

Coarse-to-Fine Multi-Resolution Diffusion Models for Trajectory Generation in Urban Systems

Wen Ye*
yewen@usc.edu
University of Southern California
Los Angeles, CA, USA

Muyan Weng*
muyanwen@usc.edu
University of Southern California
Los Angeles, CA, USA

Chuizheng Meng
chuizhem@usc.edu
University of Southern California
Los Angeles, CA, USA

Hao Niu
ha-niu@kddi.com
KDDI Research, Inc.
Saitama, Japan

Yizhou Zhang
zhangyiz@usc.edu
University of Southern California
Los Angeles, CA, USA

Yan Liu
yanliu.cs@usc.edu
University of Southern California
Los Angeles, CA, USA

Abstract

Understanding human mobility is critical for a wide range of urban applications, including traffic management, epidemic control, and urban planning. However, due to privacy concerns, the availability of large-scale public trajectory data remains limited, posing challenges for downstream mobility analysis. Existing methods for synthetic trajectory generation primarily focus on matching global distribution similarity, while often overlooking mobility patterns across different spatial and temporal resolutions that are essential for practical utility.

To address these challenges, we propose a novel multi-resolution diffusion framework, MR-Traj, for large-scale trajectory generation. MR-Traj explicitly models trajectories as compositions of coarse-grained milestones and fine-grained segments, enabling the capture of complex spatial-temporal dependencies at multiple resolutions. Experimental results demonstrate that MR-Traj achieves comparable performance to state-of-the-art methods in terms of global distribution similarity, while consistently outperforming them in modeling fine-resolution mobility patterns and supporting downstream urban mobility tasks. In addition, by introducing stochasticity at multiple resolution levels, MR-Traj generates more diverse trajectories, which empirically reduces trajectory linkage risk under a seed-guided data release setting.

CCS Concepts

• **Information systems** → **Geographic information systems**; **Location based services**; • **Computing methodologies** → **Neural networks**; **Simulation evaluation**.

Keywords

Spatial-Temporal Systems, Synthetic Data Generation, Human Mobility, Denoising Diffusion Probabilistic Model

ACM Reference Format:

Wen Ye, Muyan Weng, Chuizheng Meng, Hao Niu, Yizhou Zhang, and Yan Liu. 2026. Coarse-to-Fine Multi-Resolution Diffusion Models for Trajectory Generation in Urban Systems. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '26)*, August 09–13, 2026, Jeju Island, Republic of Korea. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3770855.3818889>

1 Introduction

Human mobility data provides critical insights into population-level movement patterns and supports a wide range of applications, including urban planning [7, 51], traffic estimation [19, 25], epidemic control [24], and location-based services [57]. In urban and transportation systems, such data plays a central role in city-scale decision making and mobility analysis. However, access to large-scale mobility data is often restricted in practice due to privacy and regulatory constraints, particularly for organizations that collect extensive trajectory records but cannot publicly release them without risking sensitive information disclosure. As a result, synthetic trajectory generation has emerged as a practical alternative, enabling data sharing while avoiding the direct exposure of individual-level records.

Early approaches to synthetic trajectory generation primarily relied on Markov Chain models. With advances in deep learning, generative models such as Variational Autoencoders (VAEs) [15] and Generative Adversarial Networks (GANs) [12] were subsequently explored, often combined with recurrent architectures such as Long Short-Term Memory networks (LSTMs) to capture temporal dependencies [38, 42, 56]. Despite their success, many existing methods depend on strong prior assumptions or handcrafted semantic features, such as home/work locations or points of interest [16, 28], and frequently struggle to reproduce fine-grained mobility patterns at high spatial resolution [35, 49]. This limitation is particularly important in urban-scale settings, where both city-wide structure and neighborhood-level movement details are necessary for reliable analysis.

More recently, diffusion models have emerged as a promising alternative, offering stable training dynamics and flexible conditioning mechanisms for controlled trajectory generation [53, 59]. Nevertheless, diffusion-based trajectory generation is typically guided by metadata from original trajectories, which may still introduce

*Both authors contributed equally to this research.



privacy risks when synthetic data is released. At the same time, accurately modeling urban mobility requires capturing patterns across multiple spatial and temporal resolutions, where global movement patterns describe long-range structure while local patterns reflect fine-grained behaviors within specific regions.

Another limitation of existing methods lies in their predominant focus on minimizing global distribution discrepancies between real and synthetic trajectories, with relatively limited consideration of downstream analytical utility and responsible data release [4, 11, 16, 22]. In practice, many mobility analysis tasks operate on partial trajectories or localized segments, making fidelity at finer resolutions equally important and motivating the need for multi-resolution modeling in which global structure and local movement are jointly represented. Hierarchical models that decompose complex processes into multiple probabilistically connected levels provide a principled framework for describing stochastic spatio-temporal phenomena [1, 8].

In parallel, recent work has explored large-scale pre-trained trajectory models that learn transferable mobility representations from massive unlabeled data, enabling improved generalization across regions and supporting downstream tasks such as prediction, completion, and generation [5, 6, 17, 50, 60].

Building on these observations, recent studies have shown that integrating diffusion models into multi-resolution or coarse-to-fine frameworks can effectively capture structured dependencies in complex data [36, 37]. For example, HGHOI generates high-level milestones before modeling detailed motions in human-object interaction [36], while HierDiff adopts a hierarchical diffusion strategy for molecular generation [37]. In the context of urban mobility, separately modeling abstract structural information (e.g., milestones) and detailed trajectory segments provides a principled way to capture multi-scale movement patterns while increasing diversity in the generated data.

Motivated by these insights, we propose a **Multi-Resolution** framework for modeling complex spatio-temporal dependencies in mobility **Trajectories** using diffusion models (MR-Traj). We focus on a two-level resolution setting, where a milestone-level diffusion model captures global trajectory structure and a segment-level diffusion model generates fine-grained movements conditioned on synthesized milestones. This formulation enables MR-Traj to preserve both large-scale mobility structure and local movement details, while reducing the risk of overly deterministic trajectory reproduction under metadata-guided generation.

In summary, our contributions are threefold:

- We propose a multi-resolution diffusion framework for synthetic urban trajectory generation, jointly modeling global milestone trajectories and local trajectory segments under metadata guidance.
- We conduct extensive experiments on large-scale real-world mobility datasets, showing that our method achieves performance comparable to state-of-the-art approaches in global distribution similarity while consistently improving local distribution fidelity.

- We demonstrate that MR-Traj supports downstream urban mobility analysis tasks, including taxi destination prediction and anomalous mobility pattern identification, while exhibiting lower empirical linkage risk than existing metadata-guided baselines.

2 Related Work

Early work on synthetic trajectory generation primarily modeled trajectories as sequences of locations governed by Markovian assumptions, where each location depends only on the previous one [33, 44]. Extensions such as Input-Output Hidden Markov Models relax uniform transition assumptions by incorporating contextual information into latent processes [54]. Despite their interpretability, Markov-based approaches rely on autoregressive sampling, which limits scalability and prevents efficient parallel generation of long trajectories.

With the rise of deep learning, generative models such as Variational Autoencoders (VAEs) [15] and Generative Adversarial Networks (GANs) [12] have been widely explored for trajectory synthesis. Recurrent architectures, including Long Short-Term Memory networks (LSTMs), are often combined with GANs to capture temporal dependencies [38, 42, 56]. Representative methods include TrajGANs [26], which first demonstrated GAN-based trajectory generation, and subsequent extensions that incorporate urban structure priors or multi-stage generation pipelines to improve realism at different spatial resolutions [16, 49]. Reinforcement learning formulations, such as SeqGAN, further model trajectory generation as a sequential decision process [55]. However, GAN-based models are often difficult to train, computationally expensive at scale, and offer limited controllability over generated trajectories, making them less suitable for large-scale urban mobility data.

VAEs have also been adapted to trajectory generation by introducing hierarchical latent structures. For example, VOLUNTEER employs a two-layer VAE that first infers user-level latent attributes and then generates trajectories conditioned on these attributes [28]. While VAEs provide a principled probabilistic framework, they often produce overly smooth trajectories and rely on restrictive distributional assumptions, which can limit their ability to capture complex and fine-grained mobility patterns.

Recently, diffusion models have emerged as a promising alternative for trajectory generation, offering stable training dynamics and flexible conditioning mechanisms [9, 10, 48]. Diffusion-based approaches have demonstrated strong performance in modeling complex, high-dimensional structured data, including spatio-temporal trajectories [52]. In particular, DiffTraj shows that diffusion probabilistic models can generate synthetic trajectories that closely match real mobility distributions under metadata guidance [59].

Beyond single-resolution generation, a growing body of work explores hierarchical and multi-resolution modeling strategies to capture the multi-scale nature of mobility data. Coarse-to-fine diffusion frameworks decompose trajectory generation into global planning and local refinement stages, enabling better modeling of long-range structure and local movement details [20]. Related

approaches combine diffusion with structured planning or multi-scale representations to model trajectories alongside other spatio-temporal signals, such as traffic flows [27, 29]. Recent studies further investigate controllable diffusion for structured or categorical trajectory data and emphasize evaluation through downstream urban analysis tasks, rather than relying solely on distribution-level similarity [14, 18, 23, 45].

Our work builds on these advances by explicitly incorporating the hierarchical structure of spatio-temporal mobility data into a multi-resolution diffusion framework. By modeling global trajectory structure and local trajectory segments separately, our approach aims to better capture multi-scale dependencies inherent in real-world urban mobility, improving both generation fidelity and downstream analytical utility.

3 Preliminaries

3.1 Problem Definition

A Trajectory that is sampled at regular time intervals is defined as $S = [s_1, \dots, s_n]$ where $s_i = (lat_i, lon_i)$ represents the location of the vehicle at timestamp i when s is on a continuous spatial scale; $s_i = gid_i$ represents the grid id that the taxi is in at timestamp i on a discretized spatial scale; n denotes the total number of points in the trajectory. This dual representation caters to different needs in spatial data processing, with latitude and longitude providing precise locations and grid ids facilitating spatial analysis on a coarser scale and evaluation concerned with distribution similarity.

The objective of Synthetic Trajectory Generation is to generate synthetic trajectories $S' = [s'_1, \dots, s'_n]$ where S' preserves the spatial-temporal characteristics and patterns observed in S . These synthetic trajectories should be highly useful for downstream spatial-temporal analysis, meaning they should mimic real-world mobility patterns and can be used to support tasks such as traffic management, and mobility behavior analysis. In general, many downstream tasks of synthetic trajectory generation involves predicting some properties of a trip, denoted as $g(S_u, S_o)$ such that the downstream model uses only a segment of the full trajectory where S_o denote the observed portion of a trajectory, while S_u denote the unobserved portion to be inferred or completed.

3.2 Theoretical Motivation

THEOREM 1. *If the target function $g(S_u, S_o)$ is L -Lipschitz and D represents the diameter of the support set of S_u , the empirical risk R of a predictor trained on synthetic data $P'(S)$ and evaluated on real data $P(S)$ is bounded as:*

$$R \leq LD^2 \|P(S_u, S_o) - P'(S_u, S_o)\|_{TV} + LD^2 \|P'(S_o) - P(S_o)\|_{TV} + C$$

where: $\|\cdot\|_{TV}$ denotes the Total Variation (TV) distance, and C is a constant.

The upper bound derived in theorem 1 indicates that minimizing the empirical risk of a downstream model requires reducing the TV distances between both the joint and marginal distributions of the synthetic and real data. While global similarity between the real and synthetic distributions $LD^2 \|P(S_u, S_o) - P'(S_u, S_o)\|_{TV}$ is often emphasized in existing methods, our theoretical results suggest

that minimizing this distance, though essential, is not sufficient for fully optimizing downstream performance. Existing methods which focus solely on reducing the global distance between P and P' may not maximize utility. The theoretical result indicates that utility is influenced not only by global distributional similarity but also by local trajectory patterns where the observed fraction of trajectories could be any segment of the complete trajectory. This motivates us to employ a multi-resolution framework to model different resolutions of a trajectory and hence modeling both complete trajectories and different resolutions of the trajectory such as milestones and segments. The details of this proof can be found in the supplementary material where we define the optimal predictor given the objective and empirical risk; leverage the Lipschitz Condition [3] and bound the Wasserstein Distance [47] between two distributions; apply triangle inequality on the Total Variance Distance [13] bound.

We further motivate our design choice from the privacy consideration perspective. Modeling trajectories in a multi-resolution fashion allows us to obtain increased diversity of generated trajectories through multi-level noise injection. By introducing noise at various resolution of the synthetic trajectory generation process, the model effectively creates a broader range of possible synthetic outcomes.

4 Methodology

In this section, we formally introduce the multi-resolution framework inspired by the motivating analysis. We employ a coarse-to-fine strategy to generate synthetic trajectories. The multi-resolution framework is composed of two models: a milestone-level diffusion model and a segment-level diffusion model. The classic Origin-Destination Matrix can be seen as the most abstract level of resolution of a trajectory; milestones then is at a finer resolution and segments provide the finest resolution. Both models leverage the U-Net architecture as their backbone but are conditioned on different levels of trajectory resolution to focus on distinct aspects of trajectory generation. Figure 1 shows the overall structure of the multi-resolution diffusion model framework. The milestone-level model is designed to learn the broader trajectory distribution, capturing high level distributions of trajectories, i.e. general direction of moving. Conversely, the segment-level model is tasked with detailing the finer aspects of the trajectory, filling in the gaps between milestones with realistic local movement patterns.

Our choice of backbone for both diffusion models is the widely acclaimed U-Net structure [40]. We adopt this structure along with a condition embedding module implemented as a straightforward feedforward neural network from prior work [59]. The condition module embeds trajectory attributes like departure time, distance traveled, start and end locations, enabling the diffusion model to learn the conditional distribution of trajectories based on these attributes. We further modify the noise schedule of diffusion model to adhere to the cosine schedule rather than the linear schedule. This choice is motivated by the cosine schedule's efficiency, allowing for a more gradual addition of noise and thus spending fewer timesteps on purely noisy transformations [34].

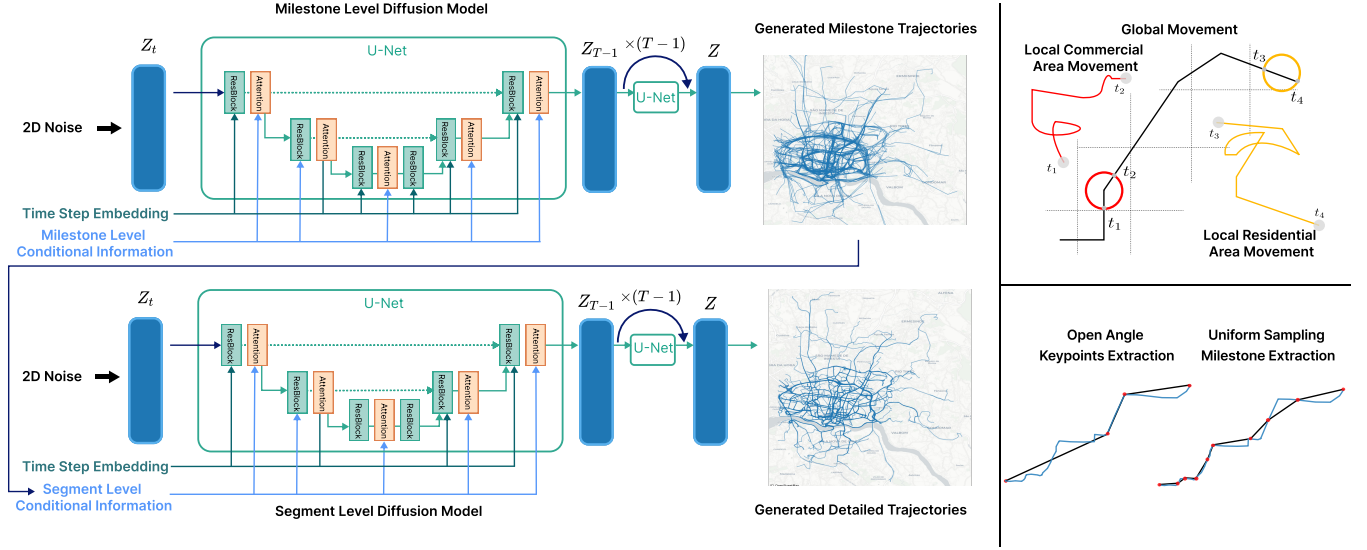


Figure 1: Left: Multi-Resolution Framework of Two-level Diffusion Model; Right top: Example of Global Movement and Local Movement Patterns; Right bottom: Illustration of two different Milestone Extraction strategies

4.1 Milestone Extraction & Learning Milestone Trajectories

We define a milestone trajectory as an abstraction of the complete trajectory at both temporal and spatial resolutions. In general, an Origin-Destination pair represents the lowest-resolution milestone, while the complete trajectory corresponds to the highest resolution. Milestone derivation is a key component of the framework, and we consider two strategies: uniform sampling and open angle extraction. Open angle extraction, such as the two-pass corner detection algorithm, is commonly used in trajectory compression tasks [46] to identify keypoints based on open angles formed by neighboring points and local curvature. A visual illustration of the two milestone extraction techniques is shown in Figure 1.

We derive milestones from original trajectories using a simple yet effective uniform sampling strategy, selecting a point every α points as a milestone. This design targets dense GPS trajectories without explicit semantic or POI annotations (e.g., taxi mobility data), where milestones act as geometric anchors rather than semantic activity boundaries. We denote the extracted milestone trajectory as

$$\mathcal{M} = [m_1, \dots, m_{\lceil n/\alpha \rceil}] = [s_1, s_\alpha, \dots, s_{\lceil n/\alpha \rceil \alpha}] \quad (1)$$

While keypoint-based extraction provides semantically meaningful compression by capturing directional changes, we find uniform sampling more effective for our setting, as demonstrated in the ablation study in Section 5. Uniform sampling preserves the overall trajectory shape and standardizes time intervals between milestones, which simplifies segment-level diffusion modeling. By enforcing equal-length segments, we reduce the complexity of modeling variable-length trajectories and enable more consistent and effective synthetic trajectory generation, ultimately improving data quality and downstream utility.

Overall, the denoising process of the milestone-level diffusion model is defined as

$$p_\theta(\mathcal{M}_{t-1} | \mathcal{M}_t, c_m) = \mathcal{N}(\mathcal{M}_{t-1}; \mu_\theta(\mathcal{M}_t, t | c_m), \Sigma_\theta(\mathcal{M}_t, t | c_m)) \quad (2)$$

where $\mathcal{M}_t$ denotes a noisy milestone trajectory at diffusion timestep t , and c_m represents milestone-level (global) conditions constructed by concatenating attributes such as travel time, travel distance, start location, and end location. The training objective of the milestone-level diffusion model is

$$\mathcal{L}(\theta) = \mathbb{E}_{t, \mathcal{M}_0, \epsilon} [\|\epsilon - \epsilon_\theta(\mathcal{M}_t, t | c_m)\|^2] \quad (3)$$

where $\epsilon \sim \mathcal{N}(0, I)$ is the Gaussian noise injected in the forward diffusion process, and $\epsilon_\theta(\cdot)$ denotes the noise prediction network parameterized by θ .

4.2 Learning Milestone Conditioned Segment

The milestone extraction operation on a single trajectory also results in $\lceil n/\alpha \rceil - 1$ segments of the original trajectory, we denote the segments as:

$$\mathcal{S} = [\mathcal{S}_1, \dots, \mathcal{S}_{\lceil n/\alpha \rceil - 1}] \quad (4)$$

$$\begin{aligned} \mathcal{S}_i &= [s_{i\alpha}, s_{(i+1)\alpha-1}, \dots, s_{(i+1)\alpha-1}, s_{(i+1)\alpha}] \\ &= [m_i, s_{i\alpha+1}, \dots, s_{(i+1)\alpha-1}, m_{i+1}] \end{aligned} \quad (5)$$

The generation of trajectory segments between milestones is conditioned on segment-level conditions as well as its corresponding milestones as start and end guidance (every segment starts and ends with a milestone point). This process ensures that the segments are not only realistic in isolation but also coherent when considered as part of the overall trajectory. The denoising process of the segment-level diffusion model can be described as

$$p_\theta(\mathcal{S}_{j,t-1} | \mathcal{S}_{j,t}, c_{s,j}) = \mathcal{N}(\mathcal{S}_{j,t-1}; \mu_\theta, \Sigma_\theta) \quad (6)$$

$$\mu_\theta \leftarrow \mu_\theta(\mathcal{S}_{j,t}, t | c_{s,j}), \Sigma_\theta \leftarrow \Sigma_\theta(\mathcal{S}_{j,t}, t | c_{s,j}) \quad (7)$$

$$c_{s,j} = [c_j, m_j, m_{j+1}] \quad (8)$$

where $S_{j,t}$ denotes the noisy j -th segment trajectory at diffusion timestep t , $c_{s,j}$ denotes the concatenated segment level conditions consisting of attribute information c_j calculated from S_j and m_j, m_{j+1} which denotes the milestones used for guidance. The subscript j specifies the segment index within a complete trajectory and m_j, m_{j+1} is the corresponding generated milestones that specify the start and end location of the j -th segment trajectory. During training, the milestones m_j are extracted from original trajectories. The training objective of the segment level diffusion model becomes

$$\mathcal{L}(\theta) = \mathbb{E}_{t, S_0, \epsilon} [\|\epsilon - \epsilon_\theta(S_{j,t}, t | c_{s,j})\|^2] \quad (9)$$

4.3 Synthetic Trajectory Generation

During inference, we use the same attribute information as in training. However, the milestones guiding segment generation are replaced with synthetic milestones produced by a milestone-level diffusion model. Details of the algorithm and variable-length trajectory handling are provided in Algorithm 4.3. Variable-length trajectories are handled via interpolation-based resampling during both training and inference, consistent with Algorithm 4.3.

We adopt an efficient sampling scheme from prior work to avoid expensive sampling at every diffusion step [43]. Specifically, for each sampled noise, we denoise every s timesteps during the backward process with stride s . As our framework conditions on multiple attributes, the generated synthetic trajectories can be viewed as seed trajectories [2], where each corresponds to a real-world trajectory. The combination of all given conditions defines a semantic class. Synthetic trajectories are generated under these conditions and can therefore replace original trajectories within the same semantic class, yielding a synthetic dataset that preserves the semantic information of the original dataset [2].

5 Experiment

In this section, we evaluate our model in comparison to various baseline model on two taxi trajectories dataset: Chengdu [58] and Porto [32]. The Porto taxi trajectory dataset was released as part of the ECML/PKDD 2015 challenge on taxi destination prediction and contains over one million trajectories collected over one year with GPS points sampled every 15 seconds. Following prior work [22], we filtered out trajectories with fewer than 10 points and focused on the central high-density kernel region of Porto, illustrated in Figure 2. The Chengdu dataset is already confined to a dense urban area and therefore does not require additional spatial filtering [58]. Table 2 summarizes the statistics of the processed datasets, including the number of trajectories, trajectory lengths, and spatial coverage.

Specifically, we evaluate the similarity in distribution between the original and synthetic trajectories and also the downstream utility of generated synthetic trajectories. We further perform privacy assessment for seed synthetic trajectories generated under the guidance from original metadata. We considered first-order Markov Model: Markov [44], Hidden Markov Model: IOHMM [54], LSTM Model: LSTM [21], GAN based Model: Movesim [16], Diffusion based Model: DiffTraj [59].

5.1 Evaluation Metrics

This section formally defines the evaluation metrics used in the main experiments.

Algorithm 1 MR-Traj algorithm

Input Trajectory T , Uniform Sampling Rate α , Attribute Information C

Output Synthetic Trajectory T'

- 1: T is an original trajectory of length n
- 2: $\mathcal{M} = [m_1, \dots, m_{n/\alpha}]$ is the extracted milestone trajectory
- 3: $\mathcal{S} = [S_1, \dots, S_{n/\alpha-1}]$ is the set of segments between consecutive milestones
- 4: $\text{length}(S_i) = \alpha + 1$
- 5: Initialize $C_m = C = [\text{total distance, total time, } \dots, \text{start location, end location}]$
- 6: Initialize $C_s = []$
- 7: **for** S_i in $\mathcal{S}$ **do**
- 8: compute attributes for S_i based on C and append to C_s
- 9: **end for**
- 10: $\mathcal{M}$ is resampled to the 0.75-th quantile length of all milestone trajectories in the dataset via interpolation
- 11: Train MDM($\mathcal{M}, C_m$) ► Milestone-Level Diffusion Model
- 12: Train SDM($S_i, C_s[i]$) ► Segment-Level Diffusion Model
- 13: $\mathcal{M}' = \text{sample from MDM}(z, C_m)$ ► z is random noise
- 14: $\mathcal{M}'$ is resampled back to length n/α by interpolation
- 15: **for** i in length of $\mathcal{S}$ **do**
- 16: $C_s[i][[-2, -1]] = [\mathcal{M}'[i], \mathcal{M}'[i+1]]$ ► replace start and end location guidance for segment i
- 17: $S'_i = \text{sample from SDM}(z, C_s[i])$ ► z is newly sampled noise
- 18: **end for**
- 19: T' is obtained by concatenating all segments in $\mathcal{S}' = [S'_1, \dots, S'_{n/\alpha-1}]$

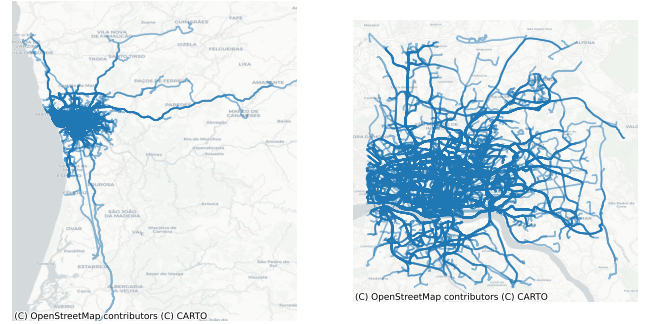


Figure 2: Left: samples from original trajectories; Right: filtered samples within the kernel region of the city of Porto.

- **Density Error** measures the discrepancy between the spatial distributions of synthetic and real trajectories over a discretized grid. It is computed using the Jensen–Shannon divergence (JSD) between the two distributions:

$$\text{Density Error} = \mathcal{JSD}(\mathcal{D}(S_{syn}), \mathcal{D}(S_{real})). \quad (10)$$

- **Distance Error** evaluates the difference in total travel distance distributions. The total distance of a trajectory is computed as the sum of distances between consecutive GPS

Table 1: Distribution Evaluation of Generated Synthetic Trajectories compared to Original Trajectories. (Bold: the best result. Underline: the 2nd best result.)'' notation stands for Error

Dataset	Model	Global				Local			
		Density. (↓)	Distance. (↓)	Trip. (↓)	Pattern Score (↑)	Density. (↓)	Distance. (↓)	Trip. (↓)	Pattern Score (↑)
Porto	MR-Traj	0.0015	0.0002	0.0079	0.91	0.0257	0.0039	0.1060	0.72
	DiffTraj	0.0010	0.0004	0.0061	0.93	0.0371	0.0044	0.1480	0.59
	Markov	0.0175	0.5250	0.1505	0.59	0.1730	0.6090	0.6560	0.26
	LSTM	0.0103	0.0257	0.0746	0.70	0.1948	0.0643	0.6124	0.25
	MoveSim	0.2816	0.0463	0.2977	0.02	0.6635	0.3210	0.6811	0.03
	IOHMM	0.0412	0.0215	0.2867	0.34	0.2060	0.0401	0.5858	0.23
Chengdu	MR-Traj	0.0028	0.0035	0.0134	0.81	0.0081	0.0096	0.0395	0.75
	DiffTraj	0.0018	0.003	0.0246	0.94	0.0182	0.0009	0.0474	0.66
	Markov	0.0356	0.0373	0.5835	0.27	0.0772	0.6930	0.5649	0.32
	LSTM	0.0658	0.335	0.266	0.14	0.0978	0.5504	0.5864	0.33
	MoveSim	0.0585	0.156	0.3852	0.13	0.1407	0.0506	0.4084	0.22
	IOHMM	0.0757	0.0725	0.46	0.11	0.2482	0.6931	0.5916	0.02

Table 2: Dataset Information.

Meta-Information	Filtered Porto	Chengdu
# of records	112991	1000000
# of taxi id	441	-
min/max/mean length of records	10/2198/44.07	180/6890/292.42
min/max latitude	41.1 / 41.25	30.65 / 30.74
min/max longitude	-8.65 / -8.5	104.03 / 104.13
sampling frequency	15s	15s
Local Case Study	Ribeira, Porto	Taikoo Li, Chengdu
min/max latitude	41.140344 / 41.143130	30.651997 / 30.661530
min/max longitude	-8.615496 / -8.606797	104.075045 / 104.089312
# of segments	99601	681038

points. The divergence between the distributions of synthetic and real trajectories is measured using JSD:

$$\text{Distance Error} = \mathcal{JSD}(\text{Dis}(S_{syn}), \text{Dis}(S_{real})). \quad (11)$$

- **Pattern Score** quantifies the similarity between frequent movement patterns extracted from synthetic and real trajectories. A pattern is defined as a frequently occurring sequence of discretized grid cells. The similarity is computed using the F1 score:

$$\text{Pattern F1} = 2 \times \frac{\text{Precision}(P_{real}, P_{syn}) \times \text{Recall}(P_{real}, P_{syn})}{\text{Precision}(P_{real}, P_{syn}) + \text{Recall}(P_{real}, P_{syn})}. \quad (12)$$

5.2 Details of Downstream Tasks

Taxi Destination Prediction. T-CONV models trajectories as two-dimensional spatial paths and applies multi-scale convolutional operations to extract movement patterns. Trajectory destinations are first clustered into arrival hotspots, and the predicted destination is computed as a weighted combination of the corresponding cluster centers. We adopt the locally enhanced variant of T-CONV, which performs higher-resolution convolution near the start and end of trajectories.

The prediction error is measured using the Haversine distance, which computes the shortest distance between two points on the

Earth's surface:

$$D(d_{real}, d_{pred}) = 2r \cdot \arctan\left(\sqrt{\frac{\alpha}{1-\alpha}}\right), \quad (13)$$

where d_{real} and d_{pred} denote the ground-truth and predicted destinations, respectively, and r is the Earth's radius.

Anomalous Mobility Pattern Identification. This task aims to identify anomalous mobility patterns from a collection of trajectories by analyzing time-dependent spatial movement distributions. Mobility vectors are constructed for each time interval over a spatial grid:

$$mv_i = \begin{pmatrix} f(t_i, g_1) \\ f(t_i, g_2) \\ \vdots \\ f(t_i, g_k) \end{pmatrix} \in \mathbb{R}^{k \times 1}, \quad (14)$$

where k is the total number of grid cells and $f(t_i, g_j)$ counts trips departing at time t_i with destinations in grid g_j . Each mobility vector is normalized by the total number of trips.

Anomalous mobility vectors are identified following the procedure in [51]: (1) constructing mobility vectors at the specified temporal scale, (2) clustering vectors using a mean-shift algorithm with a flat kernel, (3) computing an anomaly threshold based on cluster size statistics, and (4) labeling vectors in clusters smaller than the threshold as anomalous.

5.3 Distribution Similarity Evaluation

The two large-scale taxi trajectory GPS dataset we used is the Chengdu Dataset and Porto Dataset. We filtered out very short trajectories less than 10 points and focused on trajectories in the center region of the city of Porto, Portugal, also referred to as kernel region of a city [22]. This region is characterized by a higher and more uniform density of trajectories, offering a representative sample for analysis. The Chengdu dataset is already confined to a high density region and no preprocessing is performed.

To quantify the performance of MR-Traj and compare it with baseline models, we employed several metrics detailed as the following. For metrics that ask for a discretization of the spatial space, we used a 16×16 grid over the spatial scale, resulting in grid resolution less than 1km^2 per cell. The discretization is done only

Table 3: Downstream utility evaluation of generated synthetic trajectories. For Taxi Destination Prediction, HDE denotes the Haversine Distance Error. For Anomalous Mobility Pattern Similarity, GH and WH denote the cosine similarity between anomalous mobility patterns at the global-hourly and weekly-hourly scales, respectively. P in metric names indicates Porto and C indicates Chengdu.

Model	Destination Prediction		Anomalous Pattern Similarity		
	P-HDE↓	C-HDE↓	P-GH↑	P-WH↑	C-GH↑
Original	2.319	2.810	1.000	1.000	1.000
MR-Traj	2.361	2.833	0.846	0.700	0.747
DiffTraj	2.371	2.896	0.835	0.678	0.659
Markov	2.389	2.970	0.698	0.463	0.397
LSTM	2.383	3.195	0.731	0.481	0.319
MoveSim	2.447	3.082	0.419	0.330	0.202
IOHMM	2.541	2.990	0.778	0.633	0.238

for evaluation purpose, the generated trajectories are made up of precise GPS locations. For detailed explanation on the evaluation metrics, please refer to the supplementary material. In addition to the global space, we also sampled trajectory segments within two popular regions from each city: Ribeira region in the city of porto and Taikoo Li commercial region in the city of Chengdu. We chose the two regions because they are highly popular regions within the two city with either tourist attractions and commercial venues. We perform the same evaluation on the segments within the chosen region. For baselines that do not generate segments, we use uniform sampling with the same sampling rate to define milestones which in turn defines segments from a trajectory. For detailed information on bounding boxes of each region and each dataset as well as additional meta information, please refer to the supplementary material.

Table 1 shows the evaluation result of MR-Traj and other baselines. We achieve close parity with the best-performing baseline while outperforming all other baselines on global distribution similarity. The evaluation indicates that MR-Traj closely matches the performance of the best-performing baseline DiffTraj model across all metrics on the global scale. However, when evaluated on regional segment similarity, MR-Traj demonstrated superior performance on both datasets and all metrics. This result validated that MR-Traj is detail oriented and successfully captures local details.

5.4 Utility and Privacy Evaluation

In this section, we scrutinize the practical utility of the synthetic data produced by our model, MR-Traj, through its application to two critical downstream tasks: Taxi Destination Prediction and Anomalous Mobility Pattern Identification. These tasks were selected based on their relevance to the real-world applicability of synthetic trajectory data. The former represents a direct application of taxi trajectory data, crucial for traffic management and location-based services. The latter tests the synthetic data’s ability to preserve the original dataset’s unique or anomalous patterns, a vital aspect of data utility in detecting and analyzing mobility behaviors. Though normal mobility patterns are important, it’s relatively easy to learn since it constitutes the majority of the dataset. But abnormal patterns can easily be neglected while showing some

meaningful insights into atypical behaviors or events. To the best of our knowledge, the ability of a model to preserve original anomalous patterns in synthetic trajectories is highly underexplored.

Another highly considered aspect of synthetic trajectories is privacy and we further examine this crucial aspect by training a classifier on original data to link trajectory to users. By introducing noise at two levels, MR-Traj demonstrates more diversity in the generated trajectories which becomes highly beneficial when it comes to privacy preservation without compromising the utility of generated trajectories.

5.4.1 Taxi Destination Prediction. Given a partial taxi trajectory $S_p = [s_1, \dots, s_j]$, $j < n$, $S_p \subset S$ where S is defined in section 3.1, we want to predict s_n , the destination point of the trajectory. The task involves predicting the end point of a taxi trip given a partial trajectory. This challenge is not only about understanding the spatial components of the trip but also about inferring the likely destination based on observed patterns in the data. We adopt T-CONV[30] which is a convolutional neural network built for taxi trajectory prediction. Please refer to the supplementary material for details of T-CONV model. The T-CONV model is trained separately on the original and each generated synthetic datasets. The evaluation metric is the mean Haversine distance between the real destination and the predicted location for trajectories in the test set from original data.

Table 3 presents the Haversine distance errors for models trained on both original and synthetic data. MR-Traj demonstrates a remarkable capability in maintaining high quality to the original data’s spatial-temporal patterns, as evidenced by its Haversine Distance Error of 2.361. This metric, only slightly higher than the 2.319 error observed with the original data, suggests that the synthetic trajectories generated by MR-Traj closely resemble the real-world movements, capturing essential dynamics necessary for accurate destination prediction. The minimal discrepancy between the haversine distance error between MR-Traj and original data indicates that MR-Traj effectively encapsulates the variability and complexity of urban mobility patterns in its synthetic outputs. Notably, MR-Traj outperforms all baselines, further validating the utility of MR-Traj generated data.

5.4.2 Anomalous Mobility Pattern Identification. Urban mobility flows typically exhibit certain motifs or regular patterns reflecting the routine movements of people within an environment [41]. However, alongside these regular patterns, there are anomalous mobility patterns that deviate from the norm, often emerging during holidays, special events, extreme weather, or other unique circumstances [51]. Recognizing these anomalies is essential for urban planning, traffic management, and emergency response, allowing for preemptive action in anticipation of similar future events. Following prior work, we implemented a method to calculate mobility vectors and identify anomalous mobility patterns [51]. For detailed description of the procedure, please refer to the supplementary material. The number of mobility vectors depend on the temporal scale of interest. In this experiment setting, we examined global hourly scale where we calculate a mobility for each hour; and a weekly hourly scale where a mobility vector is calculated for each hour within each week. Our analysis focuses evaluating how effectively the synthetic trajectories generated by MR-Traj retain these

Table 4: Trajectory User Linking evaluation on seed synthetic trajectories. A lower metric value indicates a higher level of privacy. Other baselines are emitted in this evaluation because their synthetic generation process is not guided by specific metadata from original trajectory. The standard deviation is included in parenthesis as a result of 3 different runs with different numpy random seeds.

Metric	Original Data	DiffTraj	MR-Traj
Acc (↓)	0.9738 (0.0031)	0.8018 (0.0050)	0.7325 (0.0069)
Acc@5 (↓)	0.9986 (0.0003)	0.9092 (0.0004)	0.8668 (0.0013)
F1 Score (↓)	0.9793 (0.0031)	0.8142 (0.0043)	0.7556 (0.0038)
Precision (↓)	0.9781 (0.0035)	0.8226 (0.0060)	0.7686 (0.0042)
Recall (↓)	0.9816 (0.0021)	0.8152 (0.0024)	0.7533 (0.0034)

anomalous patterns present in the original dataset compared to other baselines. This evaluation is conducted using the pairwise cosine similarity between anomalous mobility vectors derived from both synthetic and original data.

Table 3 presents the cosine similarity scores for anomalous mobility vectors derived from synthetic and original trajectories, calculated on both global hourly and weekly hourly scales. The Chengdu dataset does not have the Weekly Hourly Scale evaluation because there is no departure date information in the dataset. So, the seed trajectories cannot be grouped by week. MR-Traj achieves the highest cosine similarity at both temporal scales. These results illustrate MR-Traj’s proficiency in capturing and preserving anomalous mobility patterns present in the original trajectory data, with its performance notably surpassing that of other models. Figures 3 visually compare mobility vectors from both normal and anomalous clusters, derived from original, MR-Traj, as well as the best performing baseline DiffTraj generated synthetic data. These visual comparisons further highlight the model’s ability to replicate both regular and anomalous urban mobility patterns effectively.

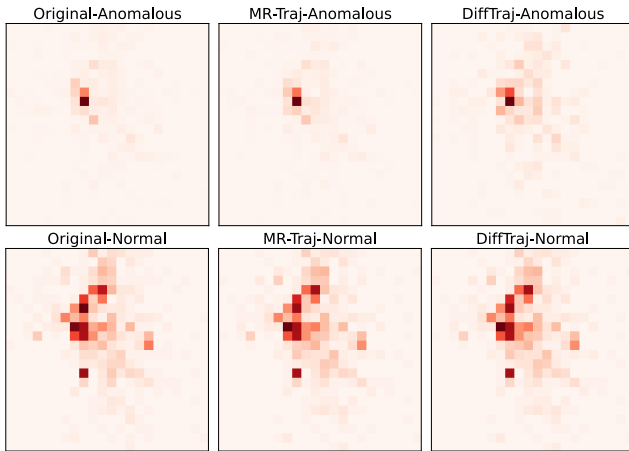


Figure 3: Comparison of normal and abnormal mobility patterns on discretized grid of Porto region from 1) Original trajectories, 2) MR-Traj generated trajectories, and 3) Diff-Traj generated trajectories

5.4.3 Privacy Evaluation. Privacy is a crucial aspect of data management, especially in trajectory data, where personal and sensitive

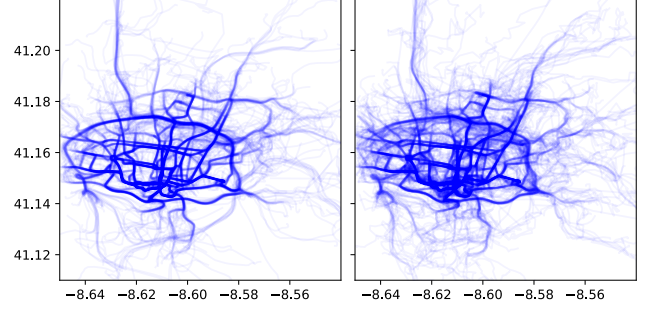


Figure 4: Left: Samples of Synthetic Trajectories by DiffTraj; Right: Samples of Synthetic Trajectories by MR-Traj . As evidenced by the comparison, MR-Traj preserves all essential mobility movement patterns while generating more diverse trajectories and hence offering more privacy protection.

information can be inferred from movement patterns. Existing methods often overlook privacy concerns by focusing solely on distribution similarity and utility. These methods tend to ensure that synthetic data maintains statistical resemblance to the original data while maximizing usability, but fail to adequately safeguard individual privacy.

In our privacy evaluation, we employ the MARC (Multiple-Aspect tTrajectory Classifier) algorithm [31], a widely recognized model for the trajectory user linking task. MARC utilizes one-hot encoded metadata, including departure day and hour, as well as geo-hashed trajectory information, to predict the probability that a given trajectory belongs to each user in the dataset. This evaluation is conducted exclusively on the Porto dataset, as it includes taxi identifiers, while the Chengdu dataset lacks user information, preventing its use in this evaluation. Moreover, the privacy evaluation is focused solely on seed synthetic trajectories generated by diffusion based models because the other baseline models generate trajectories based on their learned distribution, without leveraging metadata from the original trajectories for guidance. Randomly assigning other baseline-generated trajectories to a taxi ID does not accurately reflect the intended privacy level for this evaluation.

As shown in Table 4, MR-Traj consistently obtained a lower score across all metrics compared to DiffTraj. The key differentiator for MR-Traj is its approach to noise sampling at two distinct levels of the trajectory generation process. By injecting noise at both the higher structural level and the finer granular level, MR-Traj adds an additional layer of randomness and diversity to the synthetic trajectories. This layered noise application ensures that the synthetic trajectories deviate more significantly from the original data without compromising the realistic nature of the generated trajectories as evidenced in figure 4. As MR-Traj’s generated trajectories are a composition of segments independently sampled under the guidance of generated milestones, such formulation makes it harder for adversaries to reverse engineer the original metadata. MR-Traj strikes a balance between achieving a low distribution similarity and preserving privacy with high utility such that the synthetic trajectories remain useful in downstream applications.

Table 5: Ablation Study on variants of MR-Traj on Porto Dataset. "-" stands for Error. The composition of MR-Traj outperforms all other variants. HDE stands for Harversine Distance Error for taxi destination. For Anomalous Mobility Pattern Similarity, GH denotes the cosine similarity between anomalous mobility patterns in global hourly scale, WH denotes the cosine similarity between anomalous mobility patterns in weekly hourly scale.

Model	Density . (↓)	Distance . (↓)	Trip . (↓)	Pattern Score (↑)	HDE(↓)	GH (↑)	WH (↑)
MR-Traj	0.0015	0.0002	0.0079	0.91	2.361	0.846	0.700
open angle extraction	0.0073	0.0008	0.0178	0.79	2.388	0.826	0.656
precise location guidance	0.0039	0.0396	0.0804	0.78	2.375	0.815	0.623
global milestone	0.0038	0.0372	0.0799	0.87	2.3966	0.81	0.656
Original Data	-	-	-	-	2.319	1	1

6 Ablation study

Table 5 reports four configuration of MR-Traj evaluated on Porto Dataset: *MR-Traj* denotes the original configuration, *open angle extraction* denotes the milestones are extracted based on the open angle detection algorithm instead of uniform sampling from the original trajectory, *precise location guidance* denotes that the start and end location guidance incorporated as conditional information in both diffusion models are represented as precise location instead of being mapped to a discrete grid location, *global milestone* denotes the configuration where the whole milestone trajectory $[m_1, m_2, \dots, m_k]$ are given to generate segments instead of only using two milestones $[m_i, m_{i+1}]$ as start and end guidance to generate the i -th segment in a trajectory. We note that all variants of MR-Traj did not perform as well as MR-Traj in both distribution similarity and downstream utility, validating the choice of design for MR-Traj.

In addition, we further performed analysis on how the uniform sampling rate affects the performance of MR-Traj on distribution similarity and downstream utility. We conducted analysis on the Chengdu dataset with uniform sampling rate 15, 35, and 55 as shown in table 6. We did not include 5 in the analysis because the Chengdu trajectories are densely sampled and extracting milestones every 5 timestamps provide virtually no level of abstraction to the original trajectories. In general, there is no apparent pattern that dictates an optimal rate universally. The analysis demonstrates that a sampling rate of 15 generally results in better performance across most metrics, suggesting that the lowest level of abstraction provides an optimal balance between data abstraction and fidelity.

Table 6: Analysis on the effect of uniform sampling rate 15, 35, and 55 on the Chengdu Dataset. HDE stands for harversine distance error and GH denotes the global hourly scale. The uniform sampling rate 5 is not considered in this case because Chengdu trajectories are sampled very densely and sampling every 5 points provide virtually no level of abstraction

Metrics	MR-Traj 15	MR-Traj 35	MR-Traj 55
Density Error	0.0028	0.0337	0.0099
Distance Error	0.0035	0.0423	0.0095
Trip Error	0.0134	0.1698	0.0582
Pattern Score	0.81	0.6	0.71
HDE	2.833	2.855	2.861
GH Cosine Similarity	0.747	0.514	0.68

7 Conclusion

In this paper, we presented MR-Traj, a multi-resolution diffusion framework for synthetic human mobility trajectory generation. By jointly modeling milestone-level global structure and segment-level local refinement, MR-Traj captures multi-scale spatial-temporal dependencies and preserves both long-range movement patterns and fine-grained trajectory details. Experiments on large-scale real-world taxi datasets show that MR-Traj achieves performance comparable to state-of-the-art diffusion baselines in global distribution similarity while improving fidelity at finer spatial resolutions. The generated trajectories also demonstrate strong downstream utility on tasks such as taxi destination prediction and anomalous mobility pattern identification. Under a seed-guided release setting, multi-level stochasticity further reduces linkage success compared to single-level diffusion approaches without sacrificing utility. Overall, MR-Traj provides an effective framework for balancing realism, utility, and privacy in synthetic spatiotemporal data.

8 Limitations and Ethical Considerations

Despite the strong empirical performance of MR-Traj, several limitations remain. Although multi-level noise injection reduces linkage risk, synthetic data may still expose residual privacy information under stronger adversarial settings. In addition, training on taxi datasets from specific urban regions may introduce geographic and behavioral biases and limit generalizability to other mobility contexts. Metadata-conditioned generation may also propagate biases in the original data. Finally, while synthetic trajectories can support downstream analysis, they should not replace real-world data in safety-critical applications. We encourage responsible use of synthetic mobility data and continued research on privacy protection and robustness.

Acknowledgments

This work is partially supported by the NSF Award 2425919, and NSF Award 2413417. The funding from these sources has been a cornerstone in enabling us to bring our project to fruition. We are also deeply grateful to the anonymous reviewers for their rigorous review process. Their detailed comments and constructive suggestions have significantly contributed to the improvement of this paper.

References

- [1] Ali Arab, Mevin B Hooten, and Christopher K Wikle. 2008. Hierarchical Spatial Models. *Encyclopedia of GIS* 14, 1 (2008), 425–431.
- [2] Vincent Bindschaedler and Reza Shokri. 2016. Synthesizing plausible privacy-preserving location traces. In *2016 IEEE Symposium on Security and Privacy (SP)*. IEEE, 546–563.
- [3] Stephen Boyd and Lieven Vandenbergh. 2004. *Convex optimization*. Cambridge university press.
- [4] Defu Cao, Yousef El-Laham, Loc Trinh, Svitlana Vyetenrenko, and Yan Liu. 2022. DSLOB: a synthetic limit order book dataset for benchmarking forecasting algorithms under distributional shift. *arXiv preprint arXiv:2211.11513* (2022).
- [5] Defu Cao, Michael Gee, Jinbo Liu, Hengxuan Wang, Wei Yang, Rui Wang, and Yan Liu. 2025. Conversational Time Series Foundation Models: Towards Explainable and Effective Forecasting. *arXiv preprint arXiv:2512.16022* (2025).
- [6] Defu Cao, Furong Jia, Sercan O Arik, Tomas Pfister, Yixiang Zheng, Wen Ye, and Yan Liu. 2024. TEMPO: Prompt-based Generative Pre-trained Transformer for Time Series Forecasting. In *The Twelfth International Conference on Learning Representations*. <https://openreview.net/forum?id=YH5w12OUuU>
- [7] Defu Cao, Jiachen Li, Hengbo Ma, and Masayoshi Tomizuka. 2021. Spectral Temporal Graph Neural Network for Trajectory Prediction. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. 1839–1845. doi:10.1109/ICRA48506.2021.9561461
- [8] Defu Cao, Yujing Wang, Juanyong Duan, Ce Zhang, Xia Zhu, Conguri Huang, Yunhai Tong, Bixiong Xu, Jing Bai, Jie Tong, et al. 2021. Spectral temporal graph neural network for multivariate time-series forecasting. *Advances in neural information processing systems* (2021).
- [9] Defu Cao, Wen Ye, Yizhou Zhang, Sam Griesemer, and Yan Liu. 2026. PINFDiT: Energy-Based Physics-Informed Diffusion Transformers for General-purpose Time Series Tasks. In *The Fourteenth International Conference on Learning Representations*. <https://openreview.net/forum?id=EphTIUJ4XN>
- [10] Defu Cao, Wen Ye, Yizhou Zhang, and Yan Liu. 2024. TimeDiT: General-purpose diffusion transformers for time series foundation model. *arXiv preprint arXiv:2409.02322* (2024).
- [11] Seongjin Choi, Jiwon Kim, and Hwasoo Yeo. 2021. TrajGAIL: Generating urban vehicle trajectories using generative adversarial imitation learning. *Transportation Research Part C: Emerging Technologies* 128 (2021), 103091.
- [12] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A Bharath. 2018. Generative adversarial networks: An overview. *IEEE signal processing magazine* 35, 1 (2018), 53–65.
- [13] Luc Devroye, László Györfi, and Gábor Lugosi. 2013. *A probabilistic theory of pattern recognition*. Vol. 31. Springer Science & Business Media.
- [14] Simon Dirmeyer, Ye Hong, and Fernando Perez-Cruz. 2024. Synthetic location trajectory generation using categorical diffusion models. *arXiv preprint arXiv:2402.12242* (2024).
- [15] Carl Doersch. 2016. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908* (2016).
- [16] Jie Feng, Zeyu Yang, Fengli Xu, Haisu Yu, Mudan Wang, and Yong Li. 2020. Learning to simulate human mobility. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 3426–3433.
- [17] Umang Garg, Bowen Zhang, Anantajit Subrahmanya, Chandrakanth Gudavalli, and BS Manjunath. 2025. GPS-MTM: Capturing Pattern of Normalcy in GPS-Trajectories with self-supervised learning. *arXiv preprint arXiv:2509.24031* (2025).
- [18] Anita Graser, Anahid Jalali, Jasmin Lampert, Axel Weißenfeld, and Krzysztof Janowicz. 2025. MobilityDL: a review of deep learning from trajectory data. *Geoinformatica* 29, 1 (2025), 115–147.
- [19] Sam Griesemer, Defu Cao, Zijun Cui, Carolina Osorio, and Yan Liu. 2024. Active Sequential Posterior Estimation for Sample-Efficient Simulation-Based Inference. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*. <https://openreview.net/forum?id=fkuseU0nJs>
- [20] Baoshen Guo, Zhiqing Hong, Junyi Li, Shenhao Wang, and Jinhua Zhao. 2025. Leveraging the Spatial Hierarchy: Coarse-to-fine Trajectory Generation via Cascaded Hybrid Diffusion. *arXiv preprint arXiv:2507.13366* (2025).
- [21] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
- [22] Xu Kang, Liang Liu, Dong Zhao, and Huadong Ma. 2020. TraG: A trajectory generation technique for simulating urban crowd mobility. *IEEE Transactions on Industrial Informatics* 17, 2 (2020), 820–829.
- [23] Alexandra Kapp, Julia Hansmeyer, and Helena Mihaljević. 2023. Generative models for synthetic urban mobility data: A systematic literature review. *Comput. Surveys* 56, 4 (2023), 1–37.
- [24] Moritz UG Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M Pigott, Open COVID-19 Data Working Group†, Louis Du Plessis, Nuno R Faria, Ruoran Li, et al. 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 368, 6490 (2020), 493–497.
- [25] Li Li, Rui Jiang, Zhengbing He, Xiqun Michael Chen, and Xuesong Zhou. 2020. Trajectory data-based traffic flow studies: A revisit. *Transportation Research Part C: Emerging Technologies* 114 (2020), 225–240.
- [26] Xi Liu, Hanzhou Chen, and Clio Andris. 2018. trajGANs: Using generative adversarial networks for geo-privacy protection of trajectory data (Vision paper). In *Location privacy and security workshop*. 1–7.
- [27] Ziyi Liu, Qingyue Long, Huandong Wang, and Yong Li. 2025. Multi-scale diffusion transformer for jointly simulating user mobility and mobile traffic pattern. In *Proceedings of the 33rd ACM International Conference on Advances in Geographic Information Systems*. 522–525.
- [28] Qingyue Long, Huandong Wang, Tong Li, Lisi Huang, Kun Wang, Qiong Wu, Guangyu Li, Yanping Liang, Li Yu, and Yong Li. 2023. Practical synthetic human trajectories generation based on variational point processes. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4561–4571.
- [29] Yuxiao Luo, Songming Zhang, Kang Liu, Yang Xu, and Ling Yin. 2025. Traveller: Travel-pattern aware trajectory generation via autoregressive diffusion models. *Information Fusion* (2025), 103766.
- [30] Jianming Lv, Qing Li, Qinghui Sun, and Xintong Wang. 2018. T-CONV: A convolutional neural network for multi-scale taxi trajectory prediction. In *2018 IEEE international conference on big data and smart computing (bigcomp)*. IEEE, 82–89.
- [31] Lucas May Petry, Camila Leite Da Silva, Andrea Esuli, Chiara Renso, and Vania Bogorny. 2020. MARC: a robust method for multiple-aspect trajectory classification via space, time, and semantic embeddings. *International Journal of Geographical Information Science* 34, 7 (2020), 1428–1450.
- [32] Wendy Kan Meghan O’Connell, moreiraMatias. 2015. ECML/PKDD 15: Taxi Trajectory Prediction (I). <https://kaggle.com/competitions/pkdd-15-predict-taxi-service-trajectory-i>
- [33] Chuizheng Meng, Sam Griesemer, Defu Cao, Sungyong Seo, and Yan Liu. 2025. When physics meets machine learning: A survey of physics-informed machine learning. *Machine Learning for Computational Science and Engineering* 1, 1 (2025), 20.
- [34] Alexander Quinn Nichol and Prafulla Dhariwal. 2021. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*. PMLR, 8162–8171.
- [35] Hao Niu, Chuizheng Meng, Defu Cao, Guillaume Habault, Roberto Legaspi, Shinya Wada, Chihiro Ono, and Yan Liu. 2022. Mu2ReST: Multi-resolution Recursive Spatio-Temporal Transformer for Long-Term Prediction. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. 68–80.
- [36] Huaijin Pi, Sida Peng, Minghui Yang, Xiaowei Zhou, and Hujun Bao. 2023. Hierarchical generation of human-object interactions with diffusion probabilistic models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 15061–15073.
- [37] Bo Qiang, Yuxuan Song, Minkai Xu, Jingjing Gong, Bowen Gao, Hao Zhou, Wei-Ying Ma, and Yanyan Lan. 2023. Coarse-to-fine: a hierarchical diffusion model for molecule generation in 3d. In *International Conference on Machine Learning*. PMLR, 28277–28299.
- [38] Jinmeng Rao, Song Gao, Yuhao Kang, and Qunying Huang. 2020. LSTM-TrajGAN: A deep learning approach to trajectory privacy protection. *arXiv preprint arXiv:2006.10521* (2020).
- [39] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695.
- [40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18. Springer, 234–241.
- [41] Christian M Schneider, Vitaly Belik, Thomas Couronné, Zbigniew Smoreda, and Marta C González. 2013. Unravelling daily human mobility motifs. *Journal of The Royal Society Interface* 10, 84 (2013), 20130246.
- [42] Alex Sherstinsky. 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena* 404 (2020), 132306.
- [43] Jiaming Song, Chenlin Meng, and Stefano Ermon. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502* (2020).
- [44] Libo Song, David Kotz, Ravi Jain, and Xiaoning He. 2003. Evaluating location predictors with extensive Wi-Fi mobility data. *ACM SIGMOBILE Mobile Computing and Communications Review* 7, 4 (2003), 64–65.
- [45] Yiwen Song, Jingtao Ding, Jian Yuan, Qingmin Liao, and Yong Li. 2024. Controllable human trajectory generation using profile-guided latent diffusion. *ACM Transactions on Knowledge Discovery from Data* 19, 1 (2024), 1–25.
- [46] Penghui Sun, Shixiong Xia, Guan Yuan, Daxing Li, et al. 2016. An overview of moving object trajectory compression algorithms. *Mathematical Problems in Engineering* 2016 (2016).
- [47] Cédric Villani et al. 2009. *Optimal transport: old and new*. Vol. 338. Springer.
- [48] Xiaoda Wang, Kaiqiao Han, Yuhao Xu, Xiao Luo, Yizhou Sun, Wei Wang, and Carl Yang. 2026. SE-Diff: Simulator and Experience Enhanced Diffusion Model for Comprehensive ECG Generation. In *The Fourteenth International Conference on Learning Representations*.

- [49] Xingrui Wang, Xinyu Liu, Ziteng Lu, and Hanfang Yang. 2021. Large scale GPS trajectory generation using map based on two stage GAN. *Journal of Data Science* 19, 1 (2021), 126–141.
- [50] Xinhua Wu, Haoyu He, Yanchao Wang, and Qi Wang. 2024. Pretrained mobility transformer: A foundation model for human mobility. *arXiv preprint arXiv:2406.02578* (2024).
- [51] Zhenzhou Xu, Ge Cui, Ming Zhong, and Xin Wang. 2019. Anomalous urban mobility pattern detection based on GPS trajectories and POI data. *ISPRS International Journal of Geo-Information* 8, 7 (2019), 308.
- [52] Yiyuan Yang, Ming Jin, Haomin Wen, Chaoli Zhang, Yuxuan Liang, Lintao Ma, Yi Wang, Chenghao Liu, Bin Yang, Zenglin Xu, et al. 2024. A survey on diffusion models for time series and spatio-temporal data. *Comput. Surveys* (2024).
- [53] Wen Ye, Wei Yang, Defu Cao, Yizhou Zhang, Luminyuan Tang, Jie Cai, and Yan Liu. 2026. TS-Reasoner: Domain-Oriented Time Series Inference Agents for Reasoning and Automated Analysis. *Transactions on Machine Learning Research* (2026). <https://openreview.net/forum?id=yhy7Vigjcf>
- [54] Mogeng Yin, Madeleine Sheehan, Sidney Feygin, Jean-François Paiement, and Alexei Pozdnoukhov. 2017. A generative model of urban activities from cellular data. *IEEE Transactions on Intelligent Transportation Systems* 19, 6 (2017), 1682–1696.
- [55] Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. 2017. Seqgan: Sequence generative adversarial nets with policy gradient. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 31.
- [56] Jing Zhang, Qihan Huang, Yirui Huang, Qian Ding, and Pei-Wei Tsai. 2023. Dp-trajgan: A privacy-aware trajectory generation model with differential privacy. *Future Generation Computer Systems* 142 (2023), 25–40.
- [57] Yizhou Zhang, Defu Cao, and Yan Liu. 2022. Counterfactual neural temporal point process for estimating causal influence of misinformation on social media. *Advances in Neural Information Processing Systems* 35 (2022), 10643–10655.
- [58] Yuanshao Zhu, Yongchao Ye, Ying Wu, Xiangyu Zhao, and James Yu. 2024. Syn-Mob: Creating High-Fidelity Synthetic GPS Trajectory Dataset for Urban Mobility Analysis. *Advances in Neural Information Processing Systems* 36 (2024).
- [59] Yuanshao Zhu, Yongchao Ye, Shiyao Zhang, Xiangyu Zhao, and James Yu. 2023. DiffTraj: Generating GPS Trajectory with Diffusion Probabilistic Model. In *Thirty-seventh Conference on Neural Information Processing Systems*.
- [60] Yuanshao Zhu, James Jianqiao Yu, Xiangyu Zhao, Xun Zhou, Liang Han, Xuetao Wei, and Yuxuan Liang. 2024. Unitraj: Learning a universal trajectory foundation model from billion-scale worldwide traces. *arXiv preprint arXiv:2411.03859* (2024).

A Hyperparameter Setting

Hyperparameter configurations for training and inference are summarized in Table 7. All diffusion models are trained on an 80GB A100 GPU, while downstream task models are trained on a 10GB RTX 2080 GPU.

Table 7: Hyperparameter Details for Experiment Settings

Hyperparameter	Milestone Level Model	Segment Level Model
Epoch	200	100
Batch Size	1024	1024
Noise Schedule	Cosine	Cosine
Input Length (Porto)	11	7
Input Length (Chengdu)	23	23
Sampling Rate for Milestone (Porto)	5	-
Sampling Rate for Milestone (Chengdu)	15	-
Diffusion Timesteps	500	500

B Theoretic Analysis

The main application of synthetic trajectory generation is to train machine learning models for downstream tasks such as destination prediction, mileage prediction and anomalous mobility pattern recognition. These downstream tasks usually follow a paradigm to predict some properties (e.g., destination, mileage and mobility pattern) of a trip, denoted as $g(S_u, S_o)$, given only a fraction of the trip’s full trajectory. We denote this observable fraction of full trajectory as S_o . The rest unobservable fractions of the trajectory are denoted as S_u .

For downstream tasks, we usually use complete trajectories generated by the synthetic trajectory generator $P'(S)$ as training set to train the machine learning predictor $f(\cdot)$. The utility of this predictor on downstream tasks should be evaluated by its square empirical risk $\mathcal{R}$ on a trajectory distribution, which is defined as:

$$\mathcal{R}(f(\cdot)) = \mathbb{E}[f(S_o) - g(S_u, S_o)]^2 = \mathbb{E}_{S_o}[f(S_o) - \mathbb{E}_{S_u}g(S_u, S_o)]^2 + C \quad (15)$$

where C is a constant that is only relevant to $P(S_u, S_o)$. Thus, the optimal predictor we can acquire on the synthetic dataset is $f(S_o) = \mathbb{E}_{S_u \sim P'(S_u|S_o)}g(S_u, S_o)$.

This optimal predictor of synthetic dataset, when being applied to real dataset whose distribution is $P(S)$, its empirical risk will be:

$$\begin{aligned} \mathcal{R} &= \mathbb{E}_{S_o \sim P(S_o)}[\mathbb{E}_{S_u \sim P'(S_u|S_o)}g(S_u, S_o) \\ &\quad - \mathbb{E}_{S_u \sim P(S_u|S_o)}g(S_u, S_o)]^2 + C \end{aligned} \quad (16)$$

Suppose that $g(S_u, S_o)$ is L-lipsitz, then $\frac{g(S_u, S_o)}{L}$ is 1-lipsitz, and we have:

$$\begin{aligned} \mathcal{R} &= L\mathbb{E}_{S_o \sim P(S_o)}[\mathbb{E}_{S_u \sim P'(S_u|S_o)}\frac{g(S_u, S_o)}{L} \\ &\quad - \mathbb{E}_{S_u \sim P(S_u|S_o)}\frac{g(S_u, S_o)}{L}]^2 + C \\ &\leq L\mathbb{E}_{S_o \sim P(S_o)}[\sup_{h \text{ is 1-lipsitz}} \mathbb{E}_{S_u \sim P'(S_u|S_o)}h(S_u, S_o) \\ &\quad - \mathbb{E}_{S_u \sim P(S_u|S_o)}h(S_u, S_o)]^2 + C \\ &= L\mathbb{E}_{S_o \sim P(S_o)}[W_1(P(S_u|S_o), P'(S_u|S_o))]^2 + C \end{aligned} \quad (17)$$

where W_1 is Wasserstein distance of 1-norm. The above For Wasserstein distance, we have following upper bound:

$$W_1(P(S_u|S_o), P'(S_u|S_o)) \leq D_{S_u} \|P(S_u|S_o) - P'(S_u|S_o)\|_{TV} \quad (18)$$

where D is the diameter of the support set of S_u and $\|P(S_u|S_o) - P'(S_u|S_o)\|_{TV}$ is the Total Variance Distance between distribution $P(S_u|S_o)$ and $P'(S_u|S_o)$:

$$\|P(S_u|S_o) - P'(S_u|S_o)\|_{TV} = \int_{S_u} |p(S_u|S_o) - p'(S_u|S_o)| dS_u \quad (19)$$

Thus, when Total Variance Distance between $P(S_u, S_o)$ and $P'(S_u, S_o)$ is less than 1, i.e., the two distributions are close enough, we have:

$$\begin{aligned} \mathcal{R} &\leq L\mathbb{E}_{S_o \sim P(S_o)}[W_1(P(S_u|S_o), P'(S_u|S_o))]^2 + C \\ &\leq LD^2\mathbb{E}_{S_o \sim P(S_o)}\|P(S_u|S_o) - P'(S_u|S_o)\|_{TV} + C \end{aligned} \quad (20)$$

For Total Variance Distance we have:

$$\begin{aligned} \mathbb{E}_{S_o \sim P(S_o)} TVD &= \mathbb{E}_{S_o \sim P(S_o)} \|P(S_u|S_o) - P'(S_u|S_o)\|_{TV} \\ &= \int_{S_o} p(S_o) dS_o \int_{S_u} |p(S_u|S_o) - p'(S_u|S_o)| dS_u \\ &= \int_{S_u, S_o} |p(S_u, S_o) - p'(S_u, S_o)| dS_u dS_o \\ &= \|P(S_u, S_o) - P'(S_u, S_o)\|_{TV} \end{aligned} \quad (21)$$

Meanwhile, Total Variance Distance satisfy triangle inequality. Thus, we have:

$$\begin{aligned}
\mathcal{R} &\leq LD^2 \|P(S_u, S_o) - P'(S_u|S_o)P(S_o)\|_{TV} + C \\
&\leq LD^2 \|P(S_u, S_o) - P'(S_u, S_o)\|_{TV} \\
&\quad + LD^2 \|P'(S_u, S_o) - P'(S_u|S_o)P(S_o)\|_{TV} + C \\
&= LD^2 \|P(S_u, S_o) - P'(S_u, S_o)\|_{TV} \quad (22) \\
&\quad + LD^2 \|P'(S_u|S_o)P'(S_o) - P'(S_u|S_o)P(S_o)\|_{TV} + C \\
&= LD^2 \|P(S_u, S_o) - P'(S_u, S_o)\|_{TV} + \\
&\quad LD^2 \|P'(S_o) - P(S_o)\|_{TV} + C
\end{aligned}$$

The above upper bound shows that we should jointly minimize $\|P(S_u, S_o) - P'(S_u, S_o)\|_{TV}$ and $\|P'(S_o) - P(S_o)\|_{TV}$.

C Details on Diffusion Model Preliminaries

In recent years, denoising diffusion probabilistic model has gained tremendous success on generative tasks, particularly in image generation[39]. The allure of DDPMs lies in their capacity to model complex, high-dimensional distributions, making them an excellent choice for generating synthetic trajectories. The key idea behind diffusion is to gradually add noise to input over a series of timesteps and let the model learn the reverse process of denoising noisy input back to the realistic space. The forward process is defined as

a Markov chain where Gaussian noise is gradually added to the original input x_0 and the amount of noise added at each timestep is controlled by the noise schedule β_t .

$$q(x_{1:T}|x_0) = \prod_{t=1}^T q(x_t|x_{t-1}) \quad (23)$$

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I) \quad (24)$$

In practice, the closed form sampling of the noised version of original input at timestep t is used instead of adding noise step by step for efficiency purpose.

$$x_t = \sqrt{\bar{\alpha}_t} \cdot x_0 + \sqrt{1 - \bar{\alpha}_t} \cdot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \quad (25)$$

where $\bar{\alpha}_t = \prod_{i=1}^t 1 - \beta_i$. This closed form sampling allows for more efficient forward process. The backward process aims to learn the reverse of the forward process, learning to systematically denoise the noisy inputs. A neural network parameterized by θ is used to learn the mean and variance of the noise to remove from the noisy data at each timestep. The denoising process is described by:

$$q(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu(x_t, t), \Sigma(x_t, t)) \quad (26)$$

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \quad (27)$$

where $p_\theta(x_{t-1}|x_t)$ is learned by the neural network to approximate the ground truth conditional probability $q(x_{t-1}|x_t)$.