GeoNLF: Geometry guided Pose-Free Neural LiDAR Fields

Weiyi Xue[*]
Tongji University
xwy@tongji.edu.cn

Zehan Zheng[*]
Tongji University
zhengzehan@tongji.edu.cn

Fan Lu
Tongji University
lufan@tongji.edu.cn

Haiyun Wei
Tongji University
2311399@tongji.edu.cn

Guang Chen[†]
Tongji University
guangchen@tongji.edu.cn

Changjun Jiang
Tongji University
cjjiang@tongji.edu.cn

Abstract

Although recent efforts have extended Neural Radiance Fields (NeRF) into LiDAR point cloud synthesis, the majority of existing works exhibit a strong dependence on precomputed poses. However, point cloud registration methods struggle to achieve precise global pose estimation, whereas previous pose-free NeRFs overlook geometric consistency in global reconstruction. In light of this, we explore the geometric insights of point clouds, which provide explicit registration priors for reconstruction. Based on this, we propose **Geo**metry guided **N**eural **L**iDAR **F**ields (GeoNLF), a hybrid framework performing alternately global neural reconstruction and pure geometric pose optimization. Furthermore, NeRFs tend to overfit individual frames and easily get stuck in local minima under sparse-view inputs. To tackle this issue, we develop a selective-reweighting strategy and introduce geometric constraints for robust optimization. Extensive experiments on NuScenes and KITTI-360 datasets demonstrate the superiority of **GeoNLF** in both novel view synthesis and multi-view registration of low-frequency large-scale point clouds.

1 Introduction

Neural Radiance Fields (NeRFs) [37] have achieved tremendous achievements in image novel view synthesis (NVS). Recent studies have extended it to LiDAR point cloud synthesis [23, 51, 67, 70], mitigating the domain gap to real data and far surpassing traditional methods. Nevertheless, the majority of existing works exhibit a strong dependence on known precise poses. In the domain of images, conventional approaches rely on Structure-from-Motion algorithms like COLMAP [48] to estimate poses, which are prone to failure with sparse or textureless views. As an alternative, recent works [6, 21, 31, 41] such as BARF [31] employ bundle-adjusting techniques to achieve high-quality NVS while simultaneously enhancing the precision of pose estimation.

However, the sparse nature of LiDAR point clouds and their inherent absence of texture information distinguish them significantly from images. Trivial bundle-adjusting techniques from the image domain become less applicable in this context, encountering the following challenges: (1) Outdoor LiDAR point clouds (*e.g.*, 2Hz, 32-beam LiDAR keyframes in Nuscenes [9]) exhibit temporal and spatial sparsity. NeRF easily overfits the input views without addressing the geometric inconsistencies caused by inaccurate poses. Consequently, it fails to propagate sufficient gradients for effective pose optimization. (2) Point clouds lack texture and color information but contain explicit geometric features. However, the photometric-based optimization scheme of NeRFs overlooks these abundant geometric cues within the point cloud, which hinders geometric-based registration.

[*] Equal contribution. [†] Corresponding author. Our code is availiable at https://github.com/ispc-lab/GeoNLF.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

https://doi.org/10.52202/079017-2343

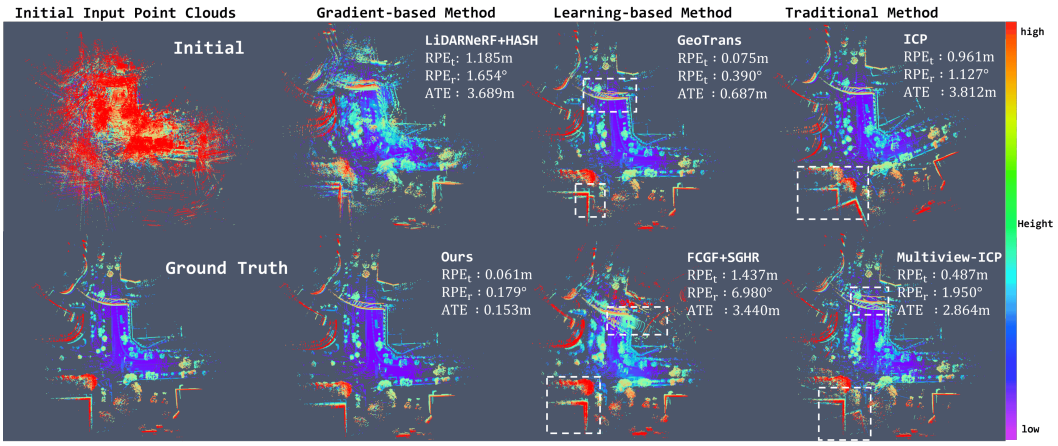

Figure 1: **Registration results.** Pairwise algorithms such as GeoTrans [44] and ICP [5] suffer from error accumulation and local mismatches. Multi-view methods like SGHR [54] and MICP [13] still manifest outlier poses. Previous gradient-based approaches LiDARNeRF-HASH [21] lack geometric consistency. Our method effectively avoids outlier frames and achieves superior registration accuracy.

An alternative to achieving pose-free LiDAR-NeRF is to employ point cloud registration (PCR) methods. Nonetheless, as the frequency of point cloud sequences decreases, the inter-frame motion escalates with a reduction in overlap. As presented in Fig. 1, pairwise and multi-view registration approaches may all trap in local optima and suffer from error accumulation, making it challenging to attain globally accurate poses. Hence, integrating local point cloud geometric features for registration with the global optimization of NeRF would be a better synergistic approach.

Furthermore, as demonstrated in [6, 53], the incorporation of geometric constraints significantly enhances the optimization of both pose and radiance fields. In the image domain, this process involves introducing additional correspondences or depth priors. However, most methods treat them solely as loss terms without fully exploiting them. In contrast, point clouds provide inter-frame correlations (*e.g.*, the closest point) for registration and explicit geometric information for reconstruction, presenting substantial advantages over images.

To this end, we propose **GeoNLF**, integrating LiDAR NVS with multi-view PCR for large-scale and low-frequency point clouds. Specifically, to address the suboptimality of global optimization and guide NeRF in the early pose optimization stage to avoid local minima, we regulate NeRF with a pure geometric optimizer. This module constructs a graph for multi-view point clouds and optimizes poses through graph-based loss. To reduce overfitting, we devised a selective-reweighting technique involving filtering out frames with outlier poses, thereby lessening their deleterious impacts throughout the optimization process. Additionally, to fully leverage the geometric attributes of point clouds, we introduced geometric constraints for point cloud modality rather than relying solely on the range map for supervision. Furthermore, our approach has demonstrated excellent performance in large-scale scenarios with sparse point cloud sequences at 2Hz, spanning hundreds of meters. To summarize, our main contributions are as follows:

(1) We propose **GeoNLF**, a novel framework for simultaneous large-scale multi-view PCR and LiDAR NVS. By exploiting geometric clues inside point clouds, **GeoNLF** couples geometric optimizer with neural reconstruction in the pose-free paradigm. (2) We introduce a selective-reweighting method to effectively alleviate overfitting, which presents excellent robustness across various scenarios. (3) Comprehensive experiments demonstrate **GeoNLF** outperforms state-of-the-art methods by a large margin on challenging large-scale and low-frequency point cloud sequences.

2 Background and Related Work

Neural Radiance Fields. NeRF [37] and related works have achieved remarkable achievements in NVS. Various neural representations [4, 10, 11, 22, 38], such as hash grids [38], triplanes [10, 22] and diverse techniques [39, 40, 55, 66] have been proposed to enhance NeRF's performance. Due to the lack of geometric information in images, some methods [16, 46, 59, 64] introduce depth prior or point clouds as auxiliary data to ensure multi-view geometric consistency. However, the geometric information and consistency encapsulated in point clouds are still not fully explored and utilized.

https://doi.org/10.52202/079017-2343

Novel View Synthesis for LiDAR. Traditional simulators [17, 27, 49] and explicit reconstruct-then-simulate [20, 28, 35] method exhibit large domain gap compared to real-world data. Very recently, a few studies have pioneered in NVS of LiDAR point clouds based on NeRF, surpassing traditional simulation methods. Among them, NeRF-LiDAR [68] and UniSim [62] require both RGB images as inputs. LiDAR-NeRF [51] and NFL [23] firstly proposed the differentiable LiDAR NVS framework, and LiDAR4D [70] further extended to dynamic scenes. However, most of these approaches still require a pre-computed pose of each point cloud frame and lack attention to geometric properties.

Point Cloud Registration. ICP [5] and its variants [45, 47, 43] are the most classic methods for registration, which rely on good initial conditions but are prone to falling into local optima. Learning-based method can be categorized into two schemes, i.e., end-to-end registration [65, 29, 24, 56, 1] and feature matching-based registration such as FCGF [14]. Recently, the specialized outdoor point cloud registration methods HRegNet [34] and HDMNet [61] have achieved excellent results. GeoTransformer [44] has achieved state-of-the-art in both indoor and outdoor point cloud registration. However, learning-based methods are data-driven and limited to specific datasets with ground truth poses, which requires costly pretraining and suffers from poor generalization.

Multiview methods are mostly designed for indoor scenes. Apart from Multiview-ICP [13, 7, 36], modern methods [2, 8, 52, 25] take global cycle consistency to optimize poses starting from an initial set of pairwise maps. Recent developments [19, 54, 3] such as SGHR [54] employ an iteratively reweighted least-squares (IRLS) scheme to adaptively downweight noisy pairwise estimates. However, their registration accuracy fundamentally depends on pairwise registration. The issues of pairwise methods for NVS still persist.

Bundle-Adjusting NeRF. iNeRF [63] and subsequent works [32, 15] demonstrated the ability of a trained NeRF to estimate novel view image poses through gradient descent. NeRFmm [58] and SCNeRF [50] extend the method to intrinsic parameter estimation. BARF [31] uses a coarse-to-fine reconstruction scheme in gradually learning positional encodings, demonstrating notable efficacy. Subsequent work HASH [21] adapts this approach on iNGP [38] through a weighted schedule of different resolution levels, further boosting performance. Besides, some studies have extended BARF to address more challenging scenarios, such as sparse input [53], dynamic scenes [33] and generalizable NeRF [12]. And [6, 53] uses monocular depth or correspondences priors for scene constraints, significantly enhancing the optimization of both pose and radiance fields. However, the aforementioned methods cannot be directly applied to point clouds or experience dramatic performance degradation when transferring. In contrast, our work is the first to introduce bundle-adjusting NeRF into LiDAR NVS task and achieve excellent results in challenging outdoor scenarios.

3 Methodology

We firstly introduce the pose-free Neural LiDAR Fields and the problem formulation of pose-free LiDAR-NVS. Following this, a detailed description of our proposed **GeoNLF** framework is provided.

Pose-Free NeRF and Neural LiDAR Fields. NeRF represents a 3D scene implicitly by encoding the density σ and color $\boldsymbol{c}$ of the scene using an implicit neural function $F_\Theta(\boldsymbol{x}, \boldsymbol{d})$, where $\boldsymbol{x}$ is the 3D coordinates and $\boldsymbol{d}$ is the view direction. When synthesizing novel views, NeRF employs volume rendering techniques to accumulate densities and colors along sampled rays. While NeRF requires precise camera parameters, pose-free NeRF only uses images $\mathcal{I} = \{I_i | i = 0, 1..., N-1\}$ and treats camera parameters $\mathcal{E} = \{\mathcal{E}_s | s = 0, 1...N-1\}$ as learnable parameters similar to Θ. Hence, the simultaneous update via gradient descent of $\mathcal{E}$ and Θ can be achieved by minimizing the error $\mathcal{L} = \sum_{i=0}^{N} \|\hat{I}_i - I_i\|_2^2$ between the rendered and ground truth image $\hat{I}, I$:

$$\Theta^*, \mathcal{E}^* = \arg\min_{\Theta, \mathcal{E}} \mathcal{L}(\hat{\mathcal{I}}, \hat{\mathcal{E}} \mid \mathcal{I}) \tag{1}$$

Following [70, 51], we project the LiDAR point clouds into range image, then cast a ray with a direction $\boldsymbol{d}$ determined by the azimuth angle θ and elevation angle ϕ under the polar coordinate system: $\mathbf{d} = (\cos\theta\cos\phi,\ \sin\theta\sin\phi,\ \cos\phi)^T$. Like pose-free NeRF, our pose-free Neural LiDAR Fields treats LiDAR poses as learnable parameters and applies neural function F_Θ to obtain a radiance depth z and a volume density value σ. Subsequently, volume rendering techniques are employed to derive the pixel depth value $\hat{\mathcal{D}}$:

$$\hat{\mathcal{D}}(\mathbf{r}) = \sum_{i=1}^{N} T_i \left(1 - e^{-\sigma_i \delta_i}\right) z_i, \quad T_i = \exp(-\sum_{j=1}^{i-1} \sigma_j \delta_j) \tag{2}$$

https://doi.org/10.52202/079017-2343

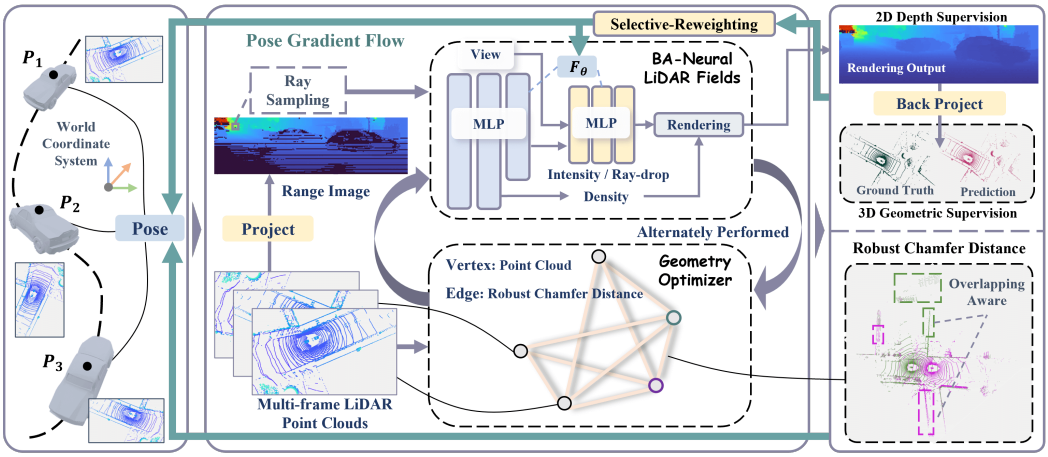

Figure 2: **Overview of our proposed GeoNLF.** We alternatively execute global optimization of bundle-adjusting neural LiDAR fields and graph-based pure geometric optimization. By integrating selective-reweighting strategy and explicit geometric constraints derived from point clouds, GeoNLF implements outlier-aware and geometry-aware mechanisms.

where δ refers to the distance between samples. We predict the intensity $\mathcal{S}$ and ray-drop probability $\mathcal{R}$ separately in the same way. Besides, our pose-free Neural LiDAR Fields adopted the Hybrid Planar-Grid representation from [70] for positional encoding $\gamma(x, y, z) = \mathbf{f}_{\text{planar}} \oplus \mathbf{f}_{\text{hash}}$.

$$\mathbf{f}_{\text{planar}} = \prod_{i=1}^{3} \text{Bilinear}(\mathcal{V}, \boldsymbol{x}), \mathcal{V} \in \mathbb{R}^{3\times M\times M\times C}, \ \mathbf{f}_{\text{hash}} = \text{TriLinear}(\mathcal{H}, \boldsymbol{x}), \mathcal{G} \in \mathbb{R}^{M\times M\times M\times C} \tag{3}$$

where $\boldsymbol{x}$ is the 3D point, $\mathcal{V}, \mathcal{H}$ store the grid features with M spatial resolution and C feature channels. This encoding method benefits the representation of large-scale scenes[70].

Problem Formulation. In the context of large-scale outdoor driving scenarios, the collected LiDAR point cloud sequence $\mathcal{P} = \{\mathcal{P}_s | s = 0, 1, ..., N-1\}$ serves as inputs with a low sampling frequency. The goal of GeoNLF is to reconstruct this scene as a continuous implicit representation based on neural fields, jointly recovering the LiDAR poses $\mathcal{E} = \{\mathcal{E}_s | s = 0, 1, ..., N-1\}$ which can align all point clouds $\mathcal{P}$ globally.

3.1 Overview of GeoNLF Framework

In contrast to prior pose-free NeRF methods, our pipeline employs a hybrid approach to optimize poses. As shown in Fig. 2, the framework can be divided into two alternately executed parts: global optimization of bundle-adjusting neural LiDAR fields (Sec. 3.2) and graph-based pure geometric optimization (Sec. 3.3) with the proposed Geo-optimizer. In the first part, we adopt a coarse-to-fine training strategy [31] and extend it to the Hybrid Planar-Grid encoding [70]. In the second part, inspired by multi-view point cloud registration, we construct a graph between multiple frame point clouds and propose a graph-based loss. The graph enables us to achieve pure geometric optimization, which encompasses both inter-frame and global optimization. Furthermore, we integrate the selective-reweighting strategy (Sec. 3.4) into the global optimization. This encourages the gradient of outliers to propagate towards pose correction while lowering the magnitude transmitted to the radiance fields, thus mitigating the adverse effects of outliers during reconstruction. To ensure geometry-aware results, we additionally incorporate explicit geometric constraints derived from point clouds in Sec. 3.5.

3.2 Bundle-Adjusting Neural LiDAR Fields for Global Optimization

In the stage of global optimization, we optimize Neural LiDAR Fields while simultaneously back-propagating gradients to the pose of each frame. By optimizing our geometry-constrained loss, which will be detailed in Sec. 3.5, the pose is individually optimized to achieve global alignment.

LiDAR Pose Representation. In previous pose-free NeRF methods, poses are often modeled by $\boldsymbol{T} = [\boldsymbol{R} \mid \boldsymbol{t}] \in SE(3)$ with a rotation $\boldsymbol{R} \in SO(3)$ and a translation $\boldsymbol{t} \in \mathbb{R}^3$. Pose updates are computed in the special Euclidean Lie algebra $\mathfrak{se}(3) = \{\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\rho} \\ \boldsymbol{\phi} \end{bmatrix}, \boldsymbol{\rho} \in \mathbb{R}^3, \boldsymbol{\phi} \in \mathfrak{so}(3)\}$ by

$\boldsymbol{\xi}' = \boldsymbol{\xi} + \Delta\boldsymbol{\xi}$, followed by the exponential map to obtain the transformation matrix $\boldsymbol{T}$:

$$\boldsymbol{T} = \exp(\boldsymbol{\xi}^\wedge) = \sum_{n=0}^{\infty} \frac{1}{n!}(\boldsymbol{\xi}^\wedge)^n = \begin{bmatrix} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\boldsymbol{\phi}^\wedge\right)^n & \sum_{n=0}^{\infty} \frac{1}{(n+1)!} \left(\boldsymbol{\phi}^\wedge\right)^n \boldsymbol{\rho} \\ \mathbf{0}^T & 1 \end{bmatrix} \tag{4}$$

where $\xi^\wedge = \begin{bmatrix} \boldsymbol{\phi}^\wedge & \boldsymbol{\rho} \\ \mathbf{0}^T & 0 \end{bmatrix}$ and $\phi^\wedge$ is the antisymmetric matrix of $\boldsymbol{\phi}$. Given a rotation vector $\boldsymbol{\phi} \in \mathfrak{so}(3)$, rotation matrix $\boldsymbol{R}$ can be obtained through the exponential map $\boldsymbol{R} = \exp(\boldsymbol{\phi}^\wedge) = \sum_{n=0}^{\infty} \frac{1}{n!} (\phi^\wedge)^n$. Simultaneously, we denote $\sum_{n=0}^{\infty} \frac{1}{(n+1)!} (\phi^\wedge)^n$ as $\boldsymbol{J}$. Then Eq. (4) can be rewritten as:

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{J}\boldsymbol{\rho} \\ \mathbf{0}^T & 1 \end{bmatrix} \tag{5}$$

Consequently, due to the coupling between $\boldsymbol{R} = \sum_{n=0}^{\infty} \frac{1}{n!} (\phi^\wedge)^n$ and $J = \sum_{n=0}^{\infty} \frac{1}{(n+1)!} (\phi^\wedge)^n$, the translation updates are influenced by rotation. Incorporating momentum may lead to non-intuitive optimization trajectories [32]. Therefore, we omit the coefficient $\boldsymbol{J}$ from the translation term. This approach enables updating the translation of the the center of mass and the rotation around the center of mass independently.

Coarse-to-Fine Positional Encoding. BARF[31]/HASH [21] propose to gradually activate high-frequency/high-resolution components within positional encoding. We further apply this approach to multi-scale planar and hash encoding [70] and found it also yields benefits in our large-scale scenarios. For the detailed formulation, we direct readers to reference [21].

3.3 Graph-based Pure Geometric Optimization

ICP [5] is a classic method for registration based on inter-frame geometric correlations. The essence of ICP lies in searching for the closest point as correspondence in another frame's point at each iteration, followed by using Singular Value Decomposition (SVD) to solve Eq. (6), then iteratively refining the solution. Nonetheless, ICP frequently converges to local optima (Fig. 1). In contrast, NeRF optimizes pose globally through the implicit radiance fields. However, it lacks geometric constraints and overlooks the strong geometric information inherent in the point cloud, leading to poor geometric consistency. As a consequence, both ICP and NeRF acting individually tend to converge to local optima. Our goal is to employ a hybrid method, utilizing NeRF for global pose optimization and integrating geometric information as an auxiliary support.

Drawing inspiration from ICP [5], we recognize that minimizing the Chamfer Distance (CD) is in line with the optimization objective of each step in ICP algorithm, as demonstrated in Eq. (7):

$$\mathbf{T}^* = \min_{\mathbf{T}} \sum_{\mathbf{p}_i \in \mathcal{P}} \min_{\mathbf{q}_i \in \mathcal{Q}} \|\mathbf{T} \cdot \mathbf{p}_i - \mathbf{q}_i\|_2^2 \tag{6}$$

$$\mathcal{L}_{(P,Q)} = \sum_{\mathbf{p}_i \in \mathbf{P}} w_i \min_{\mathbf{q}_i \in \mathbf{Q}} \|\mathbf{T}_\mathbf{P} p_i - \mathbf{T}_\mathbf{Q} q_i\|_2^2 + \sum_{\mathbf{q}_i \in \mathbf{Q}} w_i \min_{\mathbf{p}_i \in \mathbf{P}} \| \mathbf{T}_\mathbf{Q} q_i - \mathbf{T}_\mathbf{P} p_i\|_2^2 \tag{7}$$

where q, p in point cloud $\mathbf{Q}, \mathbf{P}$ are homogeneous coordinates. $\mathbf{T}_P, \mathbf{T}_Q$ represent the transformation matrix to the world coordinate system. However, minimizing the original CD does not necessarily indicate improved accuracy due to the non-overlapping regions between point clouds. To alleviate this negative impact, we weight each correspondence based on Eq. (8), whereas w_i in the original CD is normalized by the $\frac{1}{N}$, N is the number of points.

$$w_i = \frac{\exp(t/d^i_{clipped})}{\sum_{i=1}^{N} \exp(t/d^i_{clipped})}, \quad t = \text{scheduler}(t_0), \quad d^i_{clipped} = \max(\text{voxelsize}, d_i) \tag{8}$$

where d_i denotes the distance between a pair of matching nearest neighbor points, t is the temperature to sharpen the distribution of the $d^i_{clipped}$. The distance d_i is clipped to the size of the downsampled voxel grid. This soft assignment can be considered as an approximately derivable version of weighted averaging. Eq. (7) will degenerate to the original CD when $t \to 0$, degrade to considering only correspondences with the minimum distance when $t \to \infty$. Considering the distance lacks practical significance in initial optimization, the scheduler is set as a linear or exponential function to vary t from 0 to 0.5 as the optimization progresses. Building upon the above, as shown in Fig. 3, we

https://doi.org/10.52202/079017-2343

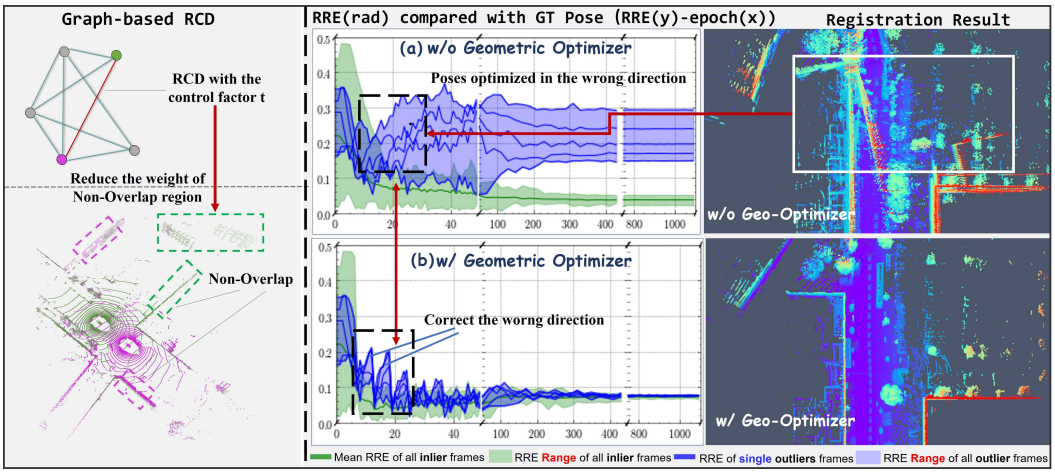

Figure 3: **Graph-based RCD (left)**. We introduce control factor t in CD to diminish the weighting of non-overlapping regions between point clouds. **Geo-optimizer and its impact on pose optimization (right)**. Pose errors are reduced after each increase caused by NeRF's incorrect optimization direction. Comparison of (a) and (b) shows Geo-optimizer prevents incorrect pose optimization of NeRF.

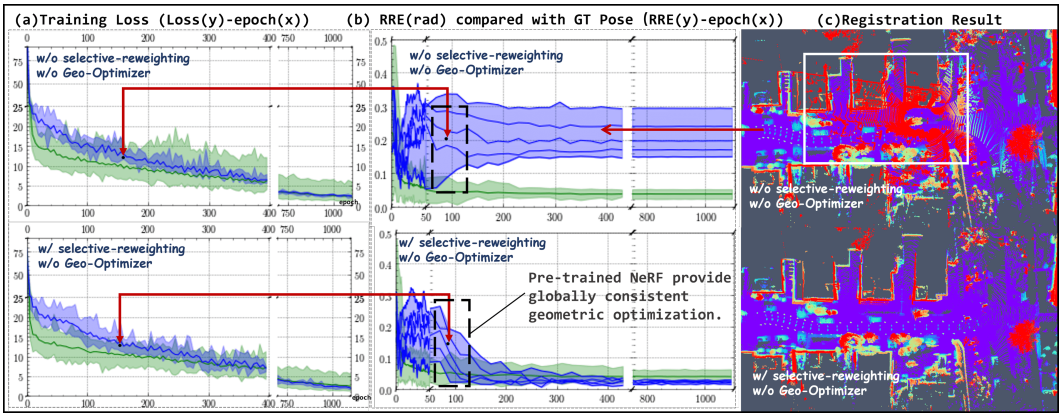

Figure 4: **Impact of selective-reweighting training strategy on pose optimization.** (a) Frames with outlier poses exhibit significantly higher losses. With selective-reweighting, outlier frames maintain a relatively higher loss without overfitting. (b) After several training iterations, the pre-trained outlier-aware NeRF can provide globally consistent geometric optimization for outlier frames.

approximate the registration objective by optimizing the Graph-based Robust Chamfer distance (G-RCD). Specifically, we construct a graph $(\mathcal{W}, \mathcal{Y})$, where each vertex $\mathcal{W}$ represents a set of points and each edge $\mathcal{Y}$ corresponds to proposed RCD via Eq. (7). We connect each frame with its temporally preceding n frames to mitigate error accumulation in ICP [5]. Then RCD is calculated for all edges as Eq. (9), and M denotes the number of frames in the sequence. Notably, in Eq. (7), G-RCD is computed using the global transform matrix, enabling direct gradient propagation of the Graph-based loss to the global transformation matrix of each frame.

$$\mathcal{L}_{graph} = \frac{1}{(nM - \frac{n(n+1)}{2})} \sum_{(i,j)\in\mathcal{E}} \mathcal{L}_{(i,j)}, \tag{9}$$

Discussion. As illustrated in Fig. 3(b), insufficient geometric guidance leads to certain frame poses being optimized in the wrong direction. Geometric optimizer can address this issue by preventing pose updates strictly following NeRF and correcting wrong optimization directions that do not conform to global geometric consistency. This method involves externally modifying pose parameters and providing effective geometric guidance early in the ill-conditioned optimization process. Consequently, few iterations of graph-based RCD computation suffice to offer ample guidance for NeRF.

3.4 Selective-Reweighting Strategy for Outlier Filtering

In bundle-adjusting optimization, as shown in Fig. 4(a), we observed that frames with outlier poses present significantly higher rendering losses during the early stages of training. However, low

https://doi.org/10.52202/079017-2343

frequency and sparsity of point clouds result in quick overfitting of individual frames including outliers (*cf.* Fig. 4(a)(b)). This leads to minimal pose updates when the overall loss decreases, resulting in incorrect poses and inferior reconstruction. Inspired by the capabilities of NeRF in pose inference [63], we decrease the learning rate (`lr`) of neural fields for the top k frames with the highest rendering losses as Eq. (10), while keeping `lr` of poses unchanged. The strategy facilitates gradient propagation towards outlier poses, while the gradient flow to the radiance fields is concurrently diminished. Consequently, it's analogous to leveraging a pre-trained NeRF for outlier pose correction and lessens the adverse effects caused by outliers during the optimization process.

$$\mathrm{lr}_{outliers} = (w_0 + l(1 - w_0))\mathrm{lr}_{inliers} \quad (w_0 > 0) \tag{10}$$

Where $l \in [0, 1]$ denotes training progress. Akin to leaky ReLU [60], we set the reweighting factor w_0 to a relatively small value. w_0 increases as the process progresses, which ensures the network's ongoing learning from these frames and avoids stagnation.

3.5 Improving Geometry Constraints for NeRF

Point clouds encapsulate rich geometric features. However, solely supervising NeRF training via range images pixel-wise fails to fully exploit their potential, *e.g.*, normal information. Furthermore, the Chamfer distance can directly supervise the synthesized point clouds from a 3D perspective. Therefore, in addition to supervising via 2D range map, we propose directly constructing a three-dimensional geometric loss function between the generated point cloud and the ground truth point cloud. Unlike our Geo-optimizer, Eq. (11) imposes constraints between synthetic point clouds $\hat{P}$ and ground truth point clouds P:

$$\mathcal{L}_{CD} = \frac{1}{N_{\hat{P}}} \sum_{\hat{p}_i \in \hat{P}} \min_{p_i \in P} \|\hat{p}_i - p_i\|_2^2 + \frac{1}{N_P} \sum_{p_i \in P} \min_{\hat{p}_i \in \hat{P}} \|p_i - \hat{p}_i\|_2^2 \tag{11}$$

Based on the point correspondences established between $\hat{P}$ and P as derived in Eq. (11), the constraint of normal can be formulated as minimizing:

$$\mathcal{L}_{normal} = \frac{1}{N_{\hat{P}}} \sum_{\hat{p}_i \in \hat{P}} \min_{p_i \in P} \|\mathcal{N}(\hat{p}_i) - \mathcal{N}(p_i)\|_1 + \frac{1}{N_P} \sum_{p_i \in P} \min_{\hat{p}_i \in \hat{P}} \|\mathcal{N}(p_i) - \mathcal{N}(\hat{p}_i)\|_1 \tag{12}$$

Thus, the normal loss is calculated between the synthetic point cloud and the ground truth point cloud to ensure more accurate normal vectors of the point cloud synthesized from NeRF. Moreover, we also employ 2D loss function to supervise NeRF as Eq. (13).

$$\mathcal{L}_r(\mathbf{r}) = \sum_{\mathbf{r} \in \mathbf{R}} \lambda_d \left\| \hat{D}(\mathbf{r}) - D(\mathbf{r}) \right\|_1 + \lambda_s \left\| \hat{\mathcal{S}}(\mathbf{r}) - \mathcal{S}(\mathbf{r}) \right\|_2^2 + \lambda_r \left\| \hat{\mathcal{R}}(\mathbf{r}) - \mathcal{R}(\mathbf{r}) \right\|_2^2 \tag{13}$$

where D represents depth and $\mathcal{S}, \mathcal{R}$ represents intensity and ray-drop probabilities. Consequently, the loss for Neural LiDAR fields is weighted combination of the depth, intensity, ray-drop loss and 3D geometry constraints, which can be formalized as $\mathcal{L} = \mathcal{L}_r + \lambda_n \mathcal{L}_{normal} + \lambda_c \mathcal{L}_{CD}$.

4 Experiment

4.1 Experimental Setup

Datasets and Experimental Settings. We conducted experiments on two public autonomous driving datasets: NuScenes [9] and KITTI-360 [30] dataset, each with five representative LiDAR point cloud sequences. We selected 36 consecutive frames at 2Hz from keyframes as a single scene for NuScenes, holding out 4 samples at 9-frame intervals for NVS evaluation. KITTI-360 has an acquisition frequency of 10Hz. We used 24 consecutive frames sampled every 5th frame to match scene sizes of Nuscenes, holding out 3 samples at 8-frame intervals for evaluation. We perturbed LiDAR poses with additive noise corresponding to a standard deviation of $20\,\mathrm{deg}$ in rotation and $3m$ in translation.

Metrics. We evaluate our method from two perspectives: pose estimation and novel view synthesis. For pose evaluation, we use standard odometry metrics, including Absolute Trajectory Error (ATE) and Relative Pose Error ($\mathrm{RPE_r}$ in rotation and $\mathrm{RPE_t}$ in translation). Following LiDAR4D [70] for NVS evaluation, we employ CD to assess the 3D geometric error and the F-score with 5cm error

https://doi.org/10.52202/079017-2343

Method	Dataset	Point Cloud		Depth					Intensity				
		CD↓	F-score↑	RMSE↓	MedAE↓	LPIPS↓	SSIM↑	PSNR↑	RMSE↓	MedAE↓	LPIPS↓	SSIM↑	PSNR↑
BARF-LN [31, 51]	Nuscenes	1.2695	0.7589	8.2414	0.1123	0.1432	0.6856	20.89	0.392	0.0144	0.1023	0.6119	26.2330
HASH-LN [21, 51]		0.9691	0.8011	7.8353	0.0441	0.1190	0.6543	20.6244	0.0459	0.0135	0.0954	0.6279	26.8870
GeoTrans [44, 51]		4.1587	0.2993	10.7899	2.1529	0.1445	0.3671	17.5885	0.0679	0.0256	0.1149	0.3476	23.6211
GeoNLF (Ours)		**0.2408**	**0.8647**	**5.8208**	**0.0281**	**0.0727**	**0.7746**	**22.9472**	**0.0378**	**0.0100**	**0.0774**	**0.7368**	**28.6078**
BAR-LN [31, 51]	KITTI-360	3.1001	0.6156	7.5767	2.0583	0.5779	0.2834	22.5759	0.2121	0.1575	0.7121	0.1468	11.9778
HASH-LN [21, 51]		2.6913	0.6082	6.3005	2.1686	0.5176	0.3752	22.6196	0.2404	**0.1502**	0.6508	0.1602	12.9286
GeoTrans [44, 51]		0.5753	0.8116	4.4291	0.2023	**0.3896**	0.5330	**25.6137**	0.2709	0.1589	0.5578	0.2578	12.9707
GeoNLF (Ours)		**0.2363**	**0.9178**	**4.0293**	**0.1009**	0.3900	**0.6272**	25.2758	**0.1495**	0.1525	**0.5379**	**0.3165**	**16.5813**

Table 1: **NVS Quantitative Comparison on Nuscenes and KITTI-360.** We compare our method to different types of approaches and color the top results as best and second best. All results are averaged over the 5 sequences.

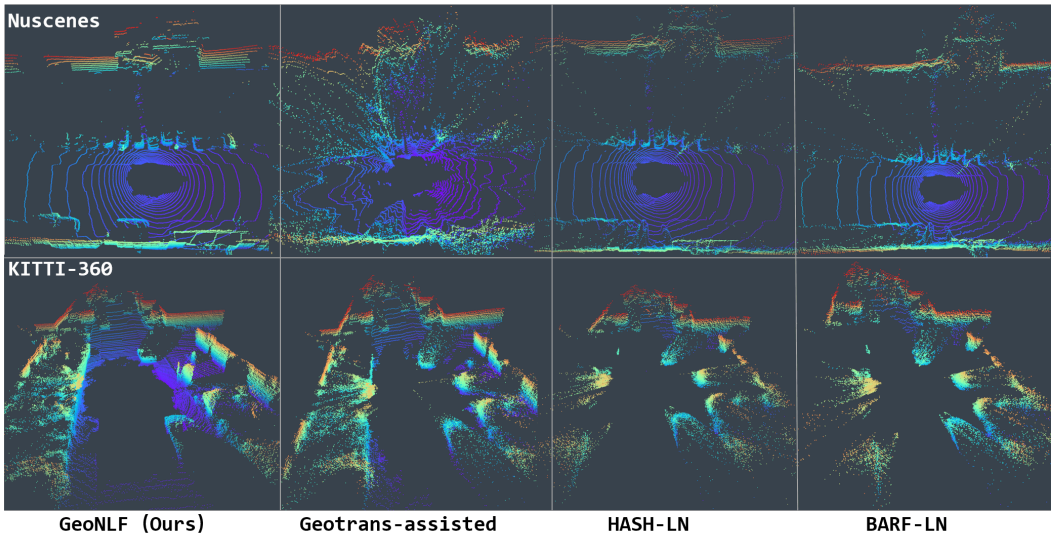

Figure 5: **Qualitative comparison of NVS.** We compared GeoNLF with other pose-free methods and GeoTrans-assisted NeRF. Especially, GeoTrans fails on Nuscenes due to the inaccurate poses.

threshold. Additionally, we use RMSE and MedAE to compute depth and intensity errors in projected range images, along with LPIPS [69], SSIM [57], and PSNR to measure overall variance.

Implementation Details. The entire point cloud scene is scaled within the unit cube space. The optimization of GeoNLF is implemented on Pytorch [42] with Adam [26] optimizer. All the sequences are trained for 60K iterations. Our Geometry optimizer's `lr` for translation and rotation is the same as the `lr` for pose in NeRF with synchronized decay. We use the coarse-to-fine strategy[31, 21], which starts from training progress 0.1 to 0.8. The reweight coefficient for the top-5 frames linearly increases from 0.15 to 1 during training. After every m_1 epoch of bundle adjusting global optimization, we proceed with m_2 epoch of pure geometric optimization, where m_2/m_1 decrease from 10 to 1.

4.2 Comparison in LiDAR NVS

We compare our model with BARF [31] and HASH [21], both of which use LiDAR-NeRF[51] as backbone. For PCR-assisted NeRF, we opt to initially estimate pose utilizing pose derived from GeoTrans [44], which is the most robust and accurate algorithm among other PCR methods in our experiments. And subsequently we leverage LiDAR-NeRF [51] for reconstruction. For all Pose-free methods, we follow NeRFmm[58] to obtain the pose of test views for rendering. The quantitative and qualitative results are in Tab. 1 and Fig. 5. Our method achieves high-precision registration and high-quality reconstruction across all sequences. However, baseline methods fail completely on certain sequences due to their lack of robustness. Please refer to Fig. 7 for details. Ultimately, our method excels in the reconstruction of depth and intensity, as evidenced by 7.9% increase in F-score on Nuscenes and 13.1% on KITTI-360 compared to the second best result.

https://doi.org/10.52202/079017-2343

Method	NuScenes			KITTI-360		
	RPE_t(cm)↓	RPE_r(deg)↓	ATE(m)↓	RPE_t(cm)↓	RPE_r(deg)↓	ATE(m)↓
ICP [5]	15.410	0.647	1.131	30.383	1.019	1.894
MICP [51]	38.84	1.101	2.519	35.584	1.419	1.483
HRegNet [34]	120.913	2.173	7.815	290.16	9.083	7.423
SGHR [54]	100.98	0.699	9.557	95.576	0.906	2.539
GeoTrans [44]	16.097	0.363	0.892	6.081	0.213	0.246
BARF-LN [51, 31]	210.331	0.819	5.244	199.74	2.203	2.763
HASH-LN [51, 21]	180.282	0.832	4.151	196.791	2.171	2.666
GeoNLF (Ours)	**7.058**	**0.103**	**0.228**	**5.449**	**0.205**	**0.170**

Table 2: **Pose estimation accuracy comparison.**

Method	Point Cloud	Depth	Intensity	Pose		
	CD↓	PSNR ↑	PSNR ↑	RPE_t(cm)↓	RPE_r(deg)↓	ATE(m)↓
w/o G-optim	0.6180	21.3211	25.8551	54.72	0.283	1.328
w/o RCD	0.2711	21.1323	26.7232	8.476	0.163	0.332
w/o SR	0.2654	21.1096	26.5269	8.124	0.156	0.264
w/o L_{3d}	0.2877	21.7128	28.5210	7.273	0.124	0.234
GeoNLF	**0.2363**	**22.9472**	**28.6078**	**7.058**	**0.103**	**0.228**

Table 3: **Ablation study on Nuscenes.**

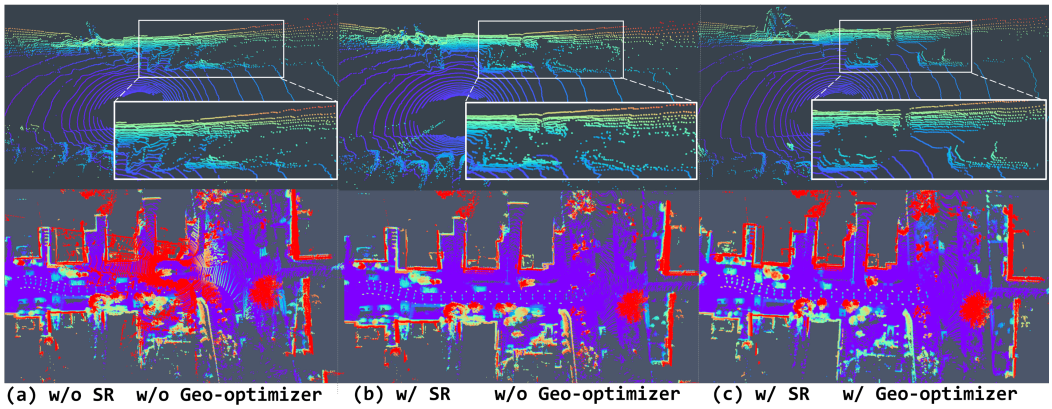

Figure 6: **Qualitative results of ablation study.** We present the NVS and Registration results in the first and second rows. Outlier frames emerged w/o SR or w/o G-optim.

4.3 Comparison in Pose Estimation

We conduct comprehensive comparisons of GeoNLF with pairwise baselines, including traditional method ICP [5], learning-based GeoTrans [44] and outdoor-specific HRegNet [34], as well as multi-view baselines MICP [13] and learning-based SGHR [54]. For pairwise methods, we perform registration between adjacent frames in an Odometry-like way. For SGHR, we utilize FCGF [14] descriptors followed by RANSAC [18] for pairwise registration. The estimated trajectory is aligned with the ground truth using $\mathrm{Sim}(3)$ with known scale.

GeoNLF outperforms both the registration and pose-free NeRF baselines. Quantitative and Qualitative results are illustrated in Tab. 2 and Fig. 1. As depicted in Fig. 1, most registration methods fail to achieve globally accurate poses and **completely fail in some scenarios**, leading to massive errors in average results. Significant generalization issues arise for learning-based registration methods due to potential disparities between testing scenarios and training data, including differences in initial pose distributions. This challenge is particularly pronounced in HRegNet [34]. While the transformer model GeoTrans [44] with its higher capacity offers some alleviation to the issue, it remains not fully resolved.

4.4 Ablation Study

In this Section, we analyze the effectiveness of each component of GeoNLF. The results of ablation studies are shown in Tab. 3. **(1) Geo-optimizer.** When training GeoNLF w/o geo-optimizer (w/o G-optim), pose optimization may initially converge towards incorrect directions. Excluding geo-optimizer in GeoNLF results in decreased pose accuracy and reconstruction quality. **(2) Control factor of graph-based RCD.** Although geo-optimizer is crucial in the early stages of optimization, we find that using the original CD limits the accuracy of pose estimation. Removing the control factor (w/o RCD) leads to decreased pose estimation accuracy due to the presence of non-overlapping regions. **(3) Selective-reweighting (SR) strategy.** As presented in Figs. 4 and 6 and Tab. 3, outlier frames cause GeoNLF w/o SR strategy to overlook multi-view consistency, adversely affecting reconstruction quality. **(4) Geometric constraints.** Removing the 3D constraints (w/o L_{3d}) results in a decline in CD due to the photometric loss's inability to adequately capture geometric information.

4.5 Limination

Despite the fact that GeoNLF has exhibited exceptional performance in PCR and LiDAR-NVS on challenging scenes, it is not designed for dynamic scenes, which is non-negligible in autonomous

https://doi.org/10.52202/079017-2343

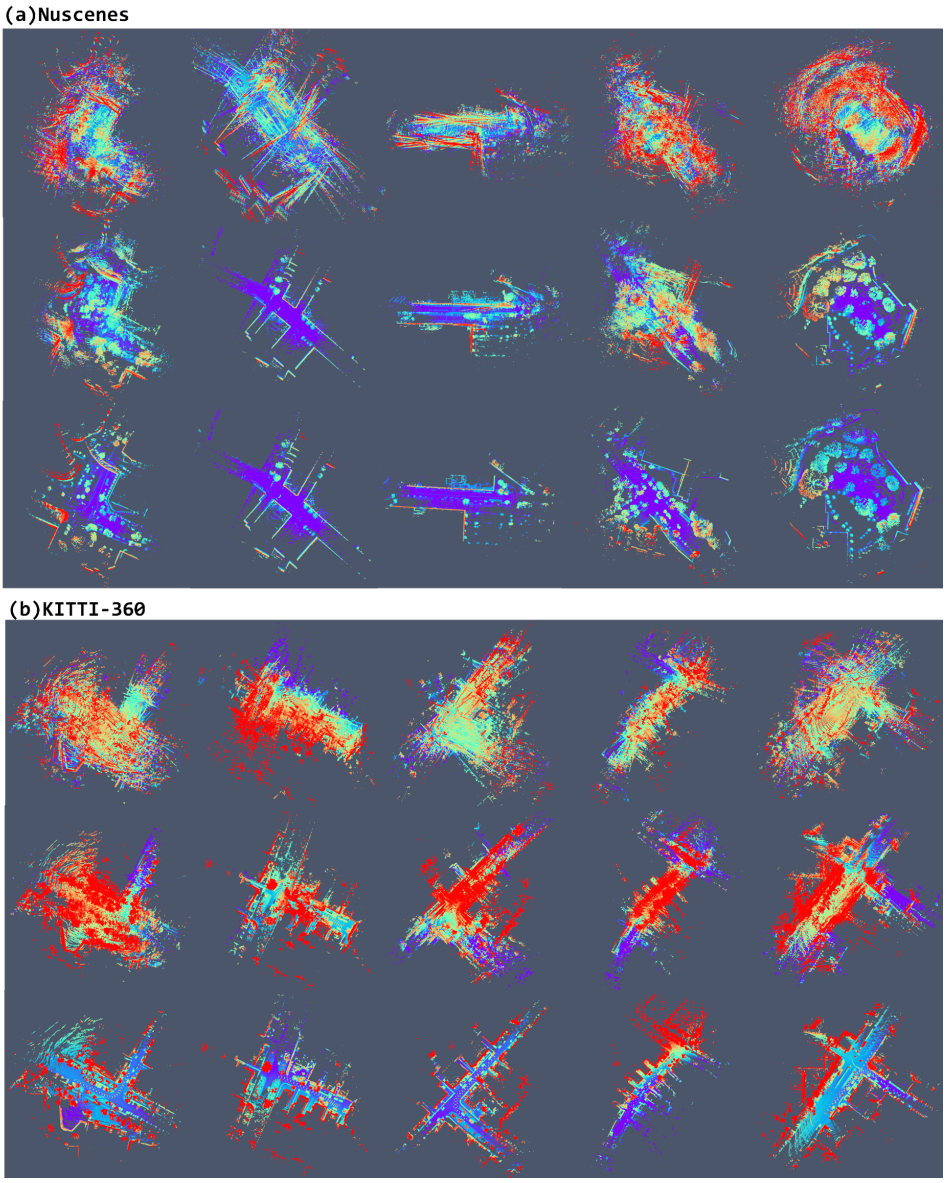

Figure 7: **Qualitative registration results of HASH-LN and GeoNLF on Nuscenes and KITTI-360 dataset.** The first row contains original inputs, the second row shows the results of HASH-LN, and the third row displays the results of GeoNLF.

driving scenarios. Additionally, GeoNLF targets point clouds within a sequence, relying on the temporal prior of the point clouds.

5 Conclusion

We introduce GeoNLF for multi-view registration and novel view synthesis from a sequence of sparsely sampled point clouds. We demonstrate the challenges encountered by previous pairwise and multi-view registration methods, as well as the difficulties faced by previous pose-free methods. Through the utilization of our Geo-Optimizer, Graph-based Robust CD, selective-reweighting strategy and geometric constraints from 3D perspective, our outlier-aware and geometry-aware GeoNLF demonstrate the promising performance in both multi-view registration and NVS tasks.

6 Acknowledgments

This work was supported by the National Key Research and Development Program of China (No.2024YFE0211000), in part by the National Natural Science Foundation of China (No. 62372329), in part by Shanghai Scientific Innovation Foundation (No.23DZ1203400), in part by Tongji-Qomolo Autonomous Driving Commercial Vehicle Joint Lab Project, and in part by Xiaomi Young Talents Program.

References

[1] Yasuhiro Aoki, Hunter Goforth, Rangaprasad Arun Srivatsan, and Simon Lucey. Pointnetlk: Robust & efficient point cloud registration using pointnet. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7163–7172, 2019.

[2] Mica Arie-Nachimson, Shahar Z Kovalsky, Ira Kemelmacher-Shlizerman, Amit Singer, and Ronen Basri. Global motion estimation from point matches. In *2012 Second international conference on 3D imaging, modeling, processing, visualization & transmission*, pages 81–88. IEEE, 2012.

[3] Federica Arrigoni, Beatrice Rossi, and Andrea Fusiello. Spectral synchronization of multiple views in se (3). *SIAM Journal on Imaging Sciences*, 9(4):1963–1990, 2016.

[4] 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.

[5] Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pages 586–606. Spie, 1992.

[6] Wenjing Bian, Zirui Wang, Kejie Li, Jia-Wang Bian, and Victor Adrian Prisacariu. Nope-nerf: Optimising neural radiance field with no pose prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4160–4169, 2023.

[7] Tolga Birdal and Slobodan Ilic. Cad priors for accurate and flexible instance reconstruction. In *Proceedings of the IEEE international conference on computer vision*, pages 133–142, 2017.

[8] Tolga Birdal and Umut Simsekli. Probabilistic permutation synchronization using the riemannian structure of the birkhoff polytope. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11105–11116, 2019.

[9] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo 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.

[10] 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.

[11] 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.

[12] Yu Chen and Gim Hee Lee. Dbarf: Deep bundle-adjusting generalizable neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24–34, 2023.

[13] Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun. Robust reconstruction of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5556–5565, 2015.

[14] Christopher Choy, Jaesik Park, and Vladlen Koltun. Fully convolutional geometric features. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8958–8966, 2019.

[15] Junyuan Deng, Qi Wu, Xieyuanli Chen, Songpengcheng Xia, Zhen Sun, Guoqing Liu, Wenxian Yu, and Ling Pei. Nerf-loam: Neural implicit representation for large-scale incremental lidar odometry and mapping. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8218–8227, 2023.

 https://doi.org/10.52202/079017-2343

[16] 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.

[17] 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.

[18] 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.

[19] Zan Gojcic, Caifa Zhou, Jan D Wegner, Leonidas J Guibas, and Tolga Birdal. Learning multiview 3d point cloud registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1759–1769, 2020.

[20] Benoît Guillard, Sai Vemprala, Jayesh K Gupta, Ondrej Miksik, 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.

[21] Hwan Heo, Taekyung Kim, Jiyoung Lee, Jaewon Lee, Soohyun Kim, Hyunwoo J Kim, and Jin-Hwa Kim. Robust camera pose refinement for multi-resolution hash encoding. In *International Conference on Machine Learning*, pages 13000–13016. PMLR, 2023.

[22] 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.

[23] 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.

[24] Xiaoshui Huang, Guofeng Mei, and Jian Zhang. Feature-metric registration: A fast semi-supervised approach for robust point cloud registration without correspondences. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11366–11374, 2020.

[25] Shengze Jin, Iro Armeni, Marc Pollefeys, and Daniel Barath. Multiway point cloud mosaicking with diffusion and global optimization. *arXiv preprint arXiv:2404.00429*, 2024.

[26] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[27] 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)*, volume 3, pages 2149–2154. Ieee, 2004.

[28] Chenqi Li, Yuan Ren, and Bingbing Liu. Pcgen: Point cloud generator for lidar simulation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11676–11682. IEEE, 2023.

[29] Jiahao Li, Changhao Zhang, Ziyao Xu, Hangning Zhou, and Chi Zhang. Iterative distance-aware similarity matrix convolution with mutual-supervised point elimination for efficient point cloud registration. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16*, pages 378–394. Springer, 2020.

[30] 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.

[31] Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. Barf: Bundle-adjusting neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5741–5751, 2021.

[32] Yunzhi Lin, Thomas Müller, Jonathan Tremblay, Bowen Wen, Stephen Tyree, Alex Evans, Patricio A Vela, and Stan Birchfield. Parallel inversion of neural radiance fields for robust pose estimation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9377–9384. IEEE, 2023.

[33] Yu-Lun Liu, Chen Gao, Andreas Meuleman, Hung-Yu Tseng, Ayush Saraf, Changil Kim, Yung-Yu Chuang, Johannes Kopf, and Jia-Bin Huang. Robust dynamic radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13–23, 2023.

[34] Fan Lu, Guang Chen, Yinlong Liu, Lijun Zhang, Sanqing Qu, Shu Liu, and Rongqi Gu. Hregnet: A hierarchical network for large-scale outdoor lidar point cloud registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16014–16023, 2021.

[35] 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.

[36] Eleonora Maset, Federica Arrigoni, and Andrea Fusiello. Practical and efficient multi-view matching. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4568–4576, 2017.

[37] 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.

[38] 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.

[39] Michael Niemeyer, Jonathan T Barron, Ben Mildenhall, Mehdi SM Sajjadi, Andreas Geiger, and Noha Radwan. Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5480–5490, 2022.

[40] Michael Oechsle, Songyou Peng, and Andreas Geiger. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5589–5599, 2021.

[41] Keunhong Park, Philipp Henzler, Ben Mildenhall, Jonathan T Barron, and Ricardo Martin-Brualla. Camp: Camera preconditioning for neural radiance fields. *ACM Transactions on Graphics (TOG)*, 42(6):1–11, 2023.

[42] 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.

[43] François Pomerleau, Francis Colas, Roland Siegwart, and Stéphane Magnenat. Comparing icp variants on real-world data sets: Open-source library and experimental protocol. *Autonomous robots*, 34:133–148, 2013.

[44] Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11143–11152, 2022.

[45] Srikumar Ramalingam and Yuichi Taguchi. A theory of minimal 3d point to 3d plane registration and its generalization. *International journal of computer vision*, 102:73–90, 2013.

[46] Barbara Roessle, 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.

 https://doi.org/10.52202/079017-2343

[47] Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the icp algorithm. In *Proceedings third international conference on 3-D digital imaging and modeling*, pages 145–152. IEEE, 2001.

[48] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016.

[49] 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.

[50] Liang Song, Guangming Wang, Jiuming Liu, Zhenyang Fu, Yanzi Miao, et al. Sc-nerf: Self-correcting neural radiance field with sparse views. *arXiv preprint arXiv:2309.05028*, 2023.

[51] 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.

[52] GK Tejus, Giacomo Zara, Paolo Rota, Andrea Fusiello, Elisa Ricci, and Federica Arrigoni. Rotation synchronization via deep matrix factorization. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2113–2119. IEEE, 2023.

[53] Prune Truong, Marie-Julie Rakotosaona, Fabian Manhardt, and Federico Tombari. Sparf: Neural radiance fields from sparse and noisy poses. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4190–4200, 2023.

[54] Haiping Wang, Yuan Liu, Zhen Dong, Yulan Guo, Yu-Shen Liu, Wenping Wang, and Bisheng Yang. Robust multiview point cloud registration with reliable pose graph initialization and history reweighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9506–9515, 2023.

[55] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*, 2021.

[56] Yue Wang and Justin M Solomon. Deep closest point: Learning representations for point cloud registration. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3523–3532, 2019.

[57] 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.

[58] Zirui Wang, Shangzhe Wu, Weidi Xie, Min Chen, and Victor Adrian Prisacariu. Nerf–: Neural radiance fields without known camera parameters. *arXiv preprint arXiv:2102.07064*, 2021.

[59] Yi Wei, Shaohui Liu, Yongming Rao, Wang Zhao, Jiwen Lu, and Jie Zhou. Nerfingmvs: Guided optimization of neural radiance fields for indoor multi-view stereo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5610–5619, 2021.

[60] Jin Xu, Zishan Li, Bowen Du, Miaomiao Zhang, and Jing Liu. Reluplex made more practical: Leaky relu. In *2020 IEEE Symposium on Computers and communications (ISCC)*, pages 1–7. IEEE, 2020.

[61] Weiyi Xue, Fan Lu, and Guang Chen. Hdmnet: A hierarchical matching network with double attention for large-scale outdoor lidar point cloud registration. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3393–3403, 2024.

[62] 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.

[63] Lin Yen-Chen, Pete Florence, Jonathan T Barron, Alberto Rodriguez, Phillip Isola, and Tsung-Yi Lin. inerf: Inverting neural radiance fields for pose estimation. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1323–1330. IEEE, 2021.

[64] Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction. *Advances in neural information processing systems*, 35:25018–25032, 2022.

[65] Wentao Yuan, Benjamin Eckart, Kihwan Kim, Varun Jampani, Dieter Fox, and Jan Kautz. Deepgmr: Learning latent gaussian mixture models for registration. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pages 733–750. Springer, 2020.

[66] Jian Zhang, Yuanqing Zhang, Huan Fu, Xiaowei Zhou, Bowen Cai, Jinchi Huang, Rongfei Jia, Binqiang Zhao, and Xing Tang. Ray priors through reprojection: Improving neural radiance fields for novel view extrapolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18376–18386, 2022.

[67] 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.

[68] Junge Zhang, Feihu Zhang, Shaochen Kuang, and Li Zhang. Nerf-lidar: Generating realistic lidar point clouds with neural radiance fields. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 7178–7186, 2024.

[69] 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.

[70] Zehan Zheng, Fan Lu, Weiyi Xue, Guang Chen, and Changjun Jiang. Lidar4d: Dynamic neural fields for novel space-time view lidar synthesis. *arXiv preprint arXiv:2404.02742*, 2024.

 https://doi.org/10.52202/079017-2343

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims made in the abstract and introduction accurately reflect the paper's contributions and scope.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: This paper does discuss the limitations of the work performed by the authors.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

Justification: This paper does not include theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: This paper fully discloses all the information needed to reproduce the main experimental results of the paper.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

https://doi.org/10.52202/079017-2343

Answer: [Yes]

Justification: This paper provide open access to the data and code.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: This paper specify all the training and test details.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: This paper report appropriate information about the statistical significance of the experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: This paper provide sufficient information on the computer resources.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This research conducted in the paper conform, in evrery respect, with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: There is no societal impact of the work performed. No harm technical paper.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

https://doi.org/10.52202/079017-2343

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: This paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We mentioned creators or original owners of assets and properly respected.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: This paper introduces new assets well documented.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing nor research with human subjects

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

https://doi.org/10.52202/079017-2343