

LiDAR4D: Dynamic Neural Fields for Novel Space-time View LiDAR Synthesis

Zehan Zheng, Fan Lu, Weiyi Xue, Guang Chen[†], Changjun Jiang
 Tongji University
 {zhengzehan, lufan, xwy, guangchen, cjiang}@tongji.edu.cn

Abstract

Although neural radiance fields (NeRFs) have achieved triumphs in image novel view synthesis (NVS), LiDAR NVS remains largely unexplored. Previous LiDAR NVS methods employ a simple shift from image NVS methods while ignoring the dynamic nature and the large-scale reconstruction problem of LiDAR point clouds. In light of this, we propose **LiDAR4D**, a differentiable LiDAR-only framework for novel space-time LiDAR view synthesis. In consideration of the sparsity and large-scale characteristics, we design a 4D hybrid representation combined with multi-planar and grid features to achieve effective reconstruction in a coarse-to-fine manner. Furthermore, we introduce geometric constraints derived from point clouds to improve temporal consistency. For the realistic synthesis of LiDAR point clouds, we incorporate the global optimization of ray-drop probability to preserve cross-region patterns. Extensive experiments on KITTI-360 and NuScenes datasets demonstrate the superiority of our method in accomplishing geometry-aware and time-consistent dynamic reconstruction. Codes are available at <https://github.com/ispc-lab/LiDAR4D>.

1. Introduction

Dynamic scene reconstruction is of crucial importance across various fields such as AR/VR, robotics and autonomous driving. Existing advanced methods in computer vision enable high-fidelity 3D scene reconstruction and novel view synthesis (NVS), which can further serve a wide range of downstream tasks and applications. For instance, we could reconstruct driving scenarios directly from collected sensor logs, allowing for scene replay and novel data generation [43]. It shows great potential for boosting data diversity, forming data closed-loop, and improving the generalizability of the autonomous driving system.

However, the majority of current research focuses on novel view synthesis for cameras, while other sensors such as LiDAR remain largely unexplored. Similar to camera images, LiDAR point clouds are also partial observations

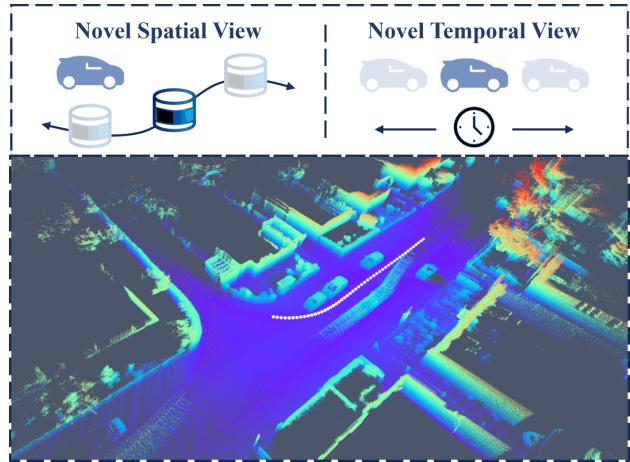


Figure 1. **Dynamic scenes of LiDAR point clouds in autonomous driving.** Large-scale vehicle motion poses a significant challenge for dynamic reconstruction and novel space-time view synthesis. White dots indicate the ego-car trajectory.

of the scene that vary across different locations and views. The reconstruction faces considerable challenges due to the sparsity, discontinuity and occlusion of LiDAR point clouds. Furthermore, as illustrated in Figure 1, dynamic scenarios combine novel spatial view and temporal view synthesis simultaneously. Meanwhile, the large motion of dynamic objects makes it difficult to align and reconstruct.

Traditional LiDAR-based 3D scene reconstruction techniques aggregate multiple sparse point cloud frames directly in the world coordinate system [19] and further convert them into explicit surface representations such as triangular meshes [25]. Subsequently, the intersection of the LiDAR beams with the mesh surface can be calculated by performing ray-casting to render novel-view LiDAR point clouds.

Nevertheless, high-quality surface reconstruction of complex large-scale scenes is challenging to accomplish, which may lead to significant geometric errors. Furthermore, the aforementioned explicit reconstruction method is limited to static scenes and struggles to accurately model the intensity or ray-drop characteristics of actual LiDAR points.

Neural radiance fields [26] implicitly reconstruct the

[†] Corresponding author.

scene and generate novel-view data through volume rendering in a continuous representation space, which also offers an alternative solution for LiDAR reconstruction. Consequently, the most recent researches [16, 39, 44] are shifting their attention towards the novel view synthesis of LiDAR. NeRF-LiDAR [44] integrates image and point cloud modalities for LiDAR synthesis, whereas LiDAR-only methods like LiDAR-NeRF [39] and NFL [16] explore the possibility of LiDAR reconstruction and generation without RGB images. Most prior methods directly apply the image NVS pipeline to LiDAR point clouds. However, LiDAR point clouds are inherently different from 2D images, which poses challenges for current LiDAR NVS methods to achieve high-quality reconstruction: (1) previous methods are limited to static scenes, ignoring the dynamic nature of autonomous driving scenarios; (2) the vast scale and high sparsity of LiDAR point clouds pose higher demands on the representations; and (3) intensity and ray-drop characteristics modeling are required for synthesis realism.

To overcome the aforementioned limitations, we propose LiDAR4D, shedding light on three pivotal insights to elevate the current LiDAR NVS pipeline. To tackle the dynamic objects, we introduce geometric constraints derived from point clouds and aggregate multi-frame dynamic features for temporal consistency. Regarding compact large-scale scene reconstruction, we design a coarse-to-fine hybrid representation combined with multi-planar and grid features to reconstruct the smooth geometry and high-frequency intensity. Additionally, we employ global optimization to preserve patterns across regions for ray-drop probability refinement. Therefore, LiDAR4D is capable of achieving geometry-aware and time-consistent reconstruction under large-scale dynamic scenarios.

We evaluate our method on diverse dynamic scenarios of KITTI-360 [23] and NuScenes [3] autonomous driving datasets. With comprehensive experiments, LiDAR4D significantly outperforms previous state-of-the-art NeRF-based implicit approaches and explicit reconstruction methods. In comparison to LiDAR-NeRF [39], we achieve 24.3% and 24.2% reduction in CD error on KITTI-360 dataset and NuScenes dataset, respectively. Similar leadership exists for other metrics of range depth and intensity.

In summary, our main contributions are three-fold:

- We propose LiDAR4D, a differentiable LiDAR-only framework for novel space-time LiDAR view synthesis, which reconstructs dynamic driving scenarios and generates realistic LiDAR point clouds end-to-end.
- We introduce 4D hybrid neural representations and motion priors derived from point clouds for geometry-aware and time-consistent large-scale scene reconstruction.
- Comprehensive experiments demonstrate the state-of-the-art performance of LiDAR4D in challenging dynamic scene reconstruction and novel view synthesis.

2. Related Work

LiDAR Simulation. Traditional simulators [7, 18, 36] such as CARLA are based on physics engines, which can generate LiDAR point clouds via ray casting within handcrafted virtual environments. However, it has diversity limitations and a heavy reliance on costly 3D assets. And there is still a large domain gap compared to real-world data. Thus, several recent works [13, 19, 25] further narrowed this gap by reconstructing the scene from real data before simulation. LiDARSim [25] reconstructs the mesh surfel representation and employs a neural network to learn the ray-drop characteristics. Besides, it is noted that there are other surface reconstruction works like NKS [15] that can convert LiDAR point clouds into mesh representations. Nonetheless, these explicit reconstruction works are troublesome for recovering precise surfaces in large-scale complex scenes, which further leads to a decrease in the accuracy of point cloud synthesis. Instead, PCGen [19] directly reconstructs from the point clouds, followed by rendering in a rasterization-like manner and first peak averaging. Although it preserves the original information better, the rendering point clouds remain relatively noisy. Moreover, all these explicit methods mentioned above are only applicable to static scenes. In contrast, our approach implicitly reconstructs the continuous representation via space-time neural radiance fields, which achieves higher-quality realistic point cloud synthesis and gets rid of static reconstruction limitations.

Neural Radiance Fields. Considerable recent research [1, 4, 5, 11, 14, 24, 26, 28, 38] based on neural radiance fields has led to breakthroughs as well as remarkable achievements in novel view synthesis (NVS) tasks. A wide variety of neural representations based on MLPs [1, 26], voxel grids [11, 24, 38], tri-planes [4, 14], vector decomposition [5], and multi-level hash grids [28] have been fully exploited for reconstruction and synthesis. Yet, most of the work focuses on object-centered reconstruction of small indoor scenes. Subsequently, several works [2, 33, 42] gradually extended it to large-scale outdoor scenarios. Despite this, neural radiance fields typically suffer from geometric ambiguity with RGB image inputs. Therefore, DS-NeRF [6] and DDP-NeRF [34] introduce the depth prior to enhancing efficiency, and URF [33] also utilizes LiDAR point clouds to facilitate reconstruction. In this paper, we employ novel hybrid representations and neural LiDAR fields to reconstruct large-scale scenarios for LiDAR NVS.

NeRF for LiDAR NVS. Very recently, a few studies [16, 39, 43, 44] have pioneered in novel view synthesis of LiDAR point clouds based on neural radiance fields, significantly surpassing traditional simulation methods. Among them, NeRF-LiDAR [44] and UniSim [43] require both RGB images and LiDAR point clouds as inputs and reconstruct the driving scene with photometric loss and depth supervision. Subsequently, novel-view LiDAR point clouds

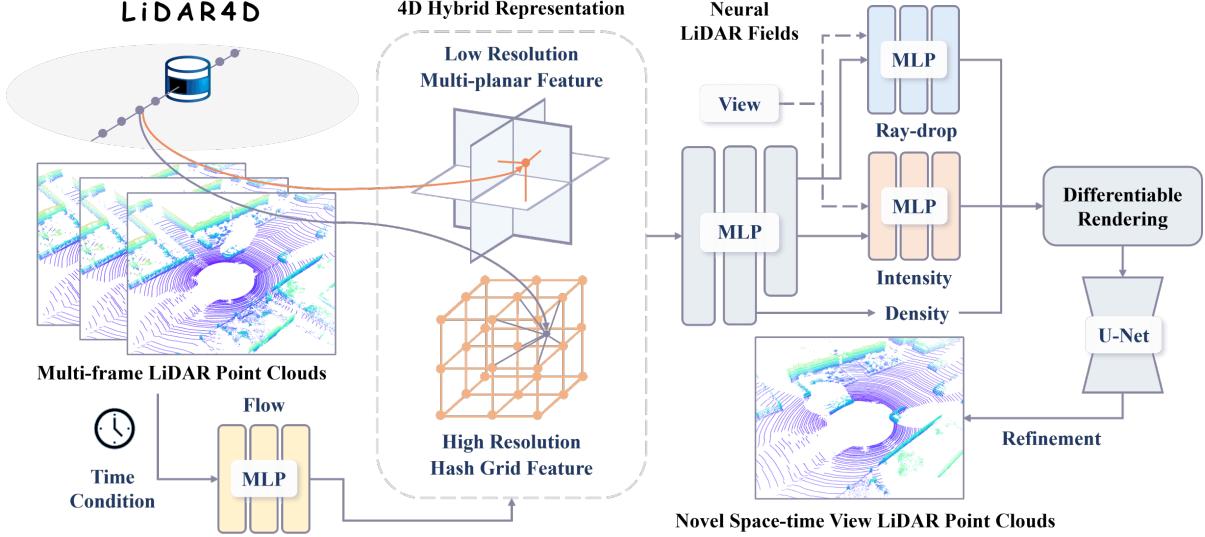


Figure 2. **Overview of our proposed LiDAR4D.** For large-scale autonomous driving scenarios, we utilize the 4D hybrid representation, which combines low-resolution multi-planar features and high-resolution hash grid features to achieve effective reconstruction. Then, multi-level spatio-temporal features aggregated by flow MLP are fed into neural LiDAR fields for density, intensity and ray-drop probability prediction. Finally, novel space-time view LiDAR point clouds are synthesized via differentiable rendering. Furthermore, we construct geometric constraints derived from point clouds for temporal consistency and the global optimization of ray-drop for generation realism.

can be generated through neural depth rendering. Among LiDAR-only methods, LiDAR-NeRF [39] and NFL [16] firstly proposed the differentiable LiDAR NVS framework, which reconstructed depth, intensity, and ray-drop probability simultaneously. Nevertheless, these approaches [16, 39, 44] are restricted to static scene reconstruction and incapable of handling dynamic objects such as moving vehicles. Although UniSim [43] does support dynamic scenes, it is largely limited by the need for ground-truth labeling of 3D object detection and decoupling the background and dynamic objects before reconstruction. Instead, our research focuses on LiDAR-only inputs for dynamic scene reconstruction and novel space-time view synthesis without the help of RGB images or ground-truth labels. And it's noteworthy that NFL [16] has contributed significantly to the detailed physical modeling of LiDAR, such as beam divergence and secondary returns, which is orthogonal to ours and could be beneficial to all LiDAR NVS works.

Dynamic Scene Reconstruction. A substantial amount of research [9, 12, 21, 22, 29, 30, 32, 37, 40] has been devoted to expanding neural radiance fields to encompass dynamic scene reconstruction. In general, dynamic NeRFs can be broadly categorized into two groups. One is the deformable neural radiance fields [9, 29, 30, 32] that map coordinates into the canonical space via continuous deformation fields. While the decoupling of deformation and radiance fields simplifies the optimization, establishing accurate long-distance correspondence remains challenging. The other is spatio-temporal neural fields [12, 21, 22, 37],

which consider time as an additional dimensional input to construct a 4D spatio-temporal representation. Thus, it is flexible to simultaneously model appearance, geometry, and motion as a continuous time-varying function. Most previous work has concentrated on relatively smaller displacements indoors, whereas large-scale vehicle movements in autonomous driving scenarios are even more challenging. Furthermore, our work is also the first to introduce dynamic neural radiance fields into the LiDAR NVS task.

3. Methodology

In this section, we start with the problem formulation of novel LiDAR view synthesis and the preliminary for NeRFs. Following this, a detailed description of our proposed LiDAR4D framework is provided.

Problem Formulation. In the dynamic driving scenario, given the collected LiDAR point cloud sequence $S = \{S_0, S_1, \dots, S_{n-1}\}$ ($S_i \in \mathbb{R}^{K \times 4}$), along with the corresponding sensor poses $P_s = \{P_0, P_1, \dots, P_{n-1}\}$ ($P_i \in SE(3)$) and timestamps $T_s = \{t_0, t_1, \dots, t_{n-1}\}$ ($t_i \in \mathbb{R}$) as inputs. Each single LiDAR frame S_i contains K points of 3D coordinates \mathbf{x} and 1D reflection intensity ρ .

The goal of LiDAR4D is to reconstruct this dynamic scene as a continuous implicit representation based on neural fields. Furthermore, given a novel sensor pose P_{novel} and any moment t_{novel} , LiDAR4D performs neural rendering to synthesize the LiDAR point cloud S_{novel} with intensities under the novel space-time view.

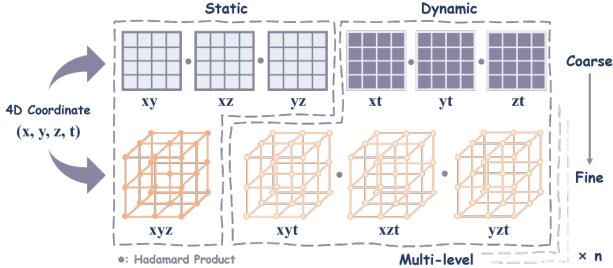


Figure 3. **4D decomposition of hybrid planar-grid representation.** Dynamic features can be further aggregated using flow MLP.

Preliminary for NeRF. Neural radiance fields, NeRFs for short, take 5D inputs of position $\mathbf{x} \in \mathbb{R}^3$ and viewing direction (θ, ϕ) as inputs and establish the mapping to the volume density σ and color \mathbf{c} . Afterward, it performs volume rendering to estimate pixel values and synthesize images in unknown novel views. In detail, it emits a light ray \mathbf{r} from the sensor center \mathbf{o} with the direction \mathbf{d} , i.e., $\mathbf{r}(t) = \mathbf{o} + td$, and then integrates the neural field outputs of N samples along this ray to approximate the pixel color \mathcal{C} . The volume rendering function can be formed as follows:

$$\hat{\mathcal{C}}(\mathbf{r}) = \sum_{i=1}^N T_i (1 - e^{-\sigma_i \delta_i}) \mathbf{c}_i, \quad T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right) \quad (1)$$

where T indicates the accumulated transmittance, σ denotes the density and δ refers to the distance between samples.

3.1. LiDAR4D Overview

Following neural radiance fields, our proposed LiDAR4D reconstructs the point cloud scene into an implicit continuous representation. Differing from original NeRFs with photometric loss for RGB images, we redefine the neural fields based on LiDAR, which are dubbed neural LiDAR fields. As depicted in Figure 2, it focuses on modeling the geometric depth, reflection intensity, and ray-drop probability of LiDAR point clouds. For large-scale dynamic driving scenarios, LiDAR4D combines coarse-resolution multi-planar features with high-resolution hash grid representation to achieve efficient and effective reconstruction. Then, we lift it to 4D and introduce temporal information encoding for novel space-time view synthesis. To ensure geometry-aware and time-consistent results, we additionally incorporate explicit geometric constraints derived from point clouds. Ultimately, we predict the ray-drop probability for each ray and perform global refinement with a runtime-optimized U-Net to improve generation realism.

3.2. 4D Hybrid Planar-Grid Representation

Figure 3 illustrates how our proposed novel hybrid representation breaks down the 4D space into *planar* and *hash grid* features, which further subdivide into *static* and

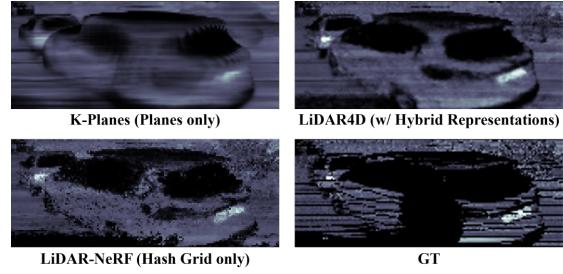


Figure 4. **Qualitative comparison for the hybrid representation.** Compared to the noisy intensity reconstruction of LiDAR-NeRF and the blurry one of K-Planes, our hybrid representation achieves more precise and smooth results.

dynamic ones. Different from the reconstruction of small indoor objects, large-scale autonomous driving scenes place higher demands on the representation ability and resolution of the features. However, the dense grid representation such as TiNeuVox [9] is unscalable for large-scale scenarios due to its cubically growing complexity. Therefore, we follow K-planes [12] and decompose the scene space into a combination of features in multiple orthogonal planes to drastically reduce the parameter quantities. The planar feature can be obtained as follows:

$$\mathbf{f}_{\text{planar}} = \mathcal{S}(\mathbf{V}, (x, y, z, t)), \quad \mathbf{V} \in \mathbb{R}^{(3M^2+3MH)C} \quad (2)$$

where \mathbf{V} stores features with M spatial resolution, H temporal resolution and C channels. \mathcal{S} refers to the sampling function that projects 4D coordinates into the corresponding planes (xy , xz , yz , xt , yt , zt) and interpolates features bilinearly. Static (xy , xz , yz) and dynamic (xt , yt , zt) features are multiplied separately by Hadamard product and multiscale features are concatenated in a coarse-to-fine manner.

Nonetheless, for scenes spanning hundreds of meters, this improvement in resolution remains inadequate, especially for the high-frequency intensity reconstruction. Owing to the hash grids proposed in Instant-NGP [28], an explicit grid structure with ultra-high resolution is possible. Additionally, the sparsity of the LiDAR point cloud scene substantially avoids the adverse effects of hash collisions.

$$\mathbf{f}_{\text{hash}} = \mathcal{S}(\mathbf{G}, (x, y, z, t)), \quad \mathbf{G} \in \mathbb{R}^{(M^3+3M^2H)C} \quad (3)$$

where the dense grid \mathbf{G} will be further compressed into limited storage via hash mapping for parameter reduction. Similarly, the 4D coordinates are projected into static (xyz) and dynamic (xyt , xzt , yzt) multi-level hash grids before trilinear interpolation and concatenation, where the dynamic features are multiplied using Hadamard product.

However, it's notable that pure hash grid representation still suffers from visual artifacts and noisy reconstruction results (as shown in Figure 4), which impede the construction of accurate object geometry. In light of this, we adopt

multi-planar features at lower resolutions for overall smooth representation and employ hash grids at higher resolutions to handle finer details, ultimately achieving high accuracy and efficiency in large-scale scene reconstruction.

3.3. Scene Flow Prior

To enhance the temporal consistency of the current 4D spatio-temporal representations, we further incorporate a flow MLP [20, 46] for motion estimation. It takes the encoded spatio-temporal coordinates as input and constructs the mapping from coordinate fields \mathbb{R}^4 to motion fields \mathbb{R}^3 .

$$\Delta \mathbf{x} = \text{MLP}_{\text{flow}}(\gamma(\mathbf{x}, t)), \quad \mathbf{x}' = \mathbf{x} + \Delta \mathbf{x} \quad (4)$$

Since the vehicle motion range may span a long distance in autonomous driving scenarios, it is extremely hard to establish long-term correspondences to the canonical space in deformable neural radiance fields. Thus, as with Li *et al.* [21, 22], we utilize the flow MLP to predict motion only between adjacent frames and aggregate multi-frame *dynamic* features to achieve time-consistent reconstruction.

In addition, explicit geometric constraints can be further derived from the input LiDAR point clouds. By feeding point clouds into the flow MLP to produce the scene flow prediction, we can regulate the chamfer distance as a geometric loss ($\mathcal{L}_{\text{flow}}$). It imposes motion prior and additional supervision on LiDAR4D, thus accomplishing the geometry-aware reconstruction. Chamfer Distance between two frames of point cloud S and \hat{S} is defined as follows:

$$\text{CD} = \frac{1}{K} \sum_{\hat{p}_i \in \hat{S}} \min_{p_i \in S} \| \hat{p}_i - p_i \|_2^2 + \frac{1}{K} \sum_{p_i \in S} \min_{\hat{p}_i \in \hat{S}} \| p_i - \hat{p}_i \|_2^2 \quad (5)$$

$$\mathcal{L}_{\text{flow}} = \sum_{j \in \pm 1} \text{CD}(S_i + \text{MLP}_{\text{flow}}(S_i), S_{i+j}), i \in (0, n-1) \quad (6)$$

3.4. Neural LiDAR Fields

LiDAR emits laser pulses and measures the time-of-flight (ToF) to determine object distance, along with the intensity of reflected lights. Spinning LiDAR has a 360-degree horizontal field of view (FOV) and a limited range of vertical field of view, which perceives the environment with certain angular resolution lasers. In the same way for neural LiDAR fields, we transmit lasers at specific angular intervals within the FOV, using the center of the LiDAR sensor as the origin \mathbf{o} . The direction \mathbf{d} of the laser is determined by the azimuth angle θ and elevation angle ϕ under the polar coordinate system, which is shown below.

$$\mathbf{d} = (\cos \theta \cos \phi, \sin \theta \cos \phi, \cos \phi)^T \quad (7)$$

Then we query 3D point coordinates sampled along the laser and feed them into neural fields to predict the density at the corresponding location. Following this, the density

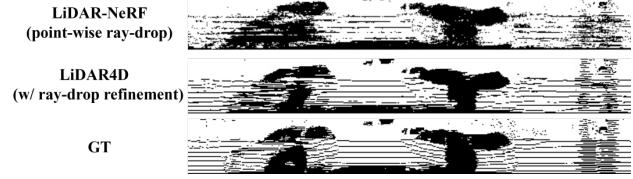


Figure 5. Qualitative comparison for the ray-drop refinement. The point-wise prediction of ray-drop probability by MLP cannot preserve global patterns well. Instead, LiDAR4D drastically improves generation realism via runtime-optimized U-Net.

along the ray is integrated to obtain the expectation of depth value \mathcal{D} , which serves as the laser beam's return distance.

$$\hat{\mathcal{D}}(\mathbf{r}) = \sum_{i=1}^N T_i (1 - e^{-\sigma_i \delta_i}) z_i, \quad \alpha_i = 1 - e^{-\sigma_i \delta_i} \quad (8)$$

where z_i is the depth value of queried points on the ray \mathbf{r} , and α is the definition of the opacity.

In addition, we predict the intensity \mathcal{I} and ray-drop probability \mathcal{P} separately for each point and similarly conduct alpha-composition along the ray.

$$\hat{\mathcal{I}}(\mathbf{r}) = \sum_{i=1}^N T_i \alpha_i i_i, \quad \hat{\mathcal{P}}(\mathbf{r}) = \sum_{i=1}^N T_i \alpha_i p_i \quad (9)$$

where i_i and p_i are the point-wise intensity and ray-drop probability outputs of LiDAR4D. We use separate MLPs to take temporal aggregated planar and hash features, as well as position-encoded viewpoints as inputs for prediction.

$$i = \text{MLP}_{\text{intensity}}(\mathbf{f}_{\text{planar}}, \mathbf{f}_{\text{hash}}, \gamma(\mathbf{d})) \quad (10)$$

$$p = \text{MLP}_{\text{raydrop}}(\mathbf{f}_{\text{planar}}, \mathbf{f}_{\text{hash}}, \gamma(\mathbf{d})) \quad (11)$$

$$\gamma(x) = (\sin(2^0 x), \cos(2^0 x), \dots, \sin(2^{L-1} x), \cos(2^{L-1} x)) \quad (12)$$

3.5. Ray-drop Refinement

During laser ranging, a portion of the emitted rays is not reflected back to the sensor, which is termed the ray-drop characteristic. In fact, the ray-drop of LiDAR is significantly impacted by various aspects, including distance, surface properties and sensor noise. As in LiDAR-NeRF [39], ray-drop prediction is directly accomplished with a point-wise MLP head, which is essentially noisy and unreliable. To address this issue, we employ the U-Net [35] with residuals to refine the ray-drop mask globally and better preserve the consistent pattern across regions. It takes the *full* ray-drop probability, depth and intensity prediction of LiDAR4D as inputs (different from previous work) and refines the final mask via binary cross-entropy loss as follows:

$$\mathcal{L}_{\text{refine}} = \text{BCELoss}(\hat{\mathcal{M}}_{\text{Pred}}, \mathcal{M}_{\text{GT}}) \quad (13)$$

Method	Type	Point Cloud				Depth				Intensity			
		CD↓	F-score↑	RMSE↓	MedAE↓	LPIPS↓	SSIM↑	PSNR↑	RMSE↓	MedAE↓	LPIPS↓	SSIM↑	PSNR↑
LiDARsim [25]	$\mathcal{E} / \mathcal{S} / \mathcal{M}$	3.2228	0.7157	6.9153	0.1279	0.2926	0.6342	21.4608	0.1666	0.0569	0.3276	0.3502	15.5853
NKSR [15]	$\mathcal{E} / \mathcal{S} / \mathcal{M}$	1.8982	0.6855	5.8403	0.0996	0.2752	0.6409	23.0368	0.1742	0.0590	0.3337	0.3517	15.2081
PCGen [19]	$\mathcal{E} / \mathcal{S}$	0.4636	0.8023	5.6583	0.2040	0.5391	0.4903	23.1675	0.1970	0.0763	0.5926	0.1351	14.1181
LiDAR-NeRF [39]	$\mathcal{I} / \mathcal{S}$	0.1438	0.9091	4.1753	0.0566	0.2797	0.6568	25.9878	0.1404	0.0443	0.3135	0.3831	17.1549
D-NeRF [32]	$\mathcal{I} / \mathcal{D}$	0.1442	0.9128	4.0194	0.0508	0.3061	0.6634	26.2344	0.1369	0.0440	0.3409	0.3748	17.3554
TiNeuVox-B [9]	$\mathcal{I} / \mathcal{D}$	0.1748	0.9059	4.1284	0.0502	0.3427	0.6514	26.0267	0.1363	0.0453	0.4365	0.3457	17.3535
K-Planes [12]	$\mathcal{I} / \mathcal{D}$	0.1302	0.9123	4.1322	0.0539	0.3457	0.6385	26.0236	0.1415	0.0498	0.4081	0.3008	17.0167
LiDAR4D (Ours)	$\mathcal{I} / \mathcal{D}$	0.1089	0.9272	3.5256	0.0404	0.1051	0.7647	27.4767	0.1195	0.0327	0.1845	0.5304	18.5561

Table 1. **Quantitative comparison on KITTI-360 dataset.** We compare our method to different types of previous approaches and color the top results as **best** and **second best**. \mathcal{E} : Explicit, \mathcal{I} : Implicit, \mathcal{S} : Static, \mathcal{D} : Dynamic, \mathcal{M} : Mesh.

Method	Type	Point Cloud				Depth				Intensity			
		CD↓	F-score↑	RMSE↓	MedAE↓	LPIPS↓	SSIM↑	PSNR↑	RMSE↓	MedAE↓	LPIPS↓	SSIM↑	PSNR↑
LiDARsim [25]	$\mathcal{E} / \mathcal{S} / \mathcal{M}$	12.1383	0.6512	10.5539	0.3572	0.1871	0.5653	17.7841	0.0659	0.0115	0.1160	0.5170	23.7791
NKSR [15]	$\mathcal{E} / \mathcal{S} / \mathcal{M}$	11.4910	0.6178	9.3731	0.5763	0.2111	0.5637	18.7774	0.0680	0.0119	0.1290	0.5031	23.4905
PCGen [19]	$\mathcal{E} / \mathcal{S}$	2.1998	0.6341	8.8364	0.4011	0.1792	0.5440	19.2799	0.0768	0.0147	0.1308	0.4410	22.4428
LiDAR-NeRF [39]	$\mathcal{I} / \mathcal{S}$	0.3225	0.8576	7.1566	0.0338	0.0702	0.7188	21.2129	0.0467	0.0076	0.0483	0.7264	26.9927
D-NeRF [32]	$\mathcal{I} / \mathcal{D}$	0.3296	0.8513	7.1089	0.0368	0.0789	0.7130	21.2594	0.0467	0.0080	0.0492	0.7180	26.9951
TiNeuVox-B [9]	$\mathcal{I} / \mathcal{D}$	0.3920	0.8627	7.2093	0.0290	0.1549	0.6873	21.0932	0.0462	0.0080	0.1294	0.7107	26.8620
K-Planes [12]	$\mathcal{I} / \mathcal{D}$	0.2982	0.8887	6.7960	0.0209	0.1218	0.7258	21.6203	0.0438	0.0076	0.1127	0.7364	27.4227
LiDAR4D (Ours)	$\mathcal{I} / \mathcal{D}$	0.2443	0.8915	6.7831	0.0258	0.0569	0.7396	21.7189	0.0426	0.0071	0.0459	0.7498	27.7977

Table 2. **Quantitative comparison on NuScenes dataset.** The notations are consistent with the KITTI-360 Table 1 above.

where \mathcal{M} indicates the global mask rendered from range view and \mathcal{M}_{GT} is calculated from the input point clouds.

We emphasize that the lightweight network is randomly initialized and optimized at runtime efficiently for reconstruction. As illustrated in Figure 5, the global optimization greatly improves the prediction results and further enhances the fidelity of the generated LiDAR point cloud.

3.6. Optimization

For the optimization of LiDAR4D, the total reconstruction loss is the weighted combination of the depth loss, intensity loss, ray-drop loss, flow loss and refinement loss, which can be formalized as follows:

$$\mathcal{L}_{\text{depth}} = \sum_{\mathbf{r} \in R} \|\hat{D}(\mathbf{r}) - D(\mathbf{r})\|_1 \quad (14)$$

$$\mathcal{L}_{\text{intensity}} = \sum_{\mathbf{r} \in R} \|\hat{I}(\mathbf{r}) - I(\mathbf{r})\|_2^2 \quad (15)$$

$$\mathcal{L}_{\text{raydrop}} = \sum_{\mathbf{r} \in R} \|\hat{P}(\mathbf{r}) - P(\mathbf{r})\|_2^2 \quad (16)$$

$$\begin{aligned} \mathcal{L} = & \lambda_{\alpha} \mathcal{L}_{\text{depth}} + \lambda_{\beta} \mathcal{L}_{\text{intensity}} + \lambda_{\gamma} \mathcal{L}_{\text{raydrop}} \\ & + \lambda_{\eta} \mathcal{L}_{\text{flow}} + \lambda_{\tau} \mathcal{L}_{\text{refine}} \end{aligned} \quad (17)$$

4. Experiments

4.1. Experimental Setup

Datasets. We conducted comprehensive experiments on the public autonomous driving datasets KITTI-360 [23] and

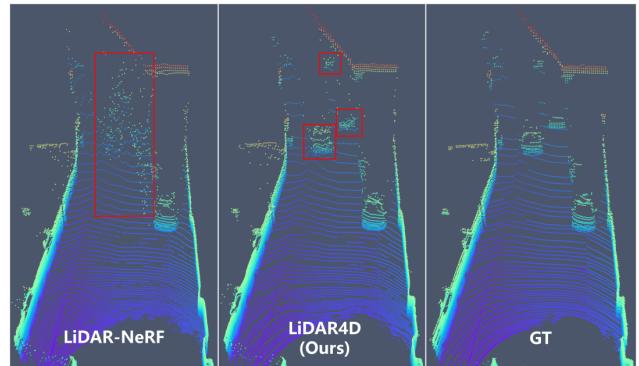


Figure 6. **Qualitative novel view LiDAR point cloud synthesis results on KITTI-360 dataset.** As highlighted in the red bounding box, LiDAR-NeRF fails to reconstruct the dynamic vehicles. In contrast, LiDAR4D generates more accurate geometry for moving cars, even in sparse point clouds far away.

NuScenes [3], from which we collected multiple dynamic point cloud sequences containing largely moving vehicles. KITTI-360 is equipped with a LIDAR of 64-beam, a 26.4-degree vertical FOV, and an acquisition frequency of 10Hz. We selected 51 consecutive frames as a single scene and held out 4 samples at a 10-frame interval for NVS evaluation. Meanwhile, Nuscenies' LiDAR has 32 beams, a 40-degree vertical FOV, and a 20-Hz acquisition frequency. To cover a larger range of reconstruction, i.e., spanning 100~200 meters, we still selected samples at a frequency of 10 Hz, which is consistent with the KITTI-360.

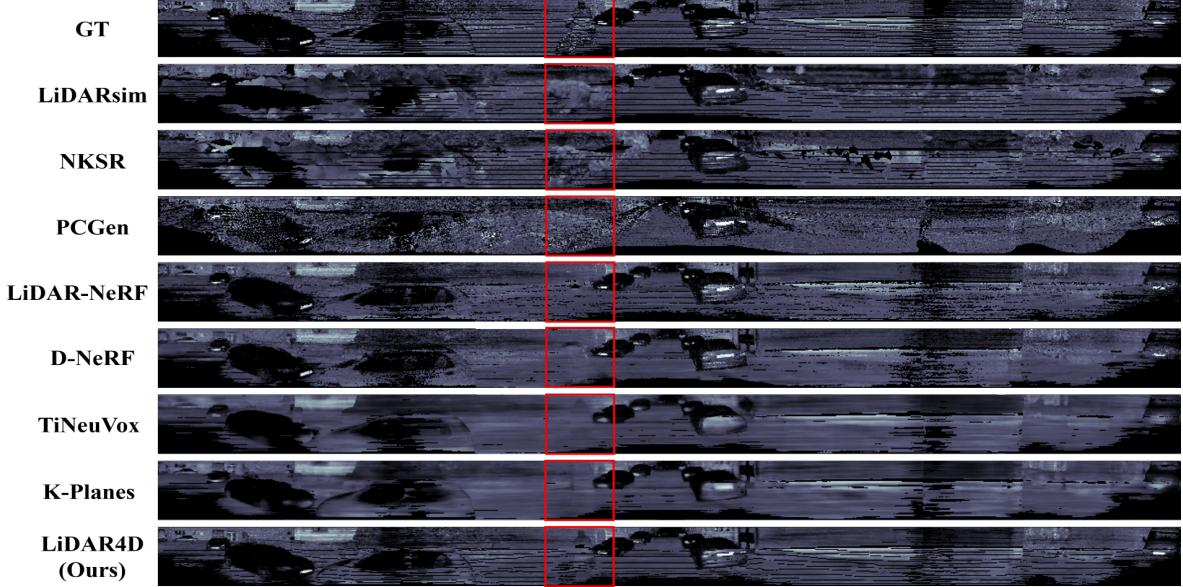


Figure 7. Qualitative comparison for LiDAR intensity reconstruction and synthesis.

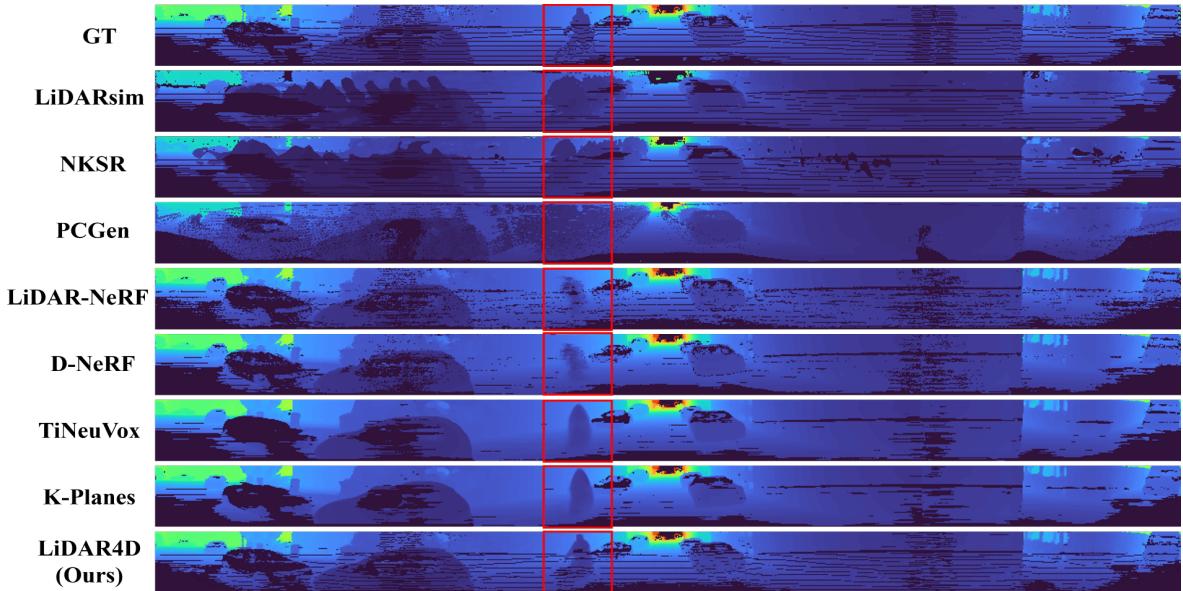


Figure 8. Qualitative comparison for LiDAR depth reconstruction and synthesis.

Baselines. We present a comprehensive comparison of LiDAR4D with different types of baselines, encompassing explicit and implicit reconstruction approaches as well as dynamic NeRFs. We reproduce the mesh-based reconstruction method LiDARsim [25] and also replace the surface reconstruction model with the state-of-the-art method NKSR [15] for convincing. It’s noted that we employ U-Net trained from original point clouds to further predict ray-drop for these two methods. Meanwhile, PCGen [19] reconstructs

directly based on the point clouds and predicts ray-drop with the MLP. LiDAR-NeRF [39] is our primary comparison, and we directly adopt the official implements. In addition, we migrate dynamic neural radiance field methods such as D-NeRF [32], K-Planes [12], and TiNeuVox [9] to LiDAR NVS pipeline for a thorough comparison.

Metrics. We offer multi-faceted metrics for evaluation. Chamfer Distance [8] measures the 3D geometric error between the generated and the ground-truth point clouds by

Method	Type	Point Cloud				Depth				Intensity			
		CD↓	F-score↑	RMSE↓	MedAE↓	LPIPS↓	SSIM↑	PSNR↑	RMSE↓	MedAE↓	LPIPS↓	SSIM↑	PSNR↑
LiDARsim [25]	$\mathcal{E} / \mathcal{S} / \mathcal{M}$.	2.2249	0.8667	6.5470	0.0759	0.2289	<u>0.7157</u>	21.7746	0.1532	0.0506	0.2502	<u>0.4479</u>	16.3045
NKSR [15]	$\mathcal{E} / \mathcal{S} / \mathcal{M}$.	0.5780	0.8685	4.6647	0.0698	0.2295	0.7052	22.5390	0.1565	<u>0.0536</u>	<u>0.2429</u>	0.4200	16.1159
PCGen [19]	$\mathcal{E} / \mathcal{S}$.	0.2090	0.8597	4.8838	0.1785	0.5210	0.5062	24.3050	0.2005	0.0818	0.6100	0.1248	13.9606
LiDAR-NeRF [39]	$\mathcal{I} / \mathcal{S}$.	<u>0.0923</u>	<u>0.9226</u>	<u>3.6801</u>	<u>0.0667</u>	<u>0.3523</u>	0.6043	<u>26.7663</u>	<u>0.1557</u>	0.0549	0.4212	0.2768	16.1683
LiDAR4D (Ours)	$\mathcal{I} / \mathcal{D}$.	0.0894	0.9264	3.2370	0.0507	0.1313	0.7218	27.8840	0.1343	0.0404	0.2127	0.4698	17.4529

Table 3. **Quantitative comparison on KITTI-360 Static Scene Sequence.** We bold the best results and underline the second-best.

nearest neighbor, and we also report the F-score value with an error threshold of 5cm. In addition, we introduce Root Mean Square Error (RMSE) and Median Absolute Error (MedAE) to calculate the pixel-by-pixel error of the projected range images, as well as LPIPS [45], SSIM [41] and PSNR to measure the overall variance. We evaluate both the depth and intensity reconstruction results.

4.2. Implementation Details

Consistent with LiDAR-NeRF [39], the entire point cloud scene is scaled within the unit cube space. And we uniformly sampled 768 points along each laser. The optimization of LiDAR4D is implemented on Pytorch [31] with Adam [17] optimizer. The maximum iteration is set to 30k for each scene, with a batch size of 1024 rays, followed by the fast ray-drop refinement with 300 epochs. We construct multi-planar representations following K-Planes [12] and hash grids built on tiny-cuda-nn [27]. The multi-level features of planes and hash grids are concatenated before feeding into MLPs. All experiments were conducted on a single NVIDIA GeForce RTX 4090 GPU. For more implementation details, please refer to *Supplementary Material*.

4.3. Evaluation of Novel-View LiDAR Synthesis

Results on KITTI-360 dataset. The quantitative comparison on KITTI-360 dataset is displayed in Table 1. Our proposed LiDAR4D exhibits remarkable performance across all metrics in comparison to prior SOTA methods, demonstrating its superiority in dynamic reconstruction. In comparison to LiDAR-NeRF, our approach has led to a 24.3% reduction in the CD error of the novel-view point cloud synthesis. As illustrated in Figure 6, LiDAR4D achieves accurate reconstruction of every dynamic vehicle, whereas LiDAR-NeRF encounters failure. As shown in Figures 7 and 8, the resolution limitation causes blurring and over-smooth results of dynamic NeRFs such as TiNeuVox and K-Planes. This demonstrates again the effectiveness of our designed hybrid representation for large-scale scenes. Additionally, we follow the setting in LiDAR-NeRF to repeat experiments in static scenarios, and Table 3 verifies that there is no performance degradation in LiDAR4D.

Results on NuScenes dataset. To further validate the generalizability of LiDAR4D, we conducted the same experi-

ment on NuScenes. As illustrated in Table 2, our method still achieves the best reconstruction quality even with a completely different LiDAR configuration. Ultimately, our method still excels in the reconstruction of depth and intensity, as evidenced by the 24.2% reduction in CD error to LiDAR-NeRF. Please refer to *Supplementary Material* for more quantitative and qualitative experimental results on KITTI-360 and NuScenes datasets, as well as the ablation study and further applications.

5. Limitations

Despite the fact that LiDAR4D has exhibited exceptional performance in a substantial number of experiments, the long-distance vehicle motion and occlusion problem of point clouds remain open questions. There is still a significant gap in the reconstruction of dynamic objects compared to static ones. In addition, foreground and background may be difficult to separate well. Furthermore, based on real-world datasets, *quantitative* evaluation of NVS is limited within the ego-car trajectory and does not allow for the de-coupling of novel spatial and temporal view synthesis.

6. Conclusion

In this paper, we revisit the limitations of existing LiDAR NVS methods and propose a novel framework to address three major challenges, namely dynamic reconstruction, large-scale scene characterization, and realistic synthesis. Our proposed method LiDAR4D proves its superiority under extensive experiments, achieving geometry-aware and time-consistent reconstruction of large-scale dynamic point cloud scenes and generating novel space-time view LiDAR point clouds closer to the real distribution. We believe that more future work will focus on combining LiDAR point clouds with neural radiance fields and explore more possibilities for dynamic scene reconstruction and synthesis.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (No.62372329), in part by the National Key Research and Development Program of China (No.2021YFB2501104), in part by Shanghai Rising Star Program (No.21QC1400900), in part by Tongji-Qomolo Autonomous Driving Commercial Vehicle Joint Lab Project, and in part by Xiaomi Young Talents Program.

References

- [1] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5855–5864, 2021. 2
- [2] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022. 2
- [3] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giacarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020. 2, 6
- [4] Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. Efficient geometry-aware 3d generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16123–16133, 2022. 2
- [5] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *European Conference on Computer Vision*, pages 333–350. Springer, 2022. 2
- [6] Kangle Deng, Andrew Liu, Jun-Yan Zhu, and Deva Ramanan. Depth-supervised nerf: Fewer views and faster training for free. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12882–12891, 2022. 2
- [7] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *Conference on robot learning*, pages 1–16. PMLR, 2017. 2
- [8] Haoqiang Fan, Hao Su, and Leonidas J Guibas. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 605–613, 2017. 7
- [9] Jiemin Fang, Taoran Yi, Xinggang Wang, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Matthias Nießner, and Qi Tian. Fast dynamic radiance fields with time-aware neural voxels. In *SIGGRAPH Asia 2022 Conference Papers*, pages 1–9, 2022. 3, 4, 6, 7, 1
- [10] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 5
- [11] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5501–5510, 2022. 2
- [12] Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. K-planes: Explicit radiance fields in space, time, and appearance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12479–12488, 2023. 3, 4, 6, 7, 8, 1
- [13] Benoît Guillard, Sai Vemprala, Jayesh K Gupta, Ondrej Mikšík, Vibhav Vineet, Pascal Fua, and Ashish Kapoor. Learning to simulate realistic lidars. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8173–8180. IEEE, 2022. 2
- [14] Wenbo Hu, Yuling Wang, Lin Ma, Bangbang Yang, Lin Gao, Xiao Liu, and Yuewen Ma. Tri-miprf: Tri-mip representation for efficient anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19774–19783, 2023. 2
- [15] Jiahui Huang, Zan Gojcic, Matan Atzmon, Or Litany, Sanja Fidler, and Francis Williams. Neural kernel surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4369–4379, 2023. 2, 6, 7, 8, 1
- [16] Shengyu Huang, Zan Gojcic, Zian Wang, Francis Williams, Yoni Kasten, Sanja Fidler, Konrad Schindler, and Or Litany. Neural lidar fields for novel view synthesis. *arXiv preprint arXiv:2305.01643*, 2023. 2, 3
- [17] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 8, 6
- [18] Nathan Koenig and Andrew Howard. Design and use paradigms for gazebo, an open-source multi-robot simulator. In *2004 IEEE/RSJ international conference on intelligent robots and systems (IROS)(IEEE Cat. No. 04CH37566)*, pages 2149–2154. IEEE, 2004. 2
- [19] Chenqi Li, Yuan Ren, and Bingbing Liu. Pgen: Point cloud generator for lidar simulation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11676–11682. IEEE, 2023. 1, 2, 6, 7, 8
- [20] Xueqian Li, Jhony Kaesemeyer Pontes, and Simon Lucey. Neural scene flow prior. *Advances in Neural Information Processing Systems*, 34:7838–7851, 2021. 5
- [21] Zhengqi Li, Simon Niklaus, Noah Snavely, and Oliver Wang. Neural scene flow fields for space-time view synthesis of dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6498–6508, 2021. 3, 5
- [22] Zhengqi Li, Qianqian Wang, Forrester Cole, Richard Tucker, and Noah Snavely. Dynibar: Neural dynamic image-based rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4273–4284, 2023. 3, 5
- [23] Yiyi Liao, Jun Xie, and Andreas Geiger. Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3292–3310, 2022. 2, 6
- [24] Lingjie Liu, Jitao Gu, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. Neural sparse voxel fields. *Advances in Neural Information Processing Systems*, 33:15651–15663, 2020. 2
- [25] Sivabalan Manivasagam, Shenlong Wang, Kelvin Wong, Wenyuan Zeng, Mikita Sazanovich, Shuhan Tan, Bin Yang,

- Wei-Chiu Ma, and Raquel Urtasun. Lidarsim: Realistic lidar simulation by leveraging the real world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11167–11176, 2020. 1, 2, 6, 7, 8
- [26] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 1, 2
- [27] Thomas Müller. tiny-cuda-nn, 2021. 8
- [28] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4):1–15, 2022. 2, 4
- [29] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5865–5874, 2021. 3
- [30] Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M Seitz. Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields. *arXiv preprint arXiv:2106.13228*, 2021. 3
- [31] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 8
- [32] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10318–10327, 2021. 3, 6, 7, 1
- [33] Konstantinos Rematas, Andrew Liu, Pratul P Srinivasan, Jonathan T Barron, Andrea Tagliasacchi, Thomas Funkhouser, and Vittorio Ferrari. Urban radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12932–12942, 2022. 2
- [34] Barbara Roessler, Jonathan T Barron, Ben Mildenhall, Pratul P Srinivasan, and Matthias Nießner. Dense depth priors for neural radiance fields from sparse input views. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12892–12901, 2022. 2
- [35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 5
- [36] Shital Shah, Debadeepta Dey, Chris Lovett, and Ashish Kapoor. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and Service Robotics: Results of the 11th International Conference*, pages 621–635. Springer, 2018. 2
- [37] Ruizhi Shao, Zerong Zheng, Hanzhang Tu, Boning Liu, Hongwen Zhang, and Yebin Liu. Tensor4d: Efficient neural 4d decomposition for high-fidelity dynamic reconstruction and rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16632–16642, 2023. 3
- [38] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5459–5469, 2022. 2
- [39] Tang Tao, Longfei Gao, Guangrun Wang, Peng Chen, Dayang Hao, Xiaodan Liang, Mathieu Salzmann, and Kaicheng Yu. Lidar-nerf: Novel lidar view synthesis via neural radiance fields. *arXiv preprint arXiv:2304.10406*, 2023. 2, 3, 5, 6, 7, 8, 1
- [40] Haithem Turki, Jason Y Zhang, Francesco Ferroni, and Deva Ramanan. Suds: Scalable urban dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12375–12385, 2023. 3
- [41] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 8
- [42] Zian Wang, Tianshang Shen, Jun Gao, Shengyu Huang, Jacob Munkberg, Jon Hasselgren, Zan Gojcic, Wenzheng Chen, and Sanja Fidler. Neural fields meet explicit geometric representations for inverse rendering of urban scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8370–8380, 2023. 2
- [43] Ze Yang, Yun Chen, Jingkang Wang, Sivabalan Manivasagam, Wei-Chiu Ma, Anqi Joyce Yang, and Raquel Urtasun. Unisim: A neural closed-loop sensor simulator. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1389–1399, 2023. 1, 2, 3
- [44] Junge Zhang, Feihu Zhang, Shaochen Kuang, and Li Zhang. Nerf-lidar: Generating realistic lidar point clouds with neural radiance fields. *arXiv preprint arXiv:2304.14811*, 2023. 2, 3
- [45] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 8, 2
- [46] Zehan Zheng, Danni Wu, Ruisi Lu, Fan Lu, Guang Chen, and Changjun Jiang. Neuralpc: Spatio-temporal neural field for 3d point cloud multi-frame non-linear interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 909–918, 2023. 5

Appendix

In this document, we start with the **ablation study** in Appendix A to demonstrate the effectiveness of the proposed key modules as a complement. Following this, we conduct a more comprehensive **analysis of the qualitative and quantitative experiments** based on Section 4 and provide additional experimental results in Appendix B. Specific **implementation details** and dataset information are subsequently presented in Appendix C for reproduction. Finally, we showcase further **applications** of LiDAR4D in Appendix D, thereby highlighting its versatility, flexibility, and great potential.

A. Ablation Study

The advantages of our method in comparison to LiDAR-NeRF [39] are illustrated in Figures 4 to 6, which corresponds to the key modules of LiDAR4D, *i.e.*, dynamic reconstruction, hybrid representation, and ray-drop refinement. In order to provide a more rigorous demonstration of the efficacy of our method, we perform the ablation study for each module and present quantitative results in Table S4.



Figure S9. Qualitative comparison for the geometric regularization of CD loss.

The results in the first row represent the basic version of LiDAR4D with only hash grid representation, which is similar to LiDAR-NeRF. The introduction of the hybrid representation (\mathcal{H}) significantly enhances the reconstruction quality, especially for the point cloud and depth metrics in Row 2. Subsequently, we further adopted time-conditioned dynamic-part representations (\mathcal{D}_T) and flow-constrained temporal feature aggregation (\mathcal{D}_F), which notably strengthened the capability of dynamic reconstruction in Row 3&4. Among them, the incorporation of CD loss as geometric regularization benefits the optimization of flow MLP and leads to more accurate results for dynamic objects, as shown in Figure S9. Ultimately, the global optimization of ray-drop (\mathcal{R}) based on U-Net assists LiDAR4D in achieving SOTA performance in the last row.

B. Additional Analysis and Experiments

B.1. Quantitative and Qualitative Comparison

Static Scenes. As shown in Figure S12, traditional explicit reconstruction methods such as LiDARsim [25] convert

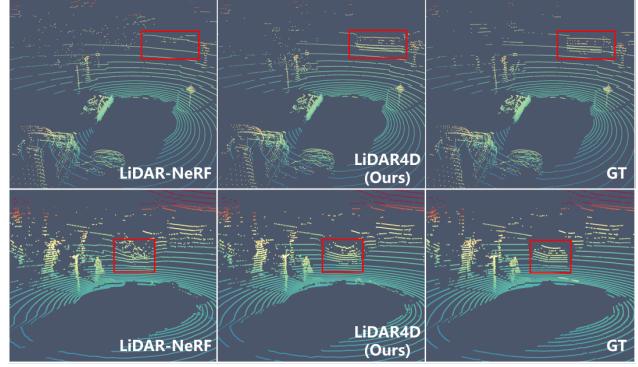


Figure S10. Qualitative novel view LiDAR point cloud synthesis results on NuScenes dataset.

point cloud scenes into mesh representations but struggle to accurately reconstruct object details in large-scale scenes (Row 2). We additionally adopt the state-of-the-art surface reconstruction algorithm NKS [15] upon LiDARsim to improve the reconstruction quality. Nevertheless, the novel-view results are still significantly different from the ground truth (Row 3). Furthermore, it is unable to establish the correlation between intensity and viewpoint. PC-Gen [19] reconstructs directly based on the point cloud, while the generated results are heavily affected by noise (Row 4). On the contrary, the implicit reconstruction method like LiDAR-NeRF [39] (Row 5) alleviates the challenges above and achieves a substantial lead. Our LiDAR4D further surpasses the previous approaches, especially in reconstruction details such as vehicle shape and window reflections (Row 6). The quantitative results illustrated in Table 3 demonstrate a similar trend. Compared to LiDAR-NeRF, the hybrid representation and ray-drop refinement of LiDAR4D lead to a 12.0% and 13.7% drop in the depth and intensity RMSE metrics.

Dynamic Scenes. Explicit reconstruction methods fail completely in dynamic scenes (Figures 7, 8, S13 and S14), which yields extremely poor validation results (Tables 1 and 2) due to the stacking of dynamic objects. In contrast, implicit reconstruction methods largely avoid the artifacts and noise of dynamic objects. However, existing methods like LiDAR-NeRF are designed for static scenes, resulting in the obscurity or absence of moving objects (Figures 6 and S10). Although D-NeRF [32] incorporates a deformation field, its impact is quite limited. The primary issue lies in the lack of constraints and the difficulty of establishing long-distance correspondence. Moreover, the state-of-the-art dynamic methods TiNeuVox [9] and K-planes [12] are limited by their representation resolution, which makes it difficult to reconstruct details in large-scale scenes, such as vehicle and pedestrian geometry (Row 7&8 in Figure 8), as well as high-frequency details in intensity (Row 7&8 in Figure S13). Our proposed LiDAR4D instead accomplishes

\mathcal{H} .	$\mathcal{D}_{\mathcal{T}}$.	$\mathcal{D}_{\mathcal{F}}$.	\mathcal{R} .	Point Cloud				Depth				Intensity			
				CD \downarrow	F-score \uparrow	RMSE \downarrow	MedAE \downarrow	LPIPS \downarrow	SSIM \uparrow	PSNR \uparrow	RMSE \downarrow	MedAE \downarrow	LPIPS \downarrow	SSIM \uparrow	PSNR \uparrow
\times	\times	\times	\times	0.1840	0.8979	4.0602	0.0639	0.2692	0.6483	26.1957	0.1398	0.0431	0.2969	0.3829	17.2018
\checkmark	\times	\times	\times	0.1429	0.9116	3.9702	0.0499	0.2586	0.6645	26.3647	0.1368	0.0411	0.2760	0.4036	17.3675
\checkmark	\checkmark	\times	\times	0.1213	0.9221	3.6947	0.0448	0.2397	0.7027	27.0285	0.1286	0.0368	0.2688	0.4553	17.8999
\checkmark	\checkmark	\checkmark	\times	0.1187	0.9260	3.6745	0.0425	0.2130	0.7104	27.1009	0.1281	0.0359	0.2426	0.4726	17.9394
\checkmark	\checkmark	\checkmark	\checkmark	0.1089	0.9272	3.5256	0.0404	0.1051	0.7647	27.4767	0.1195	0.0327	0.1845	0.5304	18.5561

Table S4. **Ablation study on KITTI-360 Dataset.** \mathcal{H} : hybrid representation, $\mathcal{D}_{\mathcal{T}}$: time-conditioned dynamic-part representations, $\mathcal{D}_{\mathcal{F}}$: flow-constrained temporal feature aggregation, \mathcal{R} : global ray-drop refinement.

w/ GT Mask	Depth				Intensity				LPIPS \downarrow
	LPIPS \downarrow	SSIM \uparrow	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow	PSNR \uparrow	LPIPS \downarrow	SSIM \uparrow	
	S.	LiDAR-NeRF	0.025	0.971	34.808	0.146	0.667	19.935	SSIM \uparrow
\mathcal{S} .	Ours	0.024	0.974	36.303	0.145	0.704	20.677		
	LiDAR-NeRF	0.126	0.843	29.361	0.192	0.583	18.891		
\mathcal{D} .	Ours	0.019	0.981	36.222	0.137	0.715	21.407		

Table S5. **Experiments with GT ray-drop mask on KITTI-360 Dataset.** \mathcal{S} : Static sequences, \mathcal{D} : Dynamic sequences.

geometric-aware and time-consistent dynamic reconstruction through 4D hybrid representation and flow-constrained temporal feature aggregation. As shown in Tables 1 and 2, LiDAR4D ranks first across almost all metrics. A considerable visualization intuitively exhibits the superior generation quality of LiDAR4D, encompassing both long-distance moving vehicles and small bicyclists (the last row in Figures 8 and S14).

Difference on Ray-drop. Existing methods differ in ray-drop modeling. PCGen [19] employs MLP to estimate ray-drop, while LiDARsim [25] adopts U-Net, which takes depth and intensity values as input. In contrast, LiDAR4D predicts the ray drop probability of each point in space through neural fields and integrates them along the ray as the inputs of U-Net. Then, the U-Net is optimized in runtime to refine the prediction for individual scenarios. As can be seen from Figure S12, the MLP-based method may handle high-frequency details, but it also results in noisy prediction (Row 4&5). The U-Net-based method preserves global patterns better (Row 2&3) and consequently achieves superior results in LPIPS [45] metrics in Tables 1 and 3 in particular. However, this data-driven paradigm is dependent on the distribution of the training samples and is difficult to predict accurately in detail, *i.e.*, the vehicle windows. LiDAR4D combines the advantages of both to achieve more realistic ray-drop modeling, as shown in Figure 5.

B.2. Experiments without Ray-drop Effect

In order to eliminate the effect of ray-drop on the evaluation metrics, we conduct supplementary experiments by

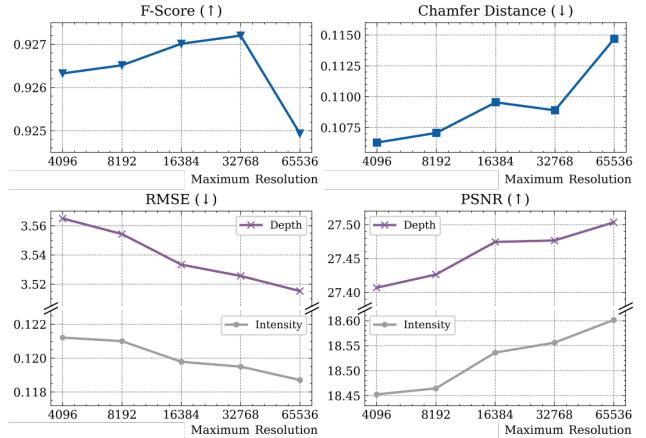


Figure S11. **Influence of maximum representation resolution.**

only calculating the results on rays that have valid values. In other words, we apply the ground-truth ray-drop mask to all results for reconstruction quality evaluation. As shown in Table S5, our method outperforms LiDAR-NeRF [39] in both static and dynamic scenarios, especially by a large margin in dynamic sequences.

B.3. Experiments on Resolution

Increasing the resolution of the representations is important for large-scale scenarios. In comparison to dense grids and planar features, hash grids can substantially raise the resolution and thus improve the accuracy of reconstruction, which has been verified in previous experiments. To determine the maximum resolution of hash grids, we further conducted additional experiments. As illustrated in Figure S11, increasing the resolution continuously alleviates the error of depth and intensity reconstruction. Considering the limited capacity, an extremely high resolution may lead to unfavorable effects, such as the degradation of point cloud metrics. Finally, we select the resolution of 2^{15} , which can adequately meet the requirements of large-scale scene reconstruction.

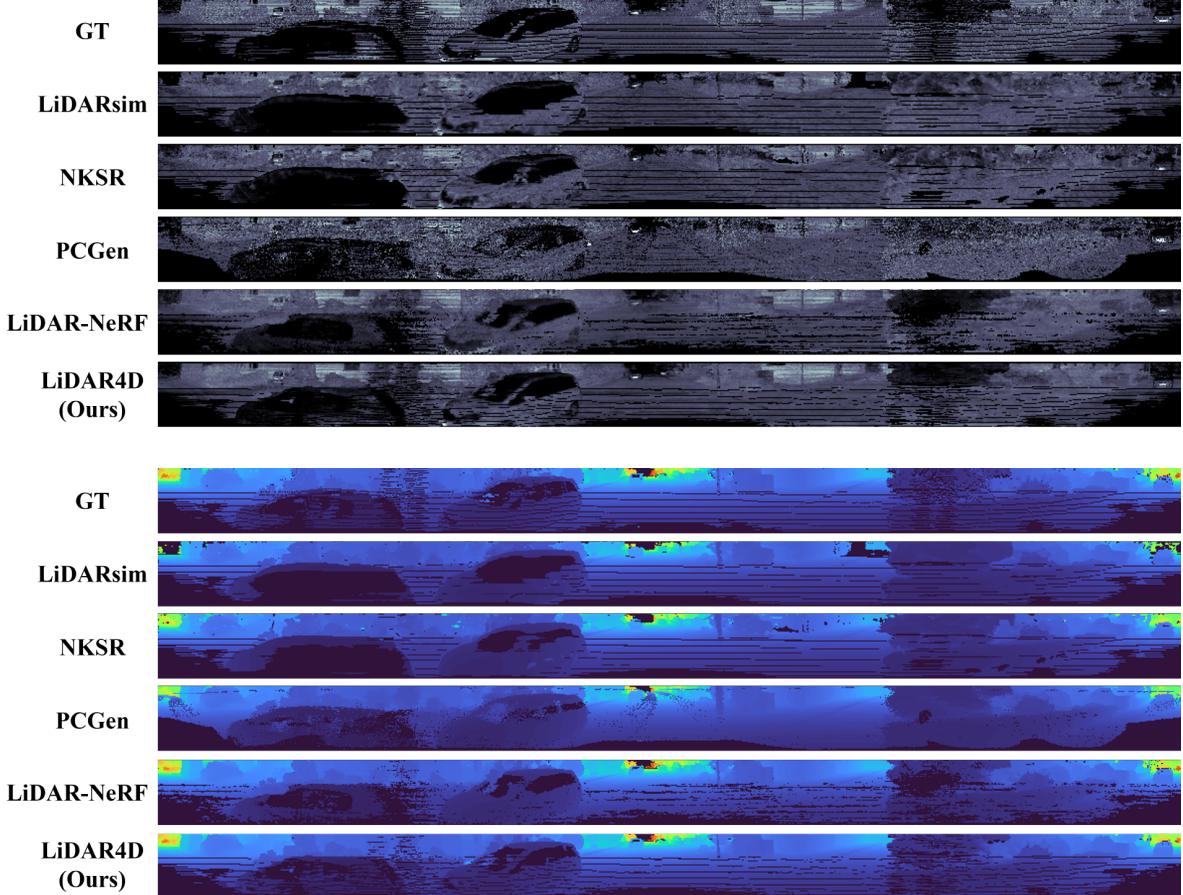


Figure S12. Qualitative comparison on KITTI-360 *Static* Scene Sequences.

KITTI-360		NuScenes	
Static	Seq 1538-1601	(ego-vehicle stationary)	Seq 450-500
	Seq 1728-1791		Seq 1250-1300
	Seq 1908-1971		
	Seq 3353-3416		
Dynamic	Seq 2350-2400	Dynamic	Seq 1600-1650
	Seq 4950-5000		
	Seq 8120-8170		Seq 2200-2250
	Seq 10200-10250		
	Seq 10750-10800		Seq 3180-3230
	Seq 11400-11450		

Table S6. Scene sequences of KITTI-360 and NuScenes.

C. Implementation Details

C.1. Dataset Visualization

As shown in Figure S17, we selected 6 representative dynamic scene sequences on KITTI-360. Each scene spans a distance of about 100–200 m and contains vehicles or pedestrians moving over long distances. Following the setup of LiDAR-NeRF [39], the same experiments were

conducted on the original 4 static scene sequences (Figure S18). The height and width of the range images are 66 and 1030. For the NuScenes dataset, we chose 5 dynamic sequences illustrated in Figure S19, including an ego-vehicle stationary scene (*Column 2*) which can be viewed as a special case of novel temporal view synthesis. The size of the range images is set to 32×1080 . The substantial variations between scenarios serve as a more accurate indicator of the reconstruction capabilities of current methods. The index number of scene sequences can be found in Table S6.

C.2. LiDAR4D

Hybrid representation. For the multi-planar features, the base resolution of the spatial plane is set to 64. The multi-scale structure has 4 levels, each doubling the spatial resolution and output 8-dimension feature, which finally yields a 32-dimension feature for both static and dynamic parts. The spatial resolution of hash grids ranges from 512 (the maximum resolution of multi-planar features) to 2^{15} . There are a total of 8 levels of hash grids, each level outputs 4-dimension features, and then the same 32-dimension fea-

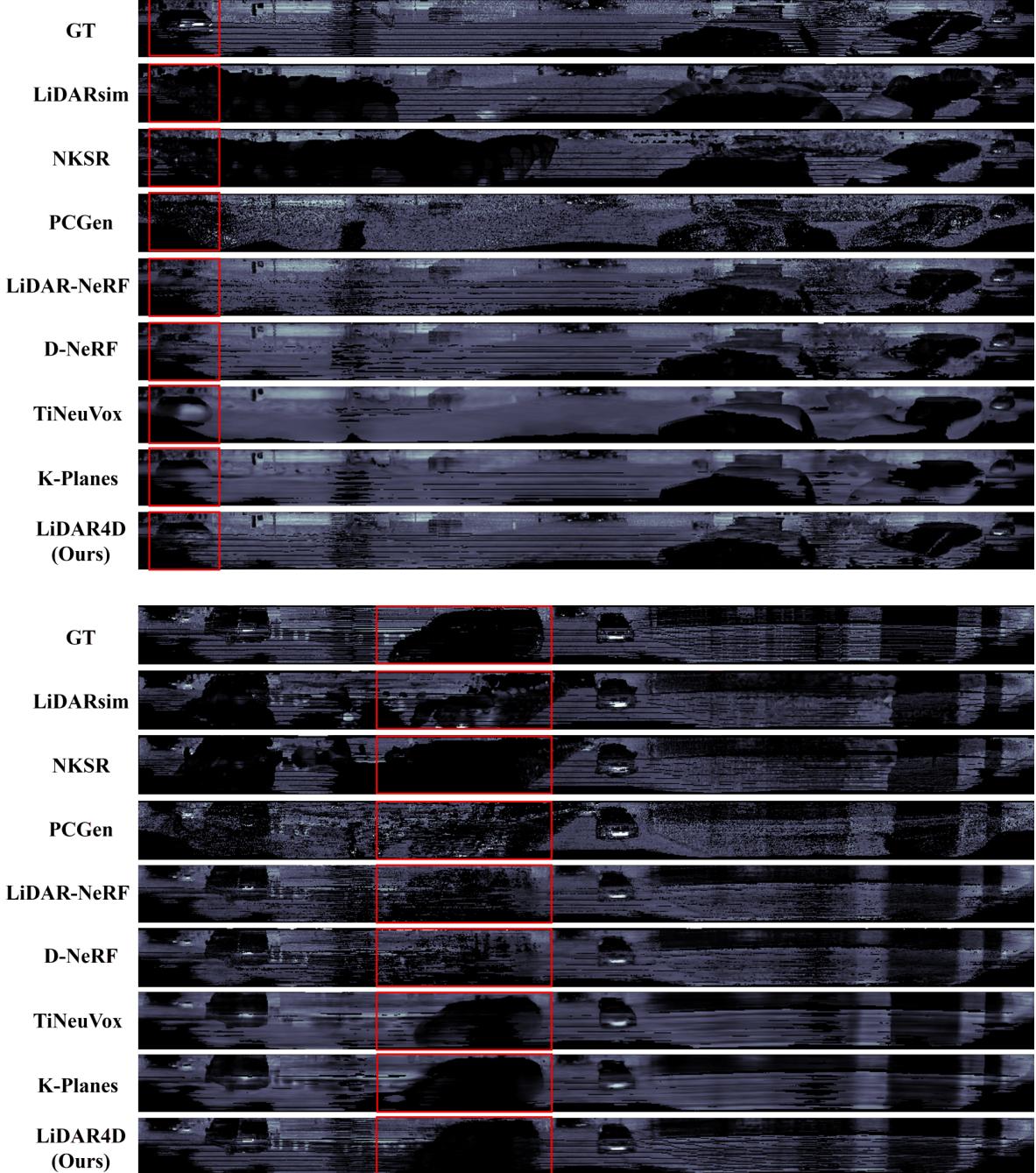


Figure S13. **Qualitative comparison for LiDAR intensity reconstruction and synthesis.** Dynamic vehicles are marked with red boxes.

tures are obtained. The grid is mapped to a hash table of 2^{19} . All the temporal resolution is fixed to 25. Ultimately, the static and dynamic features of the planes and hash grids compose a 128-dimensional latent vector.

Dynamic modules. Beyond time-conditioned multi-planar and hash grid features for dynamic reconstruction, we additionally introduce flow MLP to aggregate dynamic features

for temporal consistency. This coordinate-based MLP is an 8-layer neural field with 128 units per layer. Eventually, the dynamic features of adjacent spatio-temporal points are aggregated by weighted averaging. We incorporate the chamfer distance loss based on point clouds to effectively constrain the optimization of the flow MLP. It encourages the two adjacent frames of the point cloud transformed by the

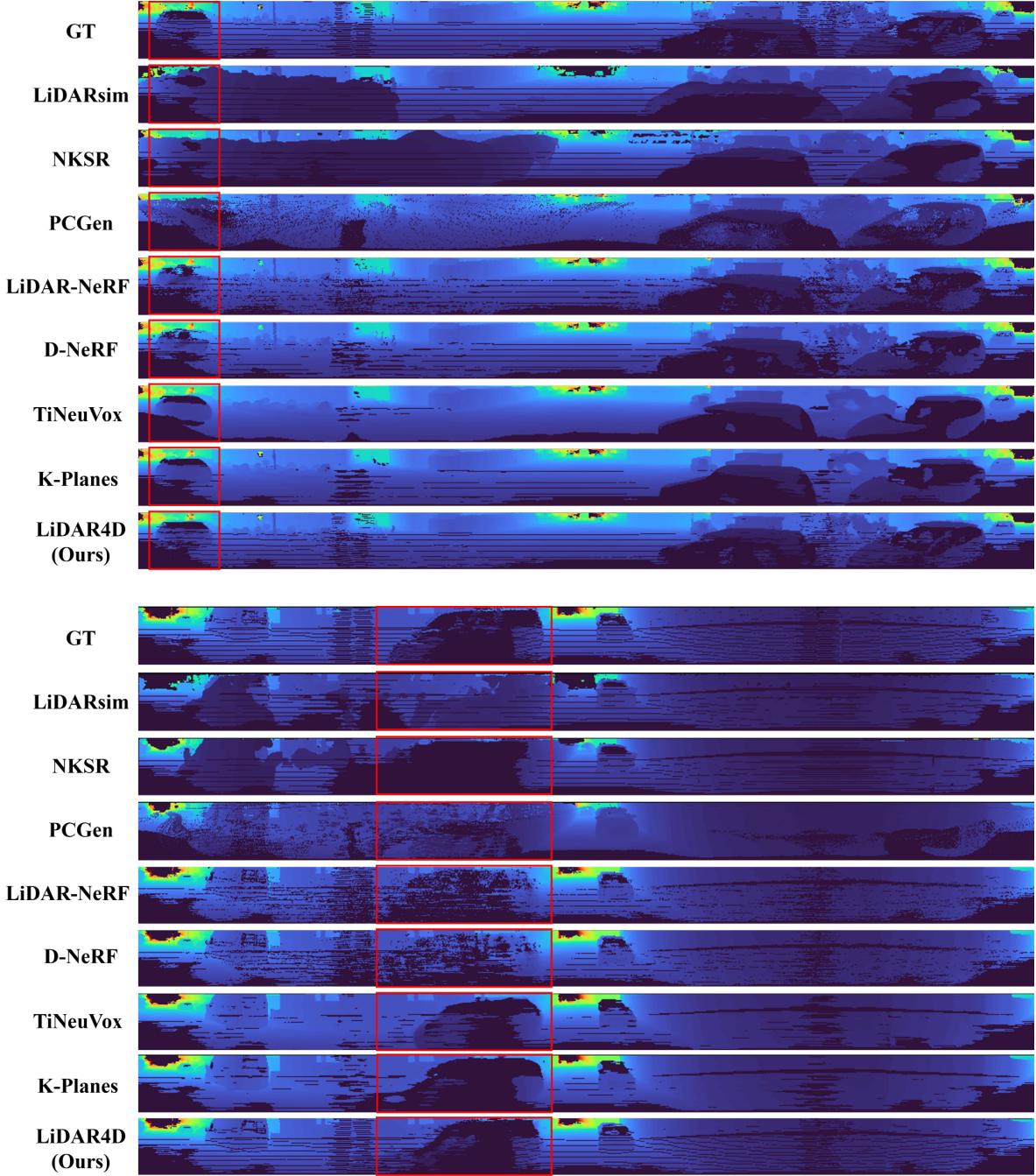


Figure S14. **Qualitative comparison for LiDAR depth reconstruction and synthesis.** Dynamic vehicles are marked with red boxes.

predicted scene flow to be as consistent as possible. According to the training process, we randomly select one moment in each epoch for optimization. In addition, we preprocess the point cloud by removing ground points using RANSAC [10] and further limiting the maximum distance within 50 meters to mitigate the adverse effect of noise.

Neural LiDAR fields. The aggregated time-conditioned

and flow-constrained dynamic features are finally fed into a 2-layer 64-dimensional MLP, which outputs the 15-dimensional geometric feature and density value. The geometric feature with the 12-band frequency-encoded viewpoint is utilized to predict intensity values and ray-drop probabilities by two independent 3-layer 64-dimensional MLPs, respectively. The expectation of the density in-

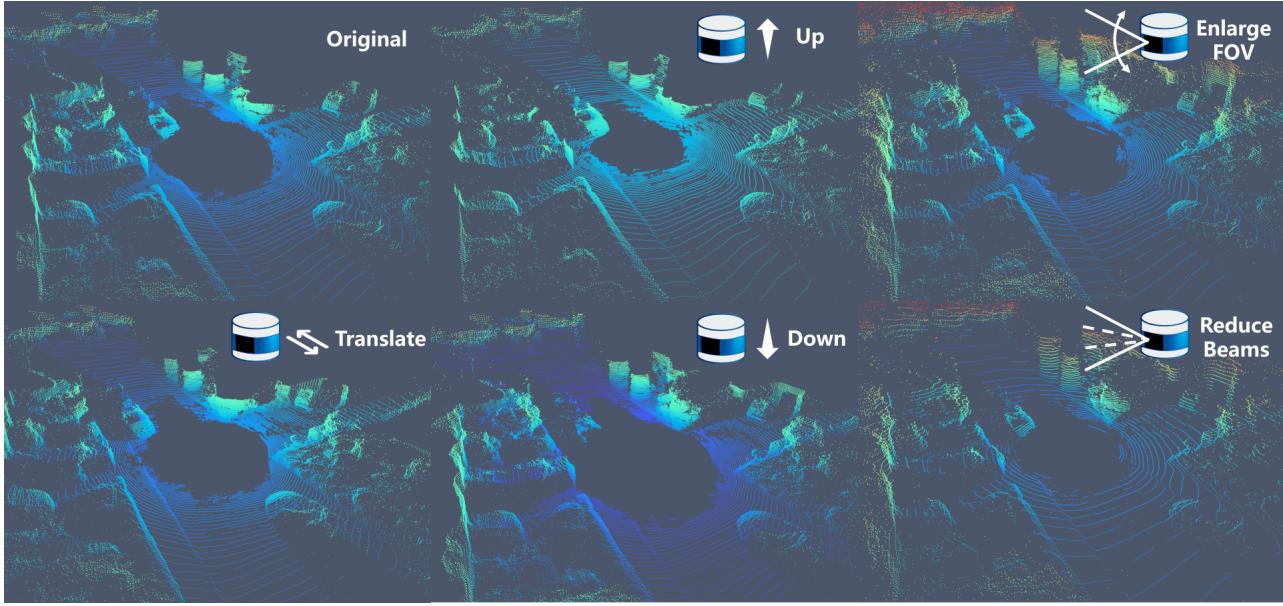


Figure S15. Novel view point cloud synthesis with different LiDAR configurations.

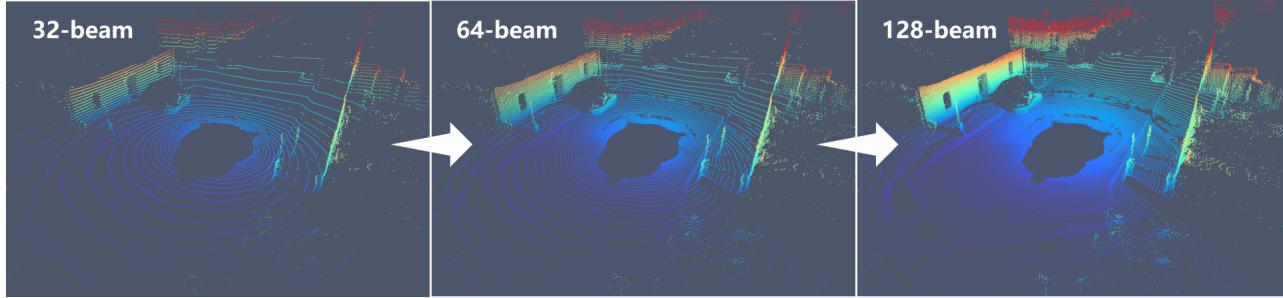


Figure S16. Increase LiDAR beams to densify the point cloud on NuScenes dataset.

tegrated along the ray serves as the depth value. Then, these initial predictions are combined as inputs to U-Net for global ray-drop optimization. The final predictions are multiplied by the ray-drop mask for synthesis.

Optimization details. The initial learning rate is set to 0.01 for the multi-planar and hash grids, and 0.001 for other MLP networks. The learning rate decreases exponentially during iterations, with a final decay coefficient of 0.1. The depth-loss weight λ_α is 1, the intensity weight λ_β is 0.1, the ray-drop weight λ_γ is 0.01, and the flow weight λ_η is 0.01. During refinement, the weight λ_r is set to 1 with other loss weights set to 0. The U-Net weights are randomly initialized and optimized with a learning rate of 0.001 by the Adam [17] optimizer. Other unmentioned optimization details are basically in line with LiDAR-NeRF [39].

Efficiency. According to experiments conducted on a single NVIDIA GeForce RTX 4090 GPU, LiDAR4D takes about 2 hours to complete the optimization of each scenario.

D. Applications

At last, we showcase the application of LiDAR4D for novel-view LiDAR synthesis with different sensor configurations. As illustrated in Figure S15, we can freely manipulate the sensor's pose, *e.g.*, moving up and down or horizontal translation. It can be observed that the LiDAR point clouds under different sensor poses vary significantly, and the accurate recovery of the scene and objects further demonstrates the high-fidelity synthesis of LiDAR4D. In addition, we can adjust LiDAR configurations, such as increasing the vertical field of view, to obtain a wider range of sensing results on the top right of Figure S15. Alternately, the modification of LiDAR beams realizes the conversion between sparse and dense point clouds. As shown on the bottom right of Figure S15, we transfer the LiDAR configuration of KITTI-360 to that of NuScenes, realizing the crossing of the domain gap. Also as shown in Figure S16, we can also densify the sparse Nuscenes data, which will also be beneficial for downstream tasks. All of this reveals the adaptability and enormous potential of LiDAR4D.

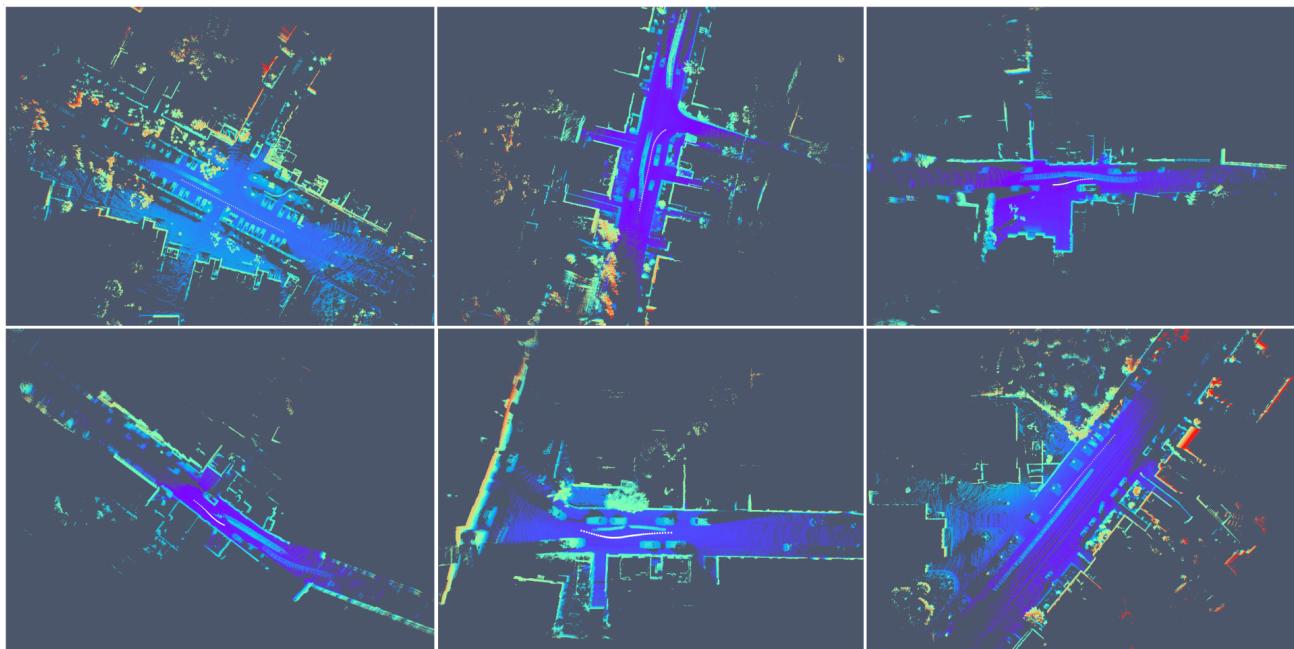


Figure S17. Visualization for the *dynamic* sequences of KITTI-360 dataset.

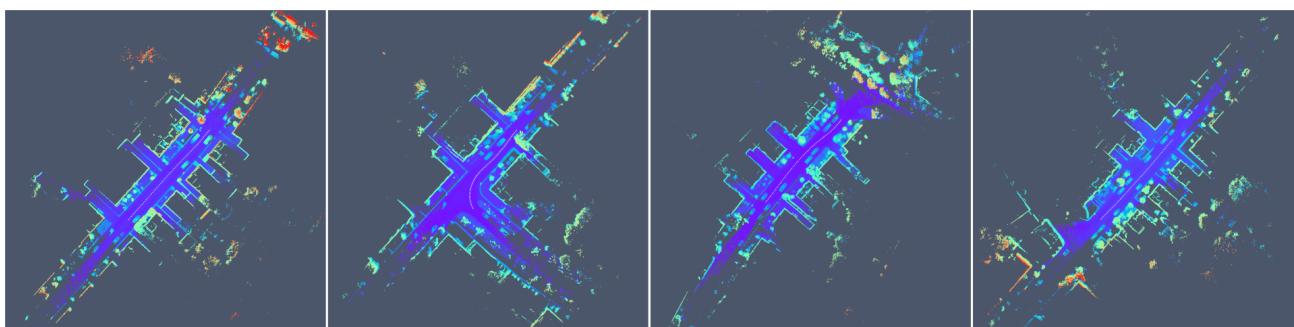


Figure S18. Visualization for the *static* sequences of KITTI-360 dataset.

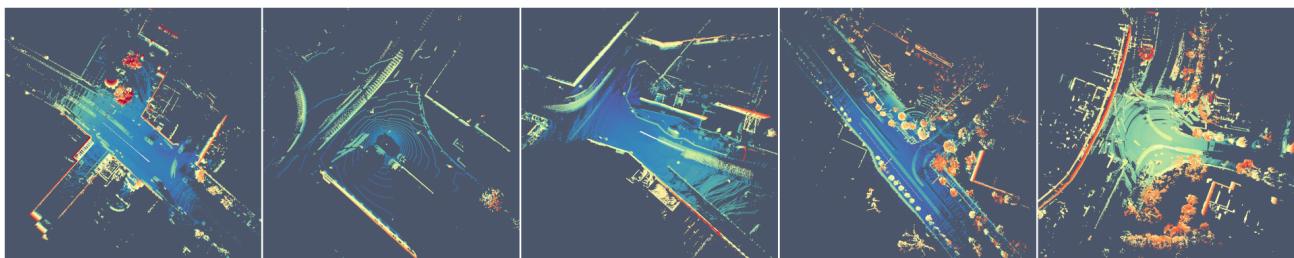


Figure S19. Visualization for the *dynamic* sequences of NuScenes dataset.