

모바일 로봇을 위한 실시간 보행자 경로 예측: 트랜스포머 기반 모델과 디퓨전 기반 모델의 비교 연구

Real-Time Pedestrian Trajectory Prediction for Mobile Robots: A Comparative Study of Transformer-Based and Diffusion-Based Models

박 하 진¹ · 나 기 인[†]
Ha Jin Park¹, Ki-In Na[†]

Abstract: Predicting pedestrian trajectories is essential for safe mobile robot navigation in human-shared environments. Most existing models are trained on Bird's-Eye-View (BEV) data from fixed sensors, which differs from robot-centric observations affected by robot motion, changing coordinate frames, limited sensing range, and real-time constraints. This study evaluates trajectory prediction models in real mobile robot environments. We compare a transformer-based model for fast single-trajectory prediction and a diffusion-based model for multi-modal prediction using the JRDB robot-centric dataset. Both models were integrated into a ROS2-based pipeline and evaluated on JRDB data and real-world experiments using a mobile robot equipped with a camera, 3D LiDAR, and an IMU. The pipeline aligns pedestrian observations using odometry and predicts future trajectories under practical sensing conditions. It achieved end-to-end computation within 100 ms, while the diffusion-based model showed average displacement error (ADE) and final displacement error (FDE) comparable to recent studies. These results show the feasibility and limitations of applying modern prediction models to egocentric robot data for real-time forecasting.

Keywords: Pedestrian Trajectory Prediction, Socially-Aware Navigation, Mobile Robot, Deep Learning

1. 서 론

일반적인 모바일 로봇 주행 시스템은 주변 보행자의 이동의 도나 보행자 간 상호작용을 명시적으로 예측하기보다는, 현재 관측된 장애물의 위치와 속도 정보를 기반으로 짧은 주기마다 경로를 재계획하는 방식에 의존해왔다^[1]. 그러나 혼잡한 보행 환경에서는 보행자들이 각자의 목적지와 이동 의도를 가지며,

주변 보행자 및 환경 구조와 지속적으로 상호작용하면서 비선형적인 이동 패턴을 보인다. 이와 같은 상황에서 단순히 현재 시점의 관측 정보에 기반한 회피 전략만으로는 잠재적 충돌 위험을 조기에 인지하기 어렵고, 결과적으로 불필요한 정지, 우회, 급격한 경로 변경이 빈번하게 발생할 수 있다. 따라서, 모바일 로봇 주행에서 주변 보행자들의 미래 이동 정보에 대한 고려는 안전한 주행을 위해서 필수적이다. 더 나아가 정확한 미래 이동 정보 예측을 위해서 보행자의 단기적인 이동 경로뿐만 아니라, 이동 의도, 보행자 간 상호작용, 보행 공간의 구조를 통합적으로 고려할 필요가 있다.

이에 따라 최근에는 보행자 이동에 영향을 미치는 복합적인 인을 실환경 데이터로부터 학습하고, 이를 바탕으로 보행자의 미래 경로를 예측하는 딥러닝 기반 연구가 활발히 수행되고 있다. BEV (Bird's-Eye View) 기반 보행자 이동 데이터셋을 활용한 LSTM (Long Short-Term Memory)^[2], 트랜스포머^[3], 디퓨전

Received : May. 29. 2026; Revised : Jul. 31. 2026; Accepted : Aug. 11. 2026

※ This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (RS-2023-00215760, Guide Dog: Development of Navigation AI Technology of a Guidance Robot for the Visually Impaired Person).

1. Graduate Student, Mechanical Design Robot Engineering, SeoulTech, Seoul, Korea (phajin1001@seoultech.ac.kr)

† Assistant Professor, Corresponding author: Mechanical Systems Engineering, SeoulTech, Seoul, Korea (kina@seoultech.ac.kr)

모델^[4] 계열의 경로 예측 기법들은 보행자의 시간적 이동 패턴과 주변 보행자와의 상호작용을 모델링함으로써 높은 예측 성능을 달성해왔다. 또한 최근 연구들은 단일 미래 경로를 예측하는 결정론적 접근을 넘어, 다중양상 예측 표현을 통해 보행자 미래 이동의 불확실성을 모델링하는 방향으로 확장되고 있다. 그러나 이러한 연구들은 대체로 고정된 관측 시점에서 수집된 BEV 좌표계 기반 오프라인 데이터셋을 중심으로 학습 및 평가되어 왔다. 따라서 실제 모바일 로봇 적용에 필수적인 로봇의 이동에 따른 관측 시점 및 좌표계 변화, 제한된 센서 관측 범위와 같은 모바일 로봇 플랫폼에 대한 고려나 실시간 추론 제약의 통합적으로 고려한 검증은 여전히 충분히 이루어지지 않았다.

본 연구는 보행자 경로 예측 결과의 실시간 모바일 로봇 주행 적용을 검증하기 위해서 기존 보행자 경로 예측 모델을 이동 로봇의 센서 기반 관측 조건에서 실행 가능한 ROS2 경로 예측 모듈로 구성하고, 해당 모듈의 예측 정확도와 처리 시간을 평가하는 데 초점을 둔다. 본 연구의 평가는 보행자 미래 경로의 예측 정확도, 좌표 정렬의 효과 및 경로 예측 모듈의 처리 시간으로 한정하였으며, 예측 결과를 local planner 또는 로봇 제어기에 연결하는 페루프 주행 실험 및 내비게이션 성능에 대한 평가는 본 연구의 범위에 포함하지 않았다.

경로 예측에 대한 비교 연구를 위해서 예측 정확도와 실시간 추론 가능성을 기준으로 트랜스포머 기반 모델인 SPU-BERT^[5]와 디퓨전 기반 모델인 LED^[6]를 평가 대상으로 선정하였다. SPU-BERT는 보행자의 위치, 속도, 주변 환경 정보를 BERT^[7] 기반 인코더로 통합하여 미래 경로를 예측하는 모델로, 본 연구에서는 기존 SPU-BERT 모델과 더불어 실시간 단일 경로 예측을 위해 단일 BERT 구조만으로 미래 경로를 직접 산출하도록 단순화하여 사용하였다. LED는 다중양상 예측의 장점을 유지하면서 디퓨전 과정의 일부 단계를 생략하여 빠른 추론을 달성하도록 설계된 모델로, 디퓨전 기반 다중 경로 예측 모델의 로봇 적용 가능성을 평가하기 위한 비교 모델로 채택하였다.

본 연구에서는 SPU-BERT와 LED를 모바일 로봇 중심에서 주변 보행자 정보를 수집한 공개 데이터셋인 JRDB (JackRabbit Dataset and Benchmark)^[8]을 활용하여 학습 및 평가하였다. 또한 실제 모바일 로봇 주행 환경에 두 모델을 적용하여, 로봇의 이동에 따른 관측 시점 및 좌표계 변화, 제한된 센서 관측 범위, 실시간 추론 제약 하에서의 정확성과 실시간성을 분석하였다. 특히 추론 시간이 100 ms 이내로 유지될 수 있는지를 평가함으로써, 실제 로봇 시스템에서의 활용 가능성과 한계를 검토하였다.

2. 보행자 경로 예측 및 로봇 적용 연구

2.1 사회적 상호작용과 불확실성을 고려한 보행자 경로 예측

보행자 경로 예측에서는 보행자 간 상호작용과 미래 행동의 불확실성을 함께 고려하는 것이 중요하다. 초기 대표 연구인 Social-LSTM^[9]은 각 보행자의 이동을 LSTM으로 모델링하고, 주변 보행자의 은닉 상태를 통합하는 Social Pooling 구조를 제안함으로써 군중 내 사회적 상호작용을 반영하였다. 이후 Social-GAN^[10]은 LSTM 기반 예측 구조에 GAN (Generative Adversarial Network)을 결합하고 Variety Loss를 활용하여, 사회적으로 타당한 다중 미래 경로를 생성하도록 학습하였다. SPU-BERT (Socio-Physically Understandable BERT)^[5]는 트랜스포머 인코더 구조를 갖는 BERT 모델을 활용하여 보행자의 위치, 속도, 주변 환경 정보를 통합적으로 인코딩하고, 이를 바탕으로 미래 경로를 예측하도록 설계되었다. 또한 SPU-BERT는 두 개의 비순환 BERT 구조와 그 사이에 배치된 CVAE를 통해 다중 목표 후보를 예측하고, 각 목표에 도달하기 위한 다중 경로를 동시에 생성한다. 본 연구에서 검증 대상으로 활용한 트랜스포머 기반 경로 예측 모델은 이러한 SPU-BERT의 구조를 모티브로 한다.

보행자의 미래 행동은 본질적으로 여러 가능한 경로로 나타날 수 있으므로, 이를 확률적으로 모델링하려는 연구도 활발히 진행되어 왔다. MID (Stochastic Trajectory Prediction via Motion Indeterminacy Diffusion)^[11]는 DDPM (Denoising Diffusion Probabilistic Model)^[4]을 경로 예측 문제에 적용하여, 관측 경로로부터 다양한 미래 경로를 생성하고 보행자 움직임의 불확실성을 표현하였다. 이후 LED (Leapfrog Diffusion Model)^[6]는 디퓨전 기반 경로 예측의 높은 표현력과 다중양상 표현 능력을 유지하면서도 추론 속도를 향상시키기 위해 leapfrog initializer를 제안하였다. 해당 구조는 다수의 초기 노이즈 제거 단계를 대체하고, 소수의 노이즈 제거 단계만으로 예측 결과를 보정할 수 있도록 설계되었다. 따라서 LED는 생성 기반 경로 예측 모델 중에서도 실시간 적용 가능성을 고려한 대표적인 모델로 볼 수 있으며, 본 연구에서는 이를 디퓨전 기반 검증 모델로 선정하였다.

2.2 로봇 중심 데이터셋 및 보행자 경로 예측 연구

고정된 카메라에서 BEV (Bird's-Eye View) 좌표계로 수집된 기존 보행자 경로 예측 데이터셋과 달리, 실제 이동하는 모바일 로봇 중심에서 수집된 보행자 이동 경로는 보행자의 움직임뿐만 아니라 로봇의 이동에 의해서도 변화한다. 따라서 실제 모바일 로봇 환경에서 보행자 경로 예측을 수행하기 위해서는 로봇의 이동에 따른 관측 시점 및 좌표계 변화를 반영하

고, 이를 일관된 좌표계로 변환할 수 있는 센서 관측 정보와 로봇의 위치 및 자세 변화 정보가 필요하다. JRDB^[8]는 스탠포드 대학에서 개발한 JackRabbit 로봇을 이용해 360° RGB 카메라, LiDAR, RGB-D 센서 등 다양한 센서로 수집한 대규모 로봇 중심 데이터셋이다. JRDB-Traj^[12]는 이를 기반으로 구축된 보행자 경로 예측 벤치마크로, 로봇 관점에서 관측된 보행자 위치, 장면 이미지, 포인트 클라우드 정보를 활용하여 보행자의 미래 위치를 로봇 기준 좌표계에서 예측하도록 정의되어 있다. 또한 End-to-End Forecasting Error (EFE)와 같은 평가 지표를 통해 인식 및 추적 모듈의 불완전성이 경로 예측 성능에 미치는 영향을 분석할 수 있도록 구성되어 있다.

HST^[13]는 JRDB 데이터에서 보행자의 과거 위치와 함께 3D skeletal keypoints, head orientation 등의 시각적 인간 특성을 활용하여, 보행자의 의도 단서를 반영한 경로 예측을 수행하였다. 이러한 기존 연구들은 주로 공개 벤치마크에서의 예측 정확도 향상에 초점을 두어 왔으며, 실제 센서 기반 추적 결과를 이용한 온라인 입력 구성과 예측 모듈의 처리 지연을 함께 분석한 사례는 제한적이다. 따라서, 본 연구는 새로운 예측 모델을 제안하기보다, 이동 로봇의 시점 변화에 일관된 경로를 구성하는 처리 절차를 정의하고, 해당 절차가 기존 트랜스포머 및 디퓨전 기반 모델의 정확도와 실행 시간에 미치는 영향을 분석한다.

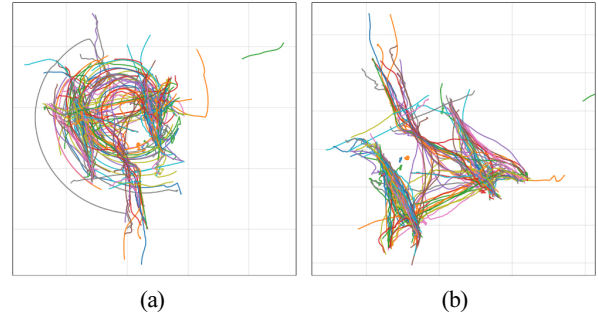
3. 실시간 경로 예측 파이프라인

3.1 데이터 전처리

본 절에서는 기존 보행자 경로 예측 모델을 Robot-centric sensing 환경에 적용하기 위한 학습 데이터 구성, 모델 입력 조건 및 ROS2 기반 온라인 추론 절차를 설명한다. 오프라인 학습 데이터와 실제 로봇의 온라인 입력에는 동일한 좌표 정렬, 관측 시퀀스 생성 및 유효 관측 처리 절차를 적용하여, 학습과 배치 단계 간 입력 표현의 일관성을 유지하였다.

로봇 중심 데이터셋의 객체 탐지 라벨은 각 시점의 로봇 좌표계를 기준으로 제공된다. 따라서 동일한 보행자의 위치를 시간 방향으로 단순히 누적하면, 로봇의 이동에 따른 좌표계 변화가 보행자의 경로에 반영되어 실제 보행자 이동과 무관한 위치 변화가 포함될 수 있다. 그 결과 [Fig. 1]과 같이 경로 왜곡이 발생하며, 이는 경로 예측 모델이 보행자의 고유한 이동 패턴이 아니라 로봇의 이동에 따른 좌표계 변화를 학습하게 만드는 원인이 될 수 있다.

이러한 문제를 완화하기 위해 본 연구에서는 동일한 ID를 갖는 보행자 탐지 라벨을 일정 시간 구간 동안 누적하고, 각 프레임의 로봇 오도메트리 정보를 이용하여 보행자 위치를 공통



[Fig. 1] Comparison of JRDB trajectories (a) before and (b) after odometry-based coordinate alignment

기준 좌표계로 정렬하였다. 경로 예측 모델의 입력에는 객체 탐지 결과에서 얻은 보행자 위치 중 지면 평면상의 x, y 성분만을 사용하였다. 현재 프레임을 $t=0$ 으로 설정하고, 시점 t 에서 로봇 좌표계 R 를 기준으로 관측된 보행자 i 의 2차원 위치를 $p_{i,t}^R \in \mathbb{R}^2$ 로 정의한다. 이때 오도메트리로부터 얻은 로봇의 위치 및 자세 정보를 이용해 로봇 좌표계 R 에서 공통 기준 좌표계 O 로의 좌표 변환 행렬 T_R^O 를 구성하면, 보행자 위치는 식 (1)을 통해 공통 기준 좌표계 O 로 변환된다.

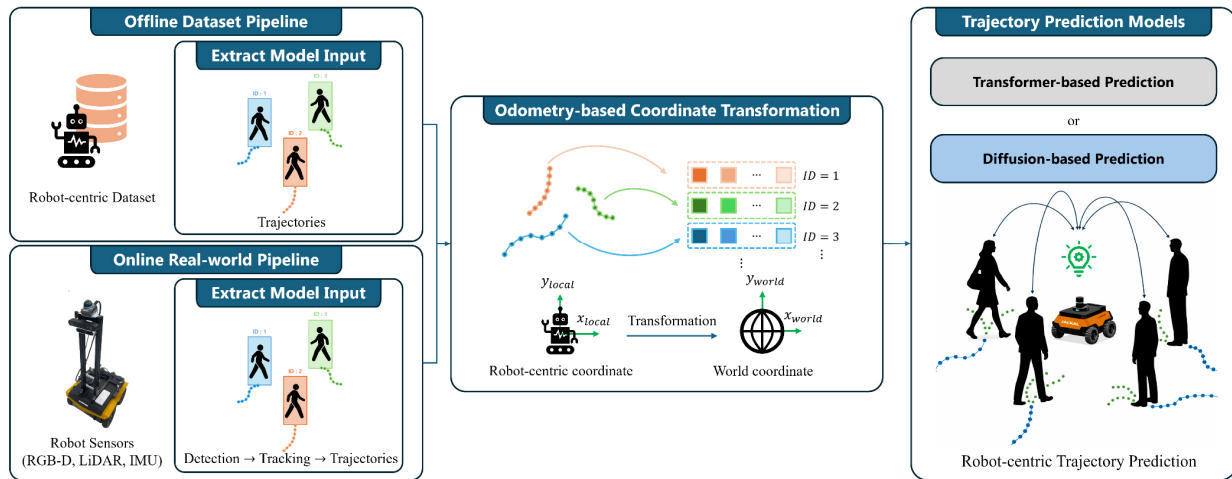
$$p_{i,t}^O = T_R^O p_{i,t}^R \quad (1)$$

O 는 시간적으로 일관된 공통 기준 좌표계로, 본 연구에서는 로봇 오도메트리 기반의 오도메트리 좌표계(odometry frame)를 기준 좌표계 O 로 정의하였다. 이러한 좌표 정렬 과정을 통해서도 다른 시점의 로봇 좌표계에서 관측된 보행자 위치를 동일한 기준 좌표계로 변환하고, 시간적으로 일관된 보행자 경로를 재구성하였다.

정렬된 보행자 경로는 일정 길이의 과거 관측 시퀀스와 미래 정답 시퀀스로 분리하여 학습 데이터로 구성하였다. 관측 프레임 누락이나 일시적 미검출이 발생한 경우 해당 위치를 유효하지 않은 값으로 처리하고, 유효 관측 여부를 나타내는 마스크를 함께 저장하였다. 동일한 좌표 정렬 및 마스크 생성 과정은 오프라인 학습 데이터 생성 단계와 ROS2 기반 실시간 추론 모듈에 공통적으로 적용되도록 구성하였다.

3.2 로봇 중심 데이터셋 기반 경로 예측 모델 적용 및 실시간 추론 모듈

본 연구는 새로운 경로 예측 모델을 제안하기보다, 기존 보행자 경로 예측 모델을 로봇 시점 데이터와 실제 모바일 로봇 환경에 적용하여 실시간 활용 가능성을 검증하는 데 초점을 둔다. 이를 위해 3.1절의 좌표 정렬 과정을 거친 JRDB 보행자 경



[Fig. 2] Overall pipeline for real-time trajectory prediction

로를 학습 데이터로 구성하였다. JRDB는 이동 로봇에서 수집된 데이터 셋이므로, 고정 카메라 시점이나 정제된 BEV 기반 데이터셋과 달리 로봇의 이동에 따른 관측 시점 및 좌표계 변화, 그리고 센서 관측 특성을 포함한다. 본 연구에서는 이러한 JRDB 기반 학습을 통해 기존 경로 예측 모델이 로봇 중심 관측 조건에서 주변 보행자의 미래 경로를 예측할 수 있는지 평가하였다.

학습 샘플은 동일한 ID를 갖는 보행자 경로를 시간 순서대로 누적하여 구성하였다. 각 샘플은 일정 길이의 과거 관측 시퀀스와 미래 예측 대상 시퀀스로 분리되며, 관측 프레임 누락이나 일시적 미검출이 발생한 경우 해당 위치를 유효하지 않은 값으로 처리하고 유효 관측 여부를 나타내는 마스크를 생성하였다. 이에 따라 모델 입력은 좌표 정렬이 완료된 보행자 위치와 각 시점의 유효 관측 정보를 함께 포함하도록 구성되었다.

경로 예측 모델로는 트랜스포머 기반 모델인 SPU-BERT^[5]와 디퓨전 기반 모델인 LED^[6]를 사용하였다. SPU-BERT는 attention 구조를 활용하여 주변 보행자와의 사회적 상호작용을 모델링하고, 대상 보행자의 미래 경로를 예측한다. 본 연구에서는 예측 표현 방식에 따른 차이를 분석하기 위해, SPU-BERT 기반 구조를 단일 경로 예측과 다중 경로 예측에 모두 활용하였다. 단일 경로 예측에서는 하나의 BERT 구조만으로 미래 경로를 직접 산출하도록 단순화하였으며, 다중 경로 예측에서는 기존 SPU-BERT의 다중 목표 및 다중 경로 생성 구조를 활용하였다. LED는 다중 양상 예측 능력을 유지하면서도 leapfrogging 방식을 통해 일부 노이즈 제거 단계를 생략하여 빠른 추론이 가능한 디퓨전 기반 모델로, 다중 경로 예측의 정확성과 실시간성을 평가하기 위한 비교 모델로 사용하였다. 두 모델 간 동일한 입력 조건에서 비교하기 위해, 본 연구에서는 주변 환경 정보를 제외하고 보행자 경로 정보만을 입력으로 사용하였다.

ROS2 기반 실시간 경로 예측 모듈은 [Fig. 2]와 같이 보행자 탐지 및 추적 결과와 로봇 오도메트리 정보를 입력으로 받아 주변 보행자의 미래 경로를 예측한다. 모듈은 추적 결과로부터 동일한 ID를 갖는 보행자의 지면 평면상 2차원 위치를 수신하고, 각 보행자에 대해 최근 10 프레임의 위치를 버퍼에 저장한다. 버퍼에 저장된 위치는 3.1절에서 설명한 오도메트리 기반 좌표 정렬을 통해 공통 기준 좌표계로 변환한 후, 경로 예측 모델의 과거 관측 경로로 사용한다. 모듈은 입력과 동일한 평면 좌표계에서 보행자의 2차원 미래 경로를 예측한다.

이후 정렬된 위치 $p_{i,t}^O$ 를 시간 순서대로 배열하여 과거 10개의 프레임에 대한 보행자 i 의 관측 시퀀스 $\mathbf{p}_i^O$ 를 구성한다. 이때 관측이 누락된 프레임의 위치 정보는 마스킹하여 입력하여 전달한다.

주변 보행자와의 상호작용을 고려하기 위해, 본 연구에서는 대상 보행자 i 의 과거 경로와 주변 보행자들의 과거 경로를 함께 모델 입력으로 사용하였다. 전처리된 보행자 경로 시퀀스는 트랜스포머 기반 모델 또는 디퓨전 기반 모델로 전달된다.

4. 실험

모든 실험은 Ubuntu 22.04 LTS와 ROS2 Humble 환경에서 수행하였다. 경로 예측 모델은 PyTorch로 구현하였으며, 학습 및 추론에는 NVIDIA GeForce RTX 5070 Ti를 사용하였다. 본 실험에서 사용한 트랜스포머 기반 모델은 4개의 self-attention layer, 256차원의 hidden representation, 4개의 attention head로 구성된 BERT 인코더를 기반으로 하였으며, feed-forward hidden dimension은 1024로 설정하였다. 예측 정확도는 예측 구간 전체에서 예측 위치와 실제 위치 간의 평균 변위 오차를 나타내는 ADE (Average Displacement Error)와 최종 예측 시점의 변위 오

[Table 1] Prediction accuracy by model type and preprocessing

Model	Preprocess	ADE (m)	FDE (m)
Transformer-based ^[5]	×	0.54	1.07
	○	0.34	0.74
Diffusion-based ^[6]	×	0.49	0.98
	○	0.33	0.70

[Table 2] Prediction accuracy and inference time by model type and number of predicted trajectories K . minADE $_K$ and minFDE $_K$ denote the minimum ADE and FDE, respectively, over K predicted trajectories. For $K = 1$, they are equivalent to the conventional ADE and FDE

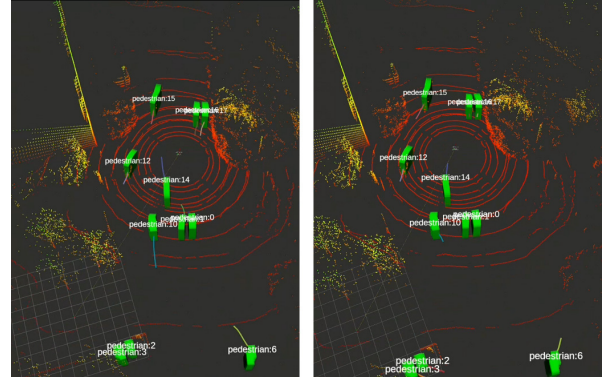
Model	K	minADE $_K$ (m)	minFDE $_K$ (m)	Inference Time (ms)
Transformer-based ^[5]	1	0.34	0.74	1.13
	20	0.15	0.26	4.78
Diffusion-based ^[6]	1	0.33	0.70	5.17
	20	0.16	0.31	9.23

차를 나타내는 FDE (Final Displacement Error)^[6]를 사용하여 평가하였다. 다중 경로 예측에서는 K 개의 후보 경로에 대해 계산된 ADE와 FDE의 최솟값을 각각 minADE $_K$ 와 minFDE $_K$ 로 정의하여 사용하였다. 실시간성은 실험 환경에 따라 JRDB에서는 10개의 경로 샘플 묶음에 대하여 모델의 평균 추론 시간을 측정하였고, 실제 주행 환경에서는 전체 ROS2 파이프라인의 처리 시간을 측정 범위로 지정하여 경로 예측 모듈의 입력 수신 이후 예측 경로 생성까지의 평균 추론 시간을 ms 단위로 측정하여 평가하였다. 본 연구에서는 로봇 주행 모듈에서 10 Hz 수준의 예측 갱신을 목표로 하여, 100 ms를 실시간 추론 기준으로 설정하였다.

JRDB 데이터는 15 Hz로 수집된 데이터 수집 특성을 반영하여 평가 시에는 30프레임의 과거 관측 구간과 60프레임의 미래 예측 구간을 구성한 뒤, 3프레임 간격으로 다운샘플링하였다. 이를 통해 최종적으로 2초 길이의 10프레임 관측 시퀀스와 4초 길이의 20프레임 예측 시퀀스를 구성하도록 하였다.

4.1 로봇 중심 좌표 정렬 유효성 검증

본 실험에서는 3.1절의 오도메트리 기반 좌표 정렬이 경로 예측 성능에 미치는 영향을 분석하였다. [Table 1]은 JRDB 객체 탐지 라벨을 로봇 기준 좌표계에서 그대로 누적한 경우와, 오도메트리 기반 좌표 정렬을 적용한 경우의 단일 경로 예측 결과이다. 좌표 정렬 이후 트랜스포머 기반 모델의 ADE/FDE는 0.54/1.07 m에서 0.34/0.74 m로 감소하였고, 디퓨전 기반 모델은



[Fig. 3] Trajectory prediction results on JRDB for the transformer-based model (left) and diffusion-based model (right)

0.49/0.98 m에서 0.33/0.70 m로 감소하였다. 이를 통해 이동 로봇의 ego-motion에 의해 발생하는 경로 왜곡을 오도메트리 정보를 통해 현재 로봇의 위치로 보정하는 과정이 모바일 로봇을 위한 보행자 경로 예측에 중요한 전처리 단계임을 확인하였다.

4.2 로봇 중심 데이터 기반 경로 예측 성능 및 추론 시간 분석

본 실험에서는 좌표 정렬이 적용된 JRDB 데이터에서 각 경로 예측 모델의 예측 정확도와 추론 시간을 동일한 입력 조건에서 [Table 2]에 비교하였다. 다중 경로 예측 모델의 경우, 20개의 서로 다른 예측 경로 중 정답과 가장 가까운 경로의 오차인 minADE $_{20}$ 와 minFDE $_{20}$ 를 평가지표로 결과를 비교하였다.

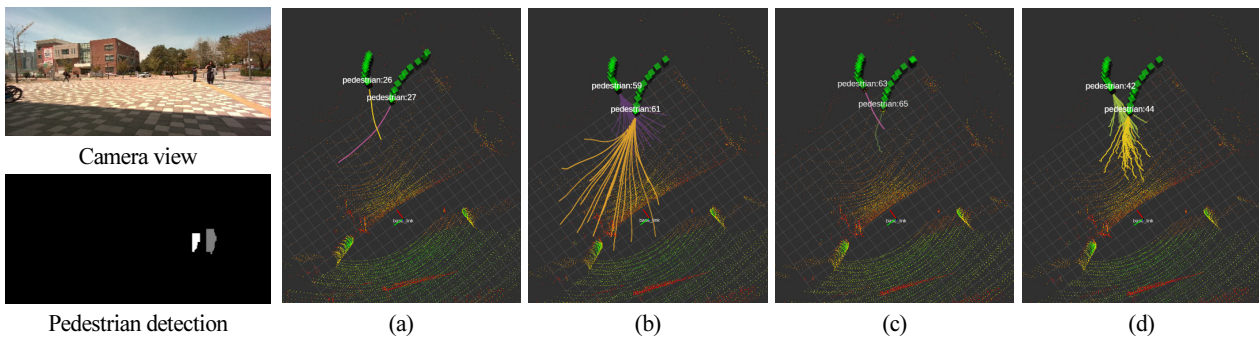
[Table 2]와 같이 단일 경로 예측 조건에서는 트랜스포머 기반 모델이 ADE/FDE 0.34/0.74 m와 1.13 ms의 추론 시간을 보였고, 디퓨전 기반 모델은 0.33/0.70 m와 5.17 ms의 추론 시간을 보였다. 즉, 단일 경로 예측에서는 두 모델 모두 동등한 수준의 오차를 보이지만 트랜스포머 기반 모델이 더 짧은 추론 시간을 달성했다. [Fig. 3]은 JRDB 환경에서 두 모델이 생성한 단일 보행자 미래 경로 예측 결과의 예를 나타낸다.

다중 경로 예측 조건에서는 트랜스포머 기반 모델이 minADE $_{20}$ /minFDE $_{20}$ 0.15/0.26 m와 4.78 ms의 추론 시간을 보였고, 디퓨전 기반 모델은 0.16/0.31 m와 9.23 ms의 추론 시간을 보였다. 두 모델은 모두 단일 경로 예측보다 정확도가 상승함을 알 수 있으며, 본 실험 조건에서는 트랜스포머 기반 모델이 minFDE $_{20}$ 에서 근소하게 낮은 오차와 짧은 추론 시간을 달성하였다.

단일 경로 예측과 다중 경로 예측 결과를 종합하면, 비교 모델 모두 100 ms 이하의 추론 시간을 보여 실시간 적용이 가능함을 확인하였다. 또한 보행자의 이동 특성에 따른 성능 차이를 분석하기 위해 TrajNet++^[14]의 분류 기준에 따라 예측 경로의 시작점과 종점 간 변위가 1 m 이상인 샘플을 이동 보행자로 분

[Table 3] Prediction accuracy by model type and pedestrian motion category for predicted trajectories. Stationary and Moving denote pedestrians classified according to their motion state, while Fast Moving denotes moving pedestrians with a ground-truth displacement greater than 6 m over the prediction horizon. Bold values indicate the lowest error in each metric column

Model	K	$\min ADE_K$ (m) (Stationary)	$\min FDE_K$ (m) (Stationary)	$\min ADE_K$ (m) (Moving)	$\min FDE_K$ (m) (Moving)	$\min FDE_K$ (m) (Fast Moving)
Transformer-based ^[5]	1	0.0918	0.1559	0.5431	1.1837	1.9311
	20	0.0523	0.0722	0.2294	0.4175	0.6702
Diffusion-based ^[6]	1	0.0898	0.1576	0.5238	1.1389	1.7056
	20	0.0508	0.0751	0.2543	0.5115	1.1497



[Fig. 4] Real-world trajectory predictions by model type and number of predicted trajectories: (a) transformer-based deterministic prediction, (b) transformer-based multi-modal prediction, (c) diffusion-based deterministic prediction, and (d) diffusion-based multi-modal prediction

류하고, 그 미만을 정지 보행자로 분류하였다. 이동 보행자로 분류된 샘플 중 변위 6 m 이상인 샘플은 빠른 이동 보행자로 별도 구분하여 단일 경로 예측과 다중 경로 예측 조건에서 트랜스포머 기반 모델과 디퓨전 기반 모델의 $\min ADE_K$, $\min FDE_K$ 를 비교한 결과를 분석하였다.

[Table 3]에 따르면 정지 보행자에서는 두 기반 모델 간 뚜렷한 차이가 없었으나, 이동 및 빠른 이동 보행자에서는 트랜스포머 기반 다중 경로 예측 모델이 전반적으로 우수하였으며, 특히 $\min FDE_K$ 에서 우수한 성능을 보였다. 이는 다중 경로 예측 비교 모델로 사용한 트랜스포머 기반 SPU-BERT가 MGP (Multi-Goal Prediction) 모듈을 통해 가능한 중점 분포를 명시적으로 학습함으로써, 예측 후반부의 FDE를 낮추고 이동 거리가 큰 조건에서도 강건한 성능을 유지한 것으로 해석된다. 반면 다중 경로 예측 비교 모델로 사용한 디퓨전 기반 LED는 원 논문의 예측 step별 loss weighting을 그대로 적용하여 초기 step에 높은 가중치가 부여되며, 이에 따라 근거리 예측에는 강점을 보이나 $K=20$ 과 같이 다양한 중점 후보를 생성해야 하는 조건에서는 horizon이 길어질수록 오차 개선폭이 제한되는 경향을 보였다.

이러한 결과는 실제 로봇 시스템에 예측 모델을 적용할 때 추론 시간과 전반적 정확도뿐만 아니라, 단일 및 다중 경로 방식 및 보행자의 이동 특성에 따른 모델 별 성능 차이를 함께 고려해야 함을 시사한다.

[Table 4] Processing time of the ROS2 trajectory prediction pipeline in the real-world environment by model type and number of predicted trajectories K

Model	K	Processing time (ms)
Transformer-based ^[5]	1	3.554
	20	9.086
Diffusion-based ^[6]	1	9.274
	20	10.303

4.3 실제 로봇 환경에서의 ROS2 실시간성 평가

실제 환경 실험에서는 Clearpath Jackal 모바일 로봇에 플랫폼에 Livox MID360 라이다와 Realsense D455 스테레오 카메라를 장착하여, 카메라-LiDAR 센서 융합을 이용한 보행자 객체 탐지 및 추적 ROS2 모듈을 구축하였다. 해당 모듈에서 출력되는 보행자의 지면 평면상 2차원 위치를 경로 예측 모듈의 입력으로 사용하여, 캠퍼스 환경에서 로봇 주행 중 주변 보행자의 미래 경로를 예측하고 추론 시간을 측정하였다.

[Table 4]와 같이 실제 환경에서 실험에 사용한 모델 모두 추론 시간 100 ms 이하의 실시간성을 달성하였으며 단일 경로 예측의 실시간성이 다중 경로 예측보다 뛰어난 것을 알 수 있다. 그러나 [Fig. 4]와 같이 다중 경로 예측에서 보행자의 미래 움직임 불확실성을 보다 우수하게 표현하는 것을 확인할 수 있다.

5. 결론 및 고찰

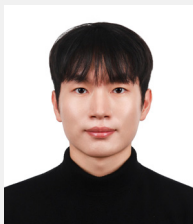
본 연구는 기존 보행자 경로 예측 모델을 로봇 시점 데이터와 실제 모바일 로봇 환경에 적용하기 위한 실시간 경로 예측 파이프라인을 구성하였다. 이를 위해 트랜스포머 기반 모델과 디퓨전 기반 모델을 JRDB 로봇 시점 보행자 경로 데이터로 학습하고, 동일한 입력 조건에서 예측 정확도와 추론 시간을 비교하였다. 실험 결과, 트랜스포머 기반 모델은 단일 경로 예측에서 짧은 추론 시간을 보였으며, 디퓨전 기반 모델은 예측 샘플 수를 증가시킬 때 최근접 예측 샘플의 ADE/FDE가 감소하는 경향을 보였다. 또한 두 모델을 ROS2 기반 실시간 추론 모듈에 통합하여 실제 모바일 로봇 환경에서 추론 시간을 측정할 결과, 설정한 100 ms 기준 이내에서 예측 결과를 생성할 수 있음을 확인하였다.

다만 본 연구는 경로 예측 모듈의 예측 정확도와 추론 시간 평가에 초점을 두었으며, 예측 결과를 실제 로봇 제어기나 local planner에 폐회로 형태로 연결한 주행 성능 검증은 수행하지 않았다. 따라서 예측된 보행자 미래 경로가 실제 로봇의 충돌 회피, 경로 효율성, 사회적 수용성 향상에 어느 정도 기여하는지는 추가적인 로봇 내비게이션 영역에서의 실험과 평가가 필요하다. 또한 실제 센서 환경에서 발생할 수 있는 객체 가림 혹은 추적 ID 끊김과 같은 탐지 및 추적 오류에 대한 강건성 분석과, 다중 예측 경로의 다양성 및 사회적 타당성에 대한 평가는 제한적으로 이루어졌다.

향후 연구에서는 본 연구에서 구성한 실시간 경로 예측 모듈을 로봇의 local planner와 연동하여 행동 양상 인지 및 예측 내비게이션 시스템으로 확장할 계획이다. 이를 통해 예측된 보행자 경로를 로봇의 속도 제어, 충돌 회피, 경로 최신화 과정에 반영하고, 실제 혼잡 환경에서의 주행 안전성, 경로 효율성, 보행자와의 사회적 거리 유지 등을 평가하고자 한다.

References

- [1] E. Jung, W. Song, and H. Myung, "High-Definition Map-based Local Path Planning for Dynamic and Static Obstacle Avoidance," *The Journal of Korea Robotics Society*, vol. 16, no. 2, pp. 112-121, 2021, DOI: 10.7746/jkros.2021.16.2.112.
- [2] S. Hochreiter and J. Schmidhuber, "Long Short-Term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997, DOI: 10.1162/neco.1997.9.8.1735.
- [3] A. Vaswani et al., "Attention is all you need," *arXiv:1706.03762*, 2017, DOI: 10.48550/arXiv.1706.03762.
- [4] J. Ho, A. Jain, and P. Abbeel, "Denosing Diffusion Probabilistic Models," *arXiv:2006.11239*, 2021, DOI: 10.48550/arXiv.2006.11239.
- [5] K.-I. Na, U.-H. Kim, and J.-H. Kim, "SPU-BERT: Faster human multi-trajectory prediction from socio-physical understanding of BERT," *Knowledge-Based Systems*, vol. 274, Art. no. 110637, 2023, DOI: 10.1016/j.knosys.2023.110637.
- [6] W. Mao, C. Xu, Q. Zhu, S. Chen, and Y. Wang, "Leapfrog Diffusion Model for Stochastic Trajectory Prediction," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, pp. 5517-5526, 2023, DOI: 10.1109/CVPR52729.2023.00534.
- [7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 1. Minneapolis, MA, USA, pp. 4171-4186, 2019, DOI: 10.18653/v1/N19-1423.
- [8] R. Martín-Martín et al., "JRDB: A Dataset and Benchmark of Egocentric Robot Visual Perception of Humans in Built Environments," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 6, pp. 6748-6765, 2023, DOI: 10.1109/TPAMI.2021.3070543.
- [9] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human Trajectory Prediction in Crowded Spaces," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 961-971, 2016, DOI: 10.1109/CVPR.2016.110.
- [10] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, pp. 2255-2264, 2018, DOI: 10.1109/CVPR.2018.00240.
- [11] T. Gu et al., "Stochastic trajectory prediction via motion indeterminacy diffusion," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, pp. 17092-17101, 2022, DOI: 10.1109/CVPR52688.2022.01660.
- [12] S. Saadatnejad, Y. Gao, H. Rezatofighi, and A. Alahi, "JRDB-Traj: A dataset and benchmark for trajectory forecasting in crowds," *arXiv:2311.02736*, 2023, DOI: 10.48550/arXiv.2311.02736.
- [13] T. Salzmann, H.-T. L. Chiang, M. Ryll, D. Sadigh, C. Parada, and A. Bewley, "Robots That Can See: Leveraging Human Pose for Trajectory Prediction," in *IEEE Robotics and Automation Letters*, vol. 8, no. 11, pp. 7090-7097, 2023, DOI: 10.1109/LRA.2023.3312035.
- [14] P. Kothari, S. Kreiss, and A. Alahi, "Human Trajectory Forecasting in Crowds: A Deep Learning Perspective," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 7386-7400, 2022, DOI: 10.1109/TITS.2021.3069362.



박 하 진

2026 서울과학기술대학교 기계시스템디자인
공학과(공학사)

2026~현재 서울과학기술대학교 기계설계로봇
공학과 석사 과정

관심분야: Mobile Robot, Trajectory Prediction, Socially-Aware Navigation



나 기 인

2009 포항공과대학교 기계공학과(공학사)

2011 한국과학기술원 로봇공학학제전공
(공학석사)

2022 한국과학기술원 로봇공학학제전공
(공학박사)

2011~2024 한국전자통신연구원 책임연구원

2024~현재 서울과학기술대학교 조교수

관심분야: Socially-Aware Navigation, VLA Model, Robotic Manipulation