# CCNDF: Curvature Constrained Neural Distance Fields from 3D LiDAR Sequences

Akshit Singh[0009−0001−5569−7170], Karan Bhakuni[0009−0003−4448−6583], and Rajendra Nagar[0000−0001−8087−0468]

Indian Institute of Technology Jodhpur, India
{singh.190,m22rm003,rn}@iitj.ac.in

**Abstract.** Neural distance fields (NDF) have emerged as a powerful tool for solving 3D computer vision and robotics downstream problems. While significant progress has been made in learning NDF from point cloud data obtained through a LiDAR scanner, a crucial aspect that demands attention is the supervision of neural fields during training, as ground-truth NDFs are not available for large-scale outdoor scenes. The existing works have approximated signed distance to guide model learning. The efficiency of the trained model heavily depends on the approximation of the signed distance. To this end, we propose a novel methodology leveraging second-order derivatives of the NDF for a better approximation of the signed distance that leads to improved neural field learning. To assess the efficacy of our methodology, we conducted comparative evaluations against prevalent methods for mapping and localization tasks. Our results demonstrate the superiority of the proposed approach compared to the state-of-the-art techniques, highlighting its potential for advancing the capabilities of neural distance fields in computer vision and graphics applications.

**Keywords:** Neural Implicit Representation · Signed Distance Field · 3D Mapping · Localization.

## 1 Introduction

The neural distance field is a fundamental 3D surface representation used in various downstream tasks of 3D computer vision and robotics, such as autonomous driving and content creation for AR-VR applications. The appeal of NDF lies in its ability to provide a continuous representation [22], unhindered by grid resolution constraints [4], making it versatile. Moreover, NDF accurately captures high-frequency information and serves as a memory-efficient approach for mapping and utilization in mobile tasks [29, 37]. Traditionally, NDF training involves expensive calculations of ground-truth signed distance values [18]. While supervising NDF with ground truth signed distances laid crucial foundations, obtaining such ground truth data for real-world scenes is challenging. The recent approaches [36, 4] predict signed distances with only point clouds as the input, eliminating the need for ground truth supervision where they leverage the properties of NDF to approximate expected ground truth values. However, some of these methods introduce geometric assumptions and train with sub-optimal approximation of the ground truth values, potentially leading to geometric artifacts in the predicted NDF [29, 4, 18].

In this work, we introduce a novel NDF supervision method based on the second-order derivative property of NDF for overcoming inherent geometric limitations and ensuring a precise alignment of NDF properties with the estimated signed distance. While higher-order derivatives have been explored for diverse applications, such as shape smoothing [34], the second-order properties of NDF have yet to be tapped in the context of supervision of NDF learning. The effectiveness of our approach is evaluated on two fundamental problems, mapping, and localization, both of which extensively utilize NDF [37, 29]. Our findings indicate an improvement over the current state-of-the-art geometrical techniques, affirming the superiority of our proposed method.

In summary, we make the following contributions:

- We present a novel neural distance field supervision approach by exploiting the second-order derivatives of the neural distance field.
- We develop a geometrical approach to ensure a precise alignment of neural distance field properties with the estimated signed distances, which enhances the accuracy and reliability of neural distance field models.
- We demonstrate that the proposed approach attains state-of-the-art (SOTA) performance for mapping and localization on benchmark challenging datasets.

The rest of the paper is organized as follows. In Section 2, we review the relevant literature and discuss the major research gaps. In Section 3, we describe the proposed approach to solve the considered problem and discuss its geometric significance. In Section 4, we present our results and compare the performance of the proposed approach with that of the state-of-the-art approaches.

## 2 Related Work

The uses of signed distance fields (SDF) have been prevalent in computer graphics and computational geometry for various applications [16, 19]. More recently, neural fields have emerged as a prominent representation for modeling three-dimensional environments and objects [13, 14, 17, 36]. The NDF extends the concept of SDFs by using neural networks to represent and learn the implicit function describing the surface of a scene or an object. Instead of using traditional mathematical formulations for surfaces, NDFs employ neural networks to model the signed distance function [4]. NDFs have demonstrated significant success in efficiently representing room-scale 3D environments and are particularly useful in real-time applications [18, 23, 38, 29]. The majority of the approaches that have employed NDF have utilised multi-layer perceptron (MLP) and supervised them either through density fields [25], normals [22, 30, 36], or directly through the distance field [4]. Recently, few approaches investigated the possibility of relating NDF directly to sensor readings [29, 37, 2] or supervise NDF directly with sensor readings of small batches while ignoring the entire geometry leading to a

distorted view [18], especially for large-scale environment representation.
Among the limited number of approaches that have directly utilised sensor readings from LiDAR for mapping and localization [35, 21, 32, 31, 27, 20, 29, 18, 33, 23, 24, 28, 2], the SHINE-Mapping [37] and LocNDF [29] approaches have used implicit surface representation for large-scale environments. NDF has also been applied in the fields of motion planning [3], localization [29], navigating in neural field [1], and learning neural fields for deformable objects [33]. The objective

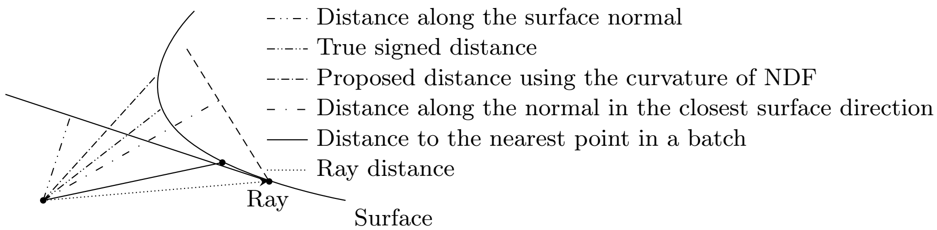



**Fig. 1.** Different methods for computing the signed distance to supervise NDF for a point on a ray beam.

of our research is to introduce a novel technique for effectively learning the NDF for large-scale environments by utilising the LiDAR sensor readings, which can subsequently be applied to a range of tasks including path planning, mapping and localization, and more. Our approach method leverages the geometric characteristics of the environment and also contributes to the current properties of NDF.
**Geometric Drawbacks of Existing Methods**: Several methodologies have been devised to supervise NDF with estimated signed distances which frequently use the ray distance as the approximated signed distance [37]. Subsequent advancements in this line of research involved refining the supervision of NDF by incorporating the ray distance computed along the surface normal as the estimated signed distance [18, 36]. The first-order derivative-based approaches calculate the ray distance along the normal pointing towards the closest surface [29]. However, no method has yet investigated the higher-order properties of NDF to determine the expected NDF values and utilise them for network supervision. In Fig. 1, we represent various approximations of the signed distances that have been used for supervising MLP where we illustrate the geometric limitations of different distances. For instance, using the nearest point to the sensor in a specific batch results in an underestimation of the ground truth signed distance values because it does not account for the geometry of the extended surface. Considering the distance along the surface normals as the true signed distance results in significant errors on a wide scale [18]. Considering the distance along the ray directly provides a simple option but is not particularly helpful for monitoring NDF as the ray distance is clearly not the minimal distance to the surface. i-SDF [18] specifically designed to map room-

scale environments directly calculated the closest distance to the surface in the selected batch and determined it as the global minimum distance to the surface. Though this technique was effective in a room-scale setting, gives sub-optimal efficiency in a large-scale environment due to the extended geometry's impact on the NDF, which may not have been captured in the chosen batch. Another seminal innovation in the estimation of the ground truth value was introduced in [29], which uses the first-order derivative of NDF to determine the gradient to the nearest surface. They then calculate the ray distance along this determined gradient, resulting in a more accurate estimation of the signed distance value. However, this method frequently produced inaccurate estimations for curvilinear or complex surfaces. As in Fig. 1, we can clearly see that the distance along the normal pointing towards the closest surface overshoots the true signed distance field, resulting in error.

All other methods utilizing NDF rely on high-resolution images or require ground truth SDF for NDF training [4, 25, 22]. These methods can further be improved, particularly real-world applications, due to the lack of ground truth SDFs or high-resolution images. Instead, we only have sensor readings available to build the NDF. So for this work, we would directly use the raw LiDAR data for constructing NDF.

In this work, we introduced a new methodology for calculating the predicted SDF values by utilising higher-order derivatives of the NDF. We then compare our approach with existing approaches used to estimate SDF values in large-scale environments. We specifically compare our results with ray distance [37] and distance along the normal pointing towards the closest surface [29] as these are the most widely used and latest approaches for determining signed distance values, moreover they also provide the tightest bound for approximating signed distances. Batch distance and normal distance along the surface normal are considered unsuitable for large-scale environments as they were originally designed for small-scale environments and are not theoretically applicable to large-scale settings.

## 3 Methodology

In this work, we learn the NDF representation of a scene from the LiDAR sensor readings using a self-supervised approach. Our approach eliminates the need for substantial processing of the sensor readings, creating specialized density fields, and finding ground truth signed distance value manually, which makes it more practical for real-world applications like mapping and localization. Our approach does not rely on point cloud normals as normal estimation has inherent dependence on the type of sensor, leading to potential errors in the process. In the following text, we will discuss how we exploited the properties of NDF and the environment's geometry to learn the NDF from the LiDAR data.

**Problem Formulation and Background**: In this work, we address the problem of finding the implicit representation of the 3D surface of a scene from its sequence of LiDAR scans of the scene. Inspired from [29], we term the implicit surface as neural distance field (NDF). The NDF $\mathsf{D}$ maps a point $\mathsf{x} \in \mathbb{R}^3$ to a scalar value $\mathsf{D}(\mathsf{x}) \in \mathbb{R}$ that represents the signed distance of the point $\mathsf{x}$ from the surface of the underlying scene. We use [29] as a backbone with a better approximation of the distance of a point from the surface that we discuss later in this section. We first discuss the approach of LocNDF [29] to create a context. LocNDF represents $\mathsf{D}$ as a multi-layer perceptron (MLP) that consumes a point and returns it signed distance from the surface, i.e., it predicts $\mathsf{D}(\mathsf{x})$. Therefore, the MLP learns a representation of the 3D geometry of the scene. Instead of passing a point directly as an input to the MLP, LocNDF uses the positional encoding of the points to learn the 3D surface geometry more efficiently as it captures the high frequency details of the scene [14]. Positional encoding maps the coordinates of a 3D point to a high-dimensional vector by utilizing periodic activation functions helping to store high frequency information of points. The positional encoding $\mathsf{P} : \mathbb{R}^3 \to \mathbb{R}^{6h+3}$ of a point is defined in Equation (1).

$$\mathsf{P}(\mathsf{x}) = \left[\mathsf{x}\ \sin(\omega_1 \mathsf{x})\ \cos(\omega_1 \mathsf{x})\ \cdots\ \sin(\omega_h \mathsf{x})\ \cos(\omega_h \mathsf{x})\right]^\top. \tag{1}$$

In order to keep the surface thin, LocNDF uses truncated signed distance field (TSDF) [15, 26] representation of the NDF. The structure of the multi-layer perceptron is inspired from the SHINE [37] to learn the NDF from LiDAR scans where the training is done in an unsupervised manner as we do not have the ground-truth NDF available. A LiDAR scanner measures the distance of a scene point from its origin and this distance is referred as the ray distance. To train the MLP, LocNDF [29] proposed a self-supervised technique where a set of points $\{\mathsf{x}_1, \ldots, \mathsf{x}_{n_i}\}$ are sampled on the ray between the sensor origin $\mathsf{o}_i$ and the ray endpoint $\mathsf{e}_i$ (the scene point where the ray intersects the scene surface) as defined in Equation (2).

$$\mathsf{x}_\ell = (1 - t_\ell)\mathsf{o}_i + t_\ell \mathsf{e}_i \ \text{ and } \ t_\ell = \frac{1}{0.9}\left(1 - 10^{\frac{\ell}{n_i - 1} - 1}\right),\ \ell \in \{1, \ldots, n_i\}. \tag{2}$$

Now, in order to train the MLP to learn the underlying surface, we require the ground-truth signed distance of each sample point $\mathsf{x}_\ell$ from the surface. Since both surface and the signed distance of each sample point is unknown, it is a chicken-and-egg problem to solve. This is a major challenge to address to solve this problem. The LocNDF framework uses the ray distance $d_\ell = \|\mathsf{e}_i - \mathsf{x}_\ell\|_2$ to approximate the signed-distance for each sample point from the surface that itself we have to learn. The LocNDF approximates the signed distance $\bar{d}_\ell$ of the $\mathsf{x}_\ell$ from the surface, defined by NDF $\mathsf{D}$, as

$$\bar{d}_\ell = \frac{\mathsf{n}_\ell^\top (\mathsf{e}_i - \mathsf{x}_\ell)}{\|\mathsf{n}_\ell\|_2}, \tag{3}$$

where $\mathsf{n}_\ell$ represents the approximation of the direction towards the closest surface and LocNDF defined it as $\mathsf{n}_\ell = -\nabla_{\mathsf{x}} \mathsf{D}(\mathsf{x}_\ell)$.

### 3.1 Curvature-Constrained Neural Distance Fields

**Proposed Approach**: In order to estimate the distance of a point from the surface, we utilize the characteristics of NDF. Given that the NDF represents the distance of a point to the closest surface, and the iso-lines of the NDFs are concentric to the surface of object, indicating that point $\mathsf{x}_\ell$ will lie on isolines of the NDF. Isolines represent the NDF and are formed by the distance to the nearest surface. Fig. 2(a) shows that the object's NDF are concentric, and the radius of curvature (ROC) of isolines increases in magnitude as one moves away from the surface. Based on this intuition, the radius of curvature of the isoline at

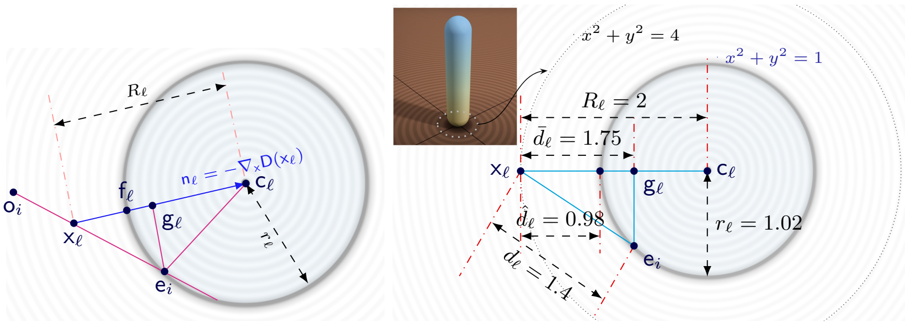



**Fig. 2.** (a) Various approximation of the signed distance of the point $\mathsf{x}_\ell$ from the surface: $d_\ell = \|\mathsf{x}_\ell - \mathsf{e}_i\|_2$ for Ray distance, $\bar{d}_\ell = \|\mathsf{x}_\ell - \mathsf{g}_\ell\|_2 = \frac{\mathsf{n}_\ell^\top(\mathsf{e}_i - \mathsf{x}_\ell)}{\|\mathsf{n}_\ell\|_2}$ for LocNDF, and proposed $\hat{d}_\ell = \|\mathsf{x}_\ell - \mathsf{f}_\ell\|_2 = R_\ell - r_\ell$ for the proposed approach, (b) An example is shown for all three approximation of the signed distance of the point $\mathsf{x}_\ell$. The true distance is 1 unit and the proposed distance is $\hat{d}_\ell = 0.98$ as we use numerical approximation of $\nabla_\mathsf{x}\mathsf{D}(\mathsf{x}_\ell)$ and $R_\ell$ to calculate $\hat{d}_\ell$ using Equation (11).

a given point is the ROC of that point. As isolines are only formed to represent the distance to the nearest surface, this would ensure that the ROC of a point on the iso-line would be concentric with the ROC of the corresponding point that is the nearest point on the surface from the query point. Consequently, if the ROC of the corresponding point on the surface could be determined, the difference between the two would provide a reliable approximation of the signed distance. While NDFs can be affected by multiple surfaces, discontinuities typically arise at significant distances. When closer to the surface, NDFs tend to be concentric to the inquiry point [5]. To mitigate discontinuity effects, we have sampled points log-linearly along the LiDAR ray, as defined in Equation (2), assigning higher weights to those closer to the surface. This approach efficiently reduces discontinuity impacts on distance calculation. Now, according to [8], the ROC for any point in the implicit representation is defined as $R_\ell = \frac{1}{|\kappa_\ell|}$. Here, $\kappa_\ell$ is the mean curvature at a point $\mathsf{x}_\ell$ on the implicit surface and is defined as:

$$\kappa_\ell = \frac{\nabla_\mathsf{x}^\top \mathsf{D}(\mathsf{x}_\ell)\mathsf{H}(\mathsf{x}_\ell)\nabla_{\mathsf{x}_\ell}\mathsf{D}(\mathsf{x}_\ell) - \|\nabla_\mathsf{x}\mathsf{D}(\mathsf{x}_\ell)\|_2^2 \mathrm{trace}(\mathsf{H}(\mathsf{x}_\ell))}{2\|\nabla_\mathsf{x}\mathsf{D}(\mathsf{x}_\ell)\|_2^3}. \tag{4}$$

The mean curvature $\kappa_\ell$ can be found by Equation (4) for calculating the ROC for implicit surfaces if we have the analytical expression of $\mathsf{D}$ available with us. The NDF is theoretically described as a scalar field representing the extrinsic properties of the scene. The formula described in Equation (4) is intricate and computationally demanding. A more computationally efficient way [8] to calculate the mean curvature is defined in Equation (5).

$$\kappa_\ell = \left| \nabla_{\mathsf{x}}^\top \left( \frac{\nabla_{\mathsf{x}} \mathsf{D}(\mathsf{x}_\ell)}{\|\nabla_{\mathsf{x}_\ell} \mathsf{D}(\mathsf{x})\|_2} \right) \right| = \frac{1}{R_\ell} \Rightarrow R_\ell = \left| \nabla_{\mathsf{x}}^\top \left( \frac{\nabla_{\mathsf{x}} \mathsf{D}(\mathsf{x}_\ell)}{\|\nabla_{\mathsf{x}_\ell} \mathsf{D}(\mathsf{x})\|_2} \right) \right|^{-1}. \tag{5}$$

Now, we establish a relationship between the ROC of a point and its closet point on the surface. As depicted in Fig. 2 (a), we observe that the NDF comprises of concentric spheres and when we apply the Equation (5) at any sampled query point $\mathsf{x}_\ell$ on the LiDAR beam having an origin at $\mathsf{o}_i$, we receive the ROC of isoline at $\mathsf{x}_\ell$. As a property of NDF, we are assured that this ROC is concentric with ROC of the point on the surface nearest to the query point. Here the corresponding point on the surface to the query point $\mathsf{x}_\ell$ is point $\mathsf{f}_\ell$. The formation of the neural field at point $\mathsf{x}_\ell$ is solely due to the point $\mathsf{f}_\ell$ on the surface, as the NDF is a representation of the distance to the nearest point.
Let the ROC of isoline at point $\mathsf{x}_\ell$ be $R_\ell$, and let $d_\ell = \|\mathsf{x}_\ell - \mathsf{e}_i\|_2$ be the known ray distance to the surface, where $\mathsf{e}_i$ is the point on the surface at which LiDAR ray intersects. We approximate the signed distance value of the point $\mathsf{x}_\ell$ from the surface as $\hat{d}_\ell = \|\mathsf{x}_\ell - \mathsf{f}_\ell\|_2$, i.e., the difference of the ROCs of the surface at the query point and its the corresponding point $\mathsf{f}_\ell$. Using Equation (5), we find the radius $R_\ell$. We exploit the geometry present to calculate the ROC $r_\ell$ at the corresponding point $\mathsf{f}_\ell$ as below. In Fig. 2 (a), we apply the cosine rule in triangle $\triangle \mathsf{x}_\ell \mathsf{e}_i \mathsf{c}_\ell$ as follows:

$$\|\mathsf{e}_i - \mathsf{c}_\ell\|_2^2 = \|\mathsf{e}_i - \mathsf{x}_\ell\|_2^2 + \|\mathsf{c}_\ell - \mathsf{x}_\ell\|_2^2 - 2\|\mathsf{e}_i - \mathsf{x}_\ell\|_2 \|\mathsf{c}_\ell - \mathsf{x}_\ell\|_2 \cos\theta \tag{6}$$

$$r_\ell^2 = d_\ell^2 + R_\ell^2 - 2 d_\ell R_\ell \frac{\mathsf{n}_\ell^\top (\mathsf{e}_i - \mathsf{x}_\ell)}{\|\mathsf{e}_i - \mathsf{x}_\ell\|_2} \tag{7}$$

$$= d_\ell^2 + R_\ell^2 - 2 R_\ell \mathsf{n}_\ell^\top (\mathsf{e}_i - \mathsf{x}_\ell) \tag{8}$$

$$= d_\ell^2 + R_\ell^2 + 2 R_\ell (\mathsf{e}_i - \mathsf{x}_\ell)^\top \nabla_{\mathsf{x}} \mathsf{D}(\mathsf{x}_\ell). \tag{9}$$

Here, $\theta = \angle \mathsf{c}_\ell \mathsf{x}_\ell \mathsf{e}_i$. Our proposed estimation $\hat{d}_\ell$ for the ground-truth signed distance of the point $\mathsf{x}_\ell$ from the surface is defined in Equation (11).

$$\hat{d}_\ell = R_\ell - r_\ell \tag{10}$$

$$\hat{d}_\ell = R_\ell - \sqrt{d_\ell^2 + R_\ell^2 + 2 R_\ell (\mathsf{e}_i - \mathsf{x}_\ell)^\top \nabla_{\mathsf{x}} \mathsf{D}(\mathsf{x}_\ell)}. \tag{11}$$

We use this $\hat{d}_\ell$ to supervise the training of a multi-layer perception to learn the NDF. In Fig. 2(b), we present an example for the approximation of $\hat{d}_\ell$ . By observing the projected NDF, which forms a circle at the inspection plane. The ray distance (SHINE mapping) is 1.41 that we directly obtain from the LiDAR measurements, the distance along the normal pointing toward the closest surface

$\bar{d}_\ell$ proposed by LocNDF [29] is 1.75 (Equation (3)), ROC of NDF is equal to 2, approximated ROC of the surface is equal to 1.02 (Equation (5)), the distance using our method is equal to 0.98, and the true signed distance is equal to 1. We observe that the proposed distance (0.98) is very close to the ground-truth distance as compared to the Ray Distance (1.41) and the distance (1.75) proposed by LocNDF.

Now, since the learning of NDF is supervised from estimated $\hat{d}_\ell$ which is estimated from the learned NDF. This is a chicken-and-egg problem to solve. Therefore, better we estimate $\hat{d}_\ell$, better we will learn NDF and vice-versa. Due to this, the training of the MLP may not be stable as we initialize the MLP with random weights leading to random SDF but neither we nor LocNDF [29] witnessed this situation. The rationale for not observing any instability is that the estimated error

$$\epsilon_\ell = \mathsf{D}(\mathsf{P}(\mathsf{x}_\ell)) - R_\ell + \sqrt{d_\ell^2 + R_\ell^2 + 2R_\ell(\mathsf{e}_i - \mathsf{x}_\ell)^\top \nabla_{\mathsf{x}} \mathsf{D}(\mathsf{P}(\mathsf{x}_\ell))} \tag{12}$$

approaches zero when the query point $\mathsf{x}_\ell$ is sampled closer to the surface, since $\mathsf{D}(\mathsf{P}(\mathsf{x}_\ell))$ converge to zero, leading $\hat{d}_\ell$ to converge to zero as well. To ensure this, we give higher weight $w_\ell = (d_{max} - \mathsf{D}(\mathsf{P}(\mathsf{x}_\ell)))^\gamma$ to the points $\mathsf{x}_\ell$ with lower $\mathsf{D}(\mathsf{P}(\mathsf{x}_\ell))$, as proposed in [29, 26]. Here, $d_{max}$ is the largest distance in the batch, while $\gamma$ is a parameter that determines the degree of weighting given to the nearby point. A larger value results in a greater influence from the nearby point. Now to supervise the MLP to learn the NDF, we define the loss function in Equation (13) where we use the proposed estimation of the signed distance $\hat{d}_\ell$ for ray points $\mathsf{x}_\ell$ along with the regularizers proposed in LocNDF [29].

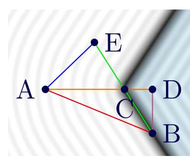
$$\mathcal{L} = \sum_\ell \left( \frac{w_\ell |\epsilon_\ell|}{\sum_j w_j} + \lambda_1 |\mathsf{D}(\mathsf{P}(\mathsf{e}_\ell))| + \lambda_2 | \|\mathsf{n}_\ell\|_2 - 1| + \lambda_3 \sum_{j \in \mathcal{N}_\ell} |\mathsf{n}_\ell^\top \mathsf{n}_j| \right). \tag{13}$$

Here, $\mathsf{n}_j = -\nabla_{\mathsf{x}} \mathsf{D}(\mathsf{x}_j)$ for the neighboring point $\mathsf{x}_j \in \mathcal{N}_\ell$ of $\mathsf{x}_\ell$, and $\lambda_1, \lambda_2$ and $\lambda_3$ are hyper-parameters that we empirically determine.

### 3.2 Geometric Advantages

Equation (11) holds even for a complex and intricate surface as this equation is defined for the differential element of NDF at a point. Previous methods for estimating signed distances, like ray distances and ray distance along the surface normal, only considered the characteristics of the endpoint on the surface ($\mathsf{e}_i$) to estimate the signed distance. Theoretically, the NDF should not be dependent on this point because it might not be the nearest point on the surface to the query point; therefore, it is problematic to have an NDF that is dependent on the characteristics of this point ($\mathsf{e}_i$). The ray distance along the normal to the nearest surface incorporates specific attributes extracted from the nearest point on the surface to the query point. By determining the direction to the closest point and estimating the signed distance, this method improves NDF supervision by making the method reliant on the nearest point on the surface. However, this approach solely depends on the directional (vector) characteristics of the closest point on the surface, resulting in the estimated signed distance affected by the directional attributes of the nearest point on the surface only. Our method estimates the signed distance as the difference between the ROC at the query point and the point closest to the query point on the surface. Using our method, the NDF became more dependent on the scalar property of the nearest point on the surface to the query point, which is directly accountable for generating the neural field at that particular location. The dependence of NDF on the scalar properties of the closest surface point to the query point enhances the reliability and accuracy of NDF supervision.

For example, in the geometry shown in the inset figure, if the ray intersects at point $B$, methods that rely on calculating signed distances based on attributes of the endpoint ($B$), such as ray distance and ray distance along the surface normal, result in significant geometric errors. Conversely, using ray distance along the normal to the closest surface overestimates the distance by presuming the geometry to be linear, as it only examines the direction to the closest surface for estimating the distance. Whereas, our method relies on the concentric properties of NDF to calculate the concentric ROC at points $A$ and $C$, resulting in a better estimate for the signed distance represented as segment $AC$. It is important to highlight that our method heavily depends on point $C$ and its magnitude properties, which are the actual cause of the neural field at point $A$, allowing us to incorporate extended geometric characteristics that would not be feasible if we only focused on properties of $B$.




## 4 Results and Evaluation

Our evaluation strategy involved comparing our approach with other relevant methodologies on two fundamental problems: mapping and localization. Given that geometric techniques within NDF are employed for diverse purposes, from motion planning to localization, and each model utilizing these approaches serves distinct objectives, direct model-to-model comparisons may not be the most suitable approach. Instead, our evaluation strategy involved comparing different geometric approaches under unified conditions. This allowed for a more appropriate and insightful assessment, considering the diverse applications and objectives of these methodologies within the NDF framework.

**Training Setup.** For all the experiments conducted in this study, consistent parameters were employed. The positional encoding parameter, denoted as $h$, was set to 30. The implementation of the MLP utilized the SIREN architecture [22], with the hidden feature dimension size set to 128. During training, $n_i$, representing the number of intervals between the sensor's starting point and endpoint, was set to 40. The coefficients for various loss components, as indicated

loss equation 13, were set as follows: $\alpha_1 = 10^{-1}$, $\alpha_2 = 10^{-4}$, $\alpha_3 = 10^{-3}$, and $\gamma = 3$. The optimization algorithm utilized is AdamW, with an initial learning rate set to $10^{-4}$. The experiments have been developed on a desktop PC with Intel @ Xeon(R) Gold 6226R CPU @ 2.90GHz $\times$ 32 and an Nvidia RTX A6000.
**Mapping**: We evaluate our method on the Apollo Southbay ColumbiaPark-3 mapping run [12] and the KITTI Sequence [7], utilizing provided poses for different batches of scans, each comprising over 700 scans. Our approach involves direct training of the MLP for the entire scene, utilizing a bounding box of size 50m. The network was trained for 10 epochs, a process that took approximately 20 minutes. Scan registration followed the same procedure as employed by LocNDF [29]. For visualization purposes, we solely utilized the mesh obtained from marching cubes [11], and not for registration. The evaluation involved a comparative analysis of our results with other geometrical distances that have been previously employed in similar contexts. In Fig. 3 and Fig. 4, we present a qualitative comparison highlighting the geometric enhancements achieved by our method on different sequences from the Apollo ColumbiaPark-3 dataset [12] and KITTI dataset [7], respectively. The results clearly illustrate the limitations of supervision by ray distance, where the representation lacks detail and exhibits significant gaps. In the mapping obtained through supervision by ray distance along the normal towards the closest surface, we observe an improvement with finer details, a clearer path, and visible surrounding buildings. The results of our approach demonstrate a significant increase in detail. Our model effectively captures finer details, such as cars and nearby trees, that were previously missing. The visual comparison across all methods highlights the effectiveness of our proposed model. Fig. 5 represents the front view mapping of the sequence showcasing geometrical features captured by different approaches on the Southbay ColumbiaPark-3 dataset [12].
**Localization**: The rationale behind evaluating NDF based on localization stems from the fact that, while mapping lacks ground truth for evaluation, localization provides a means of comparison with available ground truth locations. The premise is that if localization performs well, it implies good mapping and, by extension, well-constructed NDF. To quantitatively analyze our methods, we employed global 2D Monte Carlo Localization (MCL). In order to compare our approach, we utilized the sequence as given by Kuang *et al.* [10]. The particle filter transitions to pose tracking mode with 10,000 particles when the standard deviation of particles falls below 30 cm. The particles are reweighted and resampled if the agent moves by 5 cm or 0.1 radians. The detailed implementation of localization followed the same procedure as employed by the authors of LocNDF [29], ensuring consistency in process and technicality. For method comparison, we utilized root mean squared error (RMSE) and mean average error (MAE) between ground truth and estimated positions. The reported metrics represent averages over 5 runs, with metrics reported only if at least one run converges; otherwise, they are represented as "-". We conducted a comparative analysis of our approach against the current SOTA geometrical approach under unified conditions. Table 1 presents the comparison between the MCL results obtained by

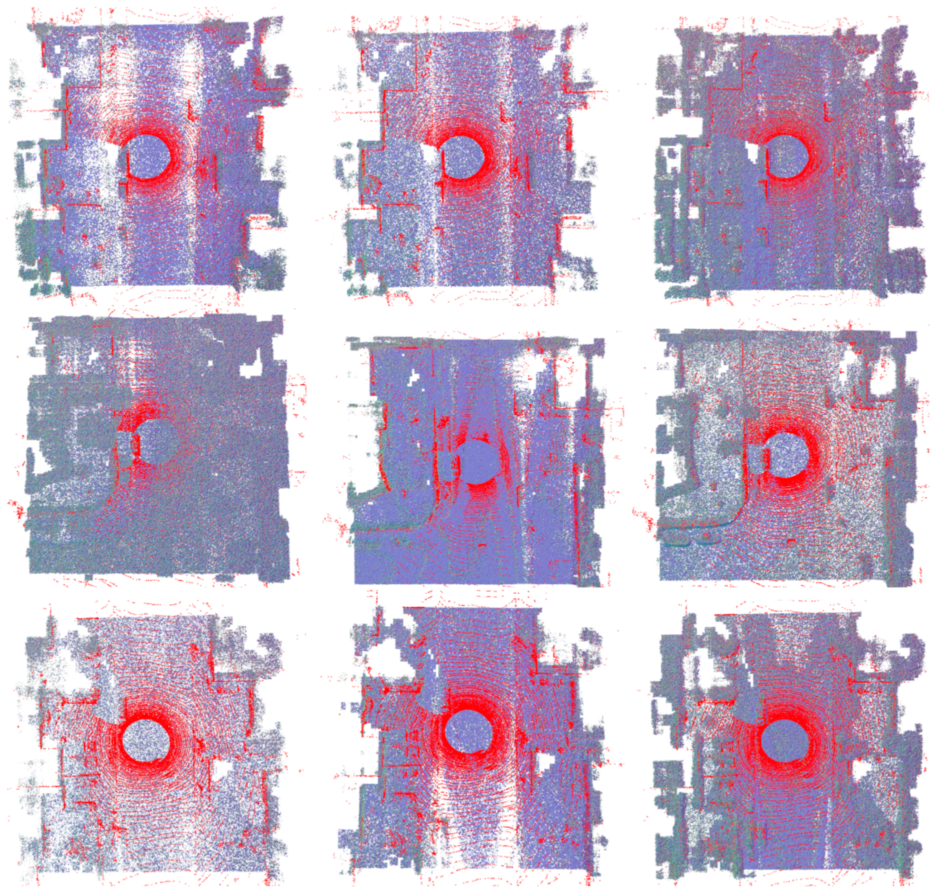
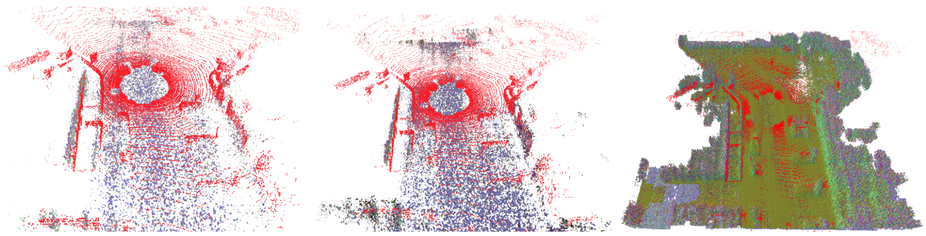


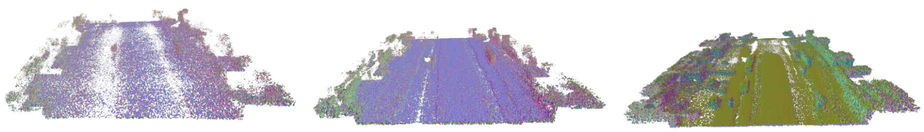

**Fig. 3.** Visual comparison of results of the proposed approach with that of SHINE-Mapping [37] and LocNDF [29] on the Apollo Southbay ColumbiaPark-3 dataset [12].

supervising the NDF with expected distances calculated through our proposed method and when using ray distance along the normal pointing towards the closest surface. The results clearly indicate that our approach outperformed the SOTA geometrical approach, as evidenced by lower values of MAE and RMSE. Furthermore, the smaller differences in MAE and RMSE values suggest that our approach generated fewer outliers. The MCL results when the NDF is supervised with ray distance showcase randomized and incoherent outcomes. In many cases, the results fail to converge to pose tracking mode at all. This outcome was anticipated due to the inherent challenges in mapping when supervised by ray distance, as seen in the previous subsection. The ineffective mapping adversely impacts localization. To comprehensively evaluate our approach, we conducted tests against various baseline localization methods, including AMCL[6], SRRG



SHINE-Mapping LocNDF Proposed Approach

**Fig. 4.** Visual comparison of results of the proposed approach with that of SHINE-Mapping [37] and LocNDF [29] on a KITTI Sequence [7].



SHINE-Mapping LocNDF Proposed Approach

**Fig. 5.** Illustration of front-view mappings obtained using SHINE-Mapping [37], LocNDF [29], and the proposed approach on the KITTI Sequence [7].

[9], IR-MCL[10], and LocNDF [29]. This comparative analysis aimed to provide a more holistic view of our model in comparison to established baselines. The results are reported in Table 2, and our findings suggest that our method outperformed the baseline models. A noteworthy observation from our results is that, unlike other models where RMSE values continue to increase with an increment in the threshold from 5 to 20 cm, our values exhibit convergence and demonstrate a saturating behavior. We attribute this phenomenon to a better continuous map representation facilitated by our avoidance of grid resolution limitations and supervision with improved geometrical distances.

## 5 Discussion, Limitations and Future work

**Geometrical Insights Leading to Improved Performance.** Our method excelled due to the effect of the incorporation of expanded geometry, as mentioned. Moreover, the observed superior performance can be attributed to a macroscopic event, which is especially advantageous in large-scale environments. This macro-event is a consequence of the formation and behavior of neural fields over substantial distances. Remarkably, the neural fields tend to approximate a circular form as the distance increases. This phenomenon is attributed to the corners of the object acting as centers for arcs within the neural field. The intensity of this event increases when the corners are rounded, making this phenomenon more observable in real-world situations. Furthermore, when the inquiry point is positioned at a sufficiently large distance from the surface, the significance

**Table 1.** MCL Results and Comparative Analysis of Different Geometrical Approaches

| Error Metric | Distance | Seq-1 | Seq-2 | Seq-3 | Seq-4 | Seq-5 |
|---|---|---|---|---|---|---|
| RMSE | Closest point along normal | 4.90 | 4.20 | 5.30 | 6.90 | 7.10 |
| | Ray Distance | 4.42 | 3.80 | 4.42 | **4.55** | **4.63** |
| | Proposed Distance | **1.40** | **1.80** | **3.40** | 5.30 | 4.90 |
| MAE | Closest point along normal | 4.4 | 3.8 | 4.6 | 5.7 | 6.5 |
| | Ray Distance | 1.31 | 1.97 | 2.93 | **4.57** | **4.57** |
| | Proposed Distance | **1.3** | **1.9** | **2.9** | 4.6 | 4.6 |

**Table 2.** MCL results and comparison of the performance of the proposed approach with that of AMCL [6], SRRG [9], IR-MCL [10], and LocNDF [29].

| Error Metric | Method | Seq-1 | Seq-2 | Seq-3 | Seq-4 | Seq-5 |
|---|---|---|---|---|---|---|
| RMSE (5cm) | AMCL [6] | – | 3.7 | 3.8 | 3.4 | – |
| | SRRG [9] | 3.4 | 3.4 | 3.3 | 3.5 | 3.5 |
| | IR-MCL [10] | 3.3 | 2.9 | 3.3 | 3.3 | 3.2 |
| | LocNDF [29] | 3.1 | 2.8 | **2.8** | **3.1** | **2.7** |
| | Proposed | **1.4** | **1.9** | 2.9 | 3.4 | 3.4 |
| RMSE (10cm) | AMCL [6] | – | 6.1 | 6.6 | 4.8 | – |
| | SRRG [9] | 4.7 | 5.9 | 5.0 | 5.2 | 5.1 |
| | IR-MCL [10] | 4.7 | 4.8 | 4.3 | 5.4 | 5.7 |
| | LocNDF [29] | 4.1 | 3.2 | 3.4 | 4.6 | **3.2** |
| | Proposed | **1.4** | **1.9** | **3.4** | **4.5** | 4.9 |
| RMSE (20cm) | AMCL [6] | – | 8.9 | 9.9 | 5.9 | – |
| | SRRG [9] | 4.9 | 6.3 | 7.5 | 5.8 | 5.5 |
| | IR-MCL [10] | 5.2 | 5.4 | 5.4 | 9.1 | 6.4 |
| | LocNDF [29] | 4.7 | 4.2 | 5.3 | 6.9 | 7.1 |
| | Proposed | **1.4** | **1.9** | **3.4** | **5.3** | **4.9** |

of the gradient diminishes while the importance of the ROC remains, providing an accurate scalar value. Notably, the gradient effectively aligns along the ray direction for any point far from the surface due to the diminishing difference between the ray's endpoint and the closest point on the surface. In contrast, the significance of the ROC persists, remaining dependent on the closest point. This characteristic ensures that the ROC provides an accurate scalar value, a crucial attribute in our context. Notably, with increased distance, neural fields tend to adopt a rounded configuration. This effect becomes more prominent in real-world scenarios, especially when the corners are rounded. Additionally, positioning the inquiry point at a considerable distance from the surface reveals an interesting dynamic. Here, the significance of the gradient diminishes while the importance of the ROC persists, providing a reliable scalar value. Specifically, the gradient effectively aligns along the ray direction for points at a significant distance from the surface due to the diminishing difference between the ray's endpoint and the closest point on the surface. In contrast, the significance of the ROC remains, maintaining dependence on the closest point. This characteristic ensures the ROC furnishes an accurate scalar value, a crucial attribute in our

context.
Despite employing a logarithmic linear sampling strategy that assigns heightened importance to distances computed near the surface, the observed significance of distances calculated from distant points remains crucial. This observed phenomenon contributes to the stability of our training process and strengthens our approach's credibility. Additionally, it underscores the pivotal role of the ROC in effectively supervising the NDF.
**Geometric Limitation.** The sole geometric constraint encountered in our methodology arises when the LiDAR ray fails to intersect the ROC at the closest point on the surface. For instance, in Fig. 2(a), if the ray from $\mathsf{o}_i$ to $\mathsf{e}_i$ does not intersect the ROC at point $\mathsf{f}_\ell$, the calculated ROC at $\mathsf{f}_\ell$ may introduce an error. However, the practical significance of this limitation is mitigated by our approach's emphasis on sampling more points in close proximity to the surface. In large-scale environments, where the ROC of objects is sufficiently large, incoming rays intersect the ROC at the closest point, minimizing the impact of this limitation. Additionally, for points farther from the surface, the NDF behaves akin to a sphere, further mitigating the consequences of this limitation. The alignment of this limitation with its predecessors is inherent, as the sampling of points near the surface is a common practice in effectively supervising the NDF across various methodologies.
**Future Work.** This work was conducted within an offline setting, and the prospect of extending this approach to an online setting presents an interesting future aspect. Additionally, the exploration of alternative deep learning models to model the NDF holds promise for further advancing the efficiency.

## 6 Conclusion

In this work, we introduced a novel geometric approach for supervising NDF. Our method was evaluated on two pivotal problems: mapping and localization, where NDF plays a fundamental role. We compared our approach with state-of-the-art geometric methods commonly used for NDF supervision. The results indicate that our method surpasses current techniques in performance and is more geometrically and mathematically aligned with the inherent properties of NDF. By leveraging higher-order properties of NDF for supervision, we established a foundation for further exploration of these properties to enhance NDF's versatility in various tasks. Our methodology advances the field and opens avenues for future research into higher-order properties of NDF.

## Acknowledgments

This work was supported by the SERB, Government of India, through the Start-up Research Grant (SRG) scheme and iHub-Drishti Foundation, IIT Jodhpur.

## References


1. Adamkiewicz, M., Chen, T., Caccavale, A., Gardner, R., Culbertson, P., Bohg, J., Schwager, M.: Vision-only robot navigation in a neural radiance world. IEEE Robotics and Automation Letters **7**(2), 4606–4613 (2022)
2. Azinović, D., Martin-Brualla, R., Goldman, D.B., Nießner, M., Thies, J.: Neural rgb-d surface reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6290–6301 (2022)
3. Chaplot, D.S., Pathak, D., Malik, J.: Differentiable spatial planning using transformers. In: International Conference on Machine Learning. pp. 1484–1495. PMLR (2021)
4. Chibane, J., Pons-Moll, G., et al.: Neural unsigned distance fields for implicit function learning. Advances in Neural Information Processing Systems **33**, 21638–21652 (2020)
5. *Takikawa et al.*: A dataset and explorer for 3d signed distance functions. Journal of Computer Graphics Techniques (2022)
6. Fox, D.: Kld-sampling: Adaptive particle filters. Advances in neural information processing systems **14** (2001)
7. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE conference on computer vision and pattern recognition. pp. 3354–3361. IEEE (2012)
8. Goldman, R.: Curvature formulas for implicit curves and surfaces. Computer Aided Geometric Design **22**(7), 632–658 (2005)
9. Grisetti, G.: srrg-localizer2d (1.6.0)" (2018)
10. Kuang, H., Chen, X., Guadagnino, T., Zimmerman, N., Behley, J., Stachniss, C.: Ir-mcl: Implicit representation-based online global localization. IEEE Robotics and Automation Letters **8**(3), 1627–1634 (2023)
11. Lewiner, T., Lopes, H., Vieira, A.W., Tavares, G.: Efficient implementation of marching cubes' cases with topological guarantees. Journal of graphics tools **8**(2), 1–15 (2003)
12. Lu, W., Zhou, Y., Wan, G., Hou, S., Song, S.: L3-net: Towards learning based lidar localization for autonomous driving. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6389–6398 (2019)
13. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: Occupancy networks: Learning 3d reconstruction in function space. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4460–4470 (2019)
14. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
15. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohi, P., Shotton, J., Hodges, S., Fitzgibbon, A.: Kinectfusion: Real-time dense surface mapping and tracking. In: 2011 10th IEEE international symposium on mixed and augmented reality. pp. 127–136. Ieee (2011)
16. Oleynikova, H., Millane, A., Taylor, Z., Galceran, E., Nieto, J., Siegwart, R.: Signed distance fields: A natural representation for both mapping and planning. In: RSS 2016 workshop: geometry and beyond-representations, physics, and scene understanding for robotics. University of Michigan (2016)
17. Oleynikova, H., Taylor, Z., Fehr, M., Siegwart, R., Nieto, J.: Voxblox: Incremental 3d euclidean signed distance fields for on-board mav planning. In: 2017 IEEE/RSJ

International Conference on Intelligent Robots and Systems (IROS). pp. 1366–1373. IEEE (2017)
18. Ortiz, J., Clegg, A., Dong, J., Sucar, E., Novotny, D., Zollhoefer, M., Mukadam, M.: isdf: Real-time neural signed distance fields for robot perception. system **30**, 39
19. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)
20. Reiser, C., Peng, S., Liao, Y., Geiger, A.: Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14335–14345 (2021)
21. Shan, T., Englot, B.: Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 4758–4765. IEEE (2018)
22. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. Advances in neural information processing systems **33**, 7462–7473 (2020)
23. Sucar, E., Liu, S., Ortiz, J., Davison, A.J.: imap: Implicit mapping and positioning in real-time. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6229–6238 (2021)
24. Vasilopoulos, V., Garg, S., Huh, J., Lee, B., Isler, V.: Hio-sdf: Hierarchical incremental online signed distance fields. arXiv preprint arXiv:2310.09463 (2023)
25. Vizzo, I., Chen, X., Chebrolu, N., Behley, J., Stachniss, C.: Poisson surface reconstruction for lidar odometry and mapping. pp. 5624–5630. IEEE (2021)
26. Vizzo, I., Guadagnino, T., Behley, J., Stachniss, C.: Vdbfusion: Flexible and efficient tsdf integration of range sensor data. Sensors **22**(3), 1296 (2022)
27. Vizzo, I., Guadagnino, T., Mersch, B., Wiesmann, L., Behley, J., Stachniss, C.: KISS-ICP: In Defense of Point-to-Point ICP – Simple, Accurate, and Robust Registration If Done the Right Way. IEEE Robotics and Automation Letters (RA-L) **8**(2), 1029–1036 (2023). https://doi.org/10.1109/LRA.2023.3236571
28. Wang, H., Wang, J., Agapito, L.: Co-slam: Joint coordinate and sparse parametric encodings for neural real-time slam. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13293–13302 (2023)
29. Wiesmann, L., Guadagnino, T., Vizzo, I., Zimmerman, N., Pan, Y., Kuang, H., Behley, J., Stachniss, C.: Locndf: Neural distance field mapping for robot localization. IEEE Robotics and Automation Letters (2023)
30. Williams, F., Trager, M., Bruna, J., Zorin, D.: Neural splines: Fitting 3d surfaces with infinitely-wide neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9949–9958 (2021)
31. Xu, W., Cai, Y., He, D., Lin, J., Zhang, F.: Fast-lio2: Fast direct lidar-inertial odometry. IEEE Transactions on Robotics **38**(4), 2053–2073 (2022)
32. Xu, W., Zhang, F.: Fast-lio: A fast, robust lidar-inertial odometry package by tightly-coupled iterated kalman filter. IEEE Robotics and Automation Letters **6**(2), 3317–3324 (2021)
33. Yan, Z., Tian, Y., Shi, X., Guo, P., Wang, P., Zha, H.: Continual neural mapping: Learning an implicit scene representation from sequential observations. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 15782–15792 (2021)
34. Yang, G., Belongie, S., Hariharan, B., Koltun, V.: Geometry processing with neural fields. Advances in Neural Information Processing Systems **34**, 22483–22497 (2021)

35. Zhang, J., Singh, S.: Loam: Lidar odometry and mapping in real-time. In: Robotics: Science and systems. vol. 2, pp. 1–9. Berkeley, CA (2014)
36. Zhang, J., Yao, Y., Quan, L.: Learning signed distance field for multi-view surface reconstruction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6525–6534 (2021)
37. Zhong, X., Pan, Y., Behley, J., Stachniss, C.: Shine-mapping: Large-scale 3d mapping using sparse hierarchical implicit neural representations. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). pp. 8371–8377. IEEE (2023)
38. Zhu, Z., Peng, S., Larsson, V., Xu, W., Bao, H., Cui, Z., Oswald, M.R., Pollefeys, M.: Nice-slam: Neural implicit scalable encoding for slam. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12786–12796 (2022)