IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

DRIM: Depth Restoration With Interference Mitigation in Multiple LiDAR Depth Cameras

Seunghui Shin, Jaeyun Jang, Sundong Park, and Hyoseok Hwang, *Associate Member, IEEE*

***Abstract*—LiDAR depth cameras are widely used for accurate depth measurement in various applications. However, when multiple cameras operate simultaneously, mutual interference causes artifacts in the captured depth data, which existing image restoration methods struggle to handle. In this letter, we propose DRIM, a novel approach for real-time depth restoration under multi-device interference. Our method begins by distinguishing interference-induced artifacts, then predicts and leverages these artifacts to guide the restoration process. Since there is no existing dataset for learning interference in multiple LiDAR depth cameras, we create and provide the first depth interference dataset. Our experiments demonstrate superior depth restoration performance compared to other image restoration methods, achieving real-time processing speeds (≈33 FPS) that are significantly faster than existing approaches while showing the capability to restore depth in challenging scenarios. These results demonstrate that our proposed method effectively restores interfered depth in multiple LiDAR depth cameras with practical real-time performance.**

***Index Terms*—RGB-D perception, deep learning for visual perception, data sets for robotic vision.**

I. Introduction

DEPTH cameras are optical devices that capture distance information of objects relative to the sensor and record this data as 2D depth images in real-time [1]. Recently, there has been a growing demand for higher accuracy depth data [2], leading to the use of LiDAR depth cameras [3], [4], [5], [6]. These cameras offer enhanced accuracy, stability, and consistent measurements [7], [8]. Moreover, these sensors can interact with information such as color and texture collected from RGB cameras, providing richer data. Therefore, depth cameras are often used in conjunction with RGB cameras [9], [10], [11], [12].

The simultaneous deployment of multiple depth cameras not only resolves occlusion issues stemming from the ambiguity inherent in single-camera setups [13] but also significantly expands the effective field of view (FoV), thereby enhancing overall spatial perception and data acquisition capabilities. However, the implementation of multiple LiDAR depth camera systems introduces significant interference issues, potentially compromising these advantages. This interference occurs because LiDAR depth cameras utilize an active sensing method that emits and receives infrared signals, leading to inaccurate depth estimations when signals from multiple cameras intersect [14]. Consequently, a critical challenge lies in obtaining depth data that maintains the accuracy and quality of single-camera measurements, even within a multi-camera setup.

Several researchers proposed synchronization methods to utilize multiple LiDAR depth cameras [15], [16]. Through these methods, it has been possible to obtain depth data resembling that captured by a single camera, even in systems with multiple cameras. However, these approaches require the connection of cables, which can limit the depth capture range and may introduce synchronization issues. Additionally, these methods sacrifice frames, making them challenging to use in dynamic environments such as robotics and autonomous navigation that require real-time performance.

The problem of restoring depth interference can also be considered an image restoration problem in the field of computer visions [17], [18]. With recent advancements in deep neural networks (DNNs), methods such as debluring [19], [20], denoising [21], [22], and deraining [23], [24] have been researched to restore images by removing artifacts. However, this approach is likely to face significant challenges in effectively recovering interfered artifacts. The primary difficulty lies in the necessity to distinguish between different types of artifacts in the interference restoration process. This means that interfered depth data exhibits two artifacts with similar distributions: the *sensor artifact* and the *interference artifact*. Sensor artifacts, which naturally occur in the LiDAR scanning method, should be preserved. Conversely, interference artifacts, arising only when multiple sensors are used, should be the focus of restoration efforts. Existing image restoration models, however, struggle to differentiate between these two types of artifacts and often require computationally expensive architectures that are unsuitable for real-time applications. Consequently, there is a critical need for two key components: firstly, a sophisticated model capable of learning to distinguish between these artifact types and accurately restore depth; and secondly, a specialized dataset tailored for training such models on these specific artifacts.

In this letter, we propose a novel approach, DRIM, for restoring depth affected by interference in multiple LiDAR depth camera systems. We classified the artifacts in the depth data acquired from these systems as sensor artifacts and interference artifacts, for the first time. Our model distinguishes these two

Received 15 June 2025; accepted 1 October 2025. Date of publication 9 October 2025; date of current version 16 October 2025. This article was recommended for publication by Associate Editor L. Shao and Editor A. Valada upon evaluation of the reviewers' comments. This work was supported in part by the National Research Foundation of Korea (NRF) funded by the Korean government (MSIT) under Grant RS-2025-00564137, in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) funded by the Korean government (MSIT) under Grant RS-2022-00155911, in part by Artificial Intelligence Convergence Innovation Human Resources Development (Kyung Hee University), and in part by Convergence Security Core Talent Training Business Support Program under Grant IITP-2023-RS-2023-00266615. *(Corresponding author: Hyoseok Hwang.)*

The authors are with the Department of Software Convergence, Kyung Hee University, Yongin-Si 17104, South Korea (e-mail: jumin1116@khu.ac.kr; yoon2926@khu.ac.kr; sundong@khu.ac.kr; hyoseok@khu.ac.kr).

Our project page is at https://sites.google.com/view/drim-dataset/.

Digital Object Identifier 10.1109/LRA.2025.3619771

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

artifacts, utilizing these artifacts to restore depth while achieving real-time performance suitable for practical applications. Due to the lack of a dataset capable of training on these artifacts generated by multiple cameras, we create and provide a new dataset for this purpose. We rigorously evaluate the performance of our proposed method through comprehensive comparisons with existing image restoration techniques. Our assessment encompasses various challenging scenarios, including interference restoration in multi-camera setups and environments with diverse types of depth cameras. The contributions presented in this study include:

- We propose DRIM for restoring depth that has been interfered with by other depth cameras. Our approach first distinguishes artifacts in the interfered depth and then designs a model to process each of these artifacts.
- We provide a *depth interference dataset* for the first time that facilitates the learning of artifacts observed in multiple LiDAR depth camera systems and is designed for further research in such environments.
- Our experimental results demonstrate that our approach effectively restores interfered depth data at real-time speeds (≈33 FPS), outperforming existing image restoration methods across challenging scenarios.

II. Related Works

A. Depth Camera

There are several depth sensing methods for depth cameras, including structured light, stereovision, time-of-flight (ToF), and LiDAR. Structured light [25], [26] works by emitting a specific pattern of infrared signals onto the scene and predicting depth based on how the infrared pattern is projected onto the scene. Stereovision [27], [28] predicts depth by matching features commonly detected in two adjacent RGB images. ToF method [29], [30] predicts depth by emitting infrared signals and measuring either the time it takes for the reflected infrared signals to return from the scene or the phase difference of the received signals. LiDAR depth camera follows the ToF principle but enhance it with an active scanning mechanism. By emitting and receiving infrared signals along a defined scan trajectory, they achieve precise spatial sampling. This scanning-based measurement enables LiDAR depth camera to deliver superior accuracy, stability, and consistency compared to other types of cameras [7], [8]. However, there is an issue of interference when using multiple LiDAR depth cameras.

B. Synchronization for Multiple LiDAR Depth Cameras

Synchronization methods have been proposed to enable the use of multiple LiDAR depth camera systems. Mulla et al. [15] addressed the interference issue by connecting a cable to a single computer and turning the cameras on and off sequentially. Similarly, Breggion et al. [16] used synchronization to perform 3D reconstruction tasks. These methods effectively prevent interference in these systems, allowing the acquisition of data similar to that obtained in a single LiDAR depth camera system. However, the synchronization approach has the drawback of sacrificing frames, making it difficult to capture the same scene in dynamic environments where objects or backgrounds are moving. Additionally, since each camera must be connected to a single computer via cables, the depth camera's capture range is inherently limited, and any synchronization issues can lead to interference.

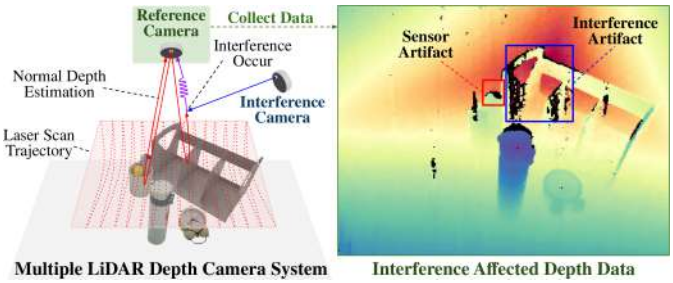

Fig. 1. Various artifacts observed in multiple LiDAR depth camera system.

C. Image Restoration

Recent studies have actively explored artifact removal for image restoration. Gao et al. [19] introduced a simple U-Net-based single-stage approach that reduces system complexity while maintaining spatial accuracy and contextual richness. Chen et al. [24] developed a multi-scale Transformer with bi-directional feedback across scales, enabling coarse-to-fine and fine-to-coarse feature exchange. Luo et al. [21] proposed a latent diffusion model operating in latent space to efficiently handle high-resolution content, achieving strong results in denoising and inpainting. Özdenizci et al. [23] presented a patch-based diffusion method for weather-degraded images, applying guided denoising over overlapping patches for size-agnostic restoration. While these methods remove artifacts, they struggle with recovering interfered artifacts.

III. Problem Analysis and Dataset Construction

Our goal is to restore depth data affected by interference in multiple LiDAR depth cameras to resemble depth data from a single one. We first identify the interference present in the multiple LiDAR depth camera systems. Next, we classify the artifacts that appear in the depth data of these systems. Since there is no available dataset to learn the artifacts occurring in these systems, we create a depth interference dataset.

A. Depth Interference

We first identify the interference that occurs in multiple LiDAR depth camera systems. These cameras estimate depth using an active method that emits and receives infrared signals along laser scan trajectory, as shown in Fig. 1. Therefore, when signals from other cameras are received, inaccurate depth estimations occur, resulting in the appearance of artifacts along the laser scan trajectory. As illustrated in Fig. 1, various artifacts appear in the depth data affected by interference. However, the causes of these artifacts differ. Some artifacts are naturally generated by the LiDAR scanning method, while others occur solely due to interference from other cameras. To effectively restore depth data, it is crucial to classify artifacts based on their distinct underlying causes, even when their distributions appear similar.

B. Definitions of Artifact Types

We classify the artifacts observed in interference affected depth data into two distinct artifacts, as shown in Fig. 1. The first

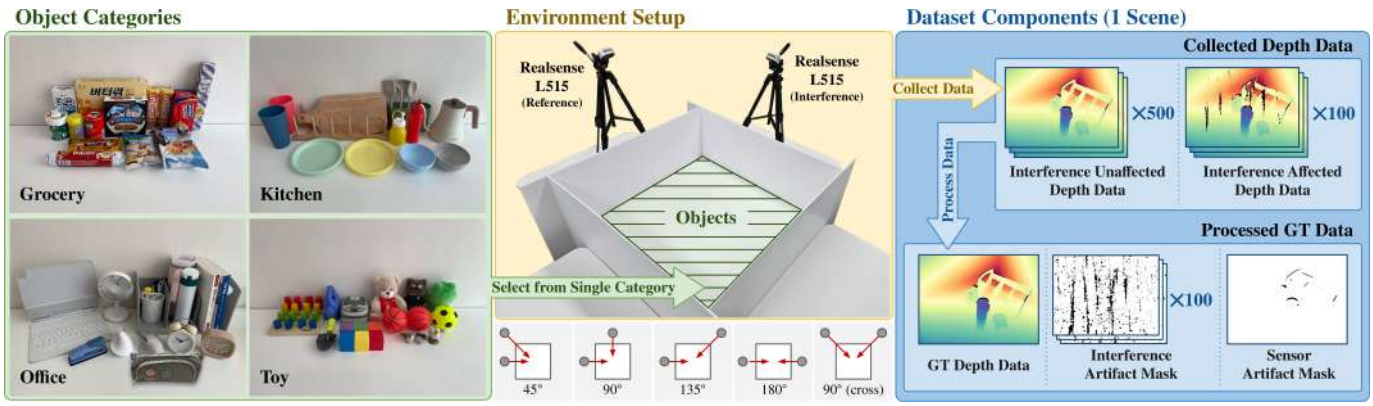

Fig. 2. Environment setup and data generation procedure for the depth interference dataset.

artifact is characterized by invalid depth values resulting from the physical limitations of the capturing equipment and the physical properties of the sensors, even in the single camera systems. We define the artifact exhibiting these characteristics as a *sensor artifact*. The second artifact is characterized by inaccurate depth values due to interference from multiple camera systems, which is defined as an *interference artifact*. Sensor artifacts can also occur in single camera systems, so, these artifacts do not require restoration. Conversely, interference artifacts, which appear only due to interference in multiple camera systems, must be restored.

C. Depth Interference Dataset

One of our main challenges is to create and provide a depth interference dataset for multiple LiDAR depth camera systems. Our dataset offers interference unaffected data obtained using a single depth camera, as well as interference affected data obtained using two cameras. Additionally, we provide ground truth (GT) masks for sensor and interference artifacts to facilitate depth restoration. To support various future learning approaches, we provide supplementary data such as RGB images of the corresponding scenes, intrinsic and extrinsic parameters between the two cameras.

1) Data Collection: We first collected interference unaffected depth data using a single Intel Realsense L515, then added Intel Realsense L515 in the same environment to capture interference affected data for the same scenes. To enable the model to learn artifacts affected by interference in various environments, scenes were created using objects from 4 different categories: Grocery, Kitchen, Office, and Toy. The environment setup involved capturing scenes under 5 distinct camera poses : 45°, 90°, 135°, 180°, and 90° (cross). For each camera pose, 10 scenes were generated per category by randomly selecting and arranging objects from that category. For each scene, 100 interference affected depth frames and 500 interference unaffected depth frames were captured, the latter was used to generate GT depth data based on the data distribution. Depth frames were captured from both cameras corresponding to each camera pose configuration. The environment setup and data generation procedure for the depth interference dataset are illustrated in Fig. 2.

2) Ground Truth Generation: We generated GT data based on the distribution of the collected data. To train our model, the dataset includes additional GTs: a sensor artifact mask, an interference artifact mask, and GT depth data.

First, the sensor artifact mask represents regions of invalid depth values that appear even in interference unaffected conditions. We calculated the distribution of 500 interference unaffected depth frames for each pixel in each scene. We defined the sensor artifact mask by selecting only pixels where the median of zero in the distribution, as follows:

$$S_{i,j} = \begin{cases} 0, & \text{if median}(d_{i,j}) = 0, \\ 1, & \text{otherwise}, \end{cases} \tag{1}$$

where S is the sensor artifact mask. The term median $(d_{i,j})$ represents the median depth value of pixel (i, j) across 500 interference unaffected depth frames, calculated as:

$$\text{median}(d) = \frac{d_{\lfloor \frac{N+1}{2} \rfloor} + d_{\lceil \frac{N+1}{2} \rceil}}{2}, \tag{2}$$

where d_i denotes the i-th value in the ordered list of pixel values and N is the total number of frames. The sensor artifact mask created using this method indicates the regions in the scene that are likely to represent invalid depth values.

Second, the interference artifact mask identifies regions of interference artifacts in each interference affected depth frame. Considering that interference typically distorts values outside the sensor's measured data distribution, we created the interference artifact mask using an interquartile range (IQR) [31] based outlier detection method on the distribution of non-zero pixels in the interference unaffected data. From the 500 interference unaffected depth frames, the lower 25% value ($Q1$) and the upper 75% value ($Q3$) for each pixel distribution were calculated, excluding zeros, as:

$$Q1 = d_{\text{non-zero}(\lfloor 0.25M \rfloor)}, \tag{3}$$

$$Q3 = d_{\text{non-zero}(\lceil 0.75M \rceil)}, \tag{4}$$

where $d_{\text{non-zero}}$ is the non-zero depth value and M is the total number of non-zero depth values for each pixel. The IQR was given by:

$$IQR = Q3 - Q1. \tag{5}$$

The normal range was defined as:

$$R = [Q1 - 1.5IQR,\ Q3 + 1.5IQR], \tag{6}$$

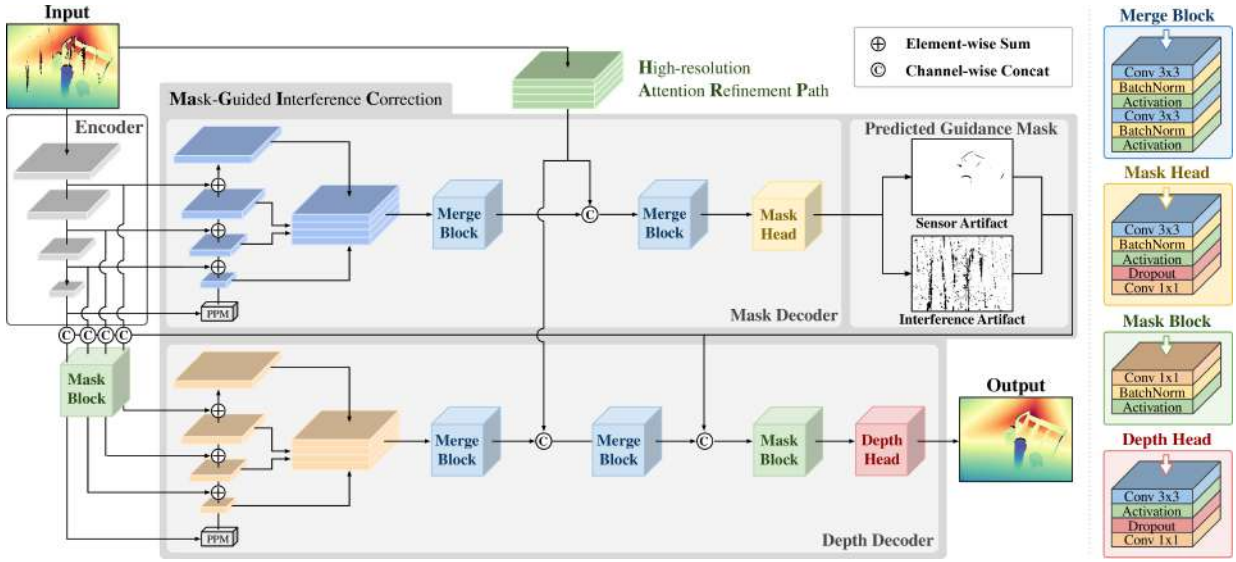

Fig. 3. Main architecture of DRIM. This architecture consists of a shared encoder, HARP, and separate task-specific decoders. The features from the HARP are fed into the mask decoder and the depth decoder to preserve fine details. The mask decoder predicts the sensor artifact mask and interference artifact mask, while the depth decoder utilizes the predicted masks to restore the interference affected depth data, a process which we term MaGIC.

where R represents normal range. Pixels in the interference affected depth data $\acute{d}$ that fell outside this normal range and did not overlap with the sensor artifact mask were defined as interference artifacts in the mask:

$$I_{i,j} = \begin{cases} 0, & \text{if } \acute{d}_{i,j} \notin R \text{ and } S_{i,j} = 1, \\ 1, & \text{otherwise}, \end{cases} \tag{7}$$

where I is interference artifact mask.

Lastly, the GT depth data represents the depth data fully restored from the effects of interference, which is the final output we aim to reconstruct. To generate the GT depth data, we computed the mean of the non-zero pixel values from the interference unaffected data. The GT depth value for each pixel was defined as:

$$D_{GT\,i,j} = \begin{cases} \frac{1}{N_{i,j}} \sum_{k=1}^{N_{i,j}} d_{i,j}^{k}, & \text{if } S_{i,j} = 1, \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

where $D_{GT\,i,j}$ represents the GT depth value, $d_{i,j}^{k}$ denotes the depth values of the non-zero pixel across the interference unaffected frames k, and $N_{i,j}$ is the number of non-zero depth values. By using the mean of the non-zero values, we ensured that the GT depth data reflects the expected values for the valid pixels. Pixels corresponding to the sensor artifact mask are initialized to zero to finalize the GT depth.

IV. Method

We propose *depth restoration with interference mitigation* (DRIM), a novel depth restoration model. As illustrated in Fig. 3, our architecture adopts a shared encoder and modified UperNet-based [32] decoders for mask prediction and depth restoration, integrating a mask-guided interference correction (MaGIC) and a high-resolution attention refinement path (HARP) as its main contributions.

A. Mask-Guided Interference Correction (MaGIC)

The MaGIC predicts a three-class segmentation mask from the interference affected depth input $D \in \mathbb{R}^{1\times H\times W}$ to provide spatial guidance for selective restoration. To guide the depth decoder, we concatenate the predicted mask probability maps (background, sensor, interference) with the decoder features at multiple scales. These three-channel mask maps are resized to match the spatial dimensions of the decoder features and concatenated along the channel axis. This multi-scale integration ensures consistent artifact-aware guidance from coarse to fine resolutions, enabling the network to focus restoration on interference artifacts while preserving sensor artifacts. The mask guidance effectively prevents unnecessary modifications to regions that should remain unchanged.

B. High-Resolution Attention Refinement Path (HARP)

To address the loss of fine spatial details during encoder downsampling, we introduce a HARP. This auxiliary branch processes the input depth data at full resolution without downsampling operations, complementing the main encoder-decoder architecture. HARP consists of shallow residual blocks combined with Convolutional Block Attention Modules (CBAM) [33] to capture fine-grained features such as edges and fine textures. The high-resolution features from HARP are integrated into both decoders through skip connections, providing critical spatial information for accurate artifact classification and depth restoration. By leveraging CBAM, HARP selectively emphasizes important spatial features while filtering out noise, which is essential for distinguishing between sensor artifacts that should be preserved and interference artifacts that must be removed, particularly in challenging regions such as thin structures and boundaries.

C. Loss Functions

We adopt a composite depth loss inspired by NeRD-Rain [24], combining Charbonnier loss,

TABLE I
QUANTITATIVE COMPARISON WITH OTHER IMAGE RESTORATION METHODS. EACH MODEL WAS TRAINED AND EVALUATED WITH MULTIPLE SEEDS USING OUR DEPTH INTERFERENCE DATASET. **BOLD** INDICATES THE BEST, WHILE UNDERLINE REPRESENTS THE SECOND BEST RESULTS

Models	Runtime (s) ↓	RMSE (m) ↓	MAE (m) ↓	Artifacts RMSE (m) ↓	Artifacts MAE (m) ↓
Input	-	0.0984	0.0122	0.3690	0.1267
M3SNet	0.0346±0.0004	0.0213±0.0001	0.0028±0.0001	0.0826±0.0008	0.0202±0.0011
NeRD-Rain	0.4468±0.0003	0.0206±0.0001	**0.0027±0.0001**	0.0795±0.0002	0.0194±0.0003
Refusion	7.7738±0.0266	0.0469±0.0007	0.0055±0.0001	0.1697±0.0015	0.0488±0.0008
WeatherDiffusion	15.8227±0.0988	0.0295±0.0001	0.0033±0.0000	0.1130±0.0002	0.0241±0.0000
DRIM w/o mask	**0.0252±0.0007**	0.0210±0.0001	0.0032±0.0001	0.0826±0.0001	0.0214±0.0003
DRIM (Ours)	0.0299±0.0014	**0.0199±0.0001**	**0.0027±0.0001**	**0.0774±0.0005**	**0.0183±0.0002**

frequency consistency loss, edge-aware loss, and L_1 term:

$$\mathcal{L}_d = \mathcal{L}_{char} + \lambda_1 \mathcal{L}_{fft} + \lambda_2 \mathcal{L}_{edge} + \lambda_3 \mathcal{L}_{\ell_1}, \tag{9}$$

with $\lambda_1 = 0.01$, $\lambda_2 = 0.05$, and $\lambda_3 = 0.1$. For segmentation, we use the standard cross-entropy loss as the mask loss:

$$\mathcal{L}_m = \mathcal{L}_{CE}(\hat{M}, M_{gt}), \tag{10}$$

and train both tasks jointly:

$$\mathcal{L}_{total} = \mathcal{L}_d + \lambda_m \mathcal{L}_m, \quad \lambda_m = 0.8. \tag{11}$$

V. EXPERIMENTAL RESULTS

Here, we conducted a comparative analysis between our method and existing image restoration methods. Then, we demonstrated the ability to restore depth in challenging scenarios, including multi-camera setups and environments with diverse types of depth cameras. Through ablation studies, we addressed the question of whether classifying and utilizing artifacts is effective for restoring depth data affected by interference. Moreover, we evaluated the performance based on the each component of the architecture.

A. Implementation

We trained DRIM for 100 epochs with AdamW optimizer [34]. The learning rate was set to 0.0005 and the weight decay was set to 0.0001. Also, the learning rate was decayed exponentially with a gamma of 0.9. We employed a Swin Transformer V2 [35] pre-trained backbone for the encoder. We trained the model with a batch size of 32 using 256 × 256 resolution images and validated with a batch size of 16 using 640 × 480 resolution images. To perform unseen testing, we constructed the training, validation, and test datasets using non-overlapping scenes. Thus, the dataset consisted of 24 k training images, 8 k validation images, and 8 k test images. The experiments were conducted on NVIDIA RTX A6000 GPU.

B. Performance Evaluation

We evaluated our method against four baseline methods for image restoration, with all models trained on our depth interference dataset. We measured both the overall RMSE and MAE, and the Artifacts RMSE and MAE, the latter of which were computed from the pixels within the GT sensor artifact mask and GT interference artifact mask. The formula for the latter evaluation metrics were as follows:

$$RMSE_{Art} = \left(\frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} \left(D_{GTi,j} - D_{Predi,j} \right)^2 \right)^{1/2}, \tag{12}$$

$$MAE_{Art} = \frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} \left| D_{GTi,j} - D_{Predi,j} \right|, \tag{13}$$

$$\Omega = \{(i,j) \,|\, I_{i,j} = 0 \vee S_{i,j} = 0\}, \tag{14}$$

where $RMSE_{Art}$ referred to the Artifact RMSE, MAE_{Art} represented the Artifact MAE, and $D_{Predi,j}$ denoted the restored depth.

As shown in Table I, DRIM achieved the best performance across all metrics, recording the lowest RMSE of 0.0199 m, MAE of 0.0027 m, artifacts RMSE of 0.0774 m, and artifacts MAE of 0.0183 m. In addition to its accuracy, ours demonstrated superior computational efficiency, processing each frame in 0.0299 seconds, which corresponded to approximately 33 FPS and enabled real-time performance. Compared to the strongest baseline, NeRD-Rain, which achieved a similar MAE of 0.0027 m but required 0.4468 seconds per frame (approximately 2.2 FPS), our method offered a 15× speedup while maintaining higher accuracy. These results showed DRIM's effectiveness for real-time depth restoration in regions affected by interference and sensor artifacts. In addition, comparison with DRIM trained without mask demonstrated the architectural advantage of restoring depth by distinguishing between sensor and interference artifacts.

The qualitative evaluation further validated our quantitative findings. Fig. 4 presented a visual comparison demonstrating our method's ability to effectively reduce interference artifacts (blue boxes) while preserving sensor artifacts (red boxes). The 3D point cloud visualization in Fig. 5 showed the superior reconstruction quality of flat surfaces such as walls and floors (blue boxes).

C. Performance Evaluation in Challenging Scenarios

We conducted additional experiments to evaluate the performance of our method in restoring depth in challenging scenarios. The first scenario involved a case of increased interference by adding a LiDAR depth camera, as seen in Fig. 6. We created a test dataset for evaluation in this scenario and ran an unseen test. As shown in Table II for scenario 1, ours achieved the lowest MAE of 0.0038 m, and Artifacts MAE of 0.0151 m. Although

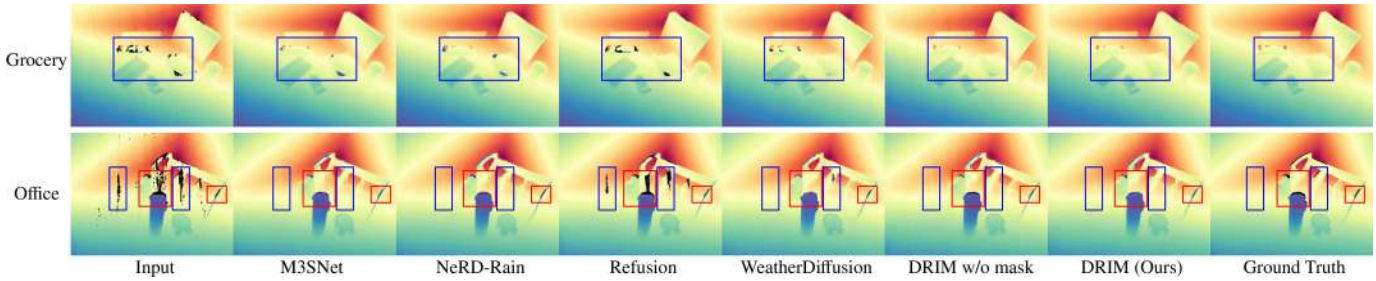

Fig. 4. Qualitative 2D depth comparison of our method with other image restoration methods.

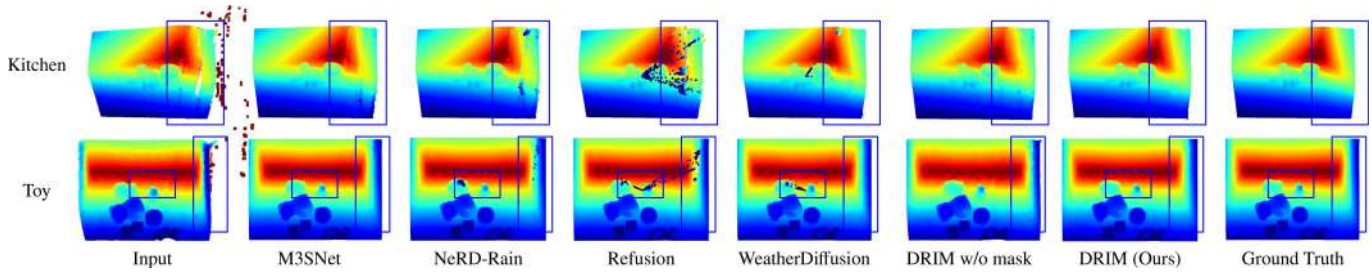

Fig. 5. Qualitative 3D point cloud comparison of our method with other image restoration methods.

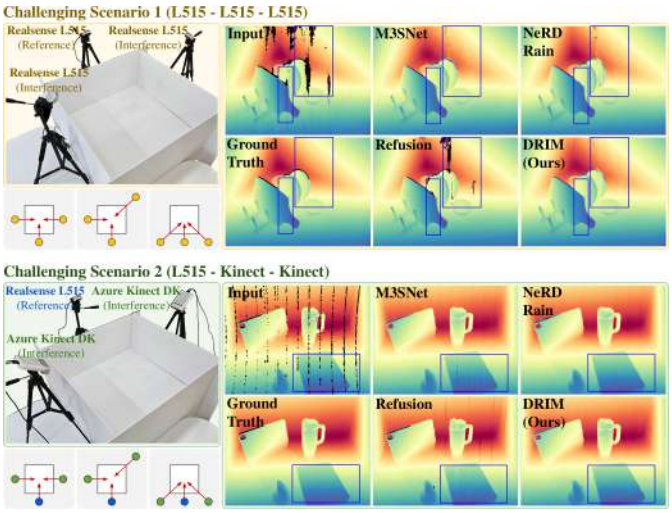

Fig. 6. Visualization results in challenging scenarios.

TABLE II
DEPTH RESTORATION RESULTS IN CHALLENGING SCENARIOS. THE SCENARIOS ARE CATEGORIZED INTO TWO TYPES BASED ON THE KIND AND NUMBER OF CAMERAS INTERFERING WITH THE LIDAR DEPTH CAMERA. EACH MODEL WAS EVALUATED WITH MULTIPLE SEEDS USING OUR CHALLENGING TEST DATASET. **BOLD** INDICATES THE BEST, WHILE UNDERLINE REPRESENTS THE SECOND BEST RESULTS

Scenarios	Models	RMSE (m) ↓	MAE (m) ↓	Artifacts RMSE (m) ↓	Artifacts MAE (m) ↓
1	Input	0.1398	0.0196	0.3613	0.1095
	M3SNet	0.0281	0.0046	0.0727	0.0192
	NeRD-Rain	**0.0251**	0.0039	**0.0649**	0.0157
	Refusion	0.0598	0.0080	0.1514	0.0388
	WeatherDiffusion	0.0382	0.0058	0.0998	0.0252
	DRIM (Ours)	0.0253	**0.0038**	0.0659	**0.0151**
2	Input	0.1819	0.0417	0.6341	0.4758
	M3SNet	0.0256	0.0037	0.0859	0.0227
	NeRD-Rain	**0.0247**	0.0036	**0.0823**	0.0215
	Refusion	0.0753	0.0102	0.2470	0.0935
	WeatherDiffusion	0.0375	0.0043	0.1245	0.0282
	DRIM (Ours)	0.0248	**0.0034**	0.0832	**0.0204**

NeRD-Rain achieved the lowest RMSE and Artifacts RMSE among all baselines, our model attained comparable RMSE and Artifacts RMSE. Even in scenario with significant interference, the interference followed the laser trajectory of the LiDAR depth camera. Thus, our method restored depth even in situations with increased interference.

The second scenario involved interference caused by cameras based on the ToF principle. We constructed a scenario with interference using two Azure Kinect DK devices, as shown in Fig. 6. We created a test dataset for evaluation in this scenario, and performed an unseen test. As shown in Table II for scenario 2, our method achieved the lowest MAE of 0.0034 m, and artifacts MAE of 0.0204 m. The Azure Kinect DK estimated depth through 9 infrared pulses. Consequently, when two Azure Kinect DK cameras were used, interference manifested as up to 18 distinct lines in the depth data captured by the LiDAR depth camera. Despite this specific and challenging interference pattern originating from ToF cameras, our method successfully restored the depth data in these challenging scenarios. To visually evaluate the two scenarios, we presented qualitative results in Fig. 6. Our method showed superior restoration of interference artifacts (blue boxes). Through these results, we demonstrated that our method is applicable even in challenging scenarios.

D. Ablation Studies

1) Analysis of Artifact Separation Strategy: We proposed a method for restoring depth data by classifying artifacts in interference affected depth data into sensor and interference artifacts. To evaluate this, we examined five settings: (1) restoring depth without predicting any artifacts, (2) predicting only the sensor artifact mask (background, sensor), (3) predicting only

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE III
DEPTH RESTORATION RESULTS BASED ON THE TYPES OF ARTIFACTS USED. EACH SETTING WAS COMPOSED ACCORDING TO THE ARTIFACTS PREDICTED DURING THE RESTORATION PROCESS

Setting	Sensor Artifact	Interference Artifact	RMSE (m) ↓	MAE (m) ↓
1			0.0210	0.0032
2		✓	0.0201	0.0027
3	✓		0.0208	0.0032
4		✓	0.0201	0.0028
5 (DRIM)	✓	✓	**0.0199**	**0.0027**

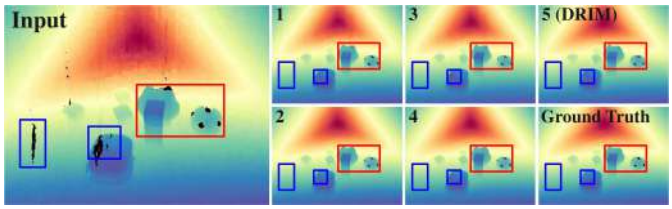

Fig. 7. Visualization of restoration results based on the types of artifacts used. The number in the top-left corner of the images corresponds to the setting number in Table III.

TABLE IV
DEPTH RESTORATION RESULTS BASED ON MODULES COMPRISING THE ARCHITECTURE

Setting	MaGIC	HARP	RMSE (m) ↓	MAE (m) ↓
Baseline			0.0217	0.0036
1	✓		0.0207	0.0033
2		✓	0.0210	0.0032
DRIM	✓	✓	**0.0199**	**0.0027**

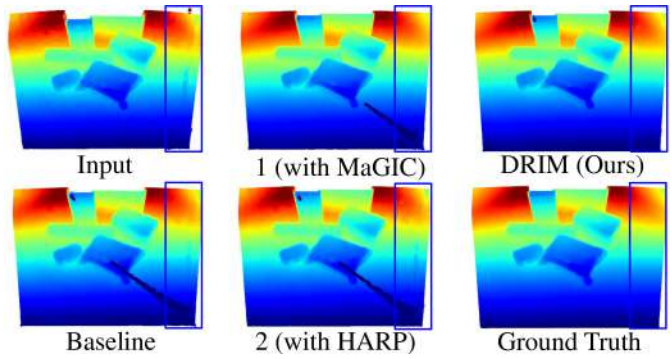

Fig. 8. Visualization of restoration results based on modules comprising the architecture.

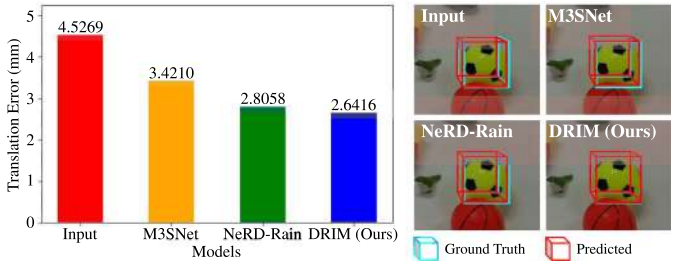

Fig. 9. 6-DoF pose estimation results using restored depth.

the interference artifact mask (background, interference), (4) predicting a merged mask of sensor and interference artifacts without distinction (background, sensor & interference), and (5) our approach with explicit classification into sensor and interference artifacts (background, sensor, interference).

As shown in Table III, our method (Setting 5) achieved the best performance with RMSE 0.0199 m and MAE 0.0027 m. Setting 2, predicting only interference artifacts, showed RMSE 0.0201 m, while Setting 3, focusing on sensor artifacts, showed RMSE 0.0208 m. These results indicate that restoring interference artifacts and preserving sensor artifacts are both essential. Setting 4, predicting both artifacts jointly, yielded RMSE 0.0201 m, inferior to our separation approach. Thus, distinguishing artifact types is important for optimal restoration. As shown in Fig. 7, ours successfully restored interference artifacts (blue boxes) while preserving sensors (red boxes).

2) Ablation on DRIM Architecture: We proposed an architecture that effectively restored interference affected depth data employing MaGIC and HARP. To evaluate the individual contributions of these key components, we conducted an ablation study as shown in Table IV. Starting from the baseline, which had RMSE of 0.0217 m and MAE of 0.0036 m, adding MaGIC alone improved to RMSE of 0.0207 m and MAE of 0.0033 m. Applying only the HARP led to RMSE of 0.0210 m and MAE of 0.0032 m. The complete DRIM model, incorporating both components, achieved the best performance with RMSE of 0.0199 m and MAE of 0.0027 m, demonstrating that the two components were essential and complementary for effectively restoring depth data affected by interference. As shown in Fig. 8, ours demonstrated superior restoration of the artifacts highlighted by the blue boxes.

E. Application of Restored Depth to Pose Estimation

We conducted an application experiment through 6-Degrees of Freedom (DoF) pose estimation. We performed model-based 6-DoF pose estimation using FoundationPose [36], which takes RGB, depth, and mask as inputs. For this, we utilized Grounded-SAM [37] to generate object masks and used the depth restored by each model. We estimated the pose of a "soccer ball" in 100 frames. Since the "soccer ball" had a symmetric property, we fixed the rotation and compared only the translation error. The pose estimated using ground truth depth was used as the ground truth pose, and the translation error was calculated as the Euclidean distance between the estimated and ground truth translations. We compared our method with M3SNet, which offers reasonable processing speed, NeRD-Rain, which provides competitive restoration quality, and interference affected depth without any restoration. As shown in Fig. 9, our method achieved the lowest translation error of 2.6416 mm. This result represented a 71.3696% improvement over the interference affected depth input. Furthermore, as demonstrated by the overlap between the predicted pose (red bounding box) and ground truth pose (blue bounding box), our method yielded the closest alignment—the two bounding boxes matched most accurately among all methods.

VI. Conclusion

We propose DRIM for restoring depth data affected by interference from multiple LiDAR depth cameras. Ours first classifies artifacts caused by interference, then predicts and leverages them to restore accurate depth. To enable this, we introduce the first depth interference dataset designed for learning interference scenarios in multi-LiDAR setups. Experimental results show that our method outperforms existing image restoration methods in both accuracy and robustness, even under challenging conditions. Importantly, ours achieves real-time performance ($\approx$33 FPS), making it suitable for practical deployment in dynamic environments. However, since this study addresses interference in a limited set of LiDAR depth cameras, future work should aim to investigate interference across a broader range of LiDAR depth cameras.

References

[1] J. Volak, D. Koniar, F. Jabloncik, L. Hargas, and S. Janisova, "Interference artifacts suppression in systems with multiple depth cameras," in *Proc. IEEE 42nd Int. Conf. Telecommun. Signal Process.*, 2019, pp. 472–476.

[2] L. Yang, L. Zhang, H. Dong, A. Alelaiwi, and A. El Saddik, "Evaluating and improving the depth accuracy of kinect for windows v2," *IEEE Sensors J.*, vol. 15, no. 8, pp. 4275–4285, Aug. 2015.

[3] Y. Ze, G. Zhang, K. Zhang, C. Hu, M. Wang, and H. Xu, "3D diffusion policy: Generalizable visuomotor policy learning via simple 3D representations," in *Proc. Robot.: Sci. Syst.*, 2024.

[4] M. Ramanathan et al., "Visual environment perception for obstacle detection and crossing of lower-limb exoskeletons," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2022, pp. 12267–12274.

[5] T. Jin et al., "Whole-body inverse kinematics and operation-oriented motion planning for robot mobile manipulation," *IEEE Trans. Ind. Inform.*, vol. 20, no. 12, pp. 14239–14248, Dec. 2024.

[6] J. Biswas and M. Veloso, "Depth camera based indoor mobile robot localization and navigation," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2012, pp. 1697–1702.

[7] F. Lourenço and H. Araujo, "Intel realsense SR305, D415 and L515: Experimental evaluation and comparison of depth estimation," in *Proc. VISIGRAPP*, 2021, pp. 362–369.

[8] A. Breitbarth, C. Hake, and G. Notni, "Measurement accuracy and practical assessment of the Lidar camera intel realsense l515," *Proc. SPIE*, 2021, vol. 11782, Art. no. 1178213.

[9] B. Huang, J. Yu, and S. Jain, "Earl: Eye-on-hand reinforcement learner for dynamic grasping with active pose estimation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2023, pp. 2963–2970.

[10] H. Zhang, A. Opipari, X. Chen, J. Zhu, Z. Yu, and O. C. Jenkins, "TransNet: Category-level transparent object pose estimation," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 148–164.

[11] Q. Yu et al., "Gamma: Generalizable articulation modeling and manipulation for articulated objects," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2024, pp. 5419–5426.

[12] C. Wang, H. Shi, W. Wang, R. Zhang, L. Fei-Fei, and C. K. Liu, "Dexcap: Scalable and portable mocap data collection system for dexterous manipulation," in *Proc. Robot.: Sci. Syst.*, 2024.

[13] L. Zhang, J. Sturm, D. Cremers, and D. Lee, "Real-time human motion tracking using multiple depth cameras," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2012, pp. 2389–2395.

[14] L. A. Seewald et al., "Toward analyzing mutual interference on infrared-enabled depth cameras," *Comput. Vis. Image Understanding*, vol. 178, pp. 1–15, 2019.

[15] O. Mulla and A. Grunnet-Jepsen, *Multi-Camera Configurations with the Intel Realsense Lidar Camera l515*. Washington, DC, USA: Slate, Apr. 2021.

[16] E. Breggion, C. Balletti, and F. Guerra, "Multi-camera lidar system for spatial and temporal preservation of the intangible cultural heritage," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 48, pp. 297–302, 2023.

[17] J. Su, B. Xu, and H. Yin, "A survey of deep learning approaches to image restoration," *Neurocomputing*, vol. 487, pp. 46–65, 2022.

[18] X. Li et al., "Diffusion models for image restoration and enhancement–a comprehensive survey," *Int. J. Comput. Vis.*, pp. 1–31, 2025.

[19] H. Gao, J. Yang, Y. Zhang, N. Wang, J. Yang, and D. Dang, "A novel single-stage network for accurate image restoration," *Vis. Comput.*, vol. 40, pp. 7385–7398, 2024.

[20] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 5728–5739.

[21] Z. Luo, F. K. Gustafsson, Z. Zhao, J. Sjölund, and T. B. Schön, "Refusion: Enabling large-size realistic image restoration with latent-space diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 1680–1691.

[22] Z. Luo, F. K. Gustafsson, Z. Zhao, J. Sjölund, and T. B. Schön, "Image restoration with mean-reverting stochastic differential equations," in *Proc. 40th Int. Conf. Mach. Learn.*, 2023, pp. 23045–23066.

[23] O. Özdenizci and R. Legenstein, "Restoring vision in adverse weather conditions with patch-based denoising diffusion models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, pp. 10346–10357, Aug. 2023.

[24] X. Chen, J. Pan, and J. Dong, "Bidirectional multi-scale implicit neural representations for image deraining," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 25627–25 636.

[25] J. Geng, "Structured-light 3D surface imaging: A tutorial," in *Proc. Adv. Opt. Photon.*, 2011, pp. 128–160.

[26] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2003, pp. 195–202.

[27] S. Mattoccia, "Stereo vision: Algorithms and applications," *Univ. Bologna*, vol. 22, 2011.

[28] D. Marr and T. Poggio, "A computational theory of human stereo vision," *Proc. Roy. Soc. London Ser. B Biol. Sci.*, vol. 204, no. 1156, pp. 301–328, 1979.

[29] C. Bamji et al., "A review of indirect time-of-flight technologies," *IEEE Trans. Electron Devices*, vol. 69, no. 6, pp. 2779–2793, Jun. 2022.

[30] P. Zanuttigh et al., "Time-of-flight and structured light depth cameras," *Technol. Appl.*, vol. 978, no. 3, 2016.

[31] X. Wan, W. Wang, J. Liu, and T. Tong, "Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range," *BMC Med. Res. Methodol.*, vol. 14, pp. 1–13, 2014.

[32] T. Xiao, Y. Liu, B. Zhou, Y. Jiang, and J. Sun, "Unified perceptual parsing for scene understanding," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 418–434.

[33] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[34] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. Int. Conf. Learn. Representations*, 2019.

[35] Z. Liu et al., "Swin transformer v2: Scaling up capacity and resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 12009–120 19.

[36] B. Wen, W. Yang, J. Kautz, and S. Birchfield, "Foundationpose: Unified 6D pose estimation and tracking of novel objects," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 17868–17879.

[37] T. Ren et al., "Grounded sam: Assembling open-world models for diverse visual tasks," 2024, *arXiv:2401.14159*.