Cost-Effective Autonomous Navigation Using Synthetic Data in a Digital Twin Environment

Azizbek Reimbaev, Sheikh Sayed Bin Rahman, Minsoo Kim, Kanita Jerin Tanha, Taesoo Jun
Department of Software Engineering, Kumoh National Institute of Technology, Gumi 39177,South Korea
(the1darken1003,sksayed,kms991022,kanitajerin17, taesoo.jun)@kumoh.ac.kr

***Abstract*—The adoption of Autonomous Mobile Robots (AMRs) in Small and Medium-sized Enterprises (SMEs) is often constrained by the high costs and risks associated with real-world data collection and physical hardware testing. Modern deep learning-based navigation systems typically require large-scale labeled datasets, which are expensive and time-consuming to acquire in real environments. To address this challenge, this paper presents a cost-effective and accessible framework for autonomous navigation based entirely on synthetic data generated within a Digital Twin environment. Using NVIDIA Isaac Sim, we propose an end-to-end pipeline that enables data generation, preprocessing, model training, and closed-loop validation without physical hardware. A lightweight model is trained using balanced synthetic lidar data and deployed back into the simulator for autonomous navigation. Experimental results demonstrate a navigation accuracy of 92.19% and collision-free performance in complex corridor environments.**

Keywords: Autonomous Mobile Robots, Synthetic Data, Digital Twin, NVIDIA Isaac Sim, LiDAR Navigation, SMEs

I. Introduction

Industry 4.0 has accelerated the adoption of intelligent automation in logistics and warehouse environments to improve efficiency, flexibility, and operational resilience. Autonomous Mobile Robots (AMRs) are increasingly used for material handling, transportation, and intralogistics tasks. However, while large enterprises can invest in extensive real-world testing and advanced sensing systems, Small and Medium-sized Enterprises (SMEs) face significant financial, technical, and safety-related barriers to AMR deployment. A key challenge in learning-based autonomous navigation is the acquisition of large-scale, diverse, and labeled sensor datasets. Collecting such data using physical robots is costly, time-consuming, and potentially disruptive to warehouse operations, while also introducing safety risks and limiting exposure to rare or hazardous scenarios.

Digital Twin technology has emerged as a powerful paradigm for virtual prototyping, system analysis, and decision support in logistics and supply chain management. Recent studies highlight the potential of Digital Twins to support monitoring, evaluation, optimization, and control of logistics systems within Industry 4.0 frameworks [1]. However, most existing Digital Twin applications in logistics focus on high-level system modeling, asset monitoring, or operational optimization, with limited emphasis on closed-loop robot navigation and learning-based control at the sensor–action level.

LiDAR sensors play a central role in industrial mobile robotics due to their robustness, wide field of view, and independence from lighting conditions. Compared to vision-based navigation systems, LiDAR-based approaches offer a cost-effective and reliable sensing modality suitable for resource-constrained SME environments. Nevertheless, real-world LiDAR data collection remains expensive and geographically limited, and datasets are often unavailable due to operational sensitivity or security concerns. As a result, synthetic LiDAR data generation has been proposed as an alternative to support autonomous navigation and perception research [2].

Despite these advances, a clear gap remains in the development of a unified, fully digital workflow that integrates synthetic LiDAR data generation, navigation algorithm training, and closed-loop evaluation within a Digital Twin environment, without reliance on physical robot experiments. In particular, such an end-to-end pipeline is critically needed for SMEs, where reducing deployment cost, development time, and operational risk is essential. To address this gap, this paper proposes a fully digital, Digital Twin–driven framework for autonomous navigation based entirely on synthetic data. The main contributions of this work are summarized as follows:

- A fully digital, Digital Twin–driven framework that integrates synthetic data generation, navigation algorithm training, and closed-loop evaluation within a unified simulation environment, eliminating the need for physical data collection.
- A systematic and repeatable simulation-based validation methodology that enables reliable assessment of autonomous navigation performance under controlled scenarios, demonstrating that simulation-only development can achieve practical navigation without physical robot experiments.

II. System Architecture and Methodology

A. Digital Twin Environment

The proposed framework is implemented using a high-fidelity robotics simulation software built on the NVIDIA Omniverse. A differential-drive mobile robot equipped with a generic lidar sensor is deployed in a maze-like environment that resembles typical SME warehouse corridors.

B. Synthetic Data Generation

Synthetic data collection is performed using a human-in-the-loop approach. The robot is manually controlled through the

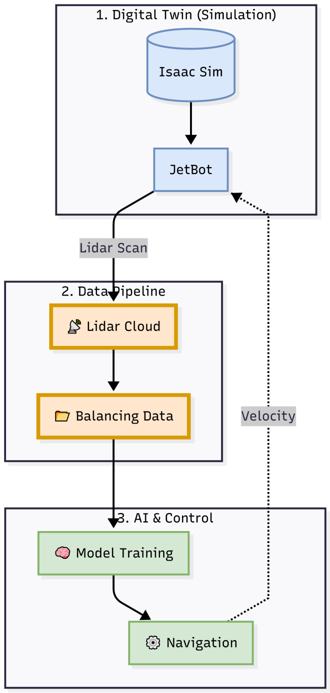

Fig. 1: Overview of the proposed synthetic data-driven navigation framework in a Digital Twin environment for autonomous navigation.

Digital Twin environment while logging synchronized sensor-action pairs. The 360-degree LiDAR point cloud is flattened into a one-dimensional vector of size $N = 200$. Each sample consists of a LiDAR observation and a corresponding discrete navigation command: Forward (F), Left (L), or Right (R).

C. Data Preprocessing and Balancing

Raw LiDAR data contain noise from floor reflections and self-occlusions. To improve data quality, distance and height-based filtering is applied. The filtered data are then standardized using a standard scaler to normalize feature distributions.

Navigation datasets often suffer from class imbalance, where forward motion dominates turning commands. To mitigate this issue, synthetic upsampling is applied to the minority classes, resulting in a balanced dataset where all navigation commands are equally represented.

D. Model Architecture

A lightweight Multi-Layer Perceptron (MLP) is designed to enable real-time inference on embedded platforms. The network architecture consists of an input layer with 200 neurons, followed by three hidden layers with 128, 64, and 32 neurons using ReLU activation functions. The output layer employs a softmax function to classify navigation commands.

III. Experimental Results

A. Training Performance

The MLP model was trained on the balanced synthetic dataset with early stopping and achieved a test accuracy of 92.19%. Forward motion was predicted reliably, while turning commands showed lower recall due to their inherent sparsity. The model nevertheless enabled stable closed-loop navigation in simulation.

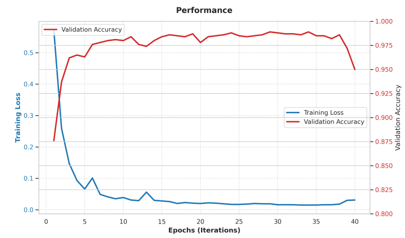

Fig. 2: Training loss and validation accuracy of the proposed model over training epochs.

B. Simulation-Based Validation

The trained model is deployed back into the Digital Twin environment for closed-loop navigation testing. The robot successfully navigates the maze without collisions. To address oscillatory behavior near decision boundaries, a temporal decision-lock mechanism of 0.5 seconds is introduced, ensuring smooth and stable navigation.

IV. Conclusion and Future Work

This paper presented an accessible and cost-effective framework for autonomous navigation tailored to SMEs. By leveraging NVIDIA Isaac Sim as a Digital Twin, we demonstrated that synthetic LiDAR data can be used to train a lightweight deep learning model capable of accurate and collision-free navigation. The proposed pipeline eliminates the need for physical data collection and lowers the entry barrier for autonomous robotics development.

Future work will focus on Sim-to-Real transfer by deploying the trained model on a physical robotic platform to evaluate real-world performance.

Acknowledgment

This research was funded by the Innovative Human Resource Development for Local Intellectualization Program (IITP-2025-RS-2020-II201612, 34%) through the Institute of Information & Communications Technology Planning & Evaluation (IITP) under the Ministry of Science and ICT (MSIT); the Basic Science Research Program (2018R1A6A1A03024003, 33%) through the National Research Foundation of Korea (NRF); and the ITRC Program (IITP-2025-RS-2024-00438430, 33%) funded by MSIT through IITP.

References

[1] Y. Liu, S. Pan, and E. Ballot, "Unveiling the potential of digital twins in logistics and supply chain management: Services, capabilities, and research opportunities," *Digital Engineering*, vol. 3, p. 100025, 11 2024.

[2] A. Phadke, S. Dipto, and P. Rana, "A workflow for generating synthetic lidar datasets in simulation environments," 06 2025.