LEARNING-BASED SOLUTIONS FOR LIDAR NAVIGATION UNDER CONSTRAINED ENVIRONMENTS

Prashant Kumar

Amar Nath and Shashi Khosla School of Information Technology

INDIAN INSTITUTE OF TECHNOLOGY DELHI

July 2025

© Indian Institute of Technology Delhi (IITD), New Delhi, 2025

LEARNING-BASED SOLUTIONS FOR LIDAR NAVIGATION UNDER CONSTRAINED ENVIRONMENTS

by

Prashant Kumar

Amar Nath and Shashi Khosla School of Information Technology

Submitted
in fulfillment of the requirements of the degree of
Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

July 2025

THESIS CERTIFICATE

This is to certify that the thesis titled, **"Learning-based Solutions for LiDAR Navigation under Constrained Environments"**, submitted by **Prashant Kumar**, to the Indian Institute of Technology, Delhi, for the award of the degree of **Doctor of Philosophy**, is a bonafide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Professor Srikanta Bedathur Jagannath
Professor and Head,
Amar Nath and Shashi Khosla School of Information Technology
Indian Institute of Technology Delhi, 110016

Prem Kumar Kalra
Professor Prem Kumar Kalra
Former Professor,
Department of Computer Science and Engineering
Indian Institute of Technology Delhi, 110016

ACKNOWLEDGEMENTS

Firstly, I want to remember the Almighty for bestowing upon me the strength and resilience to pursue my research.

I would like to express my deepest gratitude to my thesis advisors, Prof. Prem Kumar Kalra and Prof. Srikanta Bedathur at IIT Delhi. Their mentorship throughout my PhD has played a pivotal role in shaping my research. Their insightful guidance, meticulous feedback, and encouragement kept me motivated and on track.

I would like to thank my research committee consisting of Prof. Rahul Garg, Prof. Chetan Arora and Prof. Sumantra Dutta Roy for providing their continuous feedback and encouragement throughout my PhD.

I would also like to acknowledge the contributions of my research collaborators, Sabyasachi Sahoo, Dheeraj Vattikonda, Erqun Dong, Onkar Susladkar, Dhruv Makwana, Kshitij Bhat and Vedang Nadkarni. Their collaborative efforts significantly enriched my work.

I would like to thank all faculty members of the Amar Nath and Shashi Khosla School of IT (ANSKSIT), IIT Delhi for their constant support. I would like to thank Prof. Anurag Mittal for his constant support, insightful ideas and discussions that enhanced my research output.

I would also like to thank the staff and administration of IIT Delhi for supporting me throughout my journey.

I would like to thank Sahil Manchanda, Anjali Gupta, Garima Gaur, Nikhil Jangam Reddy for their constant help, motivation and assistance in my research work. They helped me get through some of the toughest parts of my journey.

I would like to thank all my peers: Rajat Singh, Samidha Verma, Kunal Dargan, Shivangi Bithel, Abhishek Panda, Mridul Gupta, Subrat Swain, Siddharth Shrivastava, Prafful Khoba, Shubham Gupta, Chaitanya Murthy and all friends at IITD.

Finally, I would like to thank my family and friends for their unwavering love and support throughout this journey. Special thanks to my parents for their constant support before and during my PhD. Their constant encouragement made this achievement possible.

Special thanks to MHRD, IIT Delhi, ANSKSIT-IIT Delhi, Google, Qualcomm, ACM India for supporting me through fellowships and travel grants for attending conferences.

I would also like to thank Prof. Chiranjib Bhattacharyya (Professor, IISc Bangalore), Dr. Rashmi Dutta Baruah (Associate Professor, IIT Guwahati) who were a source of motivation for me to pursue a PhD.

I would like to thank everyone who directly or indirectly contributed to the successful culmination of this thesis.

Prashant Kumar

ABSTRACT

Autonomous navigation using 3D LiDAR point clouds in complex real-world environments remains a challenging task despite extensive research. Uncertainty in environmental factors, such as the probabilistic motion of dynamic objects, varying levels of occlusion, and rare scenes, adds to the complexity. These challenges increase manifold in several adverse and constrained real-life situations, including the absence of labeled information, extreme sparsity, and susceptibility to adversarial attacks. Many real-world navigation scenarios lack precise labeled information due to numerous factors.

We observe that under such circumstances, standard solutions that assist 3D LiDAR navigation do not work. Segmentation, object detection, panoptic segmentation, etc., are not possible with existing solutions. New strategies need to be figured out to achieve these without labeled information while also accounting for adverse constraints. Label-averse settings limit the capacity of autonomous systems, making them susceptible to adversaries and deteriorating navigation. The thesis studies and addresses some of these challenges by proposing innovative methods for 3D LiDAR-based navigation in these challenging and restricted scenarios.

We develop a novel method to distinguish dynamic objects from static environments without requiring labeled data, addressing the limitations of existing segmentation-based approaches for accurate segmentation of dynamic objects. Our solution extracts moving and movable objects from 3D LiDAR point clouds without labels by inferring hidden scene parts occluded by dynamic objects. This inferred static background, akin to a static environment, enables object segmentation in dynamic scenes.

We also handle 3D LiDAR point cloud sparsity, where we propose a technique to augment dynamic sparse LiDAR point clouds with missing information, improving navigation accuracy in constrained environments. Higher sparsity reduces point density on objects, but the overall topology of a 3D LiDAR point cloud remains intact. We leverage tools from Algebraic Topology—particularly 0-dimensional Persistent Homology—to capture this global shape and guide point augmentation along existing static structures. This property generates attention focused on the global topology of a static 3D LiDAR point cloud while augmenting points to a sparse LiDAR point cloud.

We investigate a new paradigm that uses differentiable SLAM in a self-supervised setup to train deep learning models for 3D LiDAR-based applications, enhancing their efficiency. Our proposed approach augments standard navigation tasks with a self-supervised differentiable SLAM framework. It enables end-to-end training of deep learning models with SLAM error, which minimizes the discrepancy between model output and ground truth via trajectory estimates. We demonstrate that our approach outperforms existing methods and achieves improvements in navigation-allied deep learning tasks.

We also explore the vulnerability of 3D LiDAR-based systems to adversarial attacks. Unlike existing passive black-box attack systems, we demonstrate a very simple white-box attack that directly affects the acquired 3D LiDAR point cloud while maintaining the 3D LiDAR point cloud quality to prevent detection. We minimally augment 3D LiDAR point clouds with point injections to destabilize navigation while preserving data integrity. This white-box attack uses adversarial LiDAR point clouds to induce suboptimal trajectory estimates in SLAM, exploiting its reliance on precise 3D LiDAR data.

Our research contributes to advancing the state-of-the-art in 3D LiDAR-based autonomous navigation by addressing critical challenges in restricted settings and providing practical solutions. The proposed methods have the potential to enhance the safety, reliability, and performance of autonomous systems operating in complex and dynamic environments.

सारांश

जटिल वास्तविक दुनिया के वातावरण में **3**डी लिडार पॉइंट क्लाउड्स का उपयोग करके स्वायत्त नेविगेशन एक बड़ी चुनौती है। पर्यावरणीय कारकों जैसे गतिशील वस्तुओं की अनिश्चित गति, विभिन्न छिपी हुई स्थितियां (ऑक्लूज़न), और दुर्लभ दृश्य इसकी जटिलता को बढ़ाते हैं। लेबल रहित जानकारी, अत्यधिक विरलता, और विरोधात्मक हमलों के प्रति संवेदनशीलता जैसी प्रतिकूल परिस्थितियों में ये चुनौतियां और भी बढ़ जाती हैं, क्योंकि कई वास्तविक नेविगेशन परिदृश्यों में सटीक लेबल जानकारी नहीं होती।

ऐसी परिस्थितियों में, मौजूदा 3डी लिडार नेविगेशन समाधान (जैसे सेगमेंटेशन**,** ऑब्जेक्ट डिटेक्शन**,** पैनोप्टिक सेगमेंटेशन) काम नहीं करते। इन बाधाओं के बावजूद, लेबल रहित जानकारी के साथ इन लक्ष्यों को प्राप्त करने के लिए नई रणनीतियाँ खोजना आवश्यक है। लेबल-प्रतिकूल सेटिंग्स स्वायत्त प्रणालियों की क्षमता को सीमित करती हैं, उन्हें हमलों के प्रति संवेदनशील बनाती हैं और नेविगेशन को बाधित करती हैं। यह शोध-पत्र इन चुनौतीपूर्ण और प्रतिबंधित परिदृश्यों में 3डी लिडार-आधारित नेविगेशन के लिए अभिनव तरीकों का प्रस्ताव करके इन समस्याओं का समाधान करता है।

हमने लेबल रहित डेटा का उपयोग करके गतिशील वस्तुओं को स्थिर वातावरण से अलग करने की एक नई विधि विकसित की है। यह गतिशील वस्तुओं द्वारा छिपे हुए दृश्यों का अनुमान लगाकर चलती वस्तुओं को 3डी लिडार पॉइंट क्लाउड्स से अलग करती है, जिससे गतिशील दृश्यों में ऑब्जेक्ट सेगमेंटेशन संभव होता है।

हम 3डी लिडार पॉइंट क्लाउड की विरलता को भी संबोधित करते हैं। हमने लापता जानकारी के साथ गतिशील विरल लिडार पॉइंट क्लाउड्स को बढ़ाने की एक तकनीक प्रस्तावित की है, जिससे प्रतिबंधित वातावरण में नेविगेशन सटीकता बढ़ती है। हम बीजगणितीय टोपोलॉजी (विशेषकर **0-**आयामी परसिस्टेंट होमोलॉजी) का उपयोग करते हैं, जो 3डी लिडार पॉइंट क्लाउड की समग्र स्थलाकृति को बरकरार रखते हुए मौजूदा स्थिर संरचनाओं के साथ बिंदु वृद्धि का मार्गदर्शन करती है।

हम एक नया प्रतिमान तलाशते हैं जो 3डी लिडार-आधारित अनुप्रयोगों के लिए डीप लर्निंग मॉडल को प्रशिक्षित करने के लिए एक स्व**-**पर्यवेक्षित भिन्न **SLAM** सेटअप का उपयोग करता है। यह SLAM त्रुटि के साथ डीप लर्निंग मॉडल के एंड-टू-एंड प्रशिक्षण को सक्षम बनाता है, जिससे मॉडल आउटपुट और वास्तविक डेटा के बीच विसंगति कम होती है। हमारा दृष्टिकोण मौजूदा तरीकों से बेहतर प्रदर्शन करता है और नेविगेशन-संबंधी डीप लर्निंग कार्यों में सुधार लाता है।

हम 3डी लिडार-आधारित प्रणालियों पर विरोधात्मक हमलों की भेद्यता का भी पता लगाते हैं। मौजूदा ब्लैक-बॉक्स हमलों के विपरीत, हम एक सरल व्हाइट**-**बॉक्स हमले का प्रदर्शन करते हैं जो अधिग्रहित 3डी लिडार पॉइंट क्लाउड को सीधे प्रभावित करता है, जबकि इसकी गुणवत्ता को पहचान से बचाने के लिए बनाए रखता है। यह हमला न्यूनतम बिंदु इंजेक्शन के साथ 3डी लिडार पॉइंट क्लाउड्स को बढ़ाकर नेविगेशन को अस्थिर करता है, और

SLAM की सटीक 3डी लिडार डेटा पर निर्भरता का फायदा उठाकर उप-इष्टतम प्रक्षेपवक्र अनुमानों को प्रेरित करता है।

हमारा शोध प्रतिबंधित सेटिंग्स में महत्वपूर्ण चुनौतियों का समाधान करके और व्यावहारिक समाधान प्रदान करके 3डी लिडार-आधारित स्वायत्त नेविगेशन को आगे बढ़ाता है। प्रस्तावित तरीकों में जटिल और गतिशील वातावरण में स्वायत्त प्रणालियों की सुरक्षा, विश्वसनीयता और प्रदर्शन को बढ़ाने की क्षमता है।

Contents

Certificate

Acknowledgements **i**

Abstract **ii**

List of Tables **ix**

List of Figures **xiv**

1 Introduction **1**

1.1 Introduction to Autonomous Navigation and 3D LiDAR 1

1.2 Fundamental Challenges in 3D LiDAR-Based Autonomous Navigation: Bridging the Gap to Real-World Deployment 2

1.3 Foundational Aspects of LiDAR Sensors and Data used in the thesis 5

1.4 Thesis Contributions: A Cohesive Approach to Robust LiDAR Navigation 8

1.5 An Overall Narrative: Robust LiDAR Navigation 11

1.6 Thesis Organization 14

2 Related Work **15**

2.1 Introduction: The Landscape of LiDAR-based Autonomous Navigation 15

2.2 Dynamic Scene Understanding in Label-Constrained Environments 16

2.3 Navigating the Challenges of LiDAR Sparsity 19

2.4 Bridging the Gap: Differentiable SLAM and Deep Learning Integration 20

2.5 Adversarial Robustness in LiDAR-based Autonomous Systems 23

2.6 Conclusion: Identified Gaps and Future Research Directions 24

3 Segmentation and Navigation in Restricted Environments **27**

3.1 Existing Work 28

3.2 Solution . . . 32
3.3 Experiments & Results . . . 40
3.4 Results . . . 45
3.5 Ablation Studies . . . 54
3.6 Video Demo . . . 54
3.7 Conclusion . . . 55

4 Learning under Sparsity Constraints 57
4.1 Existing Work . . . 58
4.2 Solution . . . 59
4.3 Experiments . . . 68
4.4 Results . . . 73
4.5 Conclusion . . . 80

5 Learning using Differentiable SLAM 81
5.1 Related Work . . . 82
5.2 Contribution . . . 83
5.3 Experiments . . . 87
5.4 Results . . . 87
5.5 Conclusion . . . 89

6 Adversarial Point Injection Attacks on LiDAR Data 91
6.1 Related Work . . . 92
6.2 Solution . . . 93
6.3 Experiments . . . 99
6.4 Results . . . 102
6.5 Analysis of the results . . . 105
6.6 Conclusion . . . 107

7 Conclusion and Future Perspective 109
7.1 Summary of Contributions . . . 109
7.2 A Unified Thesis Narrative: Advancing Robust LiDAR Navigation . . . 110
7.3 Future Directions . . . 113

Publications 115

Bibliography **127**

Curriculum Vitae **128**

List of Tables

1.1 Key Challenges in LiDAR-Based Autonomous Navigation and Thesis Contributions 6

2.1 Comparative Analysis of LiDAR Generative Models for Autonomous Navigation 18

2.2 Challenges and Approaches in Integrating Deep Learning with SLAM 22

2.3 Taxonomy of Adversarial Attacks on LiDAR Perception and SLAM 24

3.1 Comparison of our model on static reconstructions with five baselines on three datasets. For all metrics, lower is better. A video demo of the reconstructions is available at Kumar (2023a). . . . 46

3.2 Comparison of SLAM Results in the absence of segmentation labels on four CARLA-64, one ARD, and ten KITTI Sequences. For all metrics, lower is better. As stated in Section 3.3.4.2 , we mainly focus on ATE as it is a global measure of trajectory accuracy and is more suited for long-term navigation, compared to RPE. 48

3.3 Variation observed in the KITTI sequences with respect to the number of loop closures and level of dynamism, associated with the number of dynamic objects and length of the sequence. 49

3.4 Detailed Comparison of MOVES (without segmentation) against segmentation baselines (KITTI-Seg) on KITTI sequences. As stated in Section 3.3.4.2, we mainly focus on ATE as it is a global measure of trajectory accuracy and is more suited for long-term navigation compared to RPE. . . 51

3.5 Detailed ablation studies to show the effect of the couple discriminator (CoD) and contrastive loss (CL) on **`MOVES`** when compared to best baseline on CARLA-64 test set. 53

3.6 Ablation studies to show the effect of the Least Squares Loss and the Hard Negative Mining (HNM) on **`MOVES`**. 53

3.7 We choose the positive scan as a LiDAR captured in a similar environment as the anchor instead of using an augmented version of the anchor. The ablation demonstrates that our approach works better. 54

3.8 Ablation studies to show the effect of the Domain Discrepancy Loss (DD) on **`MOVES-MMD`**. . . . 54

4.1 Static Augmentation Comparison of **`GLiDR`** against five baselines using five metrics for sparse and dense LiDAR. For all metrics, lower is better. 72

4.2 Navigation performance comparison of **`GLiDR`** against the best baseline in segmentation-devoid settings using ATE and RPE metrics in sparse and dense settings. As stated in Section 3.3.4.2 of Chapter 3, we mainly focus on ATE as it is a global measure of trajectory accuracy and is more suited for long-term navigation, compared to RPE. 74

4.3 `GLiDR` comparison against segmentation-assisted baseline on KITTI-Seg and against original dynamic LiDAR sequence based on the ATE and RPE metrics. 'Segmented Out' baseline numbers are not available for CARLA-64 and ARD-16. This is because these datasets do not have segmentation information available. As stated in Section 3.3.4.2 of Chapter 3, we mainly focus on ATE as it is a global measure of trajectory accuracy and is more suited for long-term navigation, compared to RPE. 74

4.4 GLiDR comparison against a segmentation-assisted baseline - KITTI-Seg and against original dynamic LiDAR sequence for eight-beam and four-beam LiDAR scans. Lower is better. `GLiDR` does not perform better for these sparsity levels. We focus on the ATE metric as it is more important of the two metrics for long-term navigation. 74

4.5 0-dim $\mathcal{PH}$ backbone results for three LiDAR scans. Every row demonstrates a LiDAR scan and its backbones at various sparsity levels. Column 1 is the original LiDAR scan. Columns 2, 3, and 4 represent backbones of 16-beam, 8-beam, and 4-beam sparse versions of the original scan, respectively. View every row independently from left to right. **Left**: Original 16-beam LiDAR scans. **Second from left**: Backbone for 16-beam LiDAR scan. **Third from left**: Backbone for 8-beam LiDAR scan. **Right**: Backbone for 4-beam LiDAR scan. While the backbone for a 16-beam LiDAR scan is precise and accurate, the backbone for an eight and four-beam LiDAR scan misses several details of the original LiDAR. Thus, `GLiDR` is unable to handle sparsity levels of 8 and 4-beam LiDAR scans. We use the top-view as it provides the most precise and clear view of the topology of the LiDAR scan. 75

4.6 0-dim $\mathcal{PH}$ backbone results for several LiDAR scans. Every row demonstrates a LiDAR scan and its backbones at various sparsity levels. Column 1 is the original LiDAR scan. Columns 2, 3, and 4 represent backbones of 16-beam, 8-beam, and 4-beam sparse versions of the original scan, respectively. View every row independently from left to right. **Left**: Original 16-beam LiDAR scans. **Second from left**: Backbone for 16-beam LiDAR scan. **Third from left**: Backbone for 8-beam LiDAR scan. **Right**: Backbone for 4-beam LiDAR scan. While the backbone for a 16-beam LiDAR scan is precise and accurate, the backbone for an eight and four-beam LiDAR scan misses several details of the original LiDAR. Thus, `GLiDR` is unable to handle sparsity levels of 8 and 4-beam LiDAR scans. We use the top-view as it provides the most precise and clear view of the topology of the LiDAR scan. 76

4.7 Ablation studies to study the effect of the LiDAR graph layers and the 0-dim $\mathcal{PH}$ constraint on `GLiDR`. 79

4.8 Ablation Result for ConvNets and `GLiDR` 79

5.1 Comparison of the Differentiable SLAM module on ground elevation estimation task and segmentation. For MSE, the lower the better; for the rest, the opposite. 88

5.2 Relative Comparison of Static Translation (using Chamfer Distance) for DSLR with and without Differentiable SLAM. 88

5.3 Relative Comparison of Static Translation for DSLR with and without Differentiable SLAM on CARLA-64 and ARD-16 dataset. We use Chamfer Distance metric. As stated in Section 3.3.4.2 of Chapter 3, we mainly focus on ATE as it is a global measure of trajectory accuracy and is more suited for long-term navigation, compared to RPE. 89

5.4 Comparison of a general-purpose generative autoencoder, AE_{lidar} with and without differentiable SLAM using Chamfer's Distance metric. 89

6.1 Comparison of our autoencoder backbone - AE_{seg} with widely used LiDAR backbone autoencoders. 102

6.2 Comparison of **SLACK** with adapted baselines for PiJ attacks. For LQI - lower is better. For DSR - the opposite. ARD-16 does not have segmentation masks available. We do not report DSR for ARD-16. Red indicates extremely bad values that cannot be used for PiJ - very high LQI or very low DSR (For a viable attack model, it is necessary that the attacked LiDAR scan has low LQI *and* high DSR. Working well on one metric while failing on the other makes the attack susceptible to failures. A high LQI would indicate that the attacked output is easily detectable as it deviates too much from the original LiDAR. A low DSR indicates that the model fails to insert new dynamism in the LiDAR scan.) 103

6.3 Comparison of PiJ attacks on SLAM. **SLACK** deteriorates SLAM more than the baselines while using the same number of injected points and maintaining better LIDAR quality. Note that the percentage of points injected by **SLACK** is dependent on the sequence and the model. As stated in Section 3.3.4.2 of Chapter 3, we mainly focus on ATE as it is a global measure of trajectory accuracy and is more suited for long-term navigation, compared to RPE. 104

6.4 Ablation studies to study the effect of segmentation-aware attention and contrastive losses for AE_{seg}. 106

6.5 Ablation studies to quantify the effect of segmentation aware LiDAR backbone - AE_{mask} on **SLACK**. For LQI - lower is better. For DSR - higher is better. 106

6.6 Ablation study to show the comparison between attack using a vanilla discriminator and **SLACK**. . 106

List of Figures

3.1 A hierarchical representation of scene components, categorizing them into stationary and non-stationary parts. 28

3.2 Block Diagram of our proposed model. The dynamic LiDAR scan from the sensor is mapped to a corresponding static LiDAR scan using **`MOVES`**. The new static structures in the reconstructed static scan, along with the original static structures, improve SLAM and segmentation. 32

3.3 **Top Row**: A corresponding dynamic-static LiDAR pair (D, S). Left - Dynamic scan, D, which has dynamic objects (in blue) and dynamic occlusions (indicated by arrows). Right - corresponding static scan, S in the same environment. **Bottom Row**: Our model in action. Given the dynamic scan D as input, our model generates a corresponding static scan $(\hat{S})$ (middle) that approximates S. We subtract D from $\hat{S}$ to extract moving and movable objects (blue boxes in the rightmost image). We use the top-view as it provides the most precise and clear view of all dynamic and static objects in the LiDAR scan. 33

3.4 Architecture Diagram for **`MOVES`**. We use LiDAR scan triplet (s_i, s_j, d_i) to create latent vector pairs - (r_{s_j}, r_{s_i}) and (r_{d_i}, r_{s_i}). The former is labeled as real, and the latter as fake. The triplet is also used to calculate a contrastive loss. Further, d_i is fed to the generator G to generate the corresponding static scan. We use the top-view as it provides the most precise and clear view of all dynamic and static objects in the LiDAR scan. 34

3.5 Formation of the LiDAR triplet for Contrastive Learning. Given a static scan s_i as the anchor, instead of taking an *augmented* version of s_i as the positive, we use another static scan, s_j, captured in a similar environment as positive. Further, for the negative, we take a semantically closer scan to anchor - d_i (corresponding dynamic to s_i) as the hard negative. This strategy improves the learning process, as shown in the Ablation Studies section. We use the top-view as it provides the most precise and clear view of all dynamic and static objects/structures in the LiDAR scan. 36

3.6 Architecture Diagram for **`MOVES-MMD`** 38

3.7 Dynamic Object Segmentation comparison on ARD-16. **Left**: A corresponding LiDAR pair. The top is dynamic (D), and the bottom is a static (S) scan. Note the presence of dynamic objects (in red) in the top dynamic scan (D). **Right**: Given dynamic scan, D as input, segmentation output generated by our model is on top. Segmentation output by the best baseline is at the bottom. **`MOVES`** precisely segments dynamic objects while the baseline is unable to capture them, which can be visualized in the Video demo Kumar (2023b). We use the top-view as it provides the most precise and clear view of all dynamic and static objects/structures in the LiDAR scan. 47

3.8 (Please view from Left to Right) **Row 1**: Top view of a contiguous series of 4 dynamic scans at time intervals *t, t+3, t+6, t+9* in a LiDAR sequence. Red denotes movable objects, while blue denotes moving objects. **Row 2**: When a dynamic scan in row 1 is given as input to **`MOVES`**, it generates a segmentation of movable objects (red) and moving objects (blue). Note that the moving object (in blue) changes position across the scans from left to right while the movable objects (in red) are currently stationary. Although stationary, these movable objects can be seen as moving in the video demo Kumar (2023c). We utilize the top-view because it clearly provides the most accurate view of moving and movable objects. 47

3.9 Comparison of the estimated trajectory using **`MOVES`** with the best baseline for 6 sequences. **Row 1**- sequences with multiple loop closures. **Row 2** - Sequences with minimal (0 or 1) loop closures. 49

3.10 SLAM error along the x, y, z coordinates for KITTI sequences. Blue denotes GT, Red denotes Ours, while Green denotes Best baseline. We observe that the best baseline consistently has high errors along the z-dimension compared to MOVES for most of the sequences. Refer to Table 6 for a quantitative analysis of the translation component of navigation error. 50

3.11 Navigation error along the rotational pose estimates (roll, pitch, yaw) for KITTI sequences. Blue denotes GT, Red denotes Ours while Green Denotes Best baseline. **`MOVES`** estimates the rotational poses closer to the ground truth consistently across the KITTI sequences. Refer to Table 6 for a quantitative analysis of the rotational component of navigation error. 52

4.1 Sparse LiDAR scans v/s dense LiDAR scans. 58

4.2 Overview of our solution to sparse LiDAR scans. SLAM results using **`GLiDR`** outperform the best baselines. Binary segmentation masks of dynamic and static points (blue denotes dynamic and red denotes static points) obtained using **`GLiDR`** are also better. We use the top-view as it provides the most precise and clear view of all dynamic and static objects/structures of the LiDAR scan. 60

4.3 **Row 1** denotes k-simplexes and simplicial complexes. These are generalizations of graphs to higher dimensions. **Row 2**: $\mathcal{PH}$ discovers the global shape of a dataset. **Row 3** denotes the progression of filtration on a point cloud. As the filtration progresses, new simplices are added to the simplicial complex. 61

4.4 Left to Right (a to c). Progression of filtration on a point cloud over different spatial resolutions as the distance threshold increases. Creation and destruction of k-dim topological features documented in (d), which is a persistence diagram representing (b_i, d_i) pairs - each point corresponds to a k-dim homology that takes birth at b_i and dies at d_iMoor et al. (2020). 62

4.5 A sample LiDAR graph layer. The input LiDAR is transformed into a graph using k-nearest neighbors (k=20). Graph convolution extracts neighborhood features followed by max-pool transform the LiDAR embedding to a higher dimensional space. 63

4.6 LiDAR Graph Generative Model using stacked LiDAR Graph layers 64

4.7 Demo representation of a LiDAR in the form of a graph. Recursive updates using k-nearest neighbors ensure global node features. 64

4.8 GLiDR network architecture with the generative LiDAR graph layers (upper branch). The lower branches constitute the 0-dim $\mathcal{PH}$ constraints and generate the topological loss. 65

4.9 LiDAR as a simplicial complex and the progress of filtration for the extraction of 0-dim Persistent Homology - a single connected component as a backbone. 66

4.10 **Left**: Input Dynamic LiDAR **Middle**: Visualization of the 0-dim $\mathcal{PH}$ based LiDAR backbone after some epochs. **right**: LiDAR backbone at the end as a single connected component. We use the top-view as it provides the most precise and clear view of the topology of the LiDAR scan. 67

4.11 A LiDAR scan is divided into eight sectors that translate to corresponding regions in the range image segmentation mask. 69

4.12 View in the clockwise direction from top right. We identify the source sector (highlighted in red in the LiDAR scan) and extract the segmentation mask (also in red) in the range image mask. The source sector mask is inserted in target sectors (0 and 3 - indicated by green) in the new mask. Overlaying the new mask over the LiDAR scan augments the LiDAR with dynamic objects (in the sectors indicated by the green.) 69

4.13 The corresponding pair pipeline results in the creation of a static-dynamic pair for KITTI. Static scan - The original scan with the dynamic object was removed using a segmentation mask. Dynamic scan - the output of the pipeline with the augmented new dynamic objects. 69

4.14 SLAM Trajectory comparison of **`GLiDR`** for sparse LiDAR scans against segmentation-assisted and segmentation-devoid baselines. 77

4.15 **Middle:** Segmentation mask generated by **`GLiDR`**. **Right:** Segmentation mask generated by the best baseline. Blue denotes dynamic points. Red denotes the static non-ground points used for SLAM. The best baseline fails to segregate the static and dynamic points. 78

5.1 Our proposed framework for integrating differentiable SLAM into deep learning task training. The framework includes a deep learning task that takes the input scans (X_i) and outputs the predicted scans ($\hat{Y_i}$). In the SLAM task, the trajectory error is calculated between trajectories x_i(calculated from predicted scans $\hat{Y_i}$) and x_{ref}(calculated from target scans Y_i). The framework aims to optimize the combined loss of both Deep Learning loss and SLAM (Absolute Trajectory Loss between trajectories) Loss. 84

5.2 Left: Setup of DSLR with Differentiable SLAM Loss. Right: Setup of GndNet with Differentiable SLAM Loss. 85

5.3 SLAM using **(1)** static output scans of DSLR (left) and **(2)** with static output from DSLR+Differentiable SLAM Error model (right). Red boxes indicate regions where the trajectory generated from our framework is better than the baseline. Differentiable SLAM helps to reduce the drift accumulated during navigation. 88

6.1 Contrastive Loss using Anchor, positive and corresponding dynamic hard negatives. We use the top-view as it provides the most precise and clear view of the dynamic and static structures of the LiDAR scan. 95

6.2 Hard Negative mining using corresponding static-dynamic pairs. Given an anchor (static), multiple hard negatives - corresponding dynamic and close-by scans are selected. We use the top-view as it provides the most precise and clear view of the all the static and dynamic structures of the LiDAR scan. 96

6.3 Segmentation Attention setup for Autoencoder backbone-AE_{mask} 97

6.4 **(a)** Adversarial Module tricks PD with an adversarial label. To classify the input pair as 1 - PD injects dynamism in s_j. **(b)** SLACK-MMD adapts CARLA-64 to KITTI. 97

6.5 Comparison of our Autoencoder module output *w.r.t* the best performing LiDAR autoencoder backbone. Column 1 represents the original KITTI scan. Column 2 represents the reconstruction using our segmentation-attention assisted autoencoder, which is coupled with contrastive learning. Column 3 represents the reconstruction using the best baseline. Our autoencoder is able to reconstruct precise static and dynamic structures. We use the top-view as it provides the most precise and clear view of the topology of the LiDAR scan. 102

6.6 Visual demonstration of an original scan v/s an attacked scan. It is very difficult to distinguish the attacked scan from the original scan. The top-view provides the most precise and clear view of all the dynamic and static structures of the LiDAR scan. Video demo. 103

6.7 **`SLACK`** attack comparison with baselines on KITTI. Dotted line: GT trajectory. Solid line: trajectory after the attack. 104

6.8 **Left:** SLAM results before attack using CARLA sequences - the map is precise. The navigation trajectory is accurate. **Right:** After adversarial attack with 0.8% points - map quality is degraded. High navigation error in estimated trajectory. 105