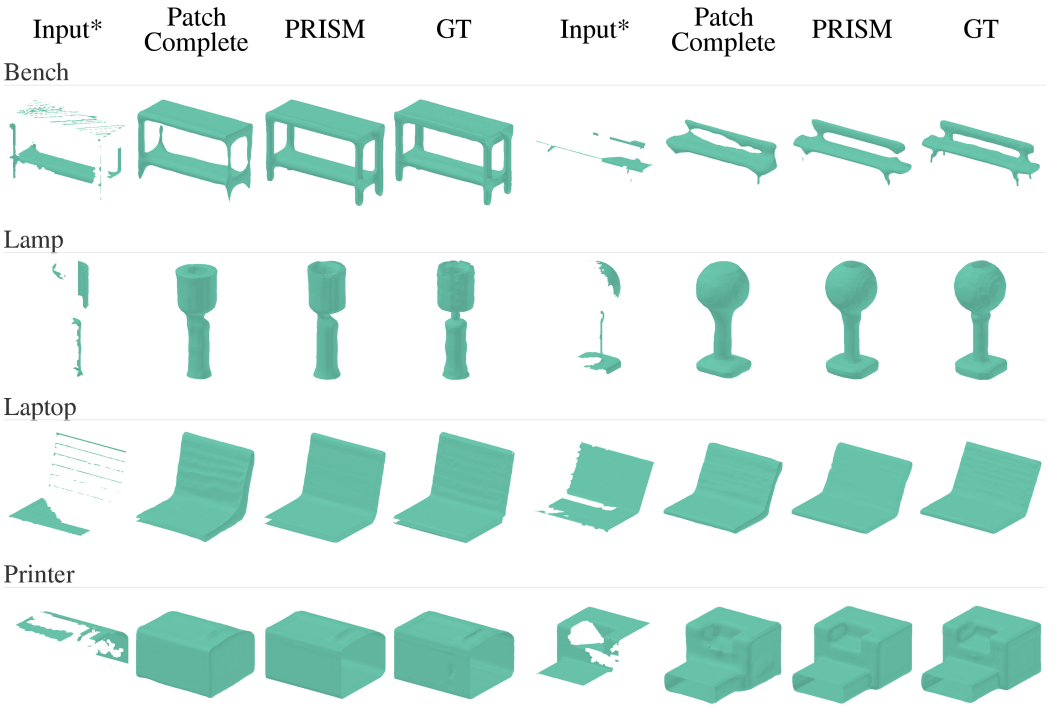

Figure 10: ShapeNet: bench, lamp, laptop, and printer.