LiNeXt: Revisiting LiDAR Completion with Efficient Non-Diffusion Architectures

Wenzhe He, Xiaojun Chen, Ruiqi Wang, Ruihui Li*, Huilong Pi*, Jiapeng Zhang, Zhuo Tang, Kenli Li

College of Computer Science and Electronic Engineering, Hunan University, China
{hewenzhe, chenxiaojun, Wrq2037825386, liruihui, phl880217, zhangjp, ztang, lkl}@hnu.edu.cn

Abstract

3D LiDAR scene completion from point clouds is a fundamental component of perception systems in autonomous vehicles. Previous methods have predominantly employed diffusion models for high-fidelity reconstruction. However, their multi-step iterative sampling incurs significant computational overhead, limiting its real-time applicability. To address this, we propose LiNeXt: a lightweight, non-diffusion network optimized for rapid and accurate point cloud completion. Specifically, LiNeXt first applies the Noise-to-Coarse (N2C) Module to denoise the input noisy point cloud in a single pass, thereby obviating the multi-step iterative sampling of diffusion-based methods. The Refine Module then takes the coarse point cloud and its intermediate features from the N2C Module to perform more precise refinement, further enhancing structural completeness. Furthermore, we observe that LiDAR point clouds exhibit a distance-dependent spatial distribution, being densely sampled at proximal ranges and sparsely sampled at distal ranges. Accordingly, we propose the Distance-aware Selected Repeat strategy to generate a more uniformly distributed noisy point cloud. On the SemanticKITTI dataset, LiNeXt achieves a 199.8 times speedup in inference, reduces Chamfer Distance by 50.7 percent, and uses only 6.1 percent of the parameters compared with LiDiff. These results demonstrate the superior efficiency and effectiveness of LiNeXt for real-time scene completion.

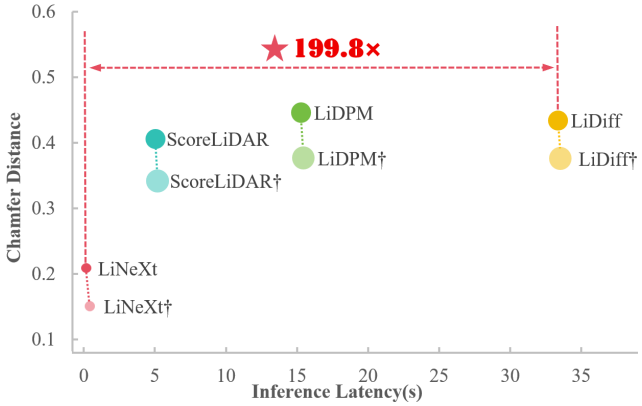

Figure 1: LiNeXt compares reconstruction performance, inference time, and model size; the area of each marker corresponds to the number of parameters, while the symbol † indicates further refinement. For instance, compared to LiDiff (Nunes et al. 2024), LiNeXt achieves 199.8× faster inference speed, a 50.7% reduction in Chamfer Distance, and uses only 6.1% of the parameters.

Introduction

Autonomous driving perception systems primarily utilize LiDAR sensors to acquire 3D point clouds of the surrounding environment, facilitating precise scene reconstruction and safe navigation. Nevertheless, the inherent sparsity of LiDAR measurements combined with frequent occlusions often leads to substantial unobserved regions in the raw point clouds. Such incompleteness hinders critical downstream tasks, including object detection (Wu et al. 2022; Lang et al. 2019; Guo et al. 2025), pose estimation (Hagelskjær and Buch 2020), and mapping (Popović et al. 2021). To overcome these limitations, scene completion methods (Vizzo et al. 2022; Zhou, Du, and Wu 2021) aim to infer and reconstruct missing spatial structures, providing complete 3D representations that enhance the robustness and reliability of autonomous driving perception.

In scene completion research, traditional grid-based representations, such as voxel grids (Li et al. 2020; Roldão, de Charette, and Verroust-Blondet 2020; Zhang, Zhu, and Du 2023) and signed distance fields (Vizzo et al. 2022; Li et al. 2023; Liu et al. 2024), have been widely employed to model 3D geometry. Voxel representations discretize scenes into regular grids, encoding geometry via occupancy or cell attributes, while SDFs, typically built on voxel grids, assign signed distances to implicitly define surfaces. However, these methods are limited by resolution trade-offs: lower resolutions fail to capture fine geometric details, whereas higher resolutions incur significant memory and computational costs. In contrast, point clouds provide a flexible, quantization-free representation that directly encodes complex geometry and fine spatial details with high fidelity. This advantage enables the preservation of the geometric fidelity of the original scene. Building on this, recent studies, including LiDiff (Nunes et al. 2024), LiDPM (Martyniuk et al. 2025), and ScoreLiDAR (Zhang

*Corresponding author.

et al. 2024), have adopted diffusion-based frameworks for LiDAR scene completion, leveraging the precision of point clouds to achieve promising results. Nevertheless, the iterative nature of diffusion sampling introduces substantial computational overhead, leading to slow inference times that impede real-time applications. Furthermore, the denoising objective in these models poses optimization challenges, especially when high-magnitude noise causes significant point displacement, thereby complicating accurate noise estimation and removal. By comparison, directly minimizing the Chamfer Distance (CD) provides a simpler and more effective approach to reconstructing the underlying geometry.

In this work, we forgo the use of cumbersome diffusion models and instead employ a lightweight network to reconstruct the scene. Specifically, we first introduce a Distance-aware Selected Repeat strategy, which replicates points according to their distance from the LiDAR sensor: closer points are repeated less frequently, while farther points are repeated more, resulting in a more uniform spatial distribution. We then introduce Gaussian noise to the replicated points. Subsequently, the Noise to Coarse (N2C) Module directly reconstructs the coarse scene structure, thereby avoiding the substantial time overhead incurred by the denoising process of diffusion models. The resulting output from N2C is then fed into the Refine Module to enhance structural completeness and geometric detail accuracy. Furthermore, we introduce a Cross-Point Attention (CPA) mechanism that dynamically aligns features and fuses complementary information between two input points, thereby strengthening the inference of missing structures and substantially improving completion accuracy and consistency. The Multi-Scale Sparse Convolution (MSSC) module extracts point features at multiple voxel resolutions via efficient sparse convolutions and fuses them into a unified descriptor, enabling the network to capture both fine-grained local geometry and coarse global context with minimal overhead.

Experimental results demonstrate that, as shown in Figure 1, on the SemanticKITTI (Behley et al. 2019) benchmark, LiNeXt achieves a $199.8\times$ inference speedup over LiDiff, reduces the Chamfer Distance by 50.7%, and requires only 6.1% of LiDiff's parameters. This combination of accelerated processing, improved reconstruction fidelity, and compact model footprint highlights LiNeXt's efficacy for real-time 3D scene completion in autonomous driving.

Our contributions are summarized as follows:

- We develop the **Cross-Point Attention (CPA)** and **Multi-Scale Sparse Convolution (MSSC)** modules to form the core of the LiNeXt architecture, enabling enriched feature alignment and multi-resolution geometric reasoning.
- We abandon traditional diffusion-based denoising and instead employ a lightweight network to directly reconstruct complete 3D scenes.
- We demonstrate state-of-the-art accuracy on both the SemanticKITTI (Behley et al. 2019) and KITTI-360 (Liao, Xie, and Geiger 2022) benchmarks, substantially outperforming existing methods.
- Experimental results demonstrate that LiNeXt achieves a $199.8\times$ inference speedup over LiDiff (Nunes et al. 2024) while using only 6.1% of its parameters, highlighting its computational efficiency and practical suitability for diverse real-world applications.

Related Work

3D LiDAR Scene Completion

Early completion techniques predominantly employed voxel-grid discretization or implicit signed distance field (SDF) representations. Voxel-based approaches include OccRWKV (Wang et al. 2024b), which achieves efficient semantic occupancy prediction through linear-complexity modeling and bird's-eye-view feature fusion, and OccFormer (Zhang, Zhu, and Du 2023), leveraging dual-path transformers for 3D volume processing. Implicit SDF methods encompass Make it Dense (Vizzo et al. 2022), a self-supervised framework completing sparse LiDAR scans to dense TSDF volumes, and SurroundSDF (Liu et al. 2024), performing implicit scene reconstruction via query-based SDF prediction. While inferring missing geometry through occupancy probabilities or continuous distance fields, these methods incur substantial computational costs and remain constrained by grid resolution, which blurs fine structures.

Single-Object Point Cloud Completion

PCN (Yuan et al. 2018) pioneered learning latent shape representations for missing point generation. Subsequent works, using a coarse-to-fine(Xie et al. 2020; Pan et al. 2021; Wen et al. 2021; Xiang et al. 2021; Zhou et al. 2022; Rong et al. 2024; Fang et al. 2025) manner, have significantly advanced robustness and detail fidelity. For example, SnowflakeNet (Xiang et al. 2021) leverages Snowflake Point Deconvolution (SPD) for hierarchical point upsampling, where skip-transformer layers learn point splitting patterns to recover fine-grained geometric details. AdaPoinTr (Yu et al. 2023) reformulates completion as set-to-set translation using a Transformer encoder-decoder structure. PointAttN (Wang et al. 2024a) refines local geometric dependency capture through structured self-attention mechanisms. Most recently, GenPC (Li, Zhu, and Wei 2025) leverages 3D generative priors with depth prompting and geometric fusion to achieve zero-shot completion on real-world scans. Existing single-object point cloud completion methods focus on virtual models and often fail in complex real-world scenarios.

Diffusion Models in Completion

Recently, an increasing number of research endeavors (Ni et al. 2025; Cao and Behnke 2024; Du et al. 2025; Zhao et al. 2025) have focused on utilizing diffusion modules for scene completion. For example, PVD (Zhou, Du, and Wu 2021) unifies shape generation and completion via a probabilistic point-voxel diffusion model, but it demonstrates limited effectiveness in outdoor scene completion. LiDiff (Nunes et al. 2024) and its variants adopt a locally guided diffusion process within the DDPM framework, yielding marked improvements in outdoor reconstruction fidelity. LiDPM (Martyniuk et al. 2025) further rethinks this paradigm, proving vanilla DDPMs with proper initialization achieve superior

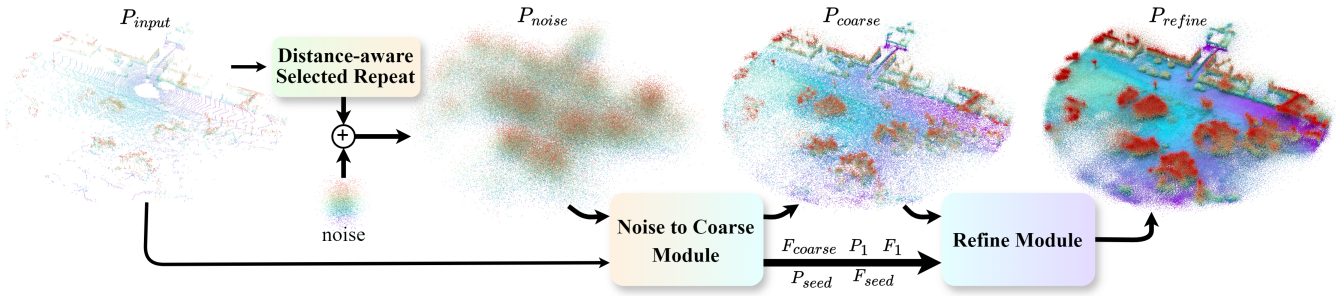

Figure 2: Overall framework of LiNeXt.

scene completion without local diffusion approximations. Critically, LiDiff and LiDPM exhibit high inference latency. ScoreLiDAR (Zhang et al. 2024) employs knowledge distillation to accelerate diffusion-based sampling, making completion 5× faster while maintaining competitive completion quality. Despite their training stability and high output fidelity, diffusion-based methods typically require slow sampling processes and complex network architectures, making them impractical for large-scale real-time perception systems.

Method

As illustrated in Figure 2, the incomplete LiDAR point cloud P_{input} is duplicated in the Distance-aware Selected Repeat strategy (less for near, more for far) and corrupted with Gaussian noise to yield P_{noise}. The Noise-to-Coarse Module denoises P_{noise} under supervision of P_{input}, producing P_{coarse}. The Refine Module then refines this output into the final high-quality point cloud P_{refine}.

Distance-aware Selected Repeat

Existing diffusion-based (Nunes et al. 2024; Zhang et al. 2024) methods replicate the input point cloud P_{input} uniformly, resulting in a noisy cloud P_{noise} that oversamples near-field points while undersampling far-field points. To address this imbalance, we propose a Distance-Aware Selected Repeat (DSR) strategy, which adjusts duplication factors based on each point's Euclidean distance from the origin, thereby yielding a more uniformly distributed P_{noise}. Formally, let $P_{input} = \{p_i\}_{i=1}^{N}$, and compute

$$d_i = \|p_i\|, \quad i = 1, \ldots, N. \tag{1}$$

Sort the points by d_i in ascending order to obtain $\{p_{(1)}, \ldots, p_{(N)}\}$, then partition them into four equal-sized groups:

$$G_k = \left\{ p_{((k-1)\frac{N}{4}+1)}, \ldots, p_{(k\frac{N}{4})} \right\}, \quad k = 1, 2, 3, 4. \tag{2}$$

Assign duplication counts of $\{r_1 = 5, r_2 = 8, r_3 = 12, r_4 = 15\}$ to groups $\{G_1, G_2, G_3, G_4\}$, respectively, forming the replicated set P_{rep}. Finally, add independent Gaussian noise to each point in P_{rep}, yielding the balanced noisy point cloud P_{noise}. This "fewer for near, more for far" replication ensures uniform coverage across distances and provides richer, evenly distributed samples for subsequent Noise-to-Coarse (N2C) modules.

Multi-Scale Sparse Convolution (MSSC)

The MSSC module hierarchically aggregates point-cloud features via parallel spatially sparse convolutions applied over multiple voxel resolutions $g_k \in \mathcal{G} = \{0.01 \times 2^{i-1} \mid i = 1, 2, \ldots, N_{vox}\}$, where N_{vox} denotes the total number of voxel scales. Given input coordinates $P \in \mathbb{R}^{N\times 3}$, initial point-wise features are computed as $X = \mathrm{MLP}_{\mathrm{init}}(P)$. For each resolution $g_k \in \mathcal{G}$, the point cloud is voxelized and embedded:

$$\hat{P}_k = \lfloor P/g_k \rfloor, \qquad F_k = \mathrm{MLP}_k(X), \tag{3}$$

where $\lfloor \cdot \rfloor$ denotes element-wise floor division. Sparse tensors $\mathcal{T}_k$ are then constructed using the grid indices $\hat{P}_k$ and features F_k. Each $\mathcal{T}_k$ undergoes dual residual sparse convolutions:

$$\begin{aligned} \mathcal{T}'_k &= \mathrm{spconv}_{k,1}(\mathcal{T}_k) + \mathcal{T}_k, \\ \mathcal{T}''_k &= \mathrm{spconv}_{k,2}(\mathcal{T}'_k) + \mathcal{T}'_k, \end{aligned} \tag{4}$$

We denote concatenation across scales by the double-bar operator. The multi-scale output for each resolution is defined as $O_k = \mathcal{T}''_k + F_k$, and these outputs are concatenated and projected to a unified descriptor: $F = \mathrm{MLP}_{end}(\mathrm{CONCAT}_{k=1}^{|\mathcal{G}|}(O_k))$. This design captures spatial hierarchies through optimized sparse 3D convolutions while preserving geometric fidelity via residual connections.

Cross-Point Attention Module

The Cross-Point Attention (CPA) Module enables robust feature fusion between the global scene and localized part representations by explicitly encoding spatial relationships and leveraging attention for correspondence. As illustrated in Figure 4, CPA operates on primary point cloud coordinates $P_{key} \in \mathbb{R}^{N\times 3}$ with key, and part coordinates $P_{query} \in \mathbb{R}^{M\times 3}$ with $query$ and $value$.

First, local correspondences are identified by performing a k-nearest neighbors search between the two coordinate sets, yielding index maps for grouping:

$$idx = \mathrm{KNN}(P_{query}, P_{key}, k). \tag{5}$$

Using these indices, the relative displacement of each primary point with respect to its neighbors in the part set is computed to obtain spatial embedding:

$$\alpha = \mathrm{MLP}_{pos}(P_{key} - \mathcal{G}(P_{key}, idx)). \tag{6}$$

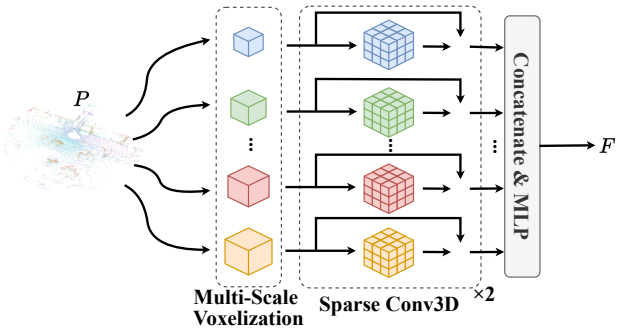

Figure 3: The structure of the Multi-Scale Sparse Convolution (MSSC) module.

This spatial embedding α augments the geometry-aware feature differences:

$$\begin{aligned} Q_{rel} &= query - \mathcal{G}(key, idx) + \alpha, \\ V_{rel} &= value - \mathcal{G}(key, idx) + \alpha. \end{aligned} \tag{7}$$

Before Serial-Segment Max-Pooling (SSMP) processing, the input point cloud is serialized into a linear sequence using randomized spatial orderings, specifically Z-order (Morton 1966) and Hilbert-order (Hilbert and Hilbert 1935), to preserve spatial locality. Next, for each point in this sequence, its local neighborhood is retrieved via k-nearest neighbors (KNN) while preserving the original sequence order within each neighborhood, producing relational feature $Q_{rel}, V_{rel} \in \mathbb{R}^{N\times C\times K}$, where N is the number of points, C is the feature dimension, and K is the number of neighbors. These tensors are partitioned along the dimension of the neighborhood into $\hat{K}$ segments of size $K/\hat{K}$, resulting in reshaped tensors of dimension $\mathbb{R}^{N\times C\times \hat{K}\times (K/\hat{K})}$. SSMP is then applied to compress each segment to its maximally activated feature:

$$\hat{Q}_{rel} = \mathrm{SSMP}(Q_{rel}), \quad \hat{V}_{rel} = \mathrm{SSMP}(V_{rel}) \tag{8}$$

where SSMP operates along the partitioned dimension $K/\hat{K}$, producing output tensors $\hat{Q}_{rel}, \hat{V}_{rel} \in \mathbb{R}^{N\times C\times \hat{K}}$.

This compression achieves critical dimensionality reduction for computational efficiency while preserving discriminative patterns, concurrently enhancing robustness to local perturbations through emphasis on dominant spatial-relationship signatures within each region.

$\hat{Q}_{rel}$ are then combined through a multi-layer perceptron and softmax to produce normalized attention weights:

$$\mathcal{A} = \mathrm{SoftMax}\left(\mathrm{MLP}_{attn}(\hat{Q}_{rel})\right). \tag{9}$$

Finally, the attention weights modulate the value embeddings to aggregate context-aware features, which are then fused with the original representation via a residual connection:

$$F_{new} = value + \sum_{j=1}^{\hat{K}} A_j \odot \hat{V}_j, \tag{10}$$

where j represents the index of the local segment and $\hat{K}$ denotes the total number of segments.

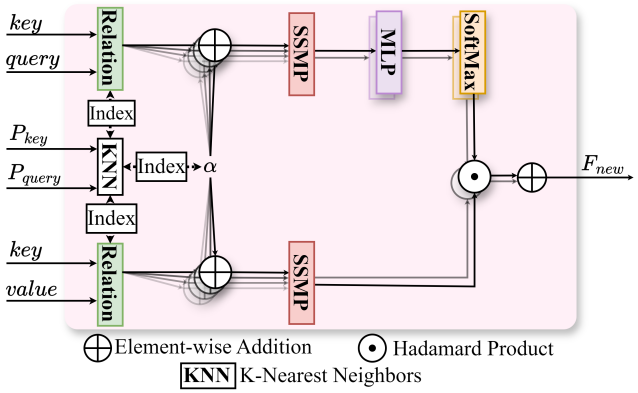

Figure 4: The structure of the Cross-Point Attention (CPA) module.

Through this ordered sequence of embedding, grouping, pooling, and attention, the CPA Module preserves geometric coherence while propagating complementary features across point sets, substantially enhancing the network's ability to infer and complete missing structures.

Noise to Coarse (N2C) Module

As illustrated in Figure 5, the N2C module generates a coarse denoised point cloud by hierarchically distilling structural priors from the input distribution. Its operational pipeline comprises three core stages.

Initial Feature Extraction: Multi-Scale Sparse Convolution (MSSC) extracts preliminary spatial features from the input point cloud P_{input} and noisy observation P_{noise}, yielding enhanced feature sets F_0 and F_{noise}, respectively:

$$F_0 = \mathrm{MSSC}(P_{input}), \quad F_{noise} = \mathrm{MSSC}(P_{noise}). \tag{11}$$

Hierarchical Seed Generation: An N-stage downsampling process iteratively condenses P_{input} into regional seed points P_{seed} with global features F_{seed}. At stage i:

$$\left(P_i, \hat{F}_{i-1}\right) = \begin{cases} (P_{input}, F_0) & i = 1, \\ \left(\mathrm{FPS}(P_{i-1}),\ F_{i-1}[index_{FPS}]\right) & i \geq 2. \end{cases} \tag{12}$$

Here $\mathrm{FPS}(\cdot)$ denotes farthest point sampling, and $[index_{FPS}]$ indexes the features of the sampled points. The initial stage ($i = 1$) skips downsampling to allow deeper feature extraction.

Subsequently, Cross-Point Attention (CPA) processes the geometric relationships between hierarchical point sets: for each stage i, P_i serves as the query points with features $\hat{F}_{i-1}$, while P_{i-1} and its features F_{i-1} provide the structural key and value. This attention mechanism outputs enhanced features F_i. The final outputs $P_{seed} = P_N$ and $F_{seed} = F_N$ encapsulate the global features of local seed regions distilled by the iterative attention refinement process.

Coarse Reconstruction Given a noisy point set P_{noise} with features F_{noise}, we first retrieve the nearest seed points $\hat{P}_{seed}$ and their associated features $\hat{F}_{seed}$. A lightweight

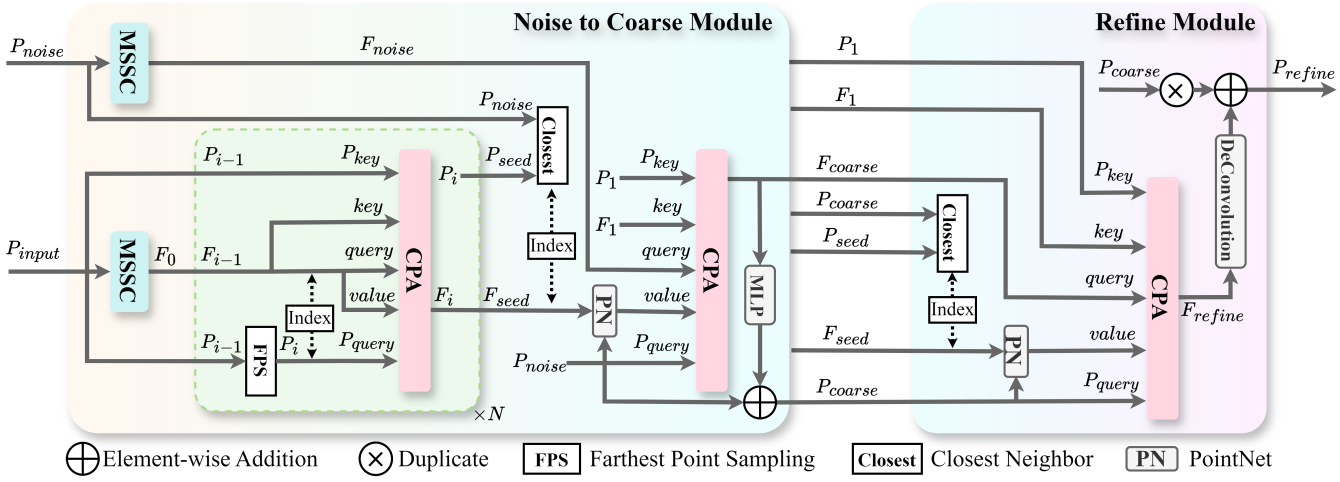

Figure 5: Detailed architectures of the Noise to Coarse Module and the Refine Module.

PointNet (Qi et al. 2016) then fuses noisy coordinates with the aggregated seed features, producing relation-aware features $value$.

In the last CPA module, we treat F_{noise}, F_1, and a relation-aware $value$ as the query, key, and value, respectively. This block jointly regresses the coarse coordinates P_{coarse} and features F_{coarse}, then forwards the intermediate tensors $\{P_1, F_1, P_{seed}, F_{seed}\}$ to the Refine Module for further detail recovery.

Refine Module

The Refine Module, as shown in Figure 5, enhances the coarse output P_{coarse} from the N2C module. For each point in P_{coarse}, we retrieve regional features $\hat{F}_{seed}$ from its nearest neighbor in the seed set P_{seed}. These seed features are combined with P_{coarse} coordinates through a lightweight PointNet (Qi et al. 2016) module to generate relationship-aware $value$. The CPA mechanism then processes P_{coarse} as query points with features F_{coarse}, using P_1 (from the N2C's first downsampling stage) as structural keys with features F_1, and $value$ as geometric-relationship values. The CPA outputs refined features F_{refine}, which are then passed through a deconvolution (Xiang et al. 2021) module to generate the upsampled point cloud P_{refine}, effectively combining local details with global context to correct residual noise while recovering geometrically consistent structures.

Training Loss

The ground-truth point cloud is downsampled via voxel grid to 180,000 points to ensure spatial uniformity and reduce computational cost. For each output point set $P \in \{P_{\text{coarse}}, P_{\text{refine}}\}$, we compute the Chamfer Distance as

$$L_{\text{CD}}(P, \widehat{P}) = \frac{1}{|P|} \sum_{x \in P} \min_{y \in \widehat{P}} \|x - y\|_2^2 + \frac{1}{|\widehat{P}|} \sum_{y \in \widehat{P}} \min_{x \in P} \|y - x\|_2^2, \tag{13}$$

where $\widehat{P}$ is the downsampled ground truth and $\|\cdot\|_2$ denotes the Euclidean norm. Each stage is trained independently using this loss.

Experiment

Experimental Settings

Dataset Preparation We train our model solely on the SemanticKITTI dataset (Behley et al. 2019). Following LiDiff's protocol (Nunes et al. 2024), scans from sequences 00–10 are concatenated using the provided ego-poses and filtered to remove all dynamic objects, resulting in dense static point clouds. Sequence 08 is reserved for validation. To assess cross-dataset generalization, the pretrained model is evaluated without fine-tuning on sequence 00 of the KITTI-360 dataset (Liao, Xie, and Geiger 2022).

Evaluation Metrics

We evaluate the completion performance using the Chamfer Distance (Akmal Butt and Maragos 1998) (CD), Jensen–Shannon Divergence (Menéndez et al. 1997) (JSD) computed in both bird's-eye view (BEV) and 3D space, as well as occupancy IoU at multiple voxel resolutions (0.5 m, 0.2 m, and 0.1 m).

Quantitative and Qualitative Comparisons

Quantitative Comparisons

LiNeXt consistently surpasses LiDiff on SemanticKITTI (Table 1): without refinement, Chamfer Distance decreases from 0.434 to 0.214, 3D JSD drops from 0.564 to 0.494, and BEV JSD from 0.444 to 0.336. The inference latency is 0.167s, which is 199.8× faster than that of LiDiff. The model size is 1.99M parameters, 16.4× smaller than LiDiff. With refinement, the CD further decreases to 0.149, the 3D JSD decreases to 0.481, and the BEV JSD decreases to 0.331. We evaluate the model trained on SemanticKITTI directly on KITTI-360: LiNeXt† retains its Chamfer Distance of 0.149 and achieves a marginal BEV-JSD reduction of 0.008, whereas LiDiff† degrades from a CD of 0.376

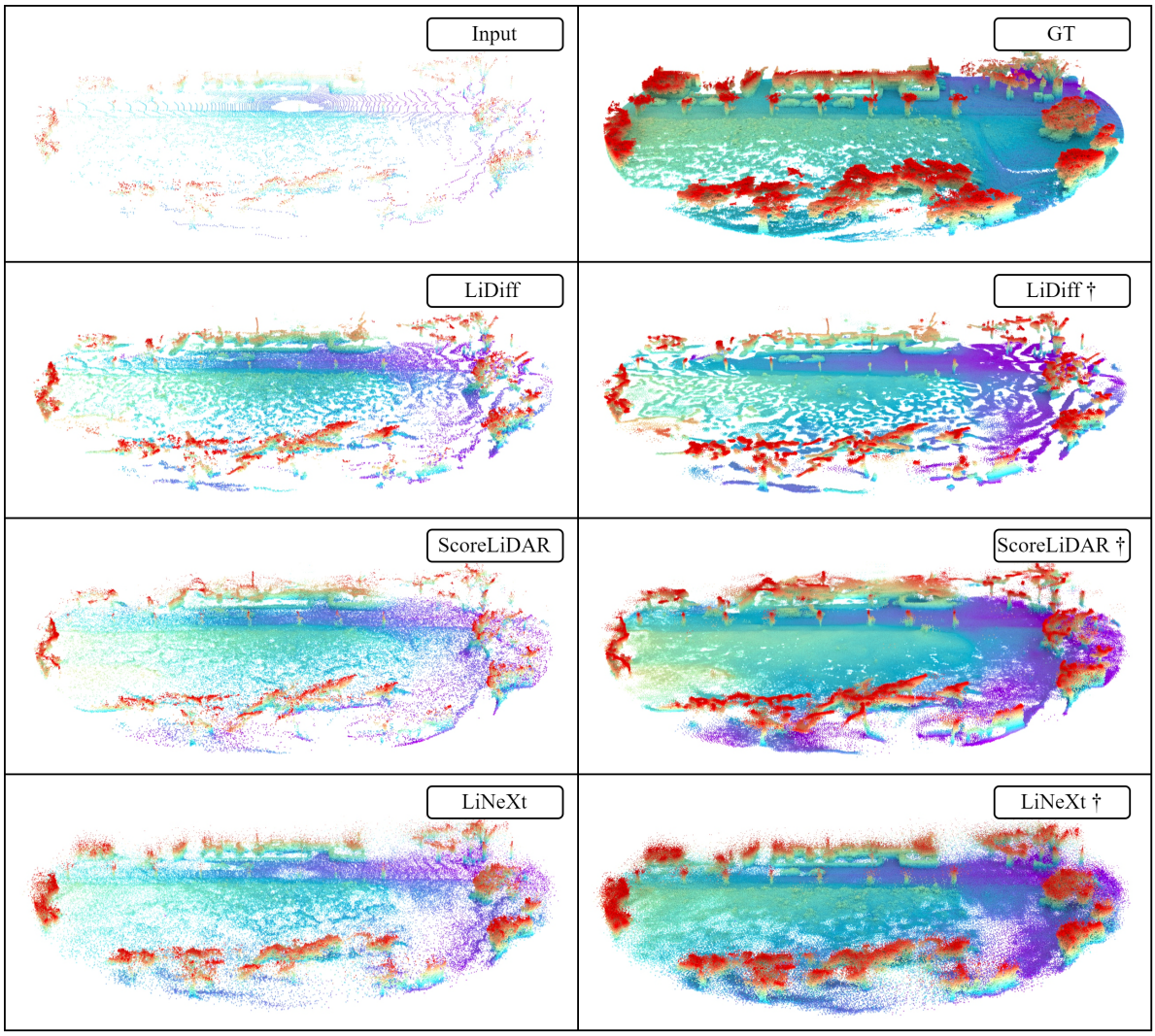

Figure 6: Visualization comparison of our method against LiDiff and ScoreLiDAR on the SemanticKITTI dataset. † indicates additional refinement.

to 0.517, and its BEV-JSD increases by 0.030. This consistency across datasets highlights LiNeXt robustness and generalizability in various LiDAR scanning scenarios, demonstrating its ability to achieve high-fidelity completion despite differing scene characteristics. As shown in table 3, LiNeXt attains an inference latency of 0.167 s (199.8× faster than LiDiff) and 0.434 s after refinement. Its model footprint is only $1.99M$ parameters (2.10 M with refinement), corresponding to 16.4× and 25.9× reductions compared to LiDiff and LiDiff†, respectively. These results underscore the exceptional trade-off of LiNeXt among accuracy, speed and compactness.

Qualitative Comparisons

Figure 6 compares scene completions on SemanticKITTI. Diffusion-based methods exhibit streak artifacts and density variations, whereas LiNeXt produces uniform, depth-consistent reconstructions. The refined LiNeXt† further suppresses noise and fills occlusions, yielding artifact-free point clouds. These qualitative gains corroborate our quantitative metrics, confirming LiNeXt superior spatial uniformity and reconstruction fidelity.

Ablation Study

Our ablation study on SemanticKITTI (Table 2) isolates the contributions of LiNeXt core modules. First, removing the Distance-Aware Selected Repeat (DSR) strategy markedly degrades all evaluation metrics except voxel IoU at 0.1m, highlighting DSR critical role in maintaining global shape coherence. Second, removing the Multi-Scale Sparse Convolution (MSSC) module degrades all voxel IoU metrics (e.g., 39.87% vs. 41.07% at 0.5 m), underscoring MSSC's role in recovering fine structural details. Finally, for the Cross-Point Attention (CPA) ablation, we replace CPA with a standard cross-attention mechanism only within the Noise-to-Coarse (N2C) Module's hierarchical seed genera-

	Method	CD↓	JSD 3D↓	JSD BEV↓	Vox.IoU(0.5 m)↑	Vox.IoU(0.2 m)↑	Vox.IoU(0.1 m)↑
SemanticKITTI	LMSCNnet	0.641	–	0.431	30.83	12.09	3.65
	LODE	1.029	–	0.451	33.81	16.39	5.00
	MID	0.503	–	0.470	31.58	22.72	13.14
	PVD	1.256	–	0.498	15.91	3.97	0.60
	LiDiff	0.434	0.564	0.444	31.47	16.79	4.67
	LiDPM	0.446	0.532	0.440	34.09	19.45	6.27
	ScoreLiDAR	0.406	–	0.425	–	–	–
	LiNeXt	<u>0.214</u>	<u>0.494</u>	<u>0.336</u>	<u>41.07</u>	19.45	6.30
	LiDiff†	0.376	0.573	0.416	32.43	22.99	13.40
	ScoreLiDAR†	0.342	–	0.399	–	–	–
	LiDPM†	0.376	0.542	0.403	36.59	<u>25.76</u>	<u>14.93</u>
	LiNeXt†	**0.149**	**0.481**	**0.331**	**41.97**	**31.25**	**15.09**
KITTI-360	LMSCNet	0.979	–	0.496	26.17	9.21	2.88
	LODE	1.565	–	0.483	33.06	15.24	4.68
	MID	0.637	–	0.476	33.05	21.32	11.30
	LiDiff	0.564	–	0.459	33.23	17.55	4.88
	ScoreLiDAR	0.472	–	0.444	–	–	–
	LiNeXt	<u>0.217</u>	<u>0.508</u>	<u>0.355</u>	<u>36.85</u>	16.91	5.73
	LiDiff†	0.517	–	0.446	33.43	<u>22.04</u>	<u>11.84</u>
	ScoreLiDAR†	0.452	–	0.437	–	–	–
	LiNeXt†	**0.149**	**0.499**	**0.339**	**41.88**	**29.34**	**13.90**

Table 1: Comparison of various methods on the scene completion task on SemanticKITTI and KITTI-360. † indicates additional refinement. Best results are highlighted in bold and second-best in underlined.

Method	CD↓	JSD 3D↓	JSD BEV↓	Vox.IoU(0.5m)↑	Vox.IoU(0.2m)↑	Vox.IoU(0.1m)↑
LiNeXt	**0.214**	**0.494**	**0.336**	**41.07**	**19.45**	<u>6.30</u>
w/o DSR strategy	<u>0.215</u>	0.508	0.352	<u>40.00</u>	<u>18.84</u>	**6.60**
w/o MSSC module	0.221	<u>0.502</u>	<u>0.350</u>	39.87	18.48	6.00
w/o CPA module	0.227	0.504	0.353	39.36	18.65	5.84

Table 2: Ablation results on SemanticKITTI. Best results are highlighted in bold and second-best in underlined.

Method	CD	Time(s)	Param(M)
LiDiff	0.434	33.359	32.67
LiDPM	0.446	15.288	32.67
ScoreLiDAR	0.406	5.047	32.67
LiNeXt	<u>0.214</u>	**0.167**	**1.99**
LiDiff†	0.376	33.531	54.40
LiDPM†	0.377	15.453	54.40
ScoreLiDAR†	0.342	5.189	54.40
LiNeXt†	**0.149**	<u>0.434</u>	<u>2.10</u>

Table 3: Quantitative comparison of reconstruction accuracy, inference speed, and model size. † denotes additional refinement. The best and second-best results are highlighted in bold and underlined, respectively. Inference speed is measured on a single NVIDIA RTX 3090 GPU.

tion stage-full replacement would incur quadratic complexity and exceed our 24 GB memory budget. This partial substitution still incurs the largest performance drop (CD rises to 0.227, +6.1%; IoU 0.5m falls to 39.36%), confirming CPA's critical role in hierarchical feature aggregation.

Conclusion

We have presented LiNeXt, a lightweight, non-diffusion framework for 3D LiDAR scene completion. By introducing a Distance-aware Selected Repeat strategy, a Noise-to-Coarse (N2C) Module, a Refine Module, a Cross-Point Attention (CPA) mechanism, and a Multi-Scale Sparse Convolution (MSSC) module, LiNeXt directly reconstructs complete point-cloud scenes with high fidelity and efficiency. Extensive experiments on SemanticKITTI and KITTI-360 demonstrate that LiNeXt achieves state-of-the-art accuracy while delivering a 199.8× inference speedup and reducing model size to 6.1% compared to LiDiff (Nunes et al. 2024). These results underscore the practical suitability of LiNeXt for real-world autonomous driving applications.

Acknowledgements

This work was supported by and the National Natural Science Foundation of China (No.U25A20421, No.62202151, No.62202152) and the National Key Research and Development Program of China (No.2025YFB3003601).

References

Akmal Butt, M.; and Maragos, P. 1998. Optimum design of chamfer distance transforms. *IEEE Transactions on Image Processing*, 7(10): 1477–1484.

Behley, J.; Garbade, M.; Milioto, A.; Quenzel, J.; Behnke, S.; Stachniss, C.; and Gall, J. 2019. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *CVPR*.

Cao, H.; and Behnke, S. 2024. DiffSSC: Semantic LiDAR Scan Completion using Denoising Diffusion Probabilistic Models. arXiv:2409.18092.

Du, Y.; Zhao, Z.; Su, S.; Golluri, S.; Zheng, H.; Yao, R.; and Wang, C. 2025. SuperPC: A Single Diffusion Model for Point Cloud Completion, Upsampling, Denoising, and Colorization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Fang, C.; Liang, J.; Liang, J.; Wang, H.; Yao, K.; and Cao, F. 2025. Multi-modal point cloud completion with interleaved attention enhanced Transformer. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 963–971. Montreal, Canada.

Guo, W.; Xu, X.; Wang, Z.; Feng, J.; Zhou, J.; and Lu, J. 2025. Text-guided Sparse Voxel Pruning for Efficient 3D Visual Grounding. *arXiv preprint arXiv:2502.10392*.

Hagelskjær, F.; and Buch, A. G. 2020. Pointvotenet: Accurate Object Detection And 6 DOF Pose Estimation In Point Clouds. In *2020 IEEE International Conference on Image Processing (ICIP)*, 2641–2645.

Hilbert, D.; and Hilbert, D. 1935. Über die stetige Abbildung einer Linie auf ein Flächenstück. *Dritter Band: Analysis· Grundlagen der Mathematik· Physik Verschiedenes: Nebst Einer Lebensgeschichte*, 1–2.

Lang, A. H.; Vora, S.; Caesar, H.; Zhou, L.; Yang, J.; and Beijbom, O. 2019. PointPillars: Fast Encoders for Object Detection From Point Clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Li, A.; Zhu, Z.; and Wei, M. 2025. GenPC: Zero-shot Point Cloud Completion via 3D Generative Priors. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, 1308–1318.

Li, J.; Han, K.; Wang, P.; Liu, Y.; and Yuan, X. 2020. Anisotropic Convolutional Networks for 3D Semantic Scene Completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Li, P.; Zhao, R.; Shi, Y.; Zhao, H.; Yuan, J.; Zhou, G.; and Zhang, Y.-Q. 2023. LODE: Locally Conditioned Eikonal Implicit Scene Completion from Sparse LiDAR. In *ICRA*.

Liao, Y.; Xie, J.; and Geiger, A. 2022. KITTI-360: A Novel Dataset and Benchmarks for Urban Scene Understanding in 2D and 3D. *T-PAMI*.

Liu, L.; Wang, B.; Xie, H.; Liu, D.; Liu, L.; Tian, Z.; Yang, K.; and Wang, B. 2024. SurroundSDF: Implicit 3D Scene Understanding Based on Signed Distance Field. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 21614–21623.

Martyniuk, T.; Puy, G.; Boulch, A.; Marlet, R.; and de Charette, R. 2025. LiDPM: Rethinking Point Diffusion for Lidar Scene Completion. In *2025 IEEE Intelligent Vehicles Symposium (IV)*.

Menéndez, M.; Pardo, J.; Pardo, L.; and Pardo, M. 1997. The Jensen-Shannon divergence. *Journal of the Franklin Institute*, 334(2): 307–318.

Morton, G. M. 1966. A computer oriented geodetic data base and a new technique in file sequencing. *physics of plasmas*.

Ni, J.; Liu, Y.; Lu, R.; Zhou, Z.; Zhu, S.-C.; Chen, Y.; and Huang, S. 2025. Decompositional Neural Scene Reconstruction with Generative Diffusion Prior. In *CVPR*.

Nunes, L.; Marcuzzi, R.; Mersch, B.; Behley, J.; and Stachniss, C. 2024. Scaling Diffusion Models to Real-World 3D LiDAR Scene Completion. In *CVPR*.

Pan, L.; Chen, X.; Cai, Z.; Zhang, J.; Zhao, H.; Yi, S.; and Liu, Z. 2021. Variational Relational Point Completion Network. *arXiv preprint arXiv:2104.10154*.

Popović, M.; Thomas, F.; Papatheodorou, S.; Funk, N.; Vidal-Calleja, T.; and Leutenegger, S. 2021. Volumetric Occupancy Mapping With Probabilistic Depth Completion for Robotic Navigation. *RA-L*.

Qi, C. R.; Su, H.; Mo, K.; and Guibas, L. J. 2016. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *arXiv preprint arXiv:1612.00593*.

Roldão, L.; de Charette, R.; and Verroust-Blondet, A. 2020. LMSCNet: Lightweight Multiscale 3D Semantic Completion. In *3DV*.

Rong, Y.; Zhou, H.; Yuan, L.; Mei, C.; Wang, J.; and Lu, T. 2024. CRA-PCN: Point Cloud Completion with Intra- and Inter-level Cross-Resolution Transformers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 4676–4685.

Vizzo, I.; Mersch, B.; Marcuzzi, R.; Wiesmann, L.; Behley, J.; and Stachniss, C. 2022. Make it Dense: Self-Supervised Geometric Scan Completion of Sparse 3D LiDAR Scans in Large Outdoor Environments. *RA-L*.

Wang, J.; Cui, Y.; Guo, D.; Li, J.; Liu, Q.; and Shen, C. 2024a. PointAttN: You Only Need Attention for Point Cloud Completion. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(6): 5472–5480.

Wang, J.; Yin, W.; Long, X.; Zhang, X.; Xing, Z.; Guo, X.; and Zhang, Q. 2024b. OccRWKV: Rethinking Efficient 3D Semantic Occupancy Prediction with Linear Complexity. *arXiv preprint arXiv:2409.19987*.

Wen, X.; Xiang, P.; Han, Z.; Cao, Y.-P.; Wan, P.; Zheng, W.; and Liu, Y.-S. 2021. PMP-Net: Point Cloud Completion by

Learning Multi-step Point Moving Paths. In *CVPR*, 7443–7452.

Wu, X.; Peng, L.; Yang, H.; Xie, L.; Huang, C.; Deng, C.; Liu, H.; and Cai, D. 2022. Sparse Fuse Dense: Towards High Quality 3D Detection with Depth Completion. In *CVPR*.

Xiang, P.; Wen, X.; Liu, Y.-S.; Cao, Y.-P.; Wan, P.; Zheng, W.; and Han, Z. 2021. Snowflake Point Deconvolution for Point Cloud Completion and Generation with Skip-Transformer. In *ICCV*, 5499–5509.

Xie, H.; Yao, H.; Zhou, S.; Mao, J.; Zhang, S.; and Sun, W. 2020. GRNet: Gridding Residual Network for Dense Point Cloud Completion. In *ECCV*.

Yu, X.; Rao, Y.; Wang, Z.; Lu, J.; and Zhou, J. 2023. AdaPoinTr: Diverse Point Cloud Completion with Adaptive Geometry-Aware Transformers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(12).

Yuan, W.; Khot, T.; Held, D.; Mertz, C.; and Hebert, M. 2018. Pcn: Point completion network. In *3DV*.

Zhang, S.; Zhao, A.; Yang, L.; Li, Z.; Meng, C.; Xu, H.; Chen, T.; Wei, A.; GU, P. P.; and Sun, L. 2024. Distilling Diffusion Models to Efficient 3D LiDAR Scene Completion. *arXiv:2412.03515*.

Zhang, Y.; Zhu, Z.; and Du, D. 2023. OccFormer: Dual-path Transformer for Vision-based 3D Semantic Occupancy Prediction. *arXiv preprint arXiv:2304.05316*.

Zhao, A.; Zhang, S.; Yang, L.; Li, Z.; Wu, J.; Xu, H.; Wei, A.; GU, P. P.; and Sun, L. 2025. Diffusion Distillation With Direct Preference Optimization For Efficient 3D LiDAR Scene Completion. arXiv:2504.11447.

Zhou, H.; Cao, Y.; Chu, W.; Zhu, J.; Lu, T.; Tai, Y.; and Wang, C. 2022. Seedformer: Patch seeds based point cloud completion with upsample transformer. In *European conference on computer vision*, 416–432. Springer.

Zhou, L.; Du, Y.; and Wu, J. 2021. 3d shape generation and completion through point-voxel diffusion. In *ICCV*.