Self-supervised Single-line LiDAR Depth Completion

Junjie Hu, *Member, IEEE,* Chenyou Fan, Xiyue Guo, Liguang Zhou, and Tin Lun Lam, *Senior Member, IEEE*

***Abstract*—Depth completion plays a crucial role in enabling real-world applications such as obstacle avoidance and SLAM for robot navigation. This paper focuses on addressing the depth completion challenge for single-line LiDAR, commonly used in conjunction with visual cameras. The sparsity of valid depth points makes supervised methods inadequate, while existing self-supervised approaches are only applicable to 64-line LiDARs. In this paper, we propose a novel self-supervised approach for single-line LiDAR depth completion. Our approach makes two key contributions. Firstly, we introduce the Relative-to-Metric (R2M) depth distillation framework, which estimates a pixel-wise metric depth map using an RGB image and its corresponding single-line depth map. This is achieved by distilling relative depth predictions from a monocular depth estimator trained on RGB images. Secondly, we propose the Line Depth Prior (LDP), a model-agnostic geometry regularization technique that promotes depth completion. Through extensive experiments, we demonstrate that our proposed method can: i) accurately reconstruct complete depth maps from single-line depth inputs without requiring additional depth supervision, except for the observed entries, and ii) facilitate downstream SLAM tasks when using single-line LiDAR.**

***Index Terms*—Depth completion, single-line lidar depth completion, self-supervised learning, robot navigation**

I. Introduction

DEPTH sensing with LiDARs plays an essential role in various real-world applications, such as 3D reconstruction [8], object tracking [40], smart farming [27], SLAM [39], robot navigation [34], and autonomous driving [21]. LiDARs provide reliable depth measurements with absolute scene scale and thus enable robots to interact and navigate in the real world. On the other hand, LiDARs do not generate high-resolution depth maps and need some enhancement and completion operations in practice. Recently, theory and technique for higher scanline LiDAR depth completion using deep learning have been well established and shown remarkable performance [10]. Unfortunately, higher scanline LiDARs are very costly and can be prohibitive for real-world deployment, e.g., in miniaturized robots. For example, the well-known Velodyne HDL-64E LiDAR is around 75,000 US dollars. Single-line LiDARs are lighter and significantly cheaper, making them a more practical option for many applications. However, single-line LiDAR depth completion (SLDC) remains underexplored and lacks practically effective methods.

Few existing works [23], [31] have formulated SLDC as a regression task and tackled it using classical data-driven supervised learning approaches with denser annotations. However, such approaches are impractical as they assume denser annotations are available and point-wise aligned. Moreover, most previous self-supervised learning approaches for depth completion [26], [42], [41] require a minimum of 64-line LiDARs, which limits their applicability. It has been demonstrated in [26] that the RMSE increases over 1000% for self-supervised depth completion when replacing 64-line depth input with a single-line depth map. In this paper, we tackle SLDC under in situ conditions where we only have a set of single-line depth maps and their corresponding RGB images collected by robots. Our goal is to develop an effective algorithm to generate a dense depth map from extremely sparse single-line depth inputs. The task is challenging since known depth measurements are too sparse to be used for supervision.

In this paper, we propose a self-supervised algorithm for single-line LiDAR depth completion, featuring two novel proposals. The first proposal is the Relative-to-Metric (R2M) depth distillation, which bridges the metric depth completion task with self-supervised monocular depth estimation (MDE). We observe that MDE methods can accurately estimate relative depth maps from only RGB images, providing strong supervision that depicts relative depths among pixels. Therefore, we propose to supervise depth completion by distilling the output of an off-the-shelf MDE model while recovering the scene scale using known depth measurements captured by a single-line LiDAR. The second proposal is a geometric regularization technique called Line Depth Prior (LDP). We observe that 3D points on a straight line of objects naturally form a linear depth relationship, providing true relative depth relations. To promote better depth completion, we impose LDP as a regularization constraint.

We conduct extensive evaluations on the widely used KITTI dataset of depth completion to demonstrate the effectiveness of our proposed method. As a result, our approach can effectively complete single-line depth maps without requiring supervision

Manuscript received 28 May 2023; revised 24 July 2023; accepted 18 September 2023. This letter was recommended for publication by Editor Sven Behnke upon evaluation of the reviewers' comments. This work was partly supported by the National Natural Science Foundation of China under Grants 62073274, Shenzhen Science and Technology Program under Grants JCYJ20220818103000001, and the Shenzhen Institute of Artificial Intelligence and Robotics for Society under Grants AC01202101103. (*Corresponding author: Tin Lun Lam.)*

J.Hu and Tin Lun Lam are with the Shenzhen Institute of Artificial Intelligence and Robotics for Society (AIRS), China. E-mail: hujunjie@cuhk.edu.cn, tllam@cuhk.edu.cn.

C.Fan is with the School of Artificial Intelligence, South China Normal University, China. E-mail: fanchenyou@scnu.edu.cn.

X.Guo is with the State Key Lab of CAD&CG, Zhejiang University, China. E-mail: gxytcrc@gmail.com.

L.Zhou and T.L.Lam are with the Chinese University of Hong Kong, Shenzhen, China. E-mail: liguangzhou@link.cuhk.edu.cn, tllam@cuhk.edu.cn.

Digital Object Identifier (DOI): see top of this page.

from denser annotations, even outperforming prior works of supervised learning. Furthermore, our experiments show that our method enables downstream SLAM tasks.

In summary, our contributions include the following:

- We introduce a novel self-supervised learning framework for single-line LiDAR depth completion, which enables single-line LiDARs to generate dense depth maps.
- We propose the Relative-to-Metric (R2M) depth distillation method that allows training a depth completion model by distilling from a monocular depth estimator.
- We introduce the Line Depth Prior (LDP), which enforces the correct depth orders of points on straight lines.
- Our self-supervised approach even outperforms prior supervised methods, demonstrating its effectiveness on downstream SLAM tasks.

The remainder of this paper is organized as follows. Section II provides an overview of related studies in the field. Section III presents our proposed self-supervised learning framework for single-line LiDAR depth completion. In Section IV, we conduct extensive numerical evaluations to demonstrate the effectiveness of our proposed method and its superior performance compared to prior supervised methods. Finally, we conclude our work in Section V.

II. Related Work

We provide an overview of previous studies on depth prediction, with a focus on monocular depth estimation and depth completion tasks. Our aim is to familiarize readers with the key technical components of existing depth sensing methods, explain why LiDARs are still necessary for real-world applications, and clarify the challenges of completing depth maps from single-line LiDARs.

A. Monocular Depth Estimation

Monocular depth estimation (MDE) has gained significant attention in recent years. MDE models can estimate depth from a single image, making them a cost-effective alternative to depth sensors. These models can be learned through data-driven supervised learning, penalizing pixel-wise differences between predicted depth maps and ground truth depth maps, as seen in [5], [19], [14], [13]. Alternatively, self-supervised learning using monocular or stereo images has also been employed, as demonstrated in [18] and [6], respectively.

Despite these advances, existing MDE methods have limitations. First, they tend to yield lower accuracy compared to depth sensors. For example, the advanced MDE method [32] yielded over 100% increase in root mean square error compared to the LiDAR completion method [22] on the KITTI benchmark [36]. Second, MDE methods suffer from poor generalizability in predicting absolute depth scales across different domains, making them unsuitable for robot applications, particularly in dynamic and complex unknown environments.

Therefore, obtaining scene depth with depth sensors, such as LiDARs, remains a practical and effective solution for robot perception. Compared to MDE methods, completing depth maps from sparse inputs is easier and more reliable in recovering scene depth scales.

B. Supervised LiDAR Depth Completion

Early methods treat depth completion as a regression task and learn by penalizing pixel-wise depth differences, with around 30% depth measurements required for supervision. With sufficient supervision, prior works proposed various numerical network architectures such as dual-encoder networks [38], [17], double encoder-decoder networks [35], [10], and residual depth networks [48] to improve representation ability. Co-attention feature interaction [47] and skip connections [20] are also commonly utilized for promoting multi-modality data fusion. Some works also integrate additional 3D cues such as surface normal [30] and 3D convolutions [1] into the networks to guide completion. To generate high-resolution depth maps, previous methods proposed a two-stage coarse-to-fine learning strategy [2] and affinity refinement [15].

The aforementioned studies primarily focus on 64-line LiDAR completion tasks, assuming approximately 5% known depth measurements as inputs, along with the assumptions of 30% denser annotations and perfect point-wise alignment between ground truth depth maps and sparse input depth maps. However, in reality, obtaining denser depth maps and ensuring perfect point-wise alignment are challenging, rendering previous supervised learning approaches incompatible with lower scanline sensors.

C. Self-supervised LiDAR Depth Completion

Self-supervised learning methods [25], [42], [41], [43] have been explored as an alternative to using denser depth annotations for depth completion, differing from previous supervised approaches for depth completion in two significant ways. Firstly, they require an additional network to estimate the poses between two consecutive frames. Secondly, they use a photometric loss based on the epipolar geometry constraint for self-supervised learning. In addition, depth consistency based on measurements from sparse inputs is used to estimate a metric depth map.

It should be noted that the prior methods are only suitable for 64-line LiDARs and their performance degrades exponentially for lower scanlines. We refer the reader to experiments varying input sparsity to 1, 2, 4, 8, 16, 32, and 64 scanlines in [25] for further details.

D. Single-line LiDAR Depth Completion

Only a few prior studies have addressed single-line LiDAR depth completion. Specifically, Lu et al. [23] proposed an RGB-guided two-branch network that estimates a global and a local depth map and uses their combination for the final prediction. Meanwhile, Ryu et al. [31] proposed a depth consistency loss that ensures a completed depth map from a low-scanline LiDAR matches that of a high-scanline LiDAR. However, since both these methods rely on denser annotations for supervision, they may not be practical for many real-world robot applications where data collection is challenging. In [24], a self-supervised approach has been introduced that first estimates a coarse depth map using stereo images and then concatenates this with the single-line depth map into a multi-layer perception for scale refinement. We regard this work as

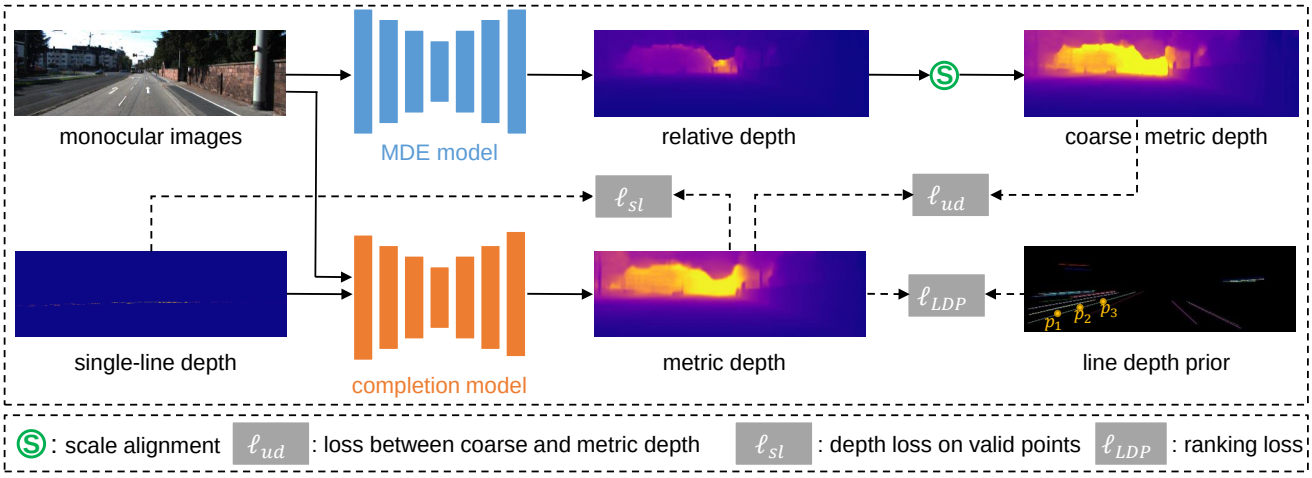

Fig. 1. The diagram of the proposed approach for single-line LiDAR depth completion. Our method uses a completion model, taking a single-line depth map and an RGB image as inputs and predicting a dense depth map. To enable the learning of the completion model, we propose i) distilling from a trained MDE model and ii) a ranking loss utilizing line depth prior. The MDE trained using monocular images predicts a relative depth map. We use valid depth points from the single-line depth map to align depth scales yielding a coarse metric depth map.

the first attempt to address SLDC in a self-supervised manner. Nonetheless, this method requires a stereo camera setup to obtain the absolute scene scale, and it only achieves a slight improvement compared to the baseline MDE models.

E. Knowledge Distillation

Knowledge distillation, initially introduced for transferring knowledge from a fixed teacher model to a student model in image recognition [9]. It has also been employed in depth estimation [29], [11], [12] and depth completion [31]. In these existing works, knowledge distillation penalizes pixel-wise depth differences between the teacher and student models, where both models are of the same type. In contrast, our approach takes a different perspective by proposing the distillation of depth predictions from a monocular depth estimation (MDE) model to a depth completion model.

III. TECHNICAL APPROACH

In this section, we present our self-supervised approach for single-line LiDAR depth completion. As illustrated in Fig 1, our approach utilizes predictions from an off-the-shelf MDE model, which was also trained in a self-supervised fashion. Additionally, we propose the line depth prior to ensure correct depth orders on straight lines.

A. Problem Formulation

Given N monocular images $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ and the corresponding single-line sparse depth maps $\mathbf{Y}' = \{\mathbf{y}'_i\}_{i=1}^N$ where $\mathbf{x}_i \in \mathbf{X}$ and $\mathbf{y}'_i \in \mathbf{Y}'$ denotes an RGB image and a sparse depth map, the depth completion task can be formulated as:

$$\hat{\mathbf{y}}_i = \mathcal{F}_{DC}\left(\mathbf{y}'_i, \mathbf{x}_i; \theta\right). \tag{1}$$

Here, $\mathcal{F}_{DC}$ represents the depth completion model with parameters θ, and $\hat{\mathbf{y}}_i$ is a dense depth map completed from $\mathbf{x}_i$ and $\mathbf{y}'_i$.

To train $\mathcal{F}_{DC}$, supervised learning methods require a set of denser annotations $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^N$ with depth points aligned with $\mathbf{Y}'$ to penalize the following depth loss:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, \mathcal{L}\left(\hat{\mathbf{Y}}, \mathbf{Y}; \theta\right), \tag{2}$$

where $\mathcal{L}$ is a loss function that penalizes the difference between valid depth points.

However, in most practical applications, $\mathbf{Y}$ is not available. Moreover, for our single-line LiDAR completion task, the known depth points are too sparse, e.g., the observed depth points are only around 0.1%. Therefore, we seek alternative depth maps $\widetilde{\mathbf{Y}} = \{\widetilde{\mathbf{y}}_i\}_{i=1}^N$ to replace $\mathbf{Y}$ in Eq. (2), as follows:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, \mathcal{L}\left(\hat{\mathbf{Y}}, \widetilde{\mathbf{Y}}; \theta\right). \tag{3}$$

B. Relative-to-Metric depth distillation

According to Eq. (3), obtaining reasonable $\widetilde{\mathbf{Y}}$ is a necessary condition for training the completion model. While completion of single-line LiDAR depth maps is unattainable, accurate estimation of relative depth maps from monocular images can still be achieved through:

$$\mathbf{y}^*_i = \mathcal{F}_{MDE}\left(\mathbf{x}_i; \theta'\right), \tag{4}$$

where $\mathcal{F}_{MDE}$ is the trained monocular depth estimation model with fixed parameters θ'.

Then, we can align $\mathbf{y}^*_i$ to the true scene depth scale by utilizing valid depth points from the single-line depth map $\mathbf{y}'$:

$$\widetilde{\mathbf{y}}_i = \mathbf{y}^*_i / median(\mathbf{M_i} \otimes \mathbf{y}^*_i) * median(\mathbf{M_i} \otimes \mathbf{y}'_i), \tag{5}$$

where $\mathbf{M_i}$ is a binary mask with the same length and width of $\mathbf{y}'_i$ and used to select valid depth points from $\mathbf{y}'_i$. $median$ is a common operation for depth alignment that selects the median value from known depth points as a scale. Finally, we can train the completion model using Eq.(3).

We consider the aligned dense depth map $\widetilde{\mathbf{y}}_i$ a coarse metric depth map and utilize it as a pseudo-label for supervision

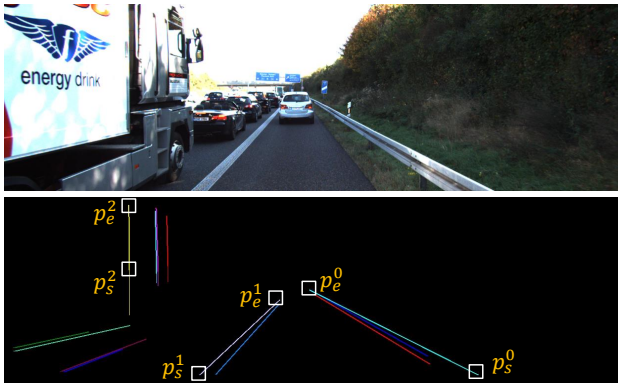

Fig. 2. An example of Line Depth Prior (LDP) where we mark the start and end points of three different lines with white boxes. p_s^l and p_e^l denote start and end points on l-th line.

to obtain a better prediction, i.e., $\widetilde{\mathbf{y}}_i$. It is intuitive that the accuracy of $\widetilde{\mathbf{y}}_i$ is directly proportional to the correctness of $\hat{\mathbf{y}}_i$. However, a non-negligible issue arises as $\widetilde{\mathbf{y}}_i$ inevitably results in large errors, particularly in distant areas due to a larger depth scale, and in object boundaries due to the depth mixing problem [16]. To address this, we utilize an uncertainty-aware depth loss [13], [48], [3] to exclude the effects of inaccurate depth points, as follows:

$$\ell_{ud} = \frac{1}{N}\sum_{i=1}^{N}(\exp(-\mathbf{s}_i)(\hat{\mathbf{y}}_i - \widetilde{\mathbf{y}}_i)^2 + \mathbf{s}_i). \tag{6}$$

Here, $\mathbf{s}_i$ is the uncertainty map for $\hat{\mathbf{y}}_i$, which indicates the depth completion model's confidence for each pixel. In Eq. (6), we note that $(\hat{\mathbf{y}}_i - \widetilde{\mathbf{y}}_i)^2$ and $\exp(-\mathbf{s}_i)$ are opposing terms. During optimization, when the difference between $\hat{\mathbf{y}}_i$ and $\widetilde{\mathbf{y}}_i$ is large, $\exp(-\mathbf{s}_i)$ tends to decrease to balance the joint loss, leading to larger values of s_i. Conversely, penalizing the additional loss term $\mathbf{s}_i$ results in smaller uncertainties when depth differences are minor.

Besides, we also penalize the absolute depth error for those valid depth points on single-line depth maps by:

$$\ell_{sl} = \frac{1}{N}\sum_{i=1}^{N}\|M_i \otimes (\hat{\mathbf{y}}_i - \mathbf{y}_i')\|_1. \tag{7}$$

C. Ranking with Line Depth Prior

As we discussed before, the pseudo-labels $\widetilde{\mathbf{y}}_i$ obtained from $\mathcal{F}_{MDE}$ after scale alignment provide an approximation to true depth maps. However, if $\widetilde{\mathbf{y}}_i$ miscalculated relative depths among pixels, $\mathcal{F}_{DC}$ is likely to generate wrong depth relations.

We observed that points on straight lines naturally form a prior regarding relative depths. Let $\mathbf{p}_s^l$ and $\mathbf{p}_e^l$ denote the start and end points of the l-th line segment, with $\mathbf{p}_s^l < \mathbf{p}_e^l$ in depth. An example is shown in Fig. 2, where $\mathbf{p}_s^0$ and $\mathbf{p}_e^0$, $\mathbf{p}_s^1$ and $\mathbf{p}_e^1$, and $\mathbf{p}_s^2$ and $\mathbf{p}_e^2$ are the start and end points of the 0-th, 1-th, and 2-th lines, respectively. If we have k points $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_k$ from $\mathbf{p}_s^l$ to $\mathbf{p}_e^l$, then we must have $\mathbf{p}_1 < \mathbf{p}_2 < \ldots < \mathbf{p}_k$ in depth.

To enforce the consistency of relative depths among pixels on lines, we propose a ranking loss based on LDP. We first detect straight lines using the Line Segment Detector [37], and then locate the start and end points in ascending depth order. However, there are some challenging cases where the start and end points have similar depth, such as $\mathbf{p}_s^2$ and $\mathbf{p}_e^2$, which may lead to errors in their relative depth relation. Furthermore, as we do not know the true depths, we can only identify this relation from the coarse depth map $\widetilde{\mathbf{y}}_i$, which may not provide reliable estimations for some lines.

To robustly distinguish the start and end points, we utilize 20 points around $\mathbf{p}_s^l$ or $\mathbf{p}_e^l$ and calculate their mean depth for comparison. Specifically, let $\{\mathbf{p}_j\}_{j=1}^{20}$ and $\{\mathbf{p}_j'\}_{j=1}^{20}$ be sets of 20 points around the two endpoints of a line segment (ordered from bottom to top in the image), and let their depth values in $\widetilde{\mathbf{y}}_i$ be denoted by $\{\mathbf{d}_j\}_{j=1}^{20}$ and $\{\mathbf{d}_j'\}_{j=1}^{20}$. Then, we can identify the start and end points, denoted by $\mathbf{p}_s$ and $\mathbf{p}_e$ (here we simplify $\mathbf{p}_s^l$ and $\mathbf{p}_e^l$ to $\mathbf{p}_s$ and $\mathbf{p}_e$ for clarity), by

$$\begin{cases} \mathbf{p}_s = \mathbf{p}_0, \mathbf{p}_e = \mathbf{p}_{20}', & \text{if } \frac{1}{20}\sum_{j=1}^{20}(\mathbf{d}_j) < \frac{1}{20}\sum_{j=1}^{20}(\mathbf{d}_j'), \\ \mathbf{p}_s = \mathbf{p}_{20}', \mathbf{p}_e = \mathbf{p}_0, & \text{otherwise.} \end{cases} \tag{8}$$

Having determined $\mathbf{p}_s^l$ and $\mathbf{p}_e^l$, we construct a set of ground truth rankings that encode the true relative depths among the points on the identified straight line. These rankings are then used to rank the points on the metric depth maps predicted by our model. Let r_j be the rank of $\mathbf{p}_j$ relative to $\mathbf{p}_{j+1}$ in depth, where $\mathbf{r}_j = -1$ if $\mathbf{p}_s = \mathbf{p}_0$ and $\mathbf{r}_j = 1$ if $\mathbf{p}_s = \mathbf{p}_{20}'$. To enforce the consistency of relative depths among the points on the identified lines, we propose the following ranking loss based on LDP:

$$\ell_{LDP}(\mathbf{p}_j, \mathbf{p}_{j+1}) = \log(1 + \exp(\mathbf{r}_j \times (\mathbf{d}_{j+1} - \mathbf{d}_j))). \tag{9}$$

It is worth noting that only when a straight line is perfectly captured parallel to an RGB camera in the world coordinate, $\mathbf{p}_s$ and $\mathbf{p}_e$ have the same depth. In reality, almost all lines can provide priors on relative depth relations.

D. Final Learning Objective

In summary, we optimize the SLDC model using joint loss terms, including an uncertainty-aware depth loss (Eq. (6)) that measures the dissimilarity between our completed depth map and the MDE model's output, a depth consistency loss (Eq. (7)) that enforces agreement between predicted and observed depths on valid points in single-line depth maps, a ranking loss (Eq. (9)) that encourages the correct depth relations of points on straight lines. Our final loss function is given as follows:

$$\mathcal{L} = \ell_{ud} + \alpha\ell_{sl} + \beta\ell_{LDP}. \tag{10}$$

where α and β are weighting coefficients and are set to 1 and 0.01, respectively, throughout the experiments.

IV. Experimental Results

A. Experimental Setup

a) Selection of MDE and SLDC models: Our goal is to propose a novel approach for SLDC, rather than focusing on improving the network architecture for depth completion. To

TABLE I
COMPARISONS OF DIFFERENT METHODS OF SINGLE-LINE LIDAR COMPLETION ON THE KITTI VALIDATION DATASET. THE METHOD OF LU ET AL.[24] WAS ORIGINALLY EVALUATED ON A DIFFERENT TEST SPLIT; WE MARK IT WITH *.

Method	Self-supervised	Scale-aware	RMSE ↓	REL ↓	$\log 10$ ↓	$\delta < 1.25$ ↑	$\delta < 1.25^2$ ↑	$\delta < 1.25^3$ ↑
Balanced DC (MDE) [45]	✗	✓	3.951	-	-	-	-	-
Balanced DC [45]	✗	✓	3.921	-	-	-	-	-
Ryu et al. (MDE) [31]	✗	✓	3.625	-	-	-	-	-
Ryu et al. [31]	✗	✓	3.616	-	-	-	-	-
Lu et al.* [24]	✓	✓	4.582	0.106	-	0.871	0.951	0.982
MDE model: MonoDepth2 [7]	✓	✗	4.198	0.134	0.054	0.854	0.974	0.993
MDE model: CADepth [44]	✓	✗	3.914	0.120	0.048	0.877	0.977	0.994
Ours (MonoDepth2 → S2D)	✓	✓	3.700	0.100	0.041	0.920	0.983	0.995
Ours (MonoDepth2 → DCVAN)	✓	✓	3.723	0.098	0.040	0.920	0.983	0.995
Ours (CADepth → S2D)	✓	✓	**3.404**	**0.088**	**0.036**	**0.933**	**0.987**	**0.996**

evaluate the effectiveness of our method, we choose different models from prior works for both MDE and SLDC. Specifically, we use MonoDepth2 [7] and CADepth [44] for MDE and S2D [25] and DCVAN [37] for depth completion. We conduct experiments with three different combinations: MonoDepth2 → S2D, MonoDepth2 → DCVAN, and CADepth → S2D.

b) Implementation Details: For the S2D, we provide a Pytorch implementation. As for the DCVAN, we utilize its original implementation. Additionally, we integrate an uncertainty estimation module consisting of two 5×5 convolutional layers into both networks for uncertainty estimation. The resulting S2D and DCVAN models have 30.7M and 2.5M parameters, respectively. During training, we train S2D for 20 epochs using the Adam optimizer with an initial learning rate of 0.0001, which is reduced by 50% every five epochs. For DCVAN, we adopt the original learning strategy, which involves training for 50 epochs with a batch size of 8. Regarding evaluation, we employ the six most widely used measures: RMSE, REL, $\log_{10}$, $\delta < 1.25$, $\delta < 1.25^2$, and $\delta < 1.25^3$. RMSE is a scale-aware measure, while the latter five are scale-invariant, capturing different aspects of depth estimation accuracy.

B. Dataset

We evaluate our method on the most popular used KITTI [36] benchmark dataset, which consists of outdoor scenes captured by a car-mounted camera and a 64-line LiDAR sensor. The depth range is from 0 to 80 meters. This dataset has been extensively used in previous studies on depth completion and MDE. To synthesize single-line LiDAR depth maps, we first convert 64-line depth maps to point clouds and extract single scanline depth maps from the corresponding 3D point clouds. We use the official KITTI dataset with the official split of scenes for training and validation, which consisted of 138 and 18 driving sequences, respectively. The resolution of most images is about 1216×352, and following the setting in [7], [44], we resize all images to 1024×320 resolution for both training and testing. It is important to note that our method is purely self-supervised, and we only used the denser annotations provided by the dataset (about 30% valid depth points) for quantitative comparisons. The valid depth points from single-line LiDAR depth maps are only around 0.1%.

C. Quantitative Comparisons

We compare our method against baselines of single-line LiDAR completion, including Ryu et al. [31], Balanced DC [45], Lu et al.[24]. The methods of Ryu et al. [31] and Balanced DC [45] are supervised learning methods evaluated using the same validation methodology as our method. They only reported the RMSE results in their papers and have not released the codes; we could not calculate the results for other metrics. For the unsupervised method of Lu et al.[24], the results are evaluated on a different split. Since the code is not publicly available, we take results from their paper for reference and mark them with *. We also quantify the performance of the unsupervised MDE models we employed in our method, namely MonoDepth2 and CADepth. For evaluation, we align their predictions with scale using corresponding single-line depth maps (coarse metric depth in Fig. 1).

The quantitative results are presented in Table I, where Balanced DC (MDE) and Balanced DC denote methods with and without single-line depth map inputs, respectively. Similarly, Ryu et al.(MDE) and Ryu et al. are models with and without single-line depth map inputs. We find that the single-line depth map inputs contribute little to their performance boost, which does not demonstrate a clear advantage from MDE. Specifically, the RMSE is reduced by 0.03 meters and 0.009 meters for Balanced DC and Ryu et al., respectively. Note that these two methods take a supervised learning approach using denser depth annotations. In contrast, the method of Lu et al. shows the worst results than other supervised approaches, although it enables unsupervised SLDC.

Our self-supervised method achieves the best performance on all metrics, as shown in Table I. Specifically, for our Monodepth2 → S2D, Monodepth2 → DCVAN, and CADepth → S2D, we reduce RMSE by 0.498, 0.475, and 0.51 meters, respectively, demonstrating clear superiority against baseline methods. Furthermore, the results also show that we can boost the performance of our method by using a better MDE model, as seen in the improved performance from MonoDepth2 → S2D to CADepth → S2D, or by using a more discriminative

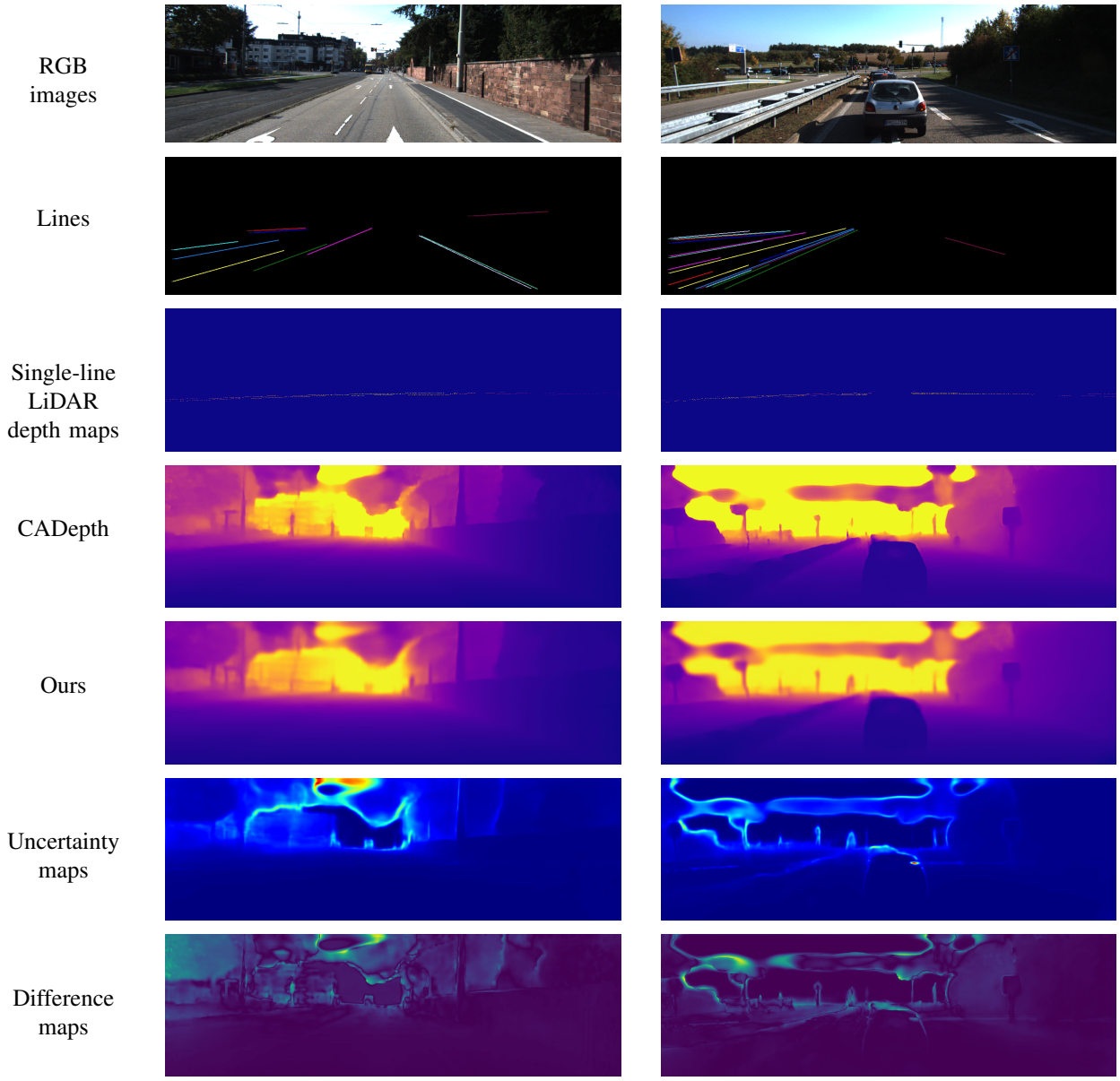

Fig. 3. The qualitative comparison between CADepth and our method on the KITTI dataset where RGB images and single-line LiDAR depth maps are inputs. We also show lines detected from RGB images used only during training. Moreover, we also show uncertainty maps estimated by our method. The difference maps are the results of absolute depth error between CADepth and our method.

depth completion model, as demonstrated from MonoDepth2 $\rightarrow$ DCVAN to MonoDepth2 $\rightarrow$ S2D.

To visualize the qualitative results of our method (CADepth $\rightarrow$ S2D), Fig. 3 shows depth maps estimated by CADepth and our method, as well as the difference maps. Notably, line maps are only used during training. We also provide uncertainty maps yielded by our method for better analysis of these results. As seen, CADepth and our method produce similar depth maps. It is because we distill predicted depth maps from CADepth. Closer observations reveal that our method captures higher attention on wrong depths estimated by CADepth in "sky" and object boundaries. This can also be validated by the difference maps, in which the highlighted regions are consistent with the uncertainty maps.

D. Ablation Studies on Loss Functions

The effect of each loss term was evaluated through experiments, using ℓ_{ud} as the baseline. Results of different loss functions, including $\ell_{ud} + \alpha\ell_{sl}$ and $\ell_{ud} + \alpha\ell_{sl} + \beta\ell_{LDP}$, are presented in Table II. Intuitively, penalizing depth differences on those measured points contribute to performance improvement. Additionally, ℓ_{LDP} contributes to a performance boost for all three different settings. Compared to the results of the used MDE models, the pixel-wise dense supervision provided by ℓ_{ud} plays a crucial role in achieving major improvement by effectively capturing estimation uncertainty. Conversely, ℓ_{sl} and ℓ_{LDP} contribute relatively slight improvements due to their supervision being limited to specific points or areas. Specifically, ℓ_{sl} penalizes depth differences based on the sparsely measured points (0.1%) acquired from the single-

TABLE II
ABLATION STUDIES ON EACH LOSS TERM. WE SHOW RMSE RESULTS FOR COMPARISON.

	MonoDepth2 → S2D	MonoDepth2 → DCVAN	CADepth → S2D
ℓ_{ud}	3.809	3.890	3.538
$\ell_{ud} + \alpha\ell_{sl}$	3.739	3.886	3.461
$\ell_{ud} + \alpha\ell_{sl} + \beta\ell_{LDP}$	**3.700**	**3.723**	**3.404**

TABLE III
RMSE OF DEPTH SCALE ALIGNMENT USING SINGLE-LINE LIDAR DEPTH AND GROUND TRUTH DEPTH, RESPECTIVELY.

	Single-line LiDAR depth		Ground truth LiDAR depth	
	MonoDepth2	CADepth	MonoDepth2	CADepth
Median	4.198	3.914	3.995	3.689
Least-squares	4.800	4.617	3.750	3.471

line LiDAR, while ℓ_{LDP} applies the ranking loss solely on straight lines. As acknowledged, these limited supervisions lead to moderate improvements, which we clearly demonstrate through the ablation studies. Nevertheless, ℓ_{sl} and ℓ_{LDP} can serve as auxiliary regularization to promote depth completion. We believe such regularization that can be applied in a self-supervised learning way is important to improve single-line LiDAR depth completion in practice.

E. Robustness of Scale Alignment

In our method, we use the median value of known depth points from single-line LiDARs for aligning relative depth maps estimated by a MDE model to metric depth maps. To measure the robustness of scale alignment for the single-line LiDAR depth completion task, we compare the median operation with another popular operation for scale alignment, namely, the least-squares fit. To identify the depth scale, we utilize single-line LiDAR depth data with 0.1% known points or ground truth depth data with 30% known points, respectively. The results are shown in Table III. We observe that 1) the accuracy of depth scale estimation improves with an increase in the number of known points, 2) the least-squares method outperforms the median operation when ground truth depths are available, offering more accurate depth scale estimation, and 3) the median operation, on the other hand, exhibits greater robustness when dealing with single-line LiDAR data. Given the above observations, we can say that the median operation is better suited for our single-line LiDAR depth completion task.

F. Application for SLAM

We apply our method to SLAM on the KITTI odometry dataset. Specifically, we selet sequences 08, 09, and 10 for comparisons following the common experimental setup. As the dataset only provides 64-line LiDAR scans, we extract single-line LiDAR depth maps from these 64-line scans. We compare our method (CADepth → S2D) with unsupervised monocular depth methods, including Zhan et al. [46], VISO2-M [33], SGANVO [4], and CADepth [44]. We use the classical ORB-SLAM2 [28] framework with estimated depth maps as input for our method and CADepth, while for other baselines, we refer to their respective papers. The proposed method provides absolute depth maps, unlike the baseline methods that align the depth scale with ground truth depth points or stereo images.

TABLE IV
QUANTITATIVE COMPARISON OF TRANSLATION AND ROTATION ERRORS BETWEEN SEVERAL UNSUPERVISED METHODS AND OUR METHOD.

Method	Sequence 08		Sequence 09		Sequence 10	
	$t_{err}(\%)$	$r_{err}(^\circ)$	$t_{err}(\%)$	$r_{err}(^\circ)$	$t_{err}(\%)$	$r_{err}(^\circ)$
Zhan et al. [46]	-	-	11.92	3.6	12.62	3.43
VISO2-M [33]	13.94	2.03	4.04	1.43	25.20	3.88
SGANVO [4]	-	-	4.95	2.37	5.89	3.56
CADepth [44]	3.964	0.584	4.132	0.458	5.496	0.818
Ours	**2.896**	**0.554**	**2.524**	**0.432**	**3.462**	**0.660**

We evaluate the performance using translation error ($t_{err}(\%)$) and rotation error ($r_{err}(^\circ)$) for every 100 meters, as shown in Table IV. Our method outperforms other approaches by a significant margin, surpassing CADepth in both metrics for all three sequences. Although our method requires additional single-line LiDAR depth maps, it can still be applied in unsupervised learning, eliminating the need for denser depth annotations and providing a practical application for robot navigation in absolute depth scale using single-line LiDARs.

V. CONCLUSION

In this paper, we presented a novel self-supervised learning framework for single-line LiDAR depth completion, addressing the challenge of generating dense depth maps from highly sparse single-line depth inputs. Our method comprises two key contributions: the Relative-to-Metric (R2M) depth distillation and the Line Depth Prior (LDP). These proposals enable the training of a depth completion model by distilling from a monocular depth estimator and enforcing correct depth orders of points on straight lines. Extensive evaluations on the KITTI dataset validate the effectiveness of our approach, surpassing the performance of prior supervised learning methods. Additionally, we demonstrate the positive impact of our method on downstream SLAM tasks. The results highlight the practicality and efficacy of self-supervised single-line LiDAR depth completion for robot navigation applications.

REFERENCES

[1] Y. Chen, B. Yang, M. Liang, and R. Urtasun, "Learning joint 2d-3d representations for depth completion," in *IEEE International Conference on Computer Vision*, 2019, pp. 10 023–10 032.

[2] Z. Chen, V. Badrinarayanan, G. Drozdov, and A. Rabinovich, "Estimating depth from rgb and sparse sensing," in *European Conference on Computer Vision*, 2018, pp. 167–182.

[3] A. Eldesokey, M. Felsberg, K. Holmquist, and M. Persson, "Uncertainty-aware cnns for depth completion: Uncertainty from beginning to end," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 014–12 023.

[4] T. Feng and D. Gu, "Sganvo: Unsupervised deep visual odometry and depth estimation with stacked generative adversarial networks," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4431–4437, 2019.

[5] H. Fu, M. Gong, C. Wang, K. Batmanghelich, and D. Tao, "Deep ordinal regression network for monocular depth estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2002–2011.

[6] C. Godard, O. M. Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6602–6611.

[7] C. Godard, O. Mac Aodha, M. Firman, and G. J. Brostow, "Digging into self-supervised monocular depth estimation," in *Proceedings of the IEEE international conference on computer vision*, 2019, pp. 3828–3838.

[8] G. He, Q. Zhang, and Y. Zhuang, "Online semantic-assisted topological map building with lidar in large-scale outdoor environments: Toward robust place recognition," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[9] G. E. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *ArXiv*, vol. abs/1503.02531, 2015.

[10] J. Hu, C. Bao, M. Ozay, C. Fan, Q. Gao, H. Liu, and T. L. Lam, "Deep depth completion from extremely sparse data: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 7, pp. 8244–8264, 2023.

[11] J. Hu, C. Fan, H. Jiang, X. Guo, Y. Gao, X. Lu, and T. L. Lam, "Boosting light-weight depth estimation via knowledge distillation," in *International Conference on Knowledge Science, Engineering and Management*, 2023, pp. 27–39.

[12] J. Hu, C. Fan, M. Ozay, H. Jiang, and T. L. Lam, "Data-free dense depth distillation," *arXiv preprint arXiv:2208.12464*, 2022.

[13] J. Hu, C. Fan, L. Zhou, Q. Gao, H. Liu, and T. L. Lam, "Lifelong-monodepth: Lifelong learning for multi-domain monocular metric depth estimation," *arXiv preprint arXiv:2303.05050*, 2023.

[14] J. Hu, M. Ozay, Y. Zhang, and T. Okatani, "Revisiting single image depth estimation: Toward higher resolution maps with accurate object boundaries," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, 2019, pp. 1043–1051.

[15] M. Hu, S. Wang, B. Li, S. Ning, L. Fan, and X. Gong, "Penet: Towards precise and efficient image guided depth completion," in *International Conference on Robotics and Automation*, 2021, pp. 13 656–13 662.

[16] S. Imran, X. Liu, and D. Morris, "Depth completion with twin surface extrapolation at occlusion boundaries," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2583–2592.

[17] M. Jaritz, R. D. Charette, E. Wirbel, X. Perrotton, and F. Nashashibi, "Sparse and dense data with cnns: Depth completion and semantic segmentation," in *International Conference on 3D Vision*, 2018, pp. 52–60.

[18] H. Jiang, L. Ding, J. Hu, and R. Huang, "Plnet: Plane and line priors for unsupervised indoor depth estimation," in *International Conference on 3D Vision*, 2021, pp. 741–750.

[19] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab, "Deeper depth prediction with fully convolutional residual networks," in *International Conference on 3D Vision*, 2016, pp. 239–248.

[20] A. Li, Z. Yuan, Y. Ling, W. Chi, C. Zhang *et al.*, "A multi-scale guided cascade hourglass network for depth completion," in *IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 32–40.

[21] X. Li, Y. Zhou, and B. Hua, "Study of a multi-beam lidar perception assessment model for real-time autonomous driving," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–15, 2021.

[22] Y. Q. Lin, T. Cheng, Q. Zhong, W. Zhou, and H. Yang, "Dynamic spatial propagation network for depth completion," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, 2022, pp. 1638–1646.

[23] H. Lu, S. Xu, and S. Cao, "Sgtbn: Generating dense depth maps from single-line lidar," *IEEE Sensors Journal*, vol. 21, no. 17, pp. 19 091–19 100, 2021.

[24] Y. Lu, Y. Wang, D. Parikh, Y. Xin, and G. Lu, "Extending single beam lidar to full resolution by fusing with single image depth estimation," in *International Conference on Pattern Recognition*, 2021, pp. 6343–6350.

[25] F. Ma, L. Carlone, U. Ayaz, and S. Karaman, "Sparse depth sensing for resource-constrained robots," *The International Journal of Robotics Research*, vol. 38, no. 8, pp. 935–980, 2019.

[26] F. Ma, G. V. Cavalheiro, and S. Karaman, "Self-supervised sparse-to-dense: Self-supervised depth completion from lidar and monocular camera," in *International Conference on Robotics and Automation*, 2019, pp. 3288–3295.

[27] F. B. Malavazi, R. Guyonneau, J.-B. Fasquel, S. Lagrange, and F. Mercier, "Lidar-only based navigation algorithm for an autonomous agricultural robot," *Computers and electronics in agriculture*, vol. 154, pp. 71–79, 2018.

[28] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[29] A. Pilzer, S. Lathuilière, N. Sebe, and E. Ricci, "Refine and distill: Exploiting cycle-inconsistency and knowledge distillation for unsupervised monocular depth estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9760–9769.

[30] J. Qiu, Z. Cui, Y. Zhang, X. Zhang, S. Liu, B. Zeng, and M. Pollefeys, "Deeplidar: Deep surface normal guided depth prediction for outdoor scene from sparse lidar data and single color image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3308–3317.

[31] K. Ryu, K.-i. Lee, J. Cho, and K.-J. Yoon, "Scanline resolution-invariant depth completion using a single image and sparse lidar point cloud," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 6961–6968, 2021.

[32] S. Shao, Z. Pei, W. Chen, R. Li, Z. Liu, and Z. Li, "Urcdc-depth: Uncertainty rectified cross-distillation with cutflip for monocular depth estimatione," *IEEE Transactions on Multimedia*, pp. 1–14, 2023.

[33] S. Song, M. Chandraker, and C. C. Guest, "High accuracy monocular sfm and scale correction for autonomous driving," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 4, pp. 730–743, 2015.

[34] H. Tang, X. Niu, T. Zhang, L. Wang, and J. Liu, "LE-VINS: A robust solid-state-lidar-enhanced visual-inertial navigation system for low-speed robots," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–13, 2023.

[35] J. Tang, F.-P. Tian, W. Feng, J. Li, and P. Tan, "Learning guided convolutional network for depth completion," *IEEE Transaction on Image Processing*, vol. 30, pp. 1116–1129, 2021.

[36] J. Uhrig, N. Schneider, L. Schneider, U. Franke, T. Brox, and A. Geiger, "Sparsity invariant cnns," in *International Conference on 3D Vision*, 2017, pp. 11–20.

[37] R. G. Von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "Lsd: A line segment detector," *Image Processing On Line*, vol. 2, pp. 35–55, 2012.

[38] A. Eldesokey, M. Felsberg, and F. S. Khan, "Confidence propagation through cnns for guided sparse depth regression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2423–2436, 2019.

[39] J. Wang, M. Xu, G. Zhao, and Z. Chen, "Feature- and distribution-based lidar SLAM with generalized feature representation and heuristic nonlinear optimization," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–15, 2023.

[40] S. Wang, R. Pi, J. Li, X. Guo, Y. Lu, T. Li, and Y. Tian, "Object tracking based on the fusion of roadside lidar and camera data," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–14, 2022.

[41] A. Wong, S. Cicek, and S. Soatto, "Learning topology from synthetic data for unsupervised depth completion," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1495–1502, 2021.

[42] A. Wong, X. Fei, S. Tsuei, and S. Soatto, "Unsupervised depth completion from visual inertial odometry," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1899–1906, 2020.

[43] A. Wong and S. Soatto, "Unsupervised depth completion with calibrated backprojection layers," in *Proceedings of the IEEE International Conference on Computer Vision*, 2021, pp. 12 747–12 756.

[44] J. Yan, H. Zhao, P. Bu, and Y. Jin, "Channel-wise attention-based network for self-supervised monocular depth estimation," in *International Conference on 3D vision*, 2021, pp. 464–473.

[45] S. Yoon and A. Kim, "Balanced depth completion between dense depth inference and sparse range measurements via kiss-gp," in *International Conference on Intelligent Robots and Systems*. IEEE, 2020, pp. 10 468–10 475.

[46] H. Zhan, R. Garg, C. S. Weerasekera, K. Li, H. Agarwal, and I. Reid, "Unsupervised learning of monocular depth estimation and visual odometry with deep feature reconstruction," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 340–349.

[47] S. Zhao, M. Gong, H. Fu, and D. Tao, "Adaptive context-aware multi-modal network for depth completion," *IEEE Transaction on Imgae Provessing*, vol. 30, pp. 5264–5276, 2021.

[48] Y. Zhu, W. Dong, L. Li, J. Wu, X. Li, and G. Shi, "Robust depth completion with uncertainty-driven loss functions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.