ADBM: Adversarial Diffusion Bridge Model for Denoising of 3D Point Cloud Data

Anonymous ICCV submission

Paper ID *****

Abstract

We address the task of point cloud denoising by leveraging a diffusion-based generative framework augmented with adversarial training. While recent diffusion models have demonstrated strong capabilities in learning complex data distributions, their effectiveness in recovering fine geometric details remains limited, especially under severe noise conditions. To mitigate this, we propose Adversarial Diffusion Bridge Model (ADBM), a novel approach for denoising 3D point cloud data by integrating diffusion bridge model with adversarial learning. ADBM incorporates a lightweight discriminator that guides the denoising process through adversarial supervision, encouraging sharper and more faithful reconstructions. The denoiser is trained using a denoising diffusion objective based on Schrödinger bridge, while the discriminator distinguishes between real clean point clouds and generated outputs, promoting perceptual realism. Experiments are conducted on the PU-Net and PC-Net datasets, with performance evaluated employing the Chamfer Distance and Point-to-Mesh metrics. Qualitative and quantitative results both highlight the effectiveness of adversarial supervision in enhancing local detail reconstruction, making our approach a promising direction for robust point cloud restoration.

1. Introduction

Point cloud denoising is critical for enhancing data quality in applications where accurate spatial representation directly impacts system performance and user accessibility. Point clouds acquired via LiDAR, depth sensors, or photogrammetry frequently contain noise from environmental interference, sensor limitations, or motion artifacts. This degradation is especially critical in accessibility applications such as assistive navigation, where noisy inputs cause errors in object detection [2, 11, 16] and scene reconstruction [14]. Also, the presence of noise can obscure fine geometric details and lead to inaccurate shape representations,

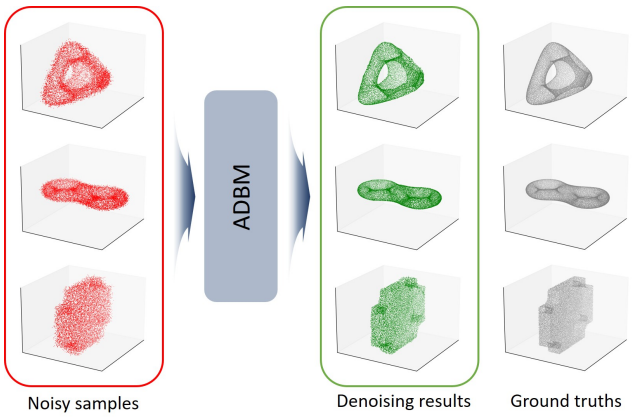

Figure 1. Visual examples of point cloud denoising results using the proposed method, ADBM. Each row represents a different object category. From left to right: noisy input point cloud, denoised output by ADBM, and clean ground-truth shape. The ADBM effectively removes noise and restores fine-grained geometric structures, producing outputs that are closely aligned with the original clean shapes.

which are especially problematic for applications requiring high-precision measurements. As the reliance on 3D point cloud data also continues to grow across diverse fields such as robotics [5, 12], urban mapping [18, 19], and medical imaging [1, 3], the demand for robust and effective denoising techniques becomes increasingly important.

Traditional 3D point cloud denoising approaches [7, 10, 24, 28] have mainly relied on geometric priors and statistical optimization. These approaches demonstrated measurable denoising efficacy under controlled conditions, particularly for Gaussian-type noise distributions. However, they consistently struggled with structural oversimplification in real-world scenarios, where rigid smoothing operators erode fine features like edges and corners, degrading geometric fidelity. Also, non-Gaussian noise from LiDAR or other sensors caused performance collapse, while iterative optimization hindered real-world deployment. These limitations have prompted a shift toward learning-based de-

noising approaches to adaptively model complex noise patterns while maintaining geometric fidelity.

Recent years have seen generative models, particularly diffusion models, emerge as powerful tools for 3D point cloud data synthesis and restoration [13, 17, 22]. By iteratively refining their understanding of complex data distributions, these models achieve high-fidelity reconstruction of noisy inputs through the structured denoising process. However, traditional diffusion approaches suffer from slow sampling speeds, sampling trajectory design inefficiencies, and instability when handling complex noise distributions. Diffusion bridges [4, 20, 21, 23] address these gaps by predicting a direct probabilistic pathway between noisy and clean data distributions, through mitigating the constraints on the prior distribution. While the direct pathway offers improved sampling efficiency and stability, achieving optimal denoising performance, particularly against complex and unknown noise patterns, necessitates a more adaptive and self-improving mechanism.

Inspired by the success of adversarial learning in generative model [8, 9, 25, 27], we propose Adversarial Diffusion Bridge Model (ADBM), which integrates adversarial supervision into the diffusion bridge framework to enhance 3D point cloud denoising. Specifically, a lightweight discriminator is incorporated into the training pipeline to compel the diffusion bridge model to generate outputs that are not only distributionally close to clean data but also perceptually realistic. As shown in Fig. 1, ADBM effectively restores clean shapes from severely noisy inputs across various object categories. The adversarial signal complements the original diffusion bridge objective, providing an additional learning signal that facilitates the recovery of fine geometric details, particularly under complex or non-Gaussian noise conditions. We validate ADBM on PC-Net [26], PU-Net [17] 3d object-level point cloud datasets. Experimental results demonstrate that ADBM consistently outperforms existing state-of-the-art denoising methods in terms of both fidelity and generalization.

2. Related work

2.1. Traditional denoising methods

Traditional methods for 3D point cloud denoising mainly leverage geometric priors and local statistics to suppress noise while preserving structural features. Han et al. [7] proposed a position-guided linear filter for 3D point cloud denoising that significantly improves computational efficiency while preserving geometric features. To preserve sharp features in noisy point clouds, Zheng et al. [28] proposed a guided filter extension that assigns multiple normals to feature points via k-medial skeleton extraction and k-means clustering. To enhance the quality of noisy point sets, Yadav et al. [24] introduced a constraint-based denoising method utilizing a vertex-based normal voting tensor and binary eigenvalue optimization. Their approach iteratively filters vertex normals and updates positions with feature-aware constraints, enabling effective noise removal while preserving geometric sharpness. To address the trade-off between noise removal and feature preservation, Liu at al. [10] developed a two-stage point cloud denoising method that decouples normal filtering from position updating. Their optimization-based framework maintains the underlying geometric structures, achieving high-quality denoising without oversmoothing sharp edges.

2.2. Deep-learning-based methods

To overcome the limitations of traditional denoising approaches, recent research has shifted toward learning-based methods that leverage neural networks to model complex noise patterns in point clouds. PointCleanNet [17] introduced supervised frameworks that learns mappings from noisy to clean point clouds using regression-based losses. They employs an acchitecture that explicitly encodes spatial features while incorporating a two-step denoising mechanism to refine predictions iteratively. Another notable approach is score-based point cloud denoising [13], which introduces a probabilistic generative framework based on score matching and Langevin dynamics. By learning a score function that estimates the gradient of the data distribution, this method can denoise corrupted point clouds through iterative updates. However, the stochastic nature and high iteration cost of score-based sampling remain key challenges. More recently, the P2P-Bridge [22] framework proposes a diffusion-bridge-based model that constructs a direct probabilistic path between noisy and clean point clouds via a Schrödinger bridge formulation [4]. This method utilizes a learnable forward diffusion and reverse denoising to generate geometrically consistent reconstructions, offering improved sample efficiency and generation quality.

While P2P-Brige demonstrates strong performance, it remains limited in adaptively learning discriminative features for real-world noise, due to the absence of an explicit adversarial signal. In this work, we integrate adversarial learning on the diffusion bridge model based on P2P-Bridge to further enhance robustness against diverse noise types.

2.3. Adversarial training approaches

Recent studies have explored adversarial training to improve the quality and realism of diffusion-based generative models. Ko et al. [8] introduces dual discriminators in time and frequency domains to enhance speech fidelity in multi-speaker TTS tasks. Zeng et al. [27] leverages semantic priors and adversarial loss for self-supervised shadow removal, enabling structure-preserving generation without paired labels. Liu et al. [9] combines adversarial learning approach

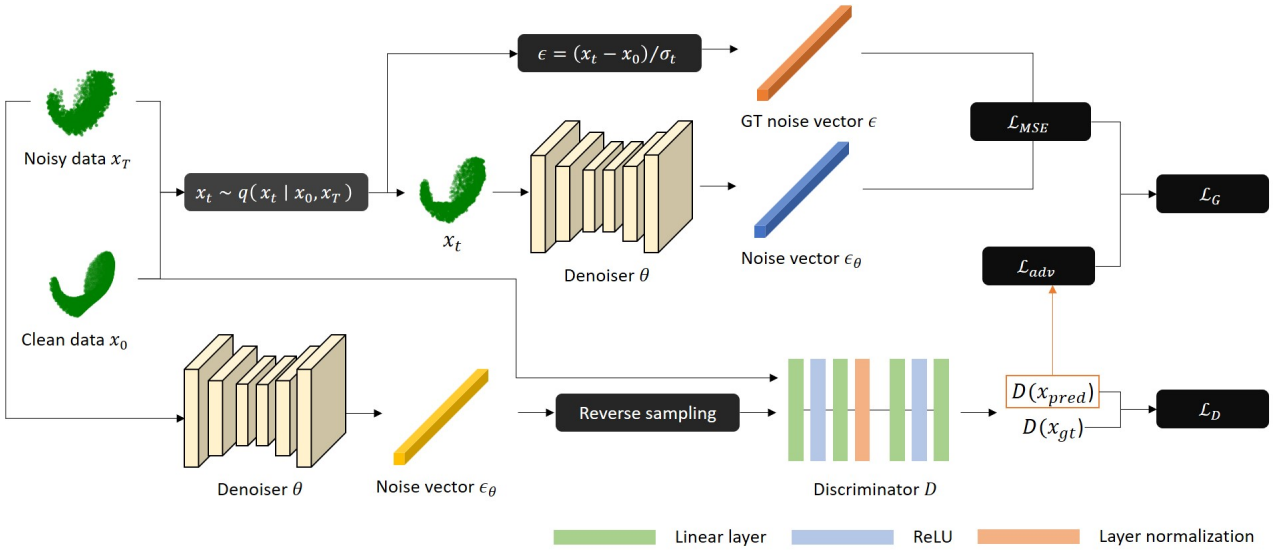

Figure 2. Overview of the proposed adversarial diffusion bridge model (ADBM) training pipeline. The model takes paired clean and noisy point clouds as input and samples an intermediate latent point cloud. A denoising network predicts the noise from this sample, which is used to reconstruct a clean version of the input through reverse sampling. The reconstructed output is evaluated by a discriminator trained to distinguish real clean point clouds from generated ones. The generator is trained with a combination of denoising loss and adversarial feedback, encouraging both accurate reconstruction and perceptual realism.

with torsion angle priors to ensure biologically valid backbones in protein structure generation. A structure-guided discriminator [25] has also been proposed to finetune diffusion models under layout constraints, improving both semantic consistency and image quality. These approaches demonstrate the effectiveness of adversarial signals in guiding diffusion models toward more realistic and task-aligned outputs.

3. Methods

We propose ADBM, an adversarial diffusion bridge model based on P2P-Bridge, which formulates point cloud denoising as a Schrödinger Bridge problem between clean and noisy distributions. This approach enables efficient sampling of intermediate states without numerically solving stochastic differential equations, by leveraging a Gaussian approximation under a paired data boundary condition. By predicting the underlying noise component, the model iteratively refines the input through a learned reverse process. To improve the perceptual quality of the denoised outputs, we further incorporate an adversarial training objective. A lightweight discriminator is trained to distinguish real clean point clouds from generated samples, providing an additional supervisory signal to guide the denoising network. Fig. 2 presents the overall framework.

3.1. Diffusion bridge training

We formulate point cloud denoising as a Schrödinger Bridge problem, which seeks a stochastic process that interpolates between two marginal distributions: the clean data distribution $p_{\text{data}}(x_0)$ and the noisy prior distribution $p_{\text{prior}}(x_T)$. The goal is to find a path measure $p^*(x_{0:T})$ that minimizes the Kullback-Leibler divergence from a reference process $p_{\text{ref}}(x_{0:T})$ while satisfying the boundary conditions:

$$p^*(x_0) = p_{\text{data}}(x_0), \quad p^*(x_T) = p_{\text{prior}}(x_T). \tag{1}$$

Following the formulation proposed in P2P-Bridge, the optimal diffusion path is modeled by a pair of forward and backward stochastic differential equations (SDE), given respectively by:

$$\begin{aligned} dx_t &= \left[f(x_t,t) + g^2(t)\nabla \log \psi_t(x_t)\right] dt + g(t)dw_t, \\ dx_t &= \left[f(x_t,t) - g^2(t)\nabla \log \hat{\psi}_t(x_t)\right] dt + g(t)d\bar{w}_t, \end{aligned} \tag{2}$$

where $f(x_t, t)$ is a vector-valued drift function, $g(t)$ is a scalar-valued diffusion coefficient controlling the noise and w_t, $\bar{w}_t$ are independent standard Wiener processes. The ψ_t and $\hat{\psi}_t$ are potential functions associated with the forward, backward processes and these two processes are coupled as follows:

$$\psi_0\,\hat{\psi}_0 = p_{\text{data}}, \quad \psi_T\,\hat{\psi}_T = p_{\text{prior}}, \quad p_t = \psi_t\,\hat{\psi}_t. \tag{3}$$

This structure ensures that the marginal density p_t interpolates the clean data distribution at $t = 0$ and the noisy prior at $t = T$, forming a time-consistent probabilistic bridge between the two distributions.

However directly solving the system of Eq. 2 is not practicable for high-dimensional data. To address this, recent works approximates this bridge under a paired data assumption $p(x_0, x_T) = p_{\text{data}}(x_0)\, p_{\text{prior}}(x_T \mid x_0)$, and assumes linear drift with zero external force, i.e., $f = 0$, yielding a tractable Gaussian posterior. Under the assumption of a linear drift $f = 0$ and a known diffusion schedule $g(t)$, the posterior of the latent process x_t conditioned on the endpoints x_0 and x_T can be written in closed form as a Gaussian distribution:

$$q(x_t \mid x_0, x_T) = \mathcal{N}(\mu_t, \Sigma_t), \tag{4}$$

where the mean μ_t and the covariance Σ_t are given by:

$$\mu_t = \frac{\bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} x_0 + \frac{\sigma_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} x_T, \quad \Sigma_t = \frac{\sigma_t^2 \bar{\sigma}_t^2}{\bar{\sigma}_t^2 + \sigma_t^2} I, \tag{5}$$

where $\sigma_t^2 = \int_0^t g^2(\tau)\, d\tau$ and $\bar{\sigma}_t^2 = \int_t^1 g^2(\tau)\, d\tau$ represent the accumulated forward and backward variances up to time t, respectively. This analytic form enables efficient sampling of intermediate states x_t without requiring numerical integration of the SDE. During training, we sample $x_t \sim q(x_t \mid x_0, x_T)$, and define the target noise as the residual between the noisy sample and the clean sample as follows:

$$\epsilon = \frac{x_t - x_0}{\sigma_t}. \tag{6}$$

The denoiser network $\epsilon_\theta(x_t, t)$ is trained to predict this noise using MSE loss:

$$\mathcal{L}_{\text{MSE}} = \left\| \epsilon_\theta(x_t, t) - \epsilon \right\|^2. \tag{7}$$

This training objective is conceptually aligned with denoising diffusion probabilistic models, but is distinct in that the noise is conditioned on paired clean and noisy samples, following the diffusion bridge model.

3.2. Adversarial training method

While the diffusion bridge framework optimizes a noise prediction loss based on the Schrödinger Bridge formulation, we further enhance the denoising performance by incorporating an adversarial learning objective. Inspired by GAN-based training schemes [6], we introduce a discriminator network that encourages the generation of samples which are indistinguishable from clean point clouds. Specifically, let x_{pred} denote the model-generated clean sample obtained via reverse diffusion, and let x_{gt} denote the corresponding ground truth clean point cloud. We define a discriminator $D(\cdot)$ that learns to assign high scores to real samples and low scores to generated samples. During each training step, we first sample $x_t \sim q(x_t \mid x_0, x_T)$ and use the denoising network ϵ_θ to estimate x_0^{pred}. We then obtain x_{pred} via reverse sampling. The discriminator is trained to distinguish real clean point clouds from those synthesized by the denoising model. Following the typical GAN formulation, the discriminator loss is defined as:

$$\begin{aligned}\mathcal{L}_D = &-\mathbb{E}_{x_{\text{gt}} \sim p_{\text{data}}} \left[\log D(x_{\text{gt}})\right] \\ &- \mathbb{E}_{x_{\text{pred}} \sim p_\theta} \left[\log\left(1 - D(x_{\text{pred}})\right)\right]. \end{aligned} \tag{8}$$

The generator (i.e., the diffusion bridge model) is trained not only to minimize the original noise prediction loss $\mathcal{L}_{\text{MSE}}$, but also to fool the discriminator by maximizing its predicted score. This adversarial objective for the generator is defined as:

$$\mathcal{L}_{\text{adv}} = -\mathbb{E}_{x_{\text{pred}} \sim p_\theta} \left[\log D(x_{\text{pred}})\right], \tag{9}$$

which encourages the generator to maximize the discriminator's belief that x_{pred} is a real sample. The adversarial signal thus acts as an additional supervisory signal, particularly effective in recovering complex geometric features that are difficult to optimize solely through point-wise regression. To balance the reconstruction and adversarial objectives, we define the final generator loss as a weighted sum:

$$\mathcal{L}_G = \mathcal{L}_{\text{MSE}} + \lambda_{\text{adv}} \mathcal{L}_{\text{adv}}, \tag{10}$$

where λ_{adv} controls the influence of the adversarial signal. This adversarial extension encourages the generator to produce denoised point clouds that not only minimize numerical reconstruction error but also align with the distribution of real clean point clouds.

The procedure of adversarial diffusion bridge training, including noise prediction, adversarial loss computation, and alternating updates of the generator and discriminator is summarized in Algorithm 1. In the training procedure, we employ λ_{adv} to 0.7 to balance the MSE and adversarial objectives.

3.3. Implementation

In this work, we adopt the point cloud denoiser network proposed in P2P-Bridge [22] as our backbone denoiser architecture. The model is designed to predict the drift vector field between clean and noisy point clouds, following the Schrödinger Bridge formulation. The denoiser network follows the encoder-decoder structure of PointNet++ [15], consisting of multi-scale set abstraction modules and feature propagation modules.

To facilitate adversarial learning, we introduce a lightweight discriminator network, which is designed to distinguish between ground-truth clean point clouds and denoised samples generated by the diffusion bridge model.

Algorithm 1: Training of Adversarial Diffusion Bridge Model

Input: Noise schedule $\{\alpha_t\}_{t=0}^T$, $\{\sigma_t\}_{t=0}^T$;
Batch size B;
Weight λ_{adv} for adversarial loss;
Output: Trained denoising network parameters θ
Initialize network parameters θ and discriminator D;
while *not converged* **do**
  Sample minibatch $\{(x_0^i, x_T^i)\}_{i=1}^B$ from $p_{\text{data}}(x_0)\, p_{\text{prior}}(x_T \mid x_0)$;
  Sample $t \sim \mathcal{U}[0,1]$, compute μ_t, Σ_t from Eq. (5);
  Sample $x_t^i \sim \mathcal{N}(\mu_t, \Sigma_t)$ for each pair (x_0^i, x_T^i);
  Compute target noise: $\epsilon^i = \frac{x_t^i - x_0^i}{\sigma_t}$;
  Predict noise: $\hat{\epsilon}_\theta^i = \epsilon_\theta(x_t^i, t)$;
  Compute $\mathcal{L}_{\text{MSE}} = \frac{1}{B}\sum_{i=1}^B \|\hat{\epsilon}_\theta^i - \epsilon^i\|^2$;
  Perform reverse sampling to obtain predicted clean sample x_{pred}^i;
  Compute adversarial loss $\mathcal{L}_{\text{adv}} = -\frac{1}{B}\sum_{i=1}^B \log D(x_{\text{pred}}^i)$;
  Update generator parameters θ using: $\mathcal{L}_G = \mathcal{L}_{\text{MSE}} + \lambda_{\text{adv}}\mathcal{L}_{\text{adv}}$;
  Freeze θ, unfreeze D;
  Compute discriminator loss: $\mathcal{L}_D = -\frac{1}{B}\sum_{i=1}^B \left[\log D(x_0^i) + \log(1 - D(x_{\text{pred}}^i))\right]$;
  Update D with gradient of $\mathcal{L}_D$;
return θ

The architecture of the discriminator first applies a point-wise encoder composed of two linear layers with ReLU activation and layer normalization, transforming each point into a latent feature. The resulting latent features are then aggregated via average pooling across the point dimension, yielding a global feature vector for each sample. This global representation is further processed by a two-layer MLP to produce a scalar output indicating the realism of the input.

4. Experiments

4.1. Datasets

We evaluate our method on two benchmark datasets: PU-Net [26] and PC-Net [17]. The PU-Net dataset contains 40 object categories for training and 20 categories for testing. For each object, ground truth point clouds are provided at three resolutions: 10000, 30000, and 50000 points. To standardize the training input size, we apply farthest point sampling [15] to extract 2048 points from each noisy input, regardless of its original resolution. This allows the model to be trained on a fixed-size representation while leveraging geometric information from diverse scales. The PC-Net dataset is used solely for testing to assess the generalization ability of the model. It consists of 10 object categories, each provided at three resolutions, totaling 30 test samples. During evaluation, the model outputs a 2048-point cloud, which is then compared to the ground truth using alignment techniques and point-wise distance metrics. This setup allows us to evaluate denoising performance of the model on both seen and unseen object distributions across varying resolutions.

4.2. Evaluation measure

To quantitatively assess the quality of denoised point clouds, we adopt two widely used metrics: Chamfer Distance (CD) and Point-to-Mesh Distance (P2M). The CD evaluates the average bidirectional proximity between predicted and ground-truth point sets. It penalizes both missing and redundant points, promoting accurate reconstruction and uniform coverage. Formally, it is defined as:

$$\text{CD}(\hat{\mathcal{P}}, \mathcal{P}) = \frac{1}{2n}\sum_{i=1}^{n} \|\hat{x}_i - \text{NN}(\hat{x}_i, \mathcal{P})\|_2^2 + \frac{1}{2m}\sum_{j=1}^{m} \left\|x_j - \text{NN}(x_j, \hat{\mathcal{P}})\right\|_2^2, \quad (11)$$

where $\hat{\mathcal{P}}$ and $\mathcal{P}$ denote the predicted and reference point clouds, and $\text{NN}(\cdot,\cdot)$ returns the nearest neighbor. To evaluate the geometric consistency with the underlying surface, we also compute the P2M distance. This metric compares points to a mesh surface, taking into account both the distance from points to the mesh and vice versa. It is defined as:

$$\text{P2M}(\hat{\mathcal{P}}, \mathcal{M}) = \frac{1}{2n}\sum_{i=1}^{n} \min_{f\in\mathcal{M}} d(\hat{x}_i, f) + \frac{1}{2|\mathcal{M}|}\sum_{f\in\mathcal{M}} \min_{\hat{x}_i\in\hat{\mathcal{P}}} d(\hat{x}_i, f). \quad (12)$$

Here, $\mathcal{M}$ denotes the ground-truth mesh, and $d(x, f)$ measures the shortest distance between a point and a mesh face. The first term captures how well the predicted points lie on the mesh surface, while the second encourages surface coverage. All point clouds and meshes are normalized to the unit sphere before evaluation to ensure scale invariance.

4.3. Training details

Training is conducted on a single NVIDIA H100 GPU 80GB with an Intel(R) Xeon(R) Platinum 8480+ CPU, running Ubuntu 22.04.2 LTS. The model is trained for a total of 650000 iterations with a batch size of 32. Automatic Mixed Precision is enabled for memory and computing efficiency, and gradient clipping with a maximum norm of 1.0 is applied to stabilize training. Both the denoiser network and

Table 1. Comparison of denoising performance (CD↓ / P2M↓) under different Gaussian noise levels and point counts.

Dataset	Number of points	$10{\cdot}10^3$ Points						$50{\cdot}10^3$ Points					
	Gaussian noise level	1%		2%		3%		1%		2%		3%	
	Method / Metric	CD	P2M	CD	P2M	CD	P2M	CD	P2M	CD	P2M	CD	P2M
PU-Net [26]	PC-Net [17]	3.52	1.15	7.47	3.97	13.1	8.74	1.05	0.35	1.45	0.61	2.29	1.29
	ScoreDenoise [13]	2.52	0.46	3.69	1.07	4.71	1.94	0.72	0.15	1.29	0.57	1.93	1.04
	P2P-Bridge [22]	2.45	0.39	3.27	0.86	4.07	1.47	0.60	0.09	0.95	0.35	1.63	0.90
	ADBM(Ours)	**2.18**	**0.34**	**3.15**	**0.77**	**3.98**	**1.40**	**0.57**	**0.08**	**0.90**	**0.32**	**1.61**	**0.88**
PC-Net [17]	PC-Net [17]	3.85	1.22	6.04	1.45	5.87	1.29	0.29	0.11	0.51	0.25	3.25	1.08
	ScoreDenoise [13]	3.37	0.95	4.52	1.16	6.78	1.94	1.07	0.17	1.66	0.35	2.49	0.66
	P2P-Bridge [22]	2.87	0.63	4.52	0.92	5.65	1.34	0.92	0.12	1.39	0.26	2.17	0.51
	ADBM(Ours)	**2.82**	**0.59**	**4.43**	**0.86**	**5.57**	**1.27**	**0.90**	**0.11**	**1.37**	**0.25**	**2.14**	**0.49**

the discriminator of ADBM are trained using the AdamW optimizer. The denoiser network training uses a constant learning rate of 0.0003, while the discriminator is trained with a learning rate of 0.0001. The exponential moving average of the denoiser network parameters is maintained with a decay factor of 0.999. We use 10 reverse diffusion steps during both adversarial training and evaluation to generate denoised point clouds.

4.4. Experimental results

We evaluate our method, ADBM on the PU-Net and PC-Net datasets under varying Gaussian noise levels and point cloud resolutions. Tab. 1 presents the quantitative comparison of denoising performance with Chamfer Distance and Point-to-Mesh distance, where lower values indicate better denoising performance. On the PU-Net dataset with 10k input points, ADBM consistently outperforms all baselines across all noise levels. At 1% noise, ADBM records a CD of 2.18 and a P2M of 0.34, outperforming P2P-Bridge which achieves 2.45 for CD and 0.39 for P2M. When the noise level increases to 2%, ADBM achieves 3.15 for CD and 0.77 for P2M, showing improvements over P2P-Bridge's 3.27 and 0.86, respectively. At the highest noise level of 3%, ADBM achieves 3.98 for CD and 1.40 for P2M, compared to 4.07 and 1.47 by P2P-Bridge. For the high-resolution setting with 50k points, ADBM continues to outperform the baselines. At 1% noise, ADBM achieves a CD of 0.57 and P2M of 0.08, showing improvements over P2P-Bridge's values of 0.60 and 0.09. At 2% noise, the CD and P2M values achieved by ADBM are 0.90 and 0.32, respectively, whereas P2P-Bridge achieves 0.95 and 0.35. At 3% noise, ADBM yields 1.61 for CD and 0.88 for P2M, outperforming P2P-Bridge's values of 1.63 and 0.90.

On the PC-Net dataset, which is used to evaluate generalization to unseen shapes, our method, ADBM also shows robust performance. At 10k input points and 1% noise, ADBM records a CD of 2.82 and a P2M of 0.59, slightly improving upon P2P-Bridge's results of 2.87 and 0.63. For 2% noise, ADBM achieves 4.43 for CD and 0.86 for P2M, again outperforming P2P-Bridge which reports 4.52 and 0.92. At 3% noise, ADBM shows a clear advantage with a CD of 5.57 and a P2M of 1.27, while P2P-Bridge reports 5.65 and 1.34. For the 50k point resolution, the same trend holds. At 1% noise, ADBM achieves a CD of 0.90 and a P2M of 0.11, whereas P2P-Bridge reports 0.92 and 0.12. With 2% noise, ADBM records 1.37 for CD and 0.25 for P2M, improving upon P2P-Bridge's 1.40 and 0.27. At 3% noise, ADBM achieves 2.14 for CD and 0.49 for P2M, while P2P-Bridge results in 2.17 and 0.51. These comprehensive results demonstrate that our proposed method not only consistently outperforms existing baselines across all noise levels and resolutions, but also generalizes effectively to unseen object categories, yielding the best performance in terms of both point-wise accuracy and surface-level fidelity.

To qualitatively evaluate the denoising performance, Fig. 3 presents visual comparisons across various object categories. The first row shows the ground-truth clean point clouds, uniformly sampled with 10k points per object. To generate the noisy inputs shown in the second row, Gaussian noise with a standard deviation of 1% unit sphere radius is added to the clean shapes. These noisy point clouds exhibit substantial structural distortion and irregular point distribution, particularly around thin or intricate regions such as the camel's legs, the chair's backrest, and the curvature of the duck shape. The third row shows the outputs produced by the P2P-Bridge baseline without adversarial learning. While the overall shapes are recovered to some extent, the results often suffer from blurring or loss of fine details. For instance, the camel's hump and legs appear less distinct, and the reconstructed chair structure lacks geomet-

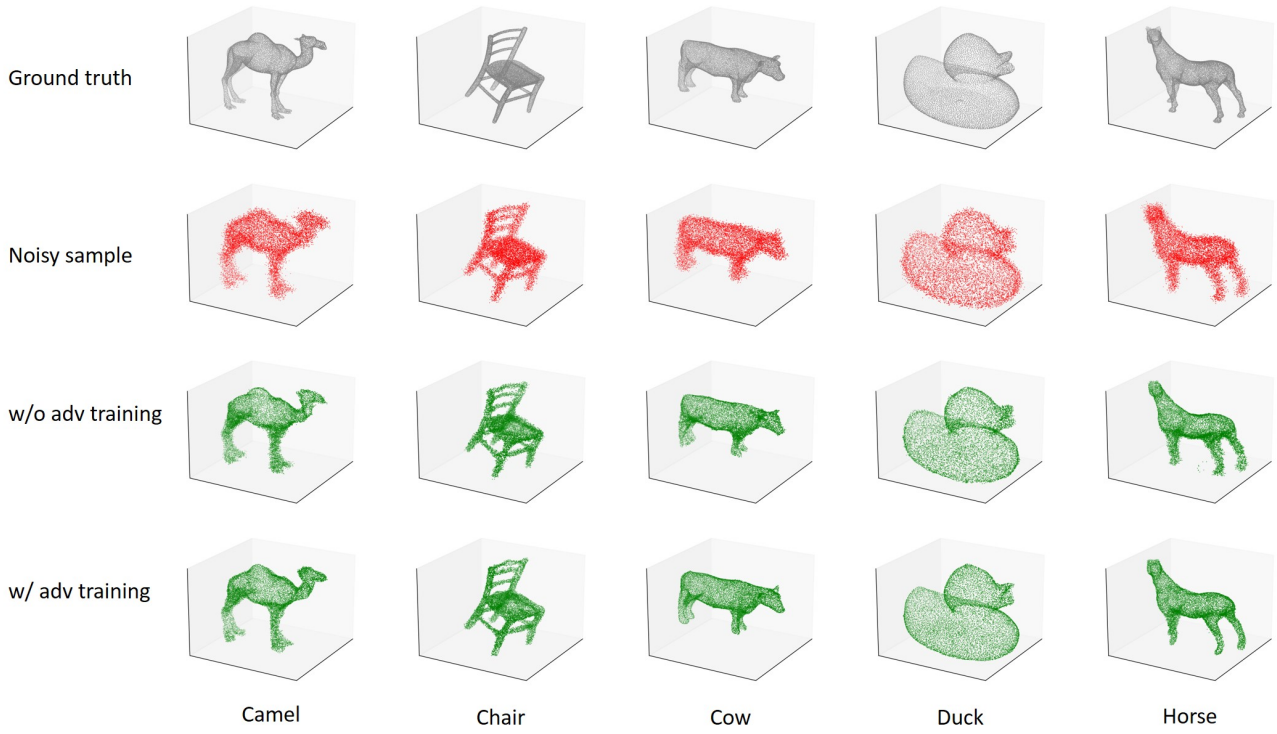

Figure 3. Qualitative comparison of point cloud denoising results. The first row shows the clean ground truth point clouds, each containing 10k points. The second row presents the noisy inputs generated by adding Gaussian noise with a standard deviation of 1% of the unit sphere. The third and fourth rows show the denoised results with and without and our proposed method, respectively.

ric sharpness and completeness. In comparison, the proposed method in fourth row, restores both global structure and fine-grained geometric details. The denoised results exhibit more faithful alignment with the ground-truth, better preserving object-specific characteristics and surface continuity. Moreover, the point distribution appears more uniform and natural, indicating improved surface coverage and sampling quality. These qualitative observations are consistent with the quantitative results, highlighting the superior denoising capability and structural fidelity of our method across diverse shapes.

Fig. 4 shows per-point error heatmaps between the denoised outputs and the ground truth shapes, where the color represents the Euclidean distance to the corresponding ground truth point. All samples consist of 10k points, and the input noise follows a Gaussian distribution with a standard deviation of 1% of the unit sphere. Overall, our method achieves low reconstruction errors across most surface regions, especially in smooth and planar areas such as the camel's torso or the cow's flank. These regions are predominantly rendered in blue, indicating accurate point-wise recovery. However, increased reconstruction errors are observed in geometrically complex areas, including thin structures and high-curvature boundaries such as the camel's legs, the edges of the chair's backrest, and the tail of the horse. These failure cases typically arise due to the local sparsity or overlapping noise in the input, which can distort fine geometric cues during denoising. To mitigate these localized failures, future work may focus on stabilizing the adversarial training process and improving the loss function to better capture fine-grained geometric discrepancies. In particular, incorporating region-aware weighting schemes or multi-scale structural constraints into the training objective could enhance the model's sensitivity to delicate features. These improvements may lead to more faithful reconstructions in challenging regions.

5. Conclusion

In this paper, we proposed an adversarial diffusion bridge training method for 3D point cloud denoising. Building on the Schrödinger bridge formulation, our method models the interpolation between noisy and clean point clouds, enabling effective restoration of fine-grained geometry. To further improve the perceptual quality and fidelity of denoised outputs, we introduced an adversarial learning scheme, where a lightweight discriminator is trained to guide the generator toward producing samples indistinguishable from real clean point clouds. The proposed

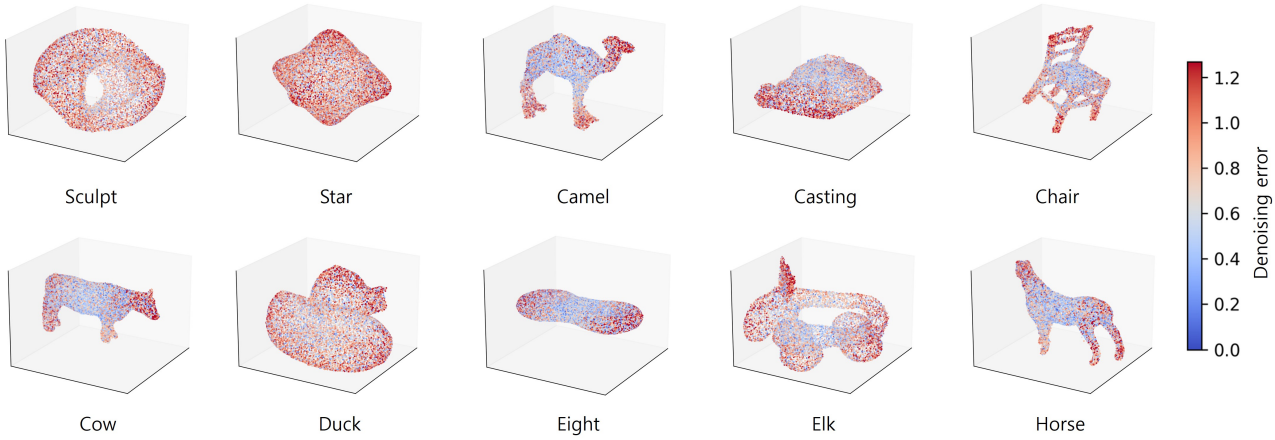

Figure 4. Visualization of per-point Euclidean errors between the denoised outputs and ground truth point clouds. All samples contain 10k points, and Gaussian noise with a standard deviation of 1% of the unit sphere is added to the inputs. Blue regions indicate low reconstruction errors, while red regions highlight areas with higher deviations

method achieves superior reconstruction fidelity, showing strong generalization performance across diverse object categories. However, as shown in Fig. 4, denoising performance in highly corrupted or geometrically complex regions remains challenging. These cases highlight the need for further refinement of the adversarial component. In future work, we aim to explore improved training stability through adversarial loss regularization and conduct systematic studies on how varying the weighting parameter (e.g., λ_{adv}) influences denoising quality and convergence behavior.

References

[1] Marcel Beetz, Abhirup Banerjee, and Vicente Grau. *Point2Mesh-Net: Combining Point Cloud and Mesh-Based Deep Learning for Cardiac Shape Reconstruction*, pages 280–290. 2022. 1

[2] Sriya Behera, Bhaskar Anand, and Rajalakshmi P. Yolov8 based novel approach for object detection on lidar point cloud. In *2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*, pages 1–5. IEEE, 2024. 1

[3] Qiangqiang Cheng, Pengyu Sun, Chunsheng Yang, Yubin Yang, and Peter Xiaoping Liu. A morphing-based 3d point cloud reconstruction framework for medical image processing. *Computer Methods and Programs in Biomedicine*, 193: 105495, 2020. 1

[4] 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*, pages 17695–17709. Curran Associates, Inc., 2021. 2

[5] Zifeng Ding, Yuxuan Sun, Sijin Xu, Yan Pan, Yanhong Peng, and Zebing Mao. Recent advances and perspectives in deep learning techniques for 3d point cloud data processing. *Robotics*, 12:100, 2023. 1

[6] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2014. 4

[7] Xian-Feng Han, Jesse S. Jin, Ming-Jie Wang, and Wei Jiang. Guided 3d point cloud filtering. *Multimedia Tools and Applications*, 77:17397–17411, 2018. 1, 2

[8] Myeongjin Ko, Euiyeon Kim, and Yong-Hoon Choi. Adversarial training of denoising diffusion model using dual discriminators for high-fidelity multi-speaker tts. *IEEE Open Journal of Signal Processing*, 5:577–587, 2024. 2

[9] Yufeng Liu, Linghui Chen, and Haiyan Liu. De novo protein backbone generation based on diffusion with structured priors and adversarial training, 2022. 2

[10] Zheng Liu, Xiaowen Xiao, Saishang Zhong, Weina Wang, Yanlei Li, Ling Zhang, and Zhong Xie. A feature-preserving framework for point cloud denoising. *Computer-Aided Design*, 127:102857, 2020. 1, 2

[11] Yuheng Lu, Chenfeng Xu, Xiaobao Wei, Xiaodong Xie, Masayoshi Tomizuka, Kurt Keutzer, and Shanghang Zhang. Open-vocabulary point-cloud object detection without 3d annotation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1190–1199. IEEE, 2023. 1

[12] Kan Luo, Hongshan Yu, Xieyuanli Chen, Zhengeng Yang, Jingwen Wang, Panfei Cheng, and Ajmal Mian. 3d point cloud-based place recognition: a survey. *Artificial Intelligence Review*, 57:83, 2024. 1

[13] Shitong Luo and Wei Hu. Score-based point cloud denoising. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4563–4572. IEEE, 2021. 2, 6

[14] Sarat Chandra Nagavarapu, Anuj Abraham, Nithish Muthuchamy Selvaraj, and Justin Dauwels. A dynamic object removal and reconstruction algorithm for point clouds. In *2023 IEEE International Conference on*

Service Operations and Logistics, and Informatics (SOLI), pages 1–8. IEEE, 2023. 1

[15] Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. 2017. 4, 5

[16] Charles R. Qi, Or Litany, Kaiming He, and Leonidas Guibas. Deep hough voting for 3d object detection in point clouds. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9276–9285. IEEE, 2019. 1

[17] Marie-Julie Rakotosaona, Vittorio La Barbera, Paul Guerrero, Niloy J. Mitra, and Maks Ovsjanikov. Pointcleannet : Learning to denoise and remove outliers from dense point clouds. *Computer Graphics Forum*, 39:185–203, 2020. 2, 5, 6

[18] Huizhe Sang. Application of uav-based 3d modeling and visualization technology in urban planning. *Advances in Engineering Technology Research*, 12:912, 2024. 1

[19] Sheraz Shamim and Syed Riaz un Nabi Jafri. Enhanced vehicle localization with low-cost sensor fusion for urban 3d mapping. *PLOS One*, 20:e0318710, 2025. 1

[20] Yuyang Shi, Valentin De Bortoli, George Deligiannidis, and Arnaud Doucet. Conditional simulation using diffusion Schrödinger bridges. In *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*, pages 1792–1802. PMLR, 2022. 2

[21] Alexander Tong, Nikolay Malkin, Kilian Fatras, Lazar Atanackovic, Yanlei Zhang, Guillaume Huguet, Guy Wolf, and Yoshua Bengio. Simulation-free schrödinger bridges via score and flow matching. 2024. 2

[22] Mathias Vogel, Keisuke Tateno, Marc Pollefeys, Federico Tombari, Marie-Julie Rakotosaona, and Francis Engelmann. *P2P-Bridge: Diffusion Bridges for 3D Point Cloud Denoising*, pages 184–201. 2025. 2, 4, 6

[23] Gefei Wang, Yuling Jiao, Qian Xu, Yang Wang, and Can Yang. Deep generative learning via schrödinger bridge. In *Proceedings of the 38th International Conference on Machine Learning*, pages 10794–10804. PMLR, 2021. 2

[24] Sunil Kumar Yadav, Ulrich Reitebuch, Martin Skrodzki, Eric Zimmermann, and Konrad Polthier. Constraint-based point set denoising using normal voting tensor and restricted quadratic error metrics. *Computers Graphics*, 74:234–243, 2018. 1, 2

[25] Ling Yang, Haotian Qian, Zhilong Zhang, Jingwei Liu, and Bin Cui. Structure-guided adversarial training of diffusion models. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7256–7266. IEEE, 2024. 2, 3

[26] Lequan Yu, Xianzhi Li, Chi-Wing Fu, Daniel Cohen-Or, and Pheng-Ann Heng. Pu-net: Point cloud upsampling network. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2790–2799. IEEE, 2018. 2, 5, 6

[27] Ziqi Zeng, Chen Zhao, Weiling Cai, and Chenyu Dong. Semantic-guided adversarial diffusion model for self-supervised shadow removal. 2024. 2

[28] Yinglong Zheng, Guiqing Li, Shihao Wu, Yuxin Liu, and Yuefang Gao. Guided point cloud denoising via sharp feature skeletons. *The Visual Computer*, 33:857–867, 2017. 1, 2