GeoGen: A Two-stage Coarse-to-Fine Framework for Fine-grained Synthetic Location-based Social Network Trajectory Generation

Rongchao Xu[1], Kunlin Cai[2], Lin Jiang[1], Zhiqing Hong[3], Yuan Tian[2], Guang Wang[1]*

[1]Florida State University, Tallahassee, Florida
[2] University of California, Los Angeles, Los Angeles, California
[3] Rutgers University, New Brunswick, New Jersey
rx21a@fsu.edu, kunlin96@g.ucla.edu, lin.jiang@fsu.edu, zh252@cs.rutgers.edu, yuant@ucla.edu, guang@cs.fsu.edu

Abstract

Location-Based Social Network (LBSN) check-in trajectory data are important for many practical applications like POI recommendation, advertising, and pandemic intervention. However, the high collection costs and ever-increasing privacy concerns prevent us from accessing large-scale LBSN trajectory data. The recent advances in synthetic data generation provide us with a new opportunity to achieve this, which utilizes generative AI to generate synthetic data that preserves the characteristics of real data while ensuring privacy protection. However, generating synthetic LBSN check-in trajectories remains challenging due to their spatially discrete, temporally irregular nature and the complex spatio-temporal patterns caused by sparse activities and uncertain human mobility. To address this challenge, we propose GeoGen, a two-stage coarse-to-fine framework for large-scale LBSN check-in trajectory generation. In the first stage, we reconstruct spatially continuous, temporally regular latent movement sequences from the original LBSN check-in trajectories and then design a Sparsity-aware Spatio-temporal Diffusion model (S^2TDiff) with an efficient denosing network to learn their underlying behavioral patterns. In the second stage, we design Coarse2FineNet, a Transformer-based Seq2Seq architecture equipped with a dynamic context fusion mechanism in the encoder and a multi-task hybrid-head decoder, which generates fine-grained LBSN trajectories based on coarse-grained latent movement sequences by modeling semantic relevance and behavioral uncertainty. Extensive experiments on four real-world datasets show that GeoGen excels state-of-the-art models for both fidelity and utility evaluation, e.g., it increases over 69% and 55% in distance and radius metrics on the FS-TKY dataset.

Code — https://github.com/Rongchao98/GeoGen

Introduction

Fine-grained location-based social network (LBSN) check-in trajectories are essential for many real-world applications such as next Point of Interest (POI) recommendation (Jiang et al. 2024), urban mobility understanding (Wang et al. 2019), business location selection (Xie et al. 2016), advertising (Jeon et al. 2021), and pandemic intervention (Hao et al. 2020). However, accessing large-scale LBSN data has become increasingly challenging due to high collection/purchasing costs and growing privacy concerns. Therefore, to support large-scale LBSN data access, it is crucial to develop methods that can generate synthetic LBSN check-in trajectories that maintain high fidelity and utility while safeguarding user privacy.

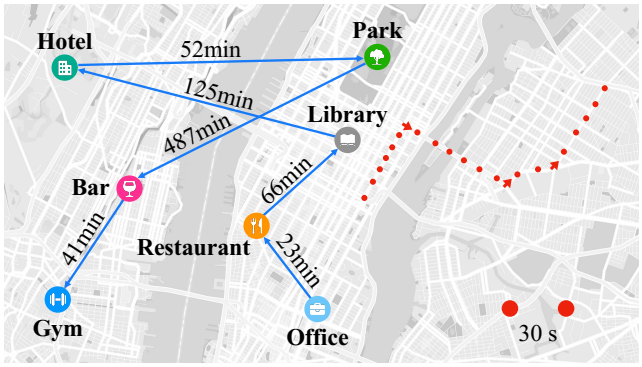


Figure 1: Example of a discrete LBSN trajectory with irregular intervals (blue); and a continuous GPS trajectory with a fixed interval of 30 seconds (red).

To date, some methods have been proposed (Isaacman et al. 2012; Jiang et al. 2016; Yin et al. 2017; Ouyang et al. 2018; Zhu et al. 2023b) to generate LBSN check-in trajectory data. For instance, Generative adversarial networks (GANs) (Ouyang et al. 2018; Feng et al. 2020; Yu et al. 2017; Yuan et al. 2022) have demonstrated substantial advances in generating LBSN check-in trajectories. However, these approaches inevitably suffer from an unstable and extremely long training procedure. Recently, the successes of diffusion models across modalities, including image (Dhariwal and Nichol 2021; Kulikov et al. 2023) and audio (Huang et al. 2022) generation, have spurred investigations into trajectory generation. Several diffusion-based models (Zhu et al. 2023b, 2024; Liu et al. 2023) have been proposed to generate spatially continuous, temporally regular GPS trajectories with fixed intervals between consecutive data points and fixed lengths (e.g., 100 data points per trajectory). These models have demonstrated strong performance, highlighting the great potential of diffusion models for gen-

*Prof. Guang Wang is the corresponding author.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

erating LBSN check-in trajectories.

However, it is nontrivial to adapt diffusion models to LBSN check-in trajectory data generation due to its two unique natures: (i) **Temporally irregular**: As shown in Figure 1, in contrast to regularly sampled GPS trajectories, LBSN check-in trajectories exhibit complex spatio-temporal patterns due to uncertain human mobility and check-in behavior, leading to *variable time intervals* between points and *unfixed trajectory lengths*. (ii) **Spatially discrete**: Diffusion models are well suited to *fixed-length* data in the *continuous* domain, whereas POIs lie in a discrete spatial space. Consequently, adapting models designed for regularly sampled trajectory generation (Zhu et al. 2023b, 2024; Liu et al. 2023) to LBSN check-in trajectories is challenging: a large sampling interval (e.g., 60 minutes) degrades fidelity, while a small interval (e.g., 1 minute) severely reduces efficiency.

To address these challenges, we propose GeoGen, a two-stage, coarse-to-fine framework for synthetic LBSN trajectory data generation. In the first stage, to adapt diffusion models to irregularly-sampled LBSN trajectory data, we reconstruct spatially continuous and temporally regular latent movement sequences from LBSN check-in trajectories through interpolation with a pre-determined coarse-grained interval. We then design a Sparsity-aware Spatio-temporal Diffusion model (S^2TDiff), which includes a novel and efficient Spatially-Aware Sparsely-Gated U-Net (SASG-UNet) as its denoising network. SASG-UNet replaces standard convolutions with hierarchical 1D blocks to capture multi-scale patterns and integrates a specialized *S^2G Attention* module to fit the unique characteristics of the latent movement sequences. In the second stage, we propose a Transformer-based Seq2Seq architecture Coarse2FineNet that auto-regressively generates fine-grained check-in points based on the coarse-grained latent movement sequences. Coarse2FineNet incorporates a POI Context-aware Encoder that captures rich contextual representations of the latent movement sequence, and a Multi-task Hybrid-Head Decoder that generates the next POI and its fine-grained timestamp. In the encoder, we introduce a *dynamic context fusion* mechanism that aligns the input latent movement sequence with POIs exhibiting diverse spatio-temporal characteristics, integrating them into unified contextual representations. In the decoder, instead of directly predicting the next check-in point, we leverage a temporal point process to model the user behavior uncertainty underlying LBSN check-in trajectories. This two-stage approach effectively resolves a key trade-off between computational efficiency and data quality, enabling the generation of large-scale synthetic LBSN check-in trajectories.

The key contributions of this paper are as follows:

- Conceptually, we focus on developing innovative diffusion models to efficiently generate large-scale fine-grained LBSN check-in trajectories, which are inherently spatially discrete and temporally irregular.
- Technically, we propose a two-stage coarse-to-fine framework called GeoGen for synthetic LBSN trajectory generation. In the first stage, we design a novel S^2TDiff model with an efficient SASG-UNet as a denoising network to generate coarse-grained latent movement sequences. In the second stage, we design a Transformer-based Seq2Seq architecture, Coarse2FineNet, which includes a POI Context-aware Encoder and a Multi-task Hybrid-head Decoder to generate fine-grained check-ins at precise, non-uniform timestamps.
- Experimentally, we conduct extensive experiments on four real-world LBSN check-in datasets. The results demonstrate that GeoGen significantly outperforms state-of-the-art models in data fidelity and show high utility for downstream tasks like next location prediction.

Preliminary

Problem Statement

Definition 1. (LBSN Check-in Trajectories). An LBSN check-in trajectory is a sequence of POI visits from an individual, denoted $s = [x_1, x_2, \ldots]$. The ith element x_i is a check-in represented as the tuple $x_i = (p_i, t_i)$, where $p_i \in P$ is a POI ID from a local finite set P with associated geographic coordinate g_{p_i}, and t_i is the check-in timestamp.

Definition 2. (Trajectory Granularity) The **granularity** of a trajectory is its smallest recorded time unit. *Coarse-grained trajectories* use large time units (e.g., hours), while *fine-grained trajectories* use smaller ones (e.g., minutes). This work aims to generate fine-grained LBSN check-in trajectories with a minute-level granularity.

Definition 3. (LBSN Check-in Trajectory Generation). Given a real-world dataset containing fine-grained LBSN check-in trajectories, denoted as $S = [s_1, s_2, \ldots]$, the LBSN check-in trajectory generation task aims to generate a synthetic trajectory dataset $\hat{S} = [\hat{s}_1, \hat{s}_2, \ldots]$ that maintains the spatio-temporal characteristics and distributions of S, as well as high data utility for downstream applications.

Denoising Diffusion Probabilistic Model

Denoising Diffusion Probabilistic Models (Ho, Jain, and Abbeel 2020) encompass two processes: (i) a forward diffusion process that gradually introduces noise into the data, and (ii) a backward diffusion process that reconstructs the original data from its noisy state. The forward diffusion process is a non-trainable Markov chain that transforms real sample $\mathbf{x}^0$ into latent variables $\mathbf{x}^1, \ldots, \mathbf{x}^N$, which can be represented as:

$$q(\mathbf{x}^n|\mathbf{x}^{n-1}) = \mathcal{N}(\mathbf{x}^n; \sqrt{1-\beta_n}\mathbf{x}^{n-1}, \beta_n\mathbf{I}), \qquad (1)$$

where $\beta_1, \ldots, \beta_N \in (0,1)$ are pre-determined variance schedules. $\mathbf{x}^n$ can be directly sampled from $\mathbf{x}^0$ with $q(\mathbf{x}^n|\mathbf{x}^0) = \mathcal{N}(\mathbf{x}^n; \sqrt{\overline{\alpha}_n}\mathbf{x}^0, (1-\overline{\alpha}_n)\mathbf{I})$, where $\alpha_n = 1-\beta_n$ and $\overline{\alpha}_n = \prod_{i=1}^{N} \alpha_i$. With the application of the reparameterization, $\mathbf{x}^n$ can be represented as $\mathbf{x}^n = \sqrt{\overline{\alpha}_n}\mathbf{x}^0 + \sqrt{1-\overline{\alpha}_n}\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$. The backward diffusion process is a trainable Markov chain that aims to recover $\mathbf{x}^0$ from $\mathbf{x}^N$, which can be formulated as:

$$p_\theta(\mathbf{x}^{n-1}|\mathbf{x}^n) = \mathcal{N}(\mathbf{x}^{n-1}; \mu_\theta(\mathbf{x}^n, n), \sigma_\theta^2(\mathbf{x}^n, n)\mathbf{I}), \qquad (2)$$

where $\mu_\theta(\mathbf{x}^n, n)$ and $\sigma_\theta(\mathbf{x}^n, n)$ are the mean and variance predicted normally by a neural network parameterized by θ. The loss function is formalized as:

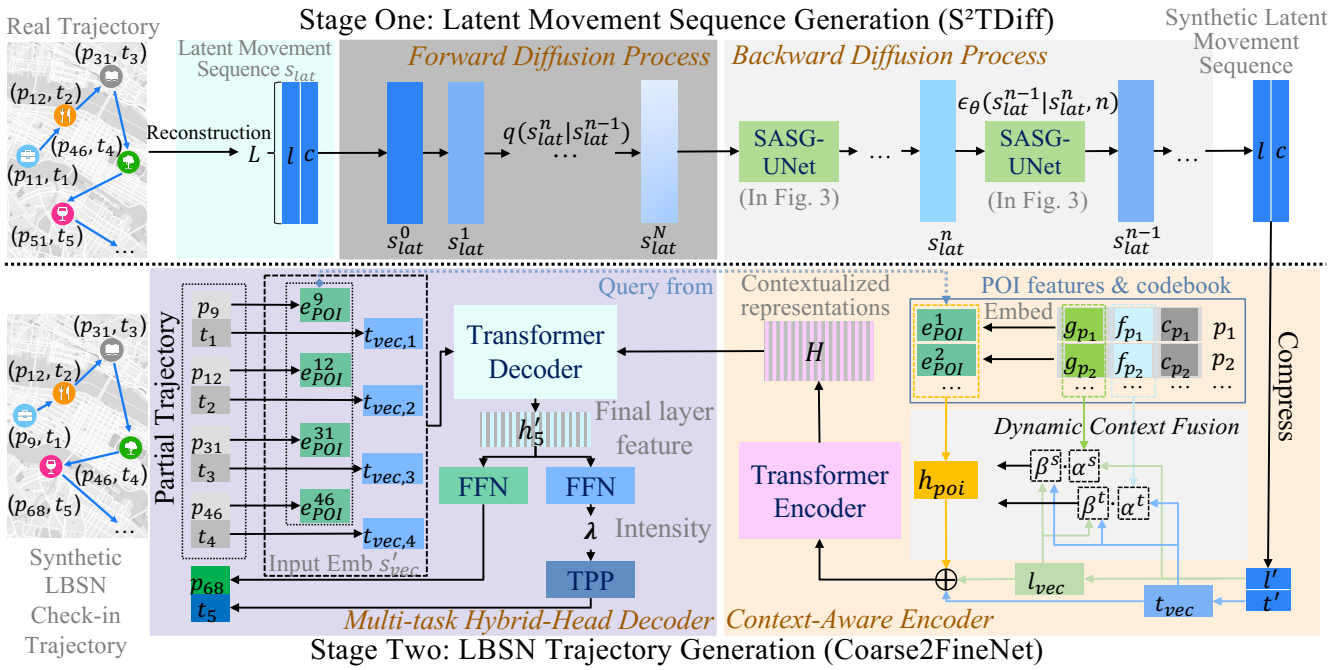


Figure 2: The overall framework of our proposed GeoGen. In the first stage, within S^{2}TDiff, each reconstructed latent movement sequence s_{lat} is first transformed into pure noise s_{lat}^N. During inference, SASG-UNet progressively denoises a sampled noise sequence to generate a synthetic latent trajectory. In the second stage, the context-aware encoder of Coarse2FineNet extracts contextualized representations from the generated latent movement sequence. The decoder then auto-regressively generates each POI and fine-grained timestamp pair, conditioned on both the contextual representations and the partially generated LBSN check-in trajectory. FFN and TPP refer to the feed-forward network and temporal point process, respectively.

$$\mathbb{E}_{\mathbf{x}^0,\boldsymbol{\epsilon},n}\left[\left\|\boldsymbol{\epsilon}-\epsilon_\theta\left(\sqrt{\overline{\alpha}_n}\mathbf{x}^0+\sqrt{1-\overline{\alpha}_n}\,\boldsymbol{\epsilon},\,n\right)\right\|^2\right]. \tag{3}$$

where ϵ_θ is a neural network for predicting sampled $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$. After training, trajectory generation is conducted by progressively sampling $\mathbf{x}^{n-1}$ from distribution $p_\theta(\mathbf{x}^{n-1}|\mathbf{x}^n)$ until reach $\mathbf{x}^0$ by computing:

$$\mathbf{x}^{n-1}=\frac{1}{\sqrt{\alpha_n}}\left(\mathbf{x}^n-\frac{\beta_n}{\sqrt{1-\overline{\alpha}_n}}\epsilon_\theta(\mathbf{x}^n,n)\right)+\sigma_\theta(\mathbf{x}^n,n)\,\mathbf{z}, \tag{4}$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ for $n \in [2, N]$, and $\mathbf{z} = \mathbf{0}$ when $n = 1$.

Methodology

In this part, we introduce the detailed design of GeoGen. An overall pipeline of GeoGen is shown in Figure 3.

Coarse-grained Latent Movement Sequence Generation

To adapt diffusion models to irregularly-sampled LBSN trajectory data, we first convert original fine-grained LBSN trajectories into coarse-grained spatially continuous and temporally regular latent movement sequences. We then design a Sparsity-aware Spatio-temporal Diffusion model (S^{2}TDiff) to generate synthetic coarse-grained latent movement sequences by learning underlying spatio-temporal distributions and physical constraints.

Coarse-grained Latent Movement Sequence Reconstruction Given a time interval I, we divide the total duration into L slots, where $L = [D/I]$, D is the duration. We reconstruct each real LBSN check-in trajectory s into a latent movement sequence, $s_{lat} = [(l_1, c_1), \ldots, (l_L, c_L)]$. For each slot i, the latent coordinate l_i is the geographic mean of all check-ins P_i falling within that time window ($l_i = \frac{1}{|P_i|}\sum_{(p_j,t_j)\in P_i} g_{p_j}$), and the intensity c_i is their count ($c_i = |P_i|$). Next, we adopt *linear interpolation* between the nearest preceding (i_p) and subsequent (i_n) time slots with no check-ins ($c_i = 0$), where the missing location is calculated as $l_i = l_{i_p} + ((i - i_p)/(i_n - i_p)) \cdot (l_{i_n} - l_{i_p})$. For empty slots at the sequence boundaries, we apply *circular interpolation*, treating the trajectory as a loop to infer locations between the first and last observed points. This reformulation transforms sparse event data into a ***spatially continuous*** and ***temporally regular*** format, enabling compatibility with diffusion models while reducing memory and computational costs via temporal compression.

Sparsity-aware Spatio-temporal Diffusion Model To capture the underlying spatio-temporal patterns, we propose a Sparsity-aware Spatio-temporal Diffusion model (S^{2}TDiff), trained with latent movement sequences $S_{\text{lat}} = [s_{\text{lat},1}, s_{\text{lat},2}, \ldots]$ as input. Although linear interpolation yields spatially continuous sequences, it poses challenges of (1) *diverse movement rates* and (2) *sparse intensity*.

To address these challenges, we design a novel and effi-

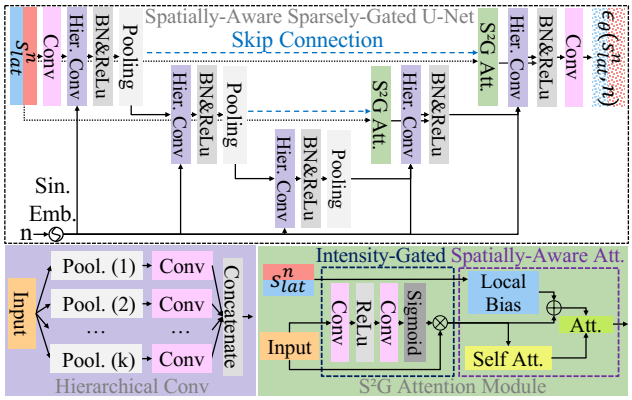


Figure 3: The architecture of the proposed Spatially-Aware Sparsely-Gated U-Net (SASG-UNet). The kernel sizes for the average pooling layers in the Hierarchical 1D Convolution layers are indicated in parentheses.

cient *Spatially-Aware Sparsely-Gated U-Net* (SASG-UNet) as the denoising network in S^{2}TDiff. S^{2}TDiff employs a U-Net backbone with hierarchical 1D convolution blocks that replace standard 2D convolutions. These blocks efficiently capture multi-resolution features using parallel branches of multi-scale average pooling and fixed-size 1D convolutions. At training/inference step n, the denoising network SASG-UNet ϵ_θ takes a noisy latent movement sequence s^n_{lat} as input and outputs the predicted noise: $\epsilon_\theta(s^n_{\text{lat}}, n)$. To further tackle sparse intensity and diverse movement rates, we introduce a novel S^{2}G Attention module within SASG-UNet's skip connections, consisting of two complementary pathways: the ***Intensity-Gated Pathway*** and the ***Spatially-Aware Attention Pathway***. Firstly, the Intensity-Gated Pathway handles sparse signals by with a lightweight convolutional sub-network to generate a dynamic gating signal $g = \sigma(\text{Conv1D}(\text{ReLU}(\text{Conv1D}(h_{\text{enc}}))))$, where h_{enc} denotes encoder features and σ is a sigmoid function. This gate adaptively modulates the features, allowing the network to prioritize learning from informative, sparse, non-zero intensity values. Secondly, the Spatially-Aware Attention Pathway accounts for diverse movement rates via a *Local Spatial Bias* b, where at index i, $b_i = \mathcal{F}_{\text{emb}}(d_{\text{H}}((s^n_{\text{lat},i}, s^n_{\text{lat},i-1}))$, and d_{H} is the Haversine distance (Maria et al. 2020) between consecutive points $s^n_{\text{lat},i}$, $s^n_{\text{lat},i-1}$ in the noisy input and $\mathcal{F}_{\text{emb}}$ is a learnable feed-forward network. Next, we add an attention mechanism, yielding the final refined features: $h_{\text{att}} = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$, where Q and K are the self-attention output features of the Intensity-Gated Pathway and V is the bias b. This mechanism provides the model with a direct and efficient understanding of real-world movement dynamics, enabling it to handle sparse, irregular mobility patterns effectively.

Coarse2FineNet for Fine-grained LBSN Trajectory Generation with Context Awareness

In the second stage, we propose **Coarse2FineNet**, a Transformer-based Seq2Seq framework that translates coarse-grained latent movement sequences generated by S^{2}TDiff into fine-grained LBSN check-in trajectories by jointly modeling the spatio-temporal features of the latent movement sequence and characteristics of POIs. Coarse2FineNet includes two key components: a Context-Aware Encoder and a Multi-task Hybrid-Head Decoder. The encoder captures a rich, contextualized representation of user mobility from coarse-grained latent movement sequences, while the decoder autoregressively generates a sequence of POI visits with fine-grained timestamps.

Context-Aware Encoder To reduce complexity in the diffusion process, we compress a latent sequence $s_{\text{latent}} = \{(l_i, c_i)\}$ into $s'_{\text{latent}} = \{(l'_i, t'_i) \mid c_i \geq \gamma\}$ by filtering points, where $l'_j = l_{i_j}$ and $t'_j = t_{i_j} = (i_j + 0.5) \cdot I$, γ denotes intensity threshold, and t_i denotes the midpoint time of slot i. We first obtain spatial embeddings $l_{vec} = Linear(l'_i)$ and temporal embedding $t_{vec} = T2V(t'_i)$, where $T2V$ denotes the Time2Vec method (Kazemi et al. 2019) that captures both linear and cyclical temporal patterns.

To establish local POI context awareness, we design a ***Dynamic Context Fusion*** (DCF) module within the encoder. In DCF, a shared POI embedding codebook is built to encode multiple attributes, including geographic coordinates g_{p_j}, temporal visit frequency vectors f_{p_j}, and category one-hot vector c_{p_j} for POI p_j: $e^{(j)}_{\text{POI}} = W_{\text{latlon}} g_{p_j} + W_{\text{freq}} f_{p_j} + W_{\text{cat}} c_{p_j}$, where $W_{\text{latlon}}, W_{\text{freq}}, W_{\text{cat}}$ are learnable weights. Then, we employ a dual attention mechanism to model spatial and temporal relationships separately. The spatial attention maps each coordinate l'_i from the filtered sequence to a distribution over POIs based on proximity. The attention weight α_j for each POI j and coordinate l_i is calculated as:

$$\alpha^{(s)}_{i,j} = \frac{\exp(-\|l'_i - g_{p_j}\|/\tau_s)}{\sum_{k=1}^{|\mathcal{P}|} \exp(-\|l'_i - g_{p_k}\|/\tau_s)} \tag{5}$$

where τ_s is a temperature parameter that controls the distribution's sharpness. Concurrently, the temporal attention captures time-specific visit patterns. The temporal embedding $t'_{vec,i}$ is projected into a query vector $w_i = W_t t'_{vec,i}$, and the attention weight is computed via dot-product similarity:

$$\alpha^{(t)}_{i,j} = \frac{\exp(f_{p_j} \cdot w_i)}{\sum_{k=1}^{|\mathcal{P}|} \exp(f_{p_k} \cdot w_i)} \tag{6}$$

Finally, we adaptively balance the spatial and temporal signals by computing fusion weights using a softmax over their average representations $(\beta_s, \beta_t) = \text{Softmax}\left(\mathbf{W}_\alpha \left[\frac{1}{L}\sum_{i=1}^{L} \mathbf{l}_i \,\|\, \frac{1}{L}\sum_{i=1}^{L} \mathbf{t}_i\right]\right)$, where $\mathbf{l}_i$ and $\mathbf{t}_i$ denote the i-th spatial and temporal feature vectors, respectively; $\mathbf{W}_\alpha$ is a learnable linear projection matrix that is used to create a unified attention distribution $\boldsymbol{\alpha}_{i,j} = \beta_s \boldsymbol{\alpha}^{(s)}_{i,j} + \beta_t \boldsymbol{\alpha}^{(t)}_{i,j}$. This fused attention is applied to the shared POI embeddings to create a context-aware representation $h_{\text{poi},i} = \sum_{j=1}^{|\mathcal{P}|} \alpha_{i,j} e^{(j)}_{\text{POI}}$. This POI-enhanced vector h_{poi} is integrated with the initial features $h_{vec} = [l_{vec}, t_{vec}]$ through a residual connection, and is processed by a Transformer Encoder to produce the contextualized representations H.

Method \ Metric	Distance	Radius	Interval	Length	Average	Distance	Radius	Interval	Length	Average
	FS-NYC					FS-TKY				
SMM (Maglaras and Katsaros 2015)	0.196	0.259	0.381	0.358	0.299	0.178	0.277	0.217	0.329	0.250
TimeGEO (Jiang et al. 2016)	0.251	0.625	0.278	0.386	0.385	0.277	0.524	0.232	0.216	0.312
Hawkes (Laub, Taimre, and Pollett 2015)	0.190	0.517	0.339	0.253	0.325	0.157	0.315	0.269	0.281	0.256
LSTM (Rossi et al. 2021)	0.166	0.441	0.175	0.235	0.254	0.184	0.292	0.230	0.257	0.241
SeqGAN (Yu et al. 2017)	0.116	0.364	0.150	0.291	0.230	0.108	0.292	0.207	0.276	0.221
MoveSim (Feng et al. 2020)	0.080	0.289	0.133	0.391	0.223	0.079	0.256	0.114	0.353	0.201
DiffTraj (Zhu et al. 2023b)	**0.051**	0.085	0.104	0.320	0.140	0.276	0.201	0.116	0.744	0.334
ControlTraj (Zhu et al. 2024)	0.193	0.356	0.160	0.337	0.262	0.049	0.096	0.191	0.223	0.140
Our Method (ours)	0.079	**0.065**	**0.067**	**0.127**	**0.085**	**0.015**	**0.043**	**0.081**	**0.120**	**0.065**
	FS-ATX					GW-STO				
SMM (Maglaras and Katsaros 2015)	0.603	0.625	0.139	0.347	0.429	0.637	0.599	0.697	0.428	0.590
TimeGEO (Jiang et al. 2016)	0.416	0.501	0.148	0.401	0.367	0.579	0.489	0.727	0.397	0.548
Hawkes (Laub, Taimre, and Pollett 2015)	0.517	0.556	0.261	0.521	0.464	0.621	0.496	0.579	0.433	0.532
LSTM (Rossi et al. 2021)	0.589	0.602	0.187	0.339	0.429	0.618	0.527	0.699	0.369	0.553
SeqGAN (Yu et al. 2017)	0.452	0.475	**0.029**	0.292	0.312	0.531	0.462	0.745	0.413	0.538
MoveSim (Feng et al. 2020)	0.306	0.382	0.078	0.275	0.260	0.181	0.265	0.318	0.138	0.226
DiffTraj (Zhu et al. 2023b)	0.061	0.104	0.241	0.702	0.277	**0.114**	0.275	0.802	0.306	0.374
ControlTraj (Zhu et al. 2024)	0.108	0.217	0.150	0.286	0.190	0.204	0.253	0.765	0.141	0.341
Our Method (ours)	**0.058**	**0.088**	0.036	**0.200**	**0.096**	0.136	**0.197**	**0.292**	**0.080**	**0.176**

Table 1: Fidelity Evaluation. The best results on each dataset are in **bold**, and the second-best results are underlined. 'Average' denotes the mean value of the four metrics.

Multi-task Hybrid-Head Decoder The Multi-task Hybrid-Head Decoder autoregressively generates a sequence of (POI, timestamp) pairs, conditioned on the contextual representations H from the encoder. To generate the n'-th check-in point, the previously generated sequence $s'_{syn} = [(p_1, t_1), \ldots, (p_{n'-1}, t_{n'-1})]$ is first embedded into a sequence of vectors s'_{vec}, where each POI is replaced by its shared embedding and each timestamp is converted using Time2Vec. Next, a Transformer-based Decoder processes this embedded sequence and the encoder's output H to produce a hidden state $h_{n'}$, which summarizes the necessary history and context. This hidden state is then fed into two specialized prediction heads to determine *where* and *when* the next check-in occurs. For POI, a linear layer computes logits over the vocabulary. For timestamps, unlike direct regression, we adopt a *Neural Temporal Point Process* (NTPP) (Shchur et al. 2021) to capture the irregular timing of human mobility. The NTPP models the probability distribution of the next timestamp through a conditional intensity function $\lambda(t) = \text{Softplus}(W_{\text{time}} h_{n'} + b_{\text{time}})$, which defines the instantaneous event rate given the current context. To sample the next timestamp, we first derive the cumulative distribution function F^{-1} corresponding to $\lambda(t)$, and then draw a sample from this distribution via inverse transform sampling:

$$t_{n'} = F^{-1}(u), \quad \text{where } u \sim \mathcal{U}(0, 1). \tag{7}$$

This enables the model to generate diverse and realistic non-uniform time intervals, capturing the stochastic nature of real-world check-in behaviors.

Training Objective Coarse2FineNet is trained with teacher forcing and causal masking (Vaswani et al. 2017). Our training objective is formulated as:

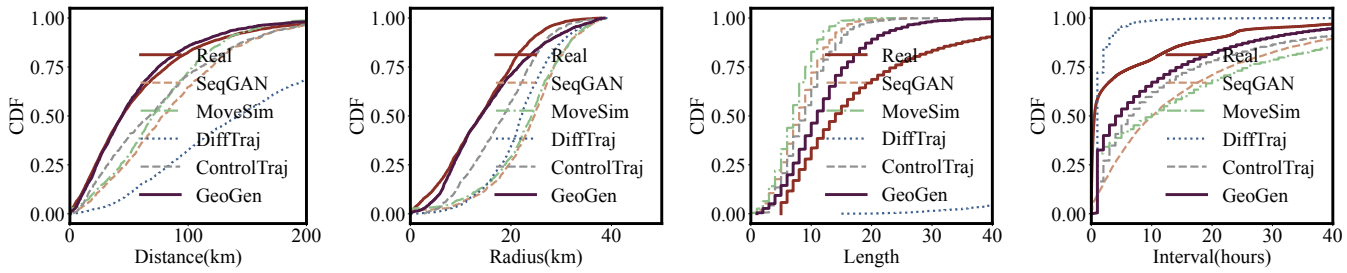
$$\mathcal{L} = \lambda_{\text{POI}} \mathcal{L}_{\text{POI}} + \lambda_{\text{time}} \mathcal{L}_{\text{time}} + \lambda_{\text{spatial}} \mathcal{L}_{\text{spatial}}. \tag{8}$$

The POI prediction loss $\mathcal{L}_{\text{POI}} = -\sum_{t=1}^{T} y_t \log(\hat{y}_t)$ enforces categorical accuracy through cross-entropy, where T represents the number of check-ins in a trajectory and y_t, $\hat{y}_t$ denote ground-truth and predicted POI ID respectively. The temporal loss is the negative log-likelihood objective for the neural temporal point processes, formulated as $\mathcal{L}_{\text{time}} = -\sum_{i=1}^{T} (\log \lambda_i - \lambda_i \Delta t_i)$, where λ_i is the intensity predicted by the network for the i-th interval, and Δt_i is the ground-truth time duration since the last check-in. This loss maximizes the probability of the observed check-in sequence by training the model to predict a high intensity just before a check-in occurs. The spatial consistency loss $\mathcal{L}_{\text{spatial}} = \frac{1}{T} \sum_{t=1}^{T} \|g_{p_t} - g_{p_{\hat{t}}}\|_2$ penalizes geographic implausibility by measuring Euclidean distance between predicted POI $p_{\hat{t}}$ and ground-truth POI p_t. The three components are balanced by three learnable weights, enabling the model to generate trajectories that are spatially plausible, temporally accurate, and categorically consistent.

Evaluation

Experiment Setup

Datasets We utilize four publicly available LBSN datasets containing POI check-in trajectories to evaluate the performance of our proposed GeoGen, including New York City (FS-NYC), Tokyo (FS-TKY), Austin (FS-ATX) from the Foursquare dataset(Yang et al. 2014; Yang, Zhang, and Qu 2016), and Stockholm (GW-STO) from the Gowalla dataset (Cho, Myers, and Leskovec 2011). We first truncate each individual LBSN check-in trajectory with a pre-defined duration (one week for the three FourSquare datasets and six weeks for GW-STO) and maintain the sequences with more than 5 visits to ensure data quality. We split each dataset into training, validation, and testing sets with a ratio of 7:1:2.



(a) CDF of distance (b) CDF of radius (c) CDF of length (d) CDF of interval

Figure 4: Performance comparison of different methods on FS-TKY dataset.

Baselines We compare our GeoGen with the following eight state-of-the-art models: SMM (Maglaras and Katsaros 2015), TimeGEO (Jiang et al. 2016), Hawkes (Laub, Taimre, and Pollett 2015), LSTM (Rossi et al. 2021), SeqGAN (Yu et al. 2017), MoveSim (Feng et al. 2020), DiffTraj (Zhu et al. 2023b), ControlTraj (Zhu et al. 2024).

Metrics We evaluate the synthetic data quality from both *fidelity* and *utility* perspectives. Following common practice (Zhu et al. 2023b; Feng et al. 2020), we evaluate data fidelity by computing the Jensen-Shannon Divergence (JSD) (Lin 1991) between the distributions of synthetic and real data across four metrics from both spatial and temporal aspects: 1) **Distance**: the sum of distance between consecutive check-ins; 2) **Radius**: the radius of gyration, which measures the trajectory's spatial range; 3) **Interval**: the time interval between consecutive check-ins; 4) **Length**: the total number of points in the trajectory. A lower JSD score indicates a higher fidelity of the synthetic data. We evaluate the utility of synthetic data by conducting data-driven downstream applications.

Overall Performance

Fidelity Evaluation The comprehensive quantitative results are presented in Table 1. We found that traditional rule-based methods (SMM, TimeGEO, Hawkes) generally fall short, as their reliance on fixed statistical assumptions prevents them from capturing the complex, long-term behaviors from POI check-in trajectories. Designed for spatially continuous and temporally regular trajectory data generation, DiffTraj shows competitive performance on spatial metrics on dense urban datasets like FS-NYC, but performs poorly on the length metric due to the lack of awareness of discrete POI locations. Although the ControlTraj can model the connection between POIs as road segments, its incompatibility with temporally irregular data degrades the LBSN check-in trajectory generation quality. In contrast, our proposed model, GeoGen, significantly outperforms baselines across the four diverse datasets, highlighting its robustness and generalizability to generate high-fidelity LBSN trajectory data. For instance, on the FS-TKY dataset, GeoGen improves the second-best method by over 69% in the Distance metric and 55% in the Radius metric. Similarly, it achieves a 38% improvement in Length on the GW-STO dataset.

The Cumulative Distribution Function (CDF) curves in Figure 4 for FS-TKY show that GeoGen's distributions most closely align with the original data, confirming its effectiveness in capturing realistic LBSN check-in behavior patterns.

Utility Evaluation As shown in Table 2, we evaluate the utility of the synthetic data through a next check-in prediction task (Yuan et al. 2023), where a prediction model is trained on synthetic data and tested on real-world data. Root Mean Square Error (RMSE) and Euclidean Distance (ED) are used to measure the prediction performance from the temporal dimension and spatial dimension, respectively. Our results demonstrate that GeoGen achieves the closest performance to the real data across both temporal and spatial metrics using FS-NYC and FS-TKY datasets, indicating its strong ability to preserve data utility. MoveSim performs better than SeqGAN and ControlTraj in the spatial dimension due to its explicit consideration of spatial structures such as distances between locations, while SeqGAN and ControlTraj treat locations as discrete categorical variables. Our proposed GeoGen consistently achieves the best performance across all metrics and datasets, thanks to its ability to effectively model spatially discrete and temporally irregular patterns of LBSN check-in trajectories.

Datasets	**FS-NYC**		**FS-TKY**	
Metrics / **Methods**	RMSE	ED	RMSE	ED
Real	0.187	7.01	0.235	13.70
SeqGAN	0.426	16.34	0.689	23.12
MoveSim	0.371	10.28	0.671	16.90
DiffTraj	0.424	15.27	0.728	18.28
ControlTraj	0.321	14.29	0.491	16.28
GeoGen	**0.225**	**9.21**	**0.352**	**15.82**

Table 2: Synthetic data utility for next check-in prediction.

In-depth Analysis

Ablation Study To validate our S^2TDiff in stage 1, we benchmark it against two other strategies: a) replication (S^2TDiff-Rep): This method fills empty time slots with the coordinates of the last known check-in. b) fixed point (S^2TDiff-Fix): This method assigns a predetermined central coordinate to all empty time slots. We also build a

variant S^2TDiff w/o S^2G where the core module S^2G Attention module is removed to test its effectiveness. For Coarse2FineNet in the second stage, we build a variant C2F w/o DCF by removing the dynamic context fusion component. The results on the FS-NYC dataset are shown in Table 3. We found that interpolation can provide richer spatio-temporal patterns and underlying physical constraints compared with replication and fix point latent movement sequence reconstruction strategies. Meanwhile, the performance decreased in both spatial and temporal domains without the S^2G Attention module, also proving its effectiveness compared with pure convolutions in the U-Net. In Coarse2FineNet, removing the dynamic context fusion and POI embedding (C2F w/o DCF) causes a dramatic performance decrease due to the lack of informative supervision for the learning of the discrete POI embeddings.

Methods \ Metrics	Distance	Radius	Interval	Length
S^2TDiff-Rep	0.084	0.071	0.074	0.128
S^2TDiff-Fix	0.097	0.073	0.079	0.130
S^2TDiff w/o S^2G	0.087	0.069	0.080	0.135
C2F w/o DCF	0.119	0.138	0.096	0.169
Complete GeoGen	0.079	0.065	0.067	0.127

Table 3: Performance comparison across different GeoGen variants on FS-NYC dataset.

Impact of Granularity on Model Performance and Efficiency We investigate the impact of time intervals I. As shown in Table 4, we found that smaller intervals (e.g., 2-4 hours) yield better temporal fidelity (Interval), whereas spatial fidelity metrics (Distance and Radius) are optimized at a larger 6-hour interval. Increasing the interval from 2 to 6 hours reduces Stage 1 memory consumption from 2.36 GB to 0.35 GB and increases throughput by more than fourfold (from $11,322$ to $50,540$ samples/sec). In the second stage, the computational cost of Coarse2FineNet remains relatively stable across different intervals. Consequently, a moderate I strikes a balance between generation quality and efficiency.

Metric \ I	2 hours	3 hours	4 hours	6 hours	8 hours	12 hours
Distance	0.454	0.183	0.136	**0.073**	0.089	0.080
Radius	0.402	0.250	0.197	**0.154**	0.153	0.154
Interval	**0.292**	**0.292**	**0.292**	0.315	0.342	0.444
Length	0.318	0.137	**0.080**	0.115	0.134	0.131
S1 Memory (GB)	2.36	1.14	0.69	0.35	0.21	0.12
S2 Memory (GB)	17.55	9.78	10.10	10.68	3.62	3.92
S1 Throughput	11322	20221	30058	50540	73794	119932
S2 Throughput	39	41	44	45	49	49

Table 4: Results under different coarse-grained interval I selection on GW-STO dataset with batch size 256 for S^2TDiff (S1) and 16 for Coarse2FineNet (S2).

Related Work

LBSN Data Mining

In recent years, LBSN data have attracted much interest from both research and industry communities (Ye, Yin, and Lee 2010; Ding et al. 2018) due to their high value for many practical applications such as POI recommendations (Luo, Liu, and Liu 2021), user behavior analysis (Ding et al. 2018), and advertising (Jeon et al. 2021). Yang et al. (Yang et al. 2024) design a self-explainable algorithm to recommend the next visiting POI with high accuracy based on LBSN data. Han et al. (Han et al. 2021) leverage LBSN data to understand human activities during COVID-19 for better pandemic control strategies. However, given the potential privacy concerns and high collection costs, it becomes highly challenging for researchers to access large-scale LBSN data (Long et al. 2024; Cai et al. 2024). Hence, in this study, we focus on generating large-scale synthetic LBSN data that shares similar data characteristics with real data and also has high utility for downstream applications.

Trajectory Data Generation

The recent advances in generative models provide a great opportunity for synthetic trajectory data generation. These models (e.g., Generative Adversarial Network (GAN) (Rao et al. 2020; Liu, Chen, and Andris 2018; Cao and Li 2021) and diffusion models (Zhu et al. 2023b)) utilize the power of deep neural networks to learn and replicate the underlying spatio-temporal distributions of real trajectory data. Once trained, these models generate synthetic trajectories by sampling from the learned distributions. For example, Feng et al. (Feng et al. 2020) propose a GAN-based framework, MoveSim, with the introduction of prior knowledge for human mobility trajectory generation. MoveSim (Feng et al. 2020) generates the next location based on the last point in the partially generated sequence and iteratively generates the rest trajectory points. Zhu et al. (Zhu et al. 2023b,a, 2024) propose a pioneering work by adopting the diffusion model to generate continuous and road network-constrained trajectories with a fixed length and time interval between two trajectory points. Different from existing works that usually focus on generating trajectories in continuous spatial spaces with fixed time intervals and fixed lengths, our work aims to generate spatially discrete, temporally irregular LBSN check-in trajectories.

Conclusion

In this paper, we propose GeoGen, a novel two-stage coarse-to-fine framework for the generation of synthetic LBSN check-in trajectories that are characterized by spatial discreteness and temporal irregularity. In the first stage, we propose a Sparsity-aware Spatio-temporal Diffusion model (S^2TDiff) with an efficient Spatially-Aware Sparsely-Gated U-Net (SASG-UNet) to learn multi-scale behavioral patterns. In the second stage, we design Coarse2FineNet, a Transformer-based Seq2Seq architecture equipped with a Dynamic Context Fusion mechanism in the encoder and a multi-task hybrid-head decoder to generate accurate next POI location and its fine-grained timestamp. We extensively evaluate our GeoGen with four datasets, and results show that GeoGen outperforms state-of-the-art models from both data fidelity and utility perspectives, e.g., it increases 69% and 55% on distance and radius metrics on the FS-TKY dataset, and enhances the next check-in prediction task.

Acknowledgments

We sincerely thank all anonymous reviewers for their insightful comments and valuable suggestions. This work is partially supported by the National Science Foundation (NSF) Awards 2411151, 2411152, 2411153, National Artificial Intelligence Research Resource (NAIRR) 240332, and FSU/AWS Computer Support Seed Fund.

References

Cai, K.; Zhang, J.; Shand, W.; Hong, Z.; Wang, G.; Zhang, D.; Chi, J.; and Tian, Y. 2024. Where have you been? A Study of Privacy Risk for Point-of-Interest Recommendation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1–12.

Cao, C.; and Li, M. 2021. Generating mobility trajectories with retained Data Utility. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2610–2620.

Cho, E.; Myers, S. A.; and Leskovec, J. 2011. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1082–1090.

Dhariwal, P.; and Nichol, A. 2021. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34: 8780–8794.

Ding, Z.; Li, X.; Jiang, C.; and Zhou, M. 2018. Objectives and State-of-the-Art of Location-Based Social Network Recommender Systems. *ACM Comput. Surv.*, 51(1).

Feng, J.; Yang, Z.; Xu, F.; Yu, H.; Wang, M.; and Li, Y. 2020. Learning to simulate human mobility. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 3426–3433.

Han, Z.; Fu, H.; Xu, F.; Tu, Z.; Yu, Y.; Hui, P.; and Li, Y. 2021. Who will survive and revive undergoing the epidemic: Analyses about POI visit behavior in Wuhan via check-in records. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(2): 1–20.

Hao, Q.; Chen, L.; Xu, F.; and Li, Y. 2020. Understanding the urban pandemic spreading of covid-19 with real world mobility data. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 3485–3492.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Huang, R.; Zhao, Z.; Liu, H.; Liu, J.; Cui, C.; and Ren, Y. 2022. Prodiff: Progressive fast diffusion model for high-quality text-to-speech. In *Proceedings of the 30th ACM International Conference on Multimedia*, 2595–2605.

Isaacman, S.; Becker, R.; Cáceres, R.; Martonosi, M.; Rowland, J.; Varshavsky, A.; and Willinger, W. 2012. Human mobility modeling at metropolitan scales. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, 239–252.

Jeon, J.; Kang, S.; Jo, M.; Cho, S.; Park, N.; Kim, S.; and Song, C. 2021. Lightmove: A lightweight next-poi recommendation for taxicab rooftop advertising. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 3857–3866.

Jiang, N.; Yuan, H.; Si, J.; Chen, M.; and Wang, S. 2024. Towards Effective Next POI Prediction: Spatial and Semantic Augmentation with Remote Sensing Data. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, 5061–5074. Los Alamitos, CA, USA: IEEE Computer Society.

Jiang, S.; Yang, Y.; Gupta, S.; Veneziano, D.; Athavale, S.; and González, M. C. 2016. The TimeGeo modeling framework for urban mobility without travel surveys. *Proceedings of the National Academy of Sciences*, 113(37): E5370–E5378.

Kazemi, S. M.; Goel, R.; Gupta, K.; McAuley, J.; Kautz, J.; and Molchanov, P. 2019. Time2Vec: Learning a Vector Representation of Time. arXiv:1907.05321.

Kulikov, V.; Yadin, S.; Kleiner, M.; and Michaeli, T. 2023. Sinddm: A single image denoising diffusion model. In *International Conference on Machine Learning*, 17920–17930. PMLR.

Laub, P. J.; Taimre, T.; and Pollett, P. K. 2015. Hawkes processes. *arXiv preprint arXiv:1507.02822*.

Lin, J. 1991. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37(1): 145–151.

Liu, M.; Huang, H.; Feng, H.; Sun, L.; Du, B.; and Fu, Y. 2023. Pristi: A conditional diffusion framework for spatiotemporal imputation. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, 1927–1939. IEEE.

Liu, X.; Chen, H.; and Andris, C. 2018. trajGANs: Using generative adversarial networks for geo-privacy protection of trajectory data (Vision paper). In *Location privacy and security workshop*, 1–7.

Long, J.; Chen, T.; Ye, G.; Zheng, K.; Nguyen, Q. V. H.; and Yin, H. 2024. Physical Trajectory Inference Attack and Defense in Decentralized POI Recommendation. In *Proceedings of the ACM Web Conference 2024*, WWW '24, 3379–3387. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701719.

Luo, Y.; Liu, Q.; and Liu, Z. 2021. STAN: Spatio-Temporal Attention Network for Next Location Recommendation. In *WWW 2021*, 2177–2185.

Maglaras, L. A.; and Katsaros, D. 2015. Social clustering of vehicles based on semi-Markov processes. *IEEE Transactions on Vehicular Technology*, 65(1): 318–332.

Maria, E.; Budiman, E.; Taruk, M.; et al. 2020. Measure distance locating nearest public facilities using Haversine and Euclidean Methods. In *Journal of Physics: Conference Series*, volume 1450, 012080. IOP Publishing.

Ouyang, K.; Shokri, R.; Rosenblum, D. S.; and Yang, W. 2018. A non-parametric generative model for human trajectories. In *IJCAI*, volume 18, 3812–3817.

Rao, J.; Gao, S.; Kang, Y.; and Huang, Q. 2020. LSTM-TrajGAN: A deep learning approach to trajectory privacy protection. *arXiv preprint arXiv:2006.10521*.

Rossi, L.; Ajmar, A.; Paolanti, M.; and Pierdicca, R. 2021. Vehicle trajectory prediction and generation using LSTM models and GANs. *Plos one*, 16(7): e0253868.

Shchur, O.; Türkmen, A. C.; Januschowski, T.; and Günnemann, S. 2021. Neural temporal point processes: A review. *arXiv preprint arXiv:2104.03528*.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. In *Advances in neural information processing systems*, volume 30.

Wang, J.; Kong, X.; Xia, F.; and Sun, L. 2019. Urban human mobility: Data-driven modeling and prediction. *ACM SIGKDD explorations newsletter*, 21(1): 1–19.

Xie, M.; Yin, H.; Wang, H.; Xu, F.; Chen, W.; and Wang, S. 2016. Learning graph-based poi embedding for location-based recommendation. In *Proceedings of the 25th ACM international on conference on information and knowledge management*, 15–24.

Yang, D.; Zhang, D.; and Qu, B. 2016. Participatory cultural mapping based on collective behavior data in location-based social networks. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 7(3): 1–23.

Yang, D.; Zhang, D.; Zheng, V. W.; and Yu, Z. 2014. Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(1): 129–142.

Yang, K.; Yang, Y.; Gao, Q.; Zhong, T.; Wang, Y.; and Zhou, F. 2024. Self-Explainable Next POI Recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '24, 2619–2623. New York, NY, USA: Association for Computing Machinery. ISBN 9798400704314.

Ye, M.; Yin, P.; and Lee, W.-C. 2010. Location recommendation for location-based social networks. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '10, 458–461. New York, NY, USA: Association for Computing Machinery. ISBN 9781450304283.

Yin, M.; Sheehan, M.; Feygin, S.; Paiement, J.-F.; and Pozdnoukhov, A. 2017. A generative model of urban activities from cellular data. *IEEE Transactions on Intelligent Transportation Systems*, 19(6): 1682–1696.

Yu, L.; Zhang, W.; Wang, J.; and Yu, Y. 2017. Seqgan: Sequence generative adversarial nets with policy gradient. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31.

Yuan, Y.; Ding, J.; Shao, C.; Jin, D.; and Li, Y. 2023. Spatio-temporal diffusion point processes. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 3173–3184.

Yuan, Y.; Ding, J.; Wang, H.; Jin, D.; and Li, Y. 2022. Activity trajectory generation via modeling spatiotemporal dynamics. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 4752–4762.

Zhu, Y.; Ye, Y.; Wu, Y.; Zhao, X.; and Yu, J. 2023a. SynMob: Creating High-Fidelity Synthetic GPS Trajectory Dataset for Urban Mobility Analysis. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

Zhu, Y.; Ye, Y.; Zhang, S.; Zhao, X.; and Yu, J. 2023b. DiffTraj: Generating GPS Trajectory with Diffusion Probabilistic Model. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Zhu, Y.; Yu, J. J.; Zhao, X.; Liu, Q.; Ye, Y.; Chen, W.; Zhang, Z.; Wei, X.; and Liang, Y. 2024. Controltraj: Controllable trajectory generation with topology-constrained diffusion model. *arXiv preprint arXiv:2404.15380*.