Towards Robust Trajectory Representations: Isolating Environmental Confounders with Causal Learning

Kang Luo[1] , **Yuanshao Zhu**[1] , **Wei Chen**[1] , **Kun Wang**[2] ,
Zhengyang Zhou[2] , **Sijie Ruan**[3] , **Yuxuan Liang**[1*]
[1]Hong Kong University of Science and Technology (Guangzhou)
[2]University of Science and Technology of China
[3]Beijing Institute of Technology
{connorluo, yuxliang}@outlook.com, yuanshaozhu@hkust-gz.edu.cn, onedeanxxx@gmail.com,
wk520529@mail.ustc.edu.cn, zzy0929@ustc.edu.cn, sjruan@bit.edu.cn

Abstract

Trajectory modeling refers to characterizing human movement behavior, serving as a pivotal step in understanding mobility patterns. Nevertheless, existing studies typically ignore the confounding effects of geospatial context, leading to the acquisition of spurious correlations and limited generalization capabilities. To bridge this gap, we initially formulate a Structural Causal Model (SCM) to decipher the trajectory representation learning process from a causal perspective. Building upon the SCM, we further present a Trajectory modeling framework (*TrajCL*) based on Causal Learning, which leverages the backdoor adjustment theory as an intervention tool to eliminate the spurious correlations between geospatial context and trajectories. Extensive experiments on two real-world datasets verify that TrajCL markedly enhances performance in trajectory classification tasks while showcasing superior generalization and interpretability.

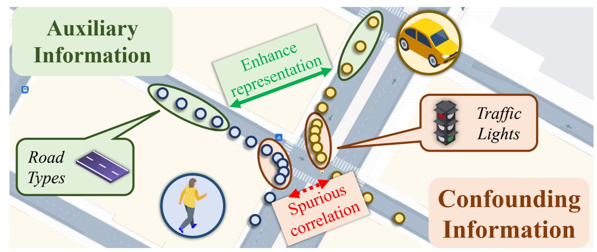

Figure 1: The impacts of geospatial context on trajectory modeling.

1 Introduction

Trajectory data has emerged as an indispensable resource for understanding human mobility patterns [Jin *et al.*, 2023]. Such data offers invaluable insights into various applications ranging from traffic management to personalized location-based services [Dai *et al.*, 2015; Chen *et al.*, 2022]. As a result, the modeling of this data is the cornerstone for transforming raw location information into mobility intelligence, thereby supporting various spatial-temporal applications, e.g., travel mode detection [Zheng *et al.*, 2008; Zhu *et al.*, 2021], next location prediction [Yin *et al.*, 2016; Lin *et al.*, 2021], and travel time estimation [Zheng, 2015].

Trajectory representation learning involves extracting useful, generalizable, and concise representations from the sequential data points of a human trajectory. Intuitively, its functionality extends to discerning the intrinsic motion properties inherent in trajectories. This pursuit is typically realized by deep sequential models, such as Recurrent Neural Networks [Wu *et al.*, 2017; Liu *et al.*, 2019] and Transformers [Liang *et al.*, 2022; Chang *et al.*, 2023], which effectively capture trajectory dynamics by encoding temporal intervals and spatio-temporal correlations [Liu and Lee, 2017; Qin *et al.*, 2019; Liu *et al.*, 2019; Liang *et al.*, 2021]. Furthermore, considering the geospatial context associated with trajectories, e.g., point of interests, road networks, recent endeavors are devoted to mining valuable insights from this auxiliary information [Guo *et al.*, 2020; Ferrero *et al.*, 2020; Pugliese *et al.*, 2023; Jiang *et al.*, 2023; Wang *et al.*, 2024].

While the integration of the geospatial context has the potential to enhance trajectory representations to a certain degree, it concurrently introduces a *confounding factor* into the learning process. This confounder poses the risk of our algorithm learning spurious correlations within the training data, leading to a degradation in performance and a compromised ability to generalize. In other words, the model is vulnerable to overfitting to specific environmental conditions. To elucidate this concern, consider an example in Figure 1. Vehicles frequently come to a halt in congested areas or traffic lights, exhibiting trajectory patterns (e.g., low speed) akin to those of pedestrians. In this scenario, there exists an increasing risk of the model displaying a pronounced inclination towards recognizing pedestrian patterns in congested areas. This, in turn, could lead to the *spurious correlation* (i.e., unwarranted association) of geospatial context with trajectory patterns.

In this paper, our target is to mitigate the impact of these confounding factors induced by the geospaital context, so as to extract *robust* and *domain-invariant* representations from human trajectories. Primarily, we present a *Structural Causal Model* (SCM) to deepen our comprehension of the trajectory

*Corresponding author. Email: yuxliang@outlook.com

representation learning process. From a causal perspective, the SCM elucidates the relationship among the environment (i.e., geospatial context[1]), the trajectory, and the representation outcome. The environment serves as a confounding factor that establishes a backdoor path between the input trajectory data and the resulting representation.

Building upon the SCM, we present a novel causal learning framework called **TrajCL** for learning robust trajectory representations. TrajCL utilizes the *backdoor adjustment* theory as an intervention tool to eliminate the spurious correlations between environment and trajectories, which is implemented by two key steps. Firstly, we design an environmental alignment module that leverages geospatial context to guide the encoders in disentangling causal and confounding representations. Secondly, we elaborately introduce a causal learning module to effectively accomplish causal intervention at the representation level, resulting in robust representations that exhibit strong generalization capabilities, particularly in few-shot learning or imbalanced sample learning scenarios.

Our major contributions can be summarized as follows:

- *A causal lens for trajectory data.* We propose a structural causal model to unravel the inherent rationale behind the learning process of trajectories. Based on this causal lens, a novel framework termed TrajCL is presented to enhance the robustness of trajectory representations.
- *Backdoor adjustment for isolating confounders.* Exploiting the backdoor adjustment theory, we initially design an environment alignment module to disentangle the causal and confounding elements from input data, and subsequently leverage causal intervention to learn robust representations.
- *Extensive empirical studies.* We conduct thorough experiments on the trajectory classification task over two mobility datasets. The results affirm that our framework enhances trajectory modeling in a plug-and-play manner, and possesses superior generalization ability and interpretability.

2 Related Work

Trajectory Modeling. Extensive research has been dedicated to trajectory modeling, aiming to uncover human mobility patterns. Early heuristic-based approaches only considered limited features. [Lee and Han, 2008] initially merged trajectory points within grid units, exploring the spatial characteristics of trajectory substructures. Building upon this, [Zheng *et al.*, 2008] and [Dodge *et al.*, 2009] respectively extracted local and global features from subgrids and trajectory points. Moreover, [Xiao *et al.*, 2017] incorporated semantic information, such as road networks, to classify vehicle trajectories. However, such works heavily rely on handcrafted features and make excessive parameter assumptions. Recently, studies in trajectory modeling have focused on deep representation learning, benefiting from its superior modeling capabilities in sequential data [Chen *et al.*, 2024]. [Liu and Lee, 2017] and [Jiang *et al.*, 2017] utilized two common RNN architectures to capture high-order movement patterns. After that, [Liu *et al.*, 2019] and [Liang *et al.*, 2021] introduced segment-wise convolutional weight mechanisms and neural differential equations to enforce RNNs, addressing the modeling of continuous temporal characteristics. [Han *et al.*, 2021] and [Yao *et al.*, 2022] integrated spatial features from road networks by graph neural networks, enabling the capture of long-term dependencies in trajectories. Furthermore, TrajFormer [Liang *et al.*, 2022] adapted an advanced transformer architecture to balance speed and accuracy in trajectory modeling. Nevertheless, due to the inherent noise in raw trajectories and environmental biases, obtaining robust trajectory representations with causal invariance remains challenging.

Causal Inference. Traditional causal inference aims to study how to learn a causal model that works under different distributions, encompasses causal mechanisms, and further employs the model for intervention or counterfactual inference [Pearl, 2009]. However, real-world observations often do not begin with basic inference units (random variables connected in a causal graph) but rather with high-dimensional raw data. Therefore, causal representation learning [Schölkopf *et al.*, 2021] seeks to integrate deep learning and causal mechanisms, widely explored in various fields such as computer vision [Lippe *et al.*, 2022], recommendation systems [Wang *et al.*, 2023], graph data mining [Sui *et al.*, 2022] and so on. Nevertheless, research in spatio-temporal data mining from a causal perspective is still in its early stages. [Li *et al.*, 2023] investigated distribution changes in time series by discovering causal structures, but they neglected to decipher spatial factors. Subsequently, [Deng *et al.*, 2023] constructed a causal graph to describe traffic prediction and analyze the causal relationships between spatio-temporal features and outcomes. Similarly, [Xia *et al.*, 2023] applied it to spatio-temporal graph forecasting, mitigating confounding effects in the temporal and spatial domains using causal inference. Still and all, they are all graph-based models considering causal effects and cannot adapt to trajectory data as spatio-temporal sequences. In this study, we adopt causal techniques to model trajectory and mitigate confounding factors in the environment.

3 A Causal View on Trajectory Modeling

3.1 Formulation

Definition 1 (Trajectory)**.** *We denote a trajectory by* $X = \{p_i \mid i = 1, 2, \ldots, n\}$ *with a sequence of spatio-temporal points* p_i *recorded in chronological order. Each point* $p = (Lon, Lat, t)$ *is a longitude, latitude, and timestamp triplet.*

Definition 2 (Geospatial Context)**.** *The geospatial context is the set* E *of environment information, where* $e_i \in \mathbb{R}^m$ *represents the attributes of the surrounding environment related to the point* p_i*, and* m *denotes the number of attributes.*

Problem Statement (Trajectory Modeling)**.** *Our target is to learn robust and high-quality trajectory representation* H *via supervised signals to support various downstream tasks. In our task, we set it up for travel mode identification.*

3.2 Structural Causal Model

Formally, we build a Structural Causal Model (SCM) to analyze the causality in trajectory modeling (Fig. 2). It comprises three variables: trajectory data X, environment E, and

[1]we use geospatial context and environment interchangeably.

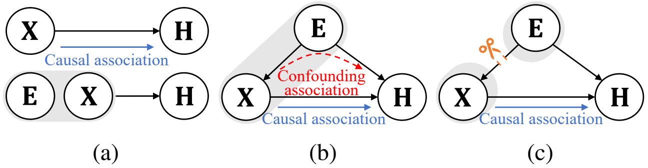

Figure 2: SCMs of trajectory modeling. The SCM (a) in traditional perspective; (b) under causal view; (c) after back-door adjustment.

trajectory representation H, with arrows denoting causal relationships. The SCM provides the following elucidation:

- $X \rightarrow H$. Trajectories encompass crucial movement characteristics such as speed, acceleration, and direction. Variations in these fundamental features give rise to discernible trajectory representations. Intuitively, trajectory data can be modeled using internal sequential patterns to learn representations: $H = \mathcal{F}(\theta; X)$,where $\mathcal{F}$ denotes the parameterized modeling function with learnable parameters θ.
- $E \rightarrow X$ and $E \rightarrow H$. Trajectory data is profoundly impacted by environmental factors. An illustrative example of this influence is the rarity of pedestrian trajectories along highways. Moreover, the dynamic nature of the environment contributes to alterations in the distribution of latent trajectory representations. Notably, distinctive motion tendencies emerge among individuals and vehicles in response to environmental cues such as traffic signals. In light of these environmental factors, the joint modeling of trajectories and environment can be expressed as $H = \mathcal{F}(\theta;\ X,\ E)$.

However, upon meticulous examination of the SCM, we identify a *backdoor path* between X and H, i.e., $X \leftarrow E \rightarrow H$, where environmental information acts as a confounding factor for both trajectory data and trajectory representation. This introduces a confounding factor set $\mathcal{E}$, signifying a confounding association between environment and trajectory representation (Fig. 2b), distinct from the causal association from X to H (Fig. 2a). Consequently, in the practical process of jointly learning trajectory-environment representations, this strong correlation comprises a blend of confounding and causal associations. Therefore, we advocate for employing causal intervention techniques on variable X to mitigate the confounding effects of variable E (Fig. 2c), facilitating the acquisition of robust representations from biased trajectory data.

3.3 Backdoor Adjustment

Based on the aforementioned causal analysis, our approach to learning trajectory representations involves eliminating backdoor paths rather than modeling the association $P(H|X)$ in Fig. 2a. We employ a powerful tool known as backdoor adjustment, rooted in causal theory [Pearl, 2009], allowing us to block the backdoor path by estimating $P(H|do(X))$, where $do(\cdot)$ denotes the *do-calculus*. Specifically, we have

$$\begin{aligned} P(H|do(X)) &= \sum\nolimits_{e\in\mathcal{E}} P(H|do(X),e)P(e|do(X)) \\ &= \sum\nolimits_{e\in\mathcal{E}} P(H|do(X),e)P(e) \\ &= \sum\nolimits_{e\in\mathcal{E}} P(H|X,e)P(e), \end{aligned} \tag{1}$$

where $\mathcal{E}$ represents the set of environment confounding factors. First, we can redefine $P(H|do(X))$ via Bayes' theorem. Then, due to the independence of variables H and X under intervention, resulting in $P(e|do(X)) = P(e)$. Simultaneously, the response of H to E and X is unrelated to the causal association between H and X, allowing us to equate the conditional probabilities $P(H|do(X), e) = P(H|X, e)$.

Nonetheless, the latent confounding factors denoted as $\mathcal{E}$ prove elusive and challenging to directly observe [Xia *et al.*, 2023]. Manipulating extensive trajectory data also poses additional complexities, necessitating adjustment at the representation level through learned strategies. To this end, we introduce two new modules in the following section to implement Eq. 1 for trajectory representation learning.

4 Model Implementation

In this section, we present the details of our TrajCL framework (see Fig. 3). We first provide an *overview of the TrajCL framework*, integrating causal techniques into the trajectory modeling. Then, we delve into the specifics of two crucial components: *environmental alignment module* and *causal learning module*. These tailored modules are designed to be compatible with any advanced trajectory model, allowing for efficient enhancement from a causal perspective.

4.1 Framework Overview

Traditional trajectory learning methods mostly involve convolutional layers to extract local features from input X at the point level. A deep sequential model is subsequently employed to capture spatio-temporal patterns at the trajectory level. Notably, the geospatial context E is regarded as auxiliary information incorporated into the encoding process. For clarity, we use the term encoder $\mathcal{F}$ to represent the aforementioned sequential encoding process in this paper.

As analyzed in section 3.2, the simple introduction of environmental information leads to confounding factors. Therefore, we extend the modeling strategy from a causal perspective. The sequence encoder $\mathcal{F}$ is duplicated into two counterparts, namely the causal encoder $\mathcal{F}_\alpha$ and the confounding encoder $\mathcal{F}_\beta$, maintaining the identical architecture but without parameter sharing. The formalized process is as follows:

$$H_\alpha = \mathcal{F}_\alpha(\theta_1;\ X,\ E),\ \ H_\beta = \mathcal{F}_\beta(\theta_2;\ X,\ E). \tag{2}$$

$\mathcal{F}_\alpha$ and $\mathcal{F}_\beta$ denotes the encoder, which can be shifted to advanced trajectory representation models, e.g., GRU. H_α and H_β denote the the embedding of original sequence.

Guided by a specially developed *environmental alignment module*, the features output by the two encoders are separately disentangled. Specifically, H_α and H_β are fed into two streams. One stream is dedicated to extracting invariant representations of the former, termed causal features (represented by the orange dashed line in Fig. 3). The other stream aims to extract the confounding effects caused by the environment, termed confounding features (as indicated by the green dashed line in Fig. 3). Finally, a carefully designed *causal learning module* intervenes at the representation level, jointly with downstream tasks, to differentiate causal and confounding features and achieve robust representation acquisition.

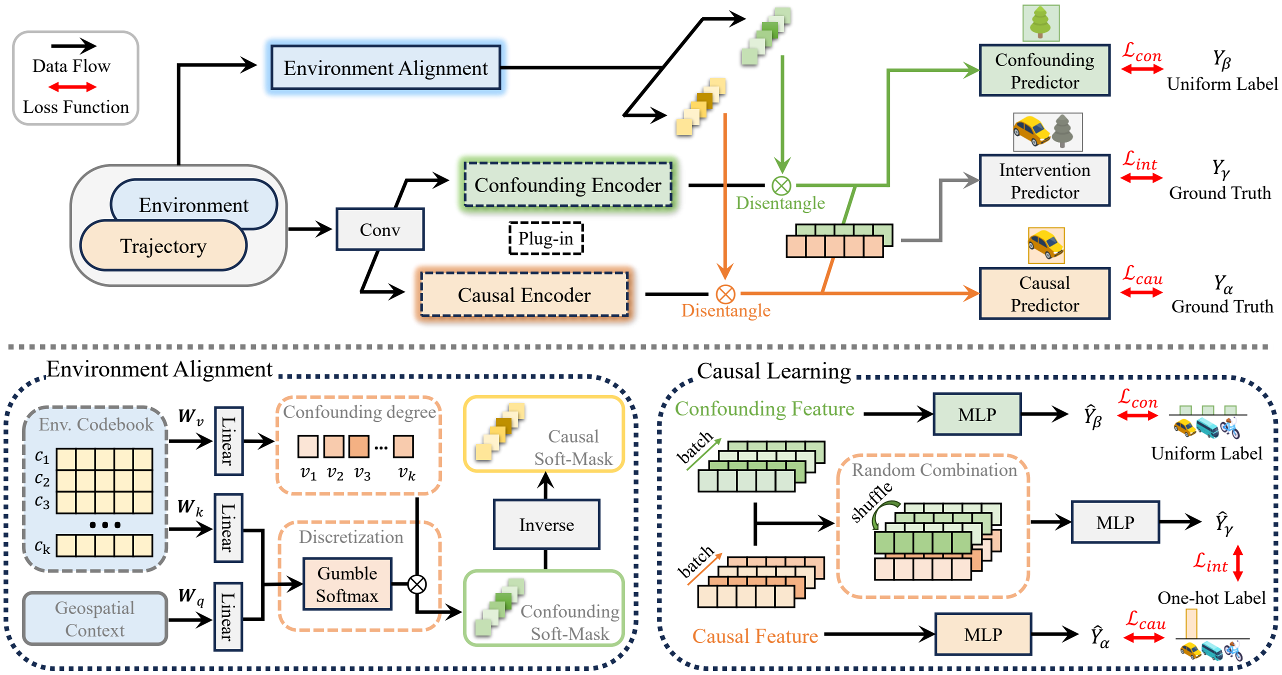

Figure 3: The architecture of the proposed TrajCL framework. Env: Environment.

4.2 Environment Alignment Module

The goal of this module is to learn two soft-masks for confounding awareness and causal awareness, guiding the two streams to disentangle causal and confounding features from the input data. It comprises two sub-modules: a *cross attention component* and a *disentanglement allocation component*.

Cross Attention Component

From an intuitive view, various types of environments inherently function as confounding factors to differing degrees. Therefore, we first create a learnable environment codebook $C = \{c_1, c_2, ..., c_k\}$, where $C \in \mathbb{R}^{k\times d}$ represents the prototype embedding of urban environments. Subsequently, each prototype is fed into a linear layer to encode its individual confounding degree:

$$V = W_v C + b_v, \tag{3}$$

where $W_v \in \mathbb{R}^{d\times 1}$ and b_v are learnable parameters, and $V \in \mathbb{R}^{k\times 1}$ denotes the confounding degrees corresponding to the environment codebook, i.e., $V = \{v_1, v_2, ..., v_k\}$.

Meanwhile, we project the environment codebook and the geospatial context $E \in \mathbb{R}^{n\times m}$ into the same latent space:

$$Q = W_q E + b_q, \quad K = W_k C + b_k, \tag{4}$$

where $W_q \in \mathbb{R}^{m\times d}$ and $W_k \in \mathbb{R}^{d\times d}$ are learnable parameter matrices for geospatial context and the codebook, respectively. b_q and b_k are bias vectors.

Next, we apply a cross-attention mechanism to assess the similarity scores among all prototypes in the environment codebook Q with geospatial context K. Specifically:

$$M_\alpha = \text{Gumbel-Softmax}(Q \cdot K^\mathsf{T}/\sqrt{d}) \cdot V, \tag{5}$$

where $M_\alpha \in \mathbb{R}^{n\times 1}$ is the adjusted confounding intensity, serving as a confounding soft-mask, and its complement is the causal soft-mask $M_\beta = 1 - M_\alpha$. Notably, allocating all prototypes to the environment contradicts our intuition. We hence employ the Gumbel-Softmax [Jang *et al.*, 2016] to select the most similar environment, expressed as:

$$\mathbf{s}_i = \frac{e^{(\log(\pi_i)+\mathcal{G}_i)/\tau)}}{\sum_{j=1}^{k} e^{(\log(\pi_j)+\mathcal{G}_j)/\tau)}}, \tag{6}$$

where $\mathbf{s}$ is the n-dimensional vector, π are class probabilities, $\mathcal{G} \sim \text{Gumbel}(0,1)$ are i.i.d samples drawn from the Gumbel distribution, and τ is the temperature. Due to orthogonal properties, it enforces discretized similarity learning.

Disentanglement Allocation Component

After completing the cross-attention process, we apply the two masks M_α and M_β to the representations from causal and confounding encoders, allocating weights to each spatio-temporal point. This process enables the encoders $\mathcal{F}_\alpha$ and $\mathcal{F}_\beta$ to learn disentangled representations. Moreover, for both branches, we perform average pooling operations along the time step dimension:

$$Z_\alpha = \text{Pooler}(H_\alpha \odot M_\alpha), \quad Z_\beta = \text{Pooler}(H_\beta \odot M_\beta), \tag{7}$$

where $\odot$ denotes the Hadamard product operation. Z_α and Z_β are trajectory-level causal and confounding features.

4.3 Causal Learning Module

We further implement diverse strategies for parameterizing backdoor adjustment. This involves a *disentangle learning strategy* that combines downstream tasks to distinguish causality from confounding, an *intervention learning strategy* to eliminate confounding, and a final *optimization* process.

Disentangle Learning Strategy

Given our objective of travel mode identification, we choose to employ category classifiers via a multilayer perceptron (MLP) across two branches. The purpose of causal features is to estimate invariant characteristics within trajectories, leading to the causal branch output being classified with actual travel mode labels. In contrast, confounding features aim to represent information unrelated to the intrinsic patterns of trajectories. Consequently, the output from the confounding branch is distanced from travel mode representations, approaching the average label across all categories. Formally, we define two classification losses:

$$\begin{gathered}\hat{Y}_\alpha = \mathrm{MLP}(Z_\alpha),\ \ \hat{Y}_\beta = \mathrm{MLP}(Z_\beta),\\ \mathcal{L}_{cau} = \mathcal{H}(Y_\alpha, \hat{Y}_\alpha),\ \ \mathcal{L}_{con} = \mathcal{H}(Y_\beta, \hat{Y}_\beta),\end{gathered} \tag{8}$$

where Y_α represents the ground-truth travel mode label, Y_β is assigned using a uniform distribution as the ground-truth label, $\mathcal{H}(\cdot,\cdot)$ refers to the conventional cross-entropy loss.

Intervention Learning Strategy

To achieve invariant predictions in a dynamic environment, the elimination of confounding factors emerges as a pivotal endeavor. To this end, we perform interventions by hierarchically manipulating the confounding features and randomly combining them with causal features to achieve backdoor adjustment in Equation 1. Specifically, we merge the causal features of one input with the confounding features of another randomly chosen input from the same batch as the intervention features. Given the inherent characteristic that confounders do not exert influence on the final outcome, the prediction results will align well with the actual travel mode as:

$$\hat{Y}_\gamma = \mathrm{MLP}(Z_\alpha + Z_\beta^{'}),\ \ \mathcal{L}_{int} = \mathcal{H}(Y_\gamma, \hat{Y}_\gamma), \tag{9}$$

where Y_γ is also the travel mode ground-truth label.

Optimization

The objective function of TrajCL can be defined as the sum of the above three losses:

$$\mathcal{L} = \lambda\mathcal{L}_{cau} + \varphi\mathcal{L}_{con} + \eta\mathcal{L}_{int}, \tag{10}$$

where λ, φ and η are weight hyperparameter. Through backpropagation optimization, we can force the causal representation to be invariant and stable, the confounder representation to be independent, and the intervention representation to be consistent. Finally, we use the causal representation as a robust trajectory representation.

5 Experiments

In this section, we evaluate our TrajCL based on an extensively studied trajectory classification task [Liang *et al.*, 2022] to answer the following research questions (RQs):

- **RQ1:** Does TrajCL really help improve travel mode classification performance, and to what extent?
- **RQ2:** How robust is TrajCL under varying conditions?
- **RQ3:** How do the key components and the hyperparameters of TrajCL affect the performance?
- **RQ4:** Does TrajCL accurately and efficiently yield explanations, and how intuitively understandable is it?

5.1 Experimental Settings

Datasets. We conduct extensive experiments on two real-world datasets GeoLife [Zheng *et al.*, 2009] and Grab-Posisi [Huang *et al.*, 2019]: GeoLife consists of human trajectories collected by 182 users in Beijing from April 2007 to August 2012. Following [Liu *et al.*, 2019], we divide trajectories into segments and classify them into four typical travel modes: walking, bus, bike, and driving. Grab-Posisi is a dataset of delivery trajectories collected by Grab, a Southeast Asian ride-hailing company. We follow [Liang *et al.*, 2021] to pick data over two weeks in Jakarta, with travel modes including car and motorcycle. In addition, we grid the city into cells of 200×200 square meters for each dataset. To capture environmental factors, we then retrieve 24 fixed geospatial features (such as the number of traffic lights, crossings, and residential areas) for each grid from OpenStreetMap.com. These environmental variables are then assigned to each GPS point. After preprocessing, we collect 26,509 trajectories for GeoLife and 507,522 for Grab-Posisi. Each trajectory comprises 20 to 50 GPS points. These trajectories are subsequently partitioned in an 8:1:1 ratio for training, validation, and test data.

Baselines & Evaluation Metrics. To validate the effectiveness and robustness of TrajCL, we implement our framework with five representative models for trajectory modeling, including vanilla GRU, BiLSTM [Liu and Lee, 2017], GRU-D [Che *et al.*, 2018], STGN [Zhao *et al.*, 2020], and TrajFormer [Liang *et al.*, 2022]. We refine them for optimal performance using the parameter settings suggested in their papers. Then, we follow [Liu and Lee, 2017; Liang *et al.*, 2022] to employ the classification accuracy (Acc.) to evaluate model performance. Each model and setting is run on every dataset thrice, and the mean accuracy is reported. Besides, the symbol Δ denotes accuracy change.

Implementation Details & Hyperparameters. Our model uses the Adam optimizer with the initial learning rate set to 0.001, reduced by 0.1 every 30 epochs. To avoid overfitting, we employ an early stopping with a patience of 20 epochs. The batch sizes for the GeoLife and Grab-Posisi datasets are 256 and 512. The default embedding dimensions are set to 64. For the predictor, we apply a 2-layer MLP uniformly. To initially merge local features from inputs, we employ two 3×1 convolutional layers. The weight parameters λ, φ, and η of the loss are 1, 0.5, and 0.5, respectively.

5.2 Overall Performance (RQ1)

Table 1 depicts the overall performance under different model settings on the two datasets. Specifically, we report the performance of the original baseline (Base), the results with incorporating environmental information (+ Env), and applied with the TrajCL framework (+ TrajCL). The experimental results demonstrate a clear overall trend: incorporating environmental information significantly enhances the performance, which underscores the importance of considering environmental influences in trajectory modeling. However, incorporating environmental information into trajectory modeling may simultaneously introduce confounding factors that negatively affect the modeling process. This phenomenon is further validated by applying TrajCL framework, where we can

Model \ Dataset	GeoLife			Grab-Posisi		
	Base	+ Env	+ TrajCL	Base	+ Env	+ TrajCL
GRU	69.15	75.78	77.74	63.46	78.76	79.47
	–	+6.63	+8.59	–	+15.30	+16.01
BiLSTM	70.20	76.58	78.02	69.07	79.81	80.70
	–	+6.38	+7.82	–	+10.74	+11.63
GRU-D	72.48	77.37	78.62	71.78	79.17	79.69
	–	+4.89	+6.14	–	+7.39	+7.91
STGN	76.75	79.03	80.27	72.94	81.24	82.74
	–	+2.28	+3.52	–	+8.30	+9.80
TrajFormer	78.40	80.05	81.22	76.30	85.34	86.53
	–	+1.65	+2.82	–	+9.04	+10.23

Table 1: Performance comparison with different backbones. Acc. report percentage (%) with 5 runs average.

observe more improvements after using TrajCL for causal intervention. The above results strongly suggest that there are indeed confounding effects within environments, and emphasize the effectiveness and necessity of our proposed TrajCL.

In addition, the environmental factors and the TrajCL present different influences across the two datasets. In the GeoLife dataset, the overall improvement of incorporating environmental factors is less significant compared to the Grab-Posisi dataset (average +4.37% in Geolife, and +10.15% in Grab-Posisi). This can be attributed to the high frequency and diversity of confounders in the Geolife dataset, which partially undermines the contribution of environmental factors. Comparably, the Grab-Posisi dataset is limited to car and motorcycle travel modes with less surrounding noise. Therefore, when the TrajCL intervention is applied to mitigate these confounding factors, the performance improvement is more substantial in the GeoLife dataset (average +1.41% in Geolife, and +0.96% in Grab-Posisi). This difference highlights the effectiveness of the TrajCL intervention, particularly in scenarios with diverse and complex environmental confounders.

5.3 Robustness Test (RQ2)

In this section, we analyze the robustness of TrajCL in the trajectory modeling task with two different scenarios.

Few-shot Learning. For few-shot learning, we divide the original data set into subsets with ratio [0.1, 0.2, 0.5], and implement on three state-of-the-art models. The results in Table 2 show that the smaller the subset, the more significant the improvement provided by TrajCL. This trend is evident across different models and datasets, demonstrating its high robustness and adaptability to few-shot learning. Notably, the better-performing models (e.g., TrajFormer) exhibit larger TrajCL boosting effects. This suggests that these models, despite having high baseline performance, also *overfit the geospatial context* and introduce more confounding factors when modeling trajectories. Similarly, the improvement in the Grab-Posisi dataset is not as significant as in the GeoLife dataset. The observation is consistent with our findings in Section 5.2, reinforcing the conclusion that the TrajCL is effective in more complex and diverse environments.

Dataset \ Model	GRU-D		STGN		TrajFormer	
	+ Env	+ TrajCL	+ Env	+ TrajCL	+ Env	+ TrajCL
GL-0.5	73.97	+0.74	75.80	+0.98	76.33	+2.51
GL-0.2	70.32	+0.94	71.40	+1.44	72.12	+2.33
GL-0.1	65.98	+1.44	68.61	+1.37	69.26	+2.94
GP-0.5	75.82	+0.85	78.70	+0.43	79.97	+1.15
GP-0.2	72.19	+0.93	73.29	+0.58	74.11	+1.08
GP-0.1	69.36	+1.09	70.51	+1.47	71.38	+1.53

Table 2: Performance in the few-shot learning setting. Accuracy is reported by percentage (%). GL: GeoLife, GP: Grab-Posisi.

Imbalanced Sample Learning. The results of imbalanced sample scenarios are summarized in Figure 4, where the X-axis represents the class imbalance ratio in the training set, and the two Y-axes indicate the accuracy and performance improvement, respectively. To focus on examining TrajCL in imbalanced data while mitigating the effects of complex model designs, we choose a relatively straightforward GRU encoder and the Grab-Posisi dataset consisting of two classes.

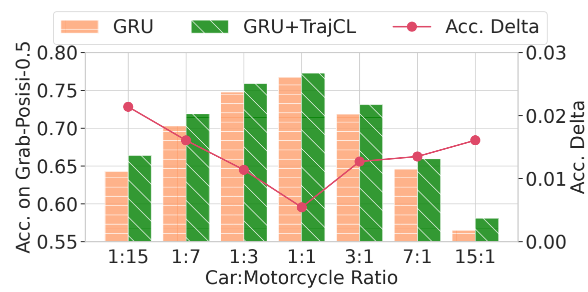

Figure 4: Exploration of imbalanced sample learning scenarios.

As depicted in Figure 4, GRU+TrajCL is consistently higher than the baseline, showing the effectiveness of our framework across all settings of imbalance. As imbalances increase (from the middle to both sides), the red line in the figure illustrates more significant TrajCL enhancements. This result further validates the robustness of TrajCL, especially as it effectively addresses the challenges under severe imbalance conditions. Interestingly, the performance gains are typically greater when Motorcycle mode is more prevalent (left) than Car mode (right). This can be intuitively attributed to the richer environmental conditions that motorcycles can encounter. This complexity introduces a higher confounding effect, which in turn is elegantly addressed by our TrajCL.

5.4 Ablation Study (RQ3)

Effects of Core Components. In the ablation study, we quantified the contribution of each component by removing them individually. Across both datasets, the results in Table 3 demonstrate performance degradation, confirming the necessity of each component in the TrajCL. Replacing the environment codebook with a random matrix (w/o EC) slightly impacts performance, highlighting the learned prototypes for specific cities can aid in better alignment. Maintaining a fixed confounding degree of 0.5 for uniform disentanglement (w/o

Dataset / Variant	GeoLife		Grab-Posisi	
	Acc.	Δ	Acc.	Δ
TrajCL	77.74	–	79.47	–
TrajCL w/o EC	76.25	-1.49	76.99	-2.48
TrajCL w/o CI	75.69	-2.05	76.39	-3.08
TrajCL w/o Dise	75.55	-2.19	76.23	-3.24
TrajCL w/o Env	70.16	-7.58	65.38	-14.09

Table 3: Component ablation results. We use GRU as backbone.

Dise) significantly affects performance, showing that soft-masks are crucial in separating confounders. Similarly, excluding random combination from the causal learning process (w/o CI) leads to a similar decrease in performance, highlighting its importance in enhancing representation. The most significant decrease occurs when geospatial features are excluded and only the raw trajectory is used to extract the confounding degree (TrajCL w/o Env). This highlights the critical role of auxiliary environmental information in providing context for disentangling. These results demonstrate that proposed components collectively contribute to the effectiveness of TrajCL. At the same time, geospatial features provide essential information for disentanglement, particularly evident in complex urban environments like the GeoLife dataset.

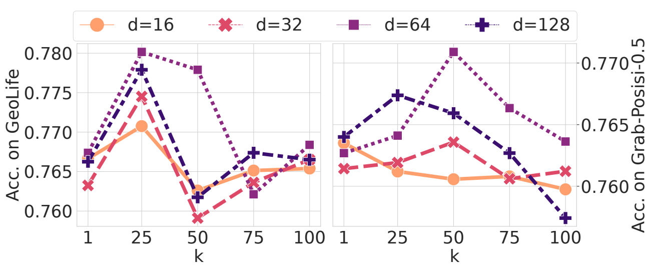

Figure 5: Effects of environment codebook size k and hidden size d.

Effects of Environment Codebook. As discussed in Section 4.2, the environment codebook uses a learnable matrix to profile environmental prototypes. Therefore, we explore the effect of the codebook parameters, analyzing the interaction between the richness of the environment (denoted by k) and the hidden dimension of each prototype (denoted by d). As shown in Figure 5, varying values of k and d lead to overall performance differences. Each dataset has an optimal k ($k = 25$ in GeoLife and $k = 50$ in Grab-Posisi). In addition, the largest hidden dimension ($d = 128$) does not necessarily yield the highest accuracy. The reason is that a too large d causes redundancy, while a small d is insufficient to learn environmental features adequately. We also find that the Geolife dataset is more sensitive to the codebook setting, which is also caused by more complex environments in this city. This suggests that we need to balance k and d based on the specific urban forms and diversity.

5.5 Interpretation Analysis (RQ4)

To further explore what the codebook has learned, we plot trajectory points and their associated environmental prototypes to maps, analyzing the interpretability through geo-visualization. As shown in Figure 6, we can clearly observe that the environment codebook can perceive different geospatial contexts. For example, the high concentration of brown (corresponding to prototype c_{24} in the codebook) portrays a typical park area. The blue (corresponding to prototype c_1) identifies a real-world airport. Meanwhile, the codebook assigns the identical environmental prototype (green area) to those with the same geospatial contexts, demonstrating its strong ability to describe and categorize various environmental features. In conclusion, the codebook can provide explanations, and also help us delve into the environmental perception of the model and its corresponding impact.

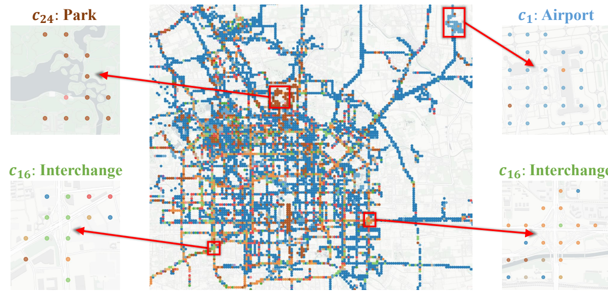

Figure 6: Visualization of environment codebook alignment on GeoLife dataset (k=25). The color represents different c_i.

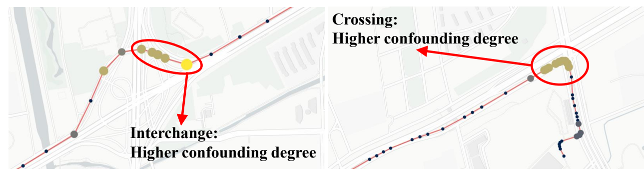

Figure 7: Visualization of the confounding soft-masks applied to two trajectories. Larger points indicate a higher confounding degree.

Furthermore, we visualize the confounding soft-mask obtained through the environment alignment module for trajectory instances. This illustrates the confounding degree assigned to each trajectory point. As shown in Figure 7, interchanges and crossings usually have higher degrees. This aligns with our intuition to focus the confounders more on environmentally complex regions for effective decoupling.

6 Conclusion and Future Work

In this paper, we propose a novel causal trajectory modeling framework (TrajCL) to facilitate the learning of robust and high-quality trajectory representations in travel mode identification task. TrajCL reexamines existing trajectory modeling processes from a causal perspective, introducing an environment alignment module and a causal learning module for invariant trajectory representation learning. Experimental results on two real-world trajectory datasets show significant advantages in performance, robustness, and interpretability. In the future, we are committed to expanding the applicability of the TrajCL to cover more diverse environments, thus enhancing its applicability to more realistic trajectory-based tasks such as travel time estimation.

Acknowledgments

This work is mainly funded by Guangzhou-HKUST (GZ) Joint Funding Program (No. 2024A03J0620). This work is also supported by Guangzhou Municiple Science and Technology Project 2023A03J0011 and the Guangzhou Industrial Information and Intelligent Key Laboratory Project (No. 2024A03J0628).

References

[Chang *et al.*, 2023] Yanchuan Chang, Jianzhong Qi, Yuxuan Liang, and Egemen Tanin. Contrastive trajectory similarity learning with dual-feature attention. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pages 2933–2945. IEEE, 2023.

[Che *et al.*, 2018] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1):6085, 2018.

[Chen *et al.*, 2022] Wei Chen, Shuzhe Li, Chao Huang, Yanwei Yu, Yongguo Jiang, and Junyu Dong. Mutual distillation learning network for trajectory-user linking. In *IJCAI*, pages 1973–1979, 2022.

[Chen *et al.*, 2024] Wei Chen, Yuxuan Liang, Yuanshao Zhu, Yanchuan Chang, Kang Luo, Haomin Wen, Lei Li, Yanwei Yu, Qingsong Wen, Chao Chen, et al. Deep learning for trajectory data management and mining: A survey and beyond. *arXiv preprint arXiv:2403.14151*, 2024.

[Dai *et al.*, 2015] Jian Dai, Bin Yang, Chenjuan Guo, and Zhiming Ding. Personalized route recommendation using big trajectory data. In *2015 IEEE 31st international conference on data engineering*, pages 543–554. IEEE, 2015.

[Deng *et al.*, 2023] Pan Deng, Yu Zhao, Junting Liu, Xiaofeng Jia, and Mulan Wang. Spatio-temporal neural structural causal models for bike flow prediction. *arXiv preprint arXiv:2301.07843*, 2023.

[Dodge *et al.*, 2009] Somayeh Dodge, Robert Weibel, and Ehsan Forootan. Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems*, 33(6):419–434, 2009.

[Ferrero *et al.*, 2020] Carlos Andres Ferrero, Lucas May Petry, Luis Otavio Alvares, Camila Leite da Silva, Willian Zalewski, and Vania Bogorny. Mastermovelets: discovering heterogeneous movelets for multiple aspect trajectory classification. *Data Mining and Knowledge Discovery*, 34(3):652–680, 2020.

[Guo *et al.*, 2020] Chenjuan Guo, Bin Yang, Jilin Hu, Christian S Jensen, and Lu Chen. Context-aware, preference-based vehicle routing. *The VLDB Journal*, 29:1149–1170, 2020.

[Han *et al.*, 2021] Peng Han, Jin Wang, Di Yao, Shuo Shang, and Xiangliang Zhang. A graph-based approach for trajectory similarity computation in spatial networks. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 556–564, 2021.

[Huang *et al.*, 2019] Xiaocheng Huang, Yifang Yin, Simon Lim, Guanfeng Wang, Bo Hu, Jagannadan Varadarajan, Shaolin Zheng, Ajay Bulusu, and Roger Zimmermann. Grab-posisi: An extensive real-life gps trajectory dataset in southeast asia. In *Proceedings of the 3rd ACM SIGSPATIAL international workshop on prediction of human mobility*, pages 1–10, 2019.

[Jang *et al.*, 2016] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*, 2016.

[Jiang *et al.*, 2017] Xiang Jiang, Erico N de Souza, Ahmad Pesaranghader, Baifan Hu, Daniel L. Silver, and Stan Matwin. Trajectorynet: An embedded gps trajectory representation for point-based classification using recurrent neural networks. In *Proceedings of the 27th Annual International Conference on Computer Science and Software Engineering*, CASCON '17, page 192–200, USA, 2017. IBM Corp.

[Jiang *et al.*, 2023] Jiawei Jiang, Dayan Pan, Houxing Ren, Xiaohan Jiang, Chao Li, and Jingyuan Wang. Self-supervised trajectory representation learning with temporal regularities and travel semantics. In *2023 IEEE 39th international conference on data engineering (ICDE)*, pages 843–855. IEEE, 2023.

[Jin *et al.*, 2023] Guangyin Jin, Yuxuan Liang, Yuchen Fang, Zezhi Shao, Jincai Huang, Junbo Zhang, and Yu Zheng. Spatio-temporal graph neural networks for predictive learning in urban computing: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 2023.

[Lee and Han, 2008] Jae-Gil Lee and Jiawei Han. Traclass: trajectory classification using hierarchical region-based and trajectory-based clustering. *Proceedings of the VLDB Endowment*, 1(1):1081–1094, 2008.

[Li *et al.*, 2023] Zijian Li, Ruichu Cai, Tom ZJ Fu, Zhifeng Hao, and Kun Zhang. Transferable time-series forecasting under causal conditional shift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[Liang *et al.*, 2021] Yuxuan Liang, Kun Ouyang, Hanshu Yan, Yiwei Wang, Zekun Tong, and Roger Zimmermann. Modeling trajectories with neural ordinary differential equations. In *IJCAI*, pages 1498–1504, 2021.

[Liang *et al.*, 2022] Yuxuan Liang, Kun Ouyang, Yiwei Wang, Xu Liu, Hongyang Chen, Junbo Zhang, Yu Zheng, and Roger Zimmermann. Trajformer: Efficient trajectory classification with transformers. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 1229–1237, 2022.

[Lin *et al.*, 2021] Yan Lin, Huaiyu Wan, Shengnan Guo, and Youfang Lin. Pre-training context and time aware location embeddings from spatial-temporal trajectories for user next location prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4241–4248, 2021.

[Lippe *et al.*, 2022] Phillip Lippe, Sara Magliacane, Sindy Löwe, Yuki M Asano, Taco Cohen, and Stratis Gavves.

Citris: Causal identifiability from temporal intervened sequences. In *International Conference on Machine Learning*, pages 13557–13603. PMLR, 2022.

[Liu and Lee, 2017] Hongbin Liu and Ickjai Lee. End-to-end trajectory transportation mode classification using bi-lstm recurrent neural network. In *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pages 1–5. IEEE, 2017.

[Liu *et al.*, 2019] Hongbin Liu, Hao Wu, Weiwei Sun, and Ickjai Lee. Spatio-temporal gru for trajectory classification. In *2019 IEEE International Conference on Data Mining (ICDM)*, pages 1228–1233. IEEE, 2019.

[Pearl, 2009] Judea Pearl. Causal inference in statistics: An overview. *Statistics Surveys*, 3:96 – 146, 2009.

[Pugliese *et al.*, 2023] Chiara Pugliese, Francesco Lettich, Fabio Pinelli, and Chiara Renso. Summarizing trajectories using semantically enriched geographical context. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems*, pages 1–10, 2023.

[Qin *et al.*, 2019] Yanjun Qin, Haiyong Luo, Fang Zhao, Chenxing Wang, Jiaqi Wang, and Yuexia Zhang. Toward transportation mode recognition using deep convolutional and long short-term memory recurrent neural networks. *IEEE Access*, 7:142353–142367, 2019.

[Schölkopf *et al.*, 2021] Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021.

[Sui *et al.*, 2022] Yongduo Sui, Xiang Wang, Jiancan Wu, Min Lin, Xiangnan He, and Tat-Seng Chua. Causal attention for interpretable and generalizable graph classification. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1696–1705, 2022.

[Wang *et al.*, 2023] Wenjie Wang, Yang Zhang, Haoxuan Li, Peng Wu, Fuli Feng, and Xiangnan He. Causal recommendation: Progresses and future directions. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3432–3435, 2023.

[Wang *et al.*, 2024] Yu Wang, Tongya Zheng, Yuxuan Liang, Shunyu Liu, and Mingli Song. Cola: Cross-city mobility transformer for human trajectory simulation. *arXiv preprint arXiv:2403.01801*, 2024.

[Wu *et al.*, 2017] Hao Wu, Ziyang Chen, Weiwei Sun, Baihua Zheng, and Wei Wang. Modeling trajectories with recurrent neural networks. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 3083–3090, 2017.

[Xia *et al.*, 2023] Yutong Xia, Yuxuan Liang, Haomin Wen, Xu Liu, Kun Wang, Zhengyang Zhou, and Roger Zimmermann. Deciphering spatio-temporal graph forecasting: A causal lens and treatment. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[Xiao *et al.*, 2017] Zhibin Xiao, Yang Wang, Kun Fu, and Fan Wu. Identifying different transportation modes from trajectory data using tree-based ensemble classifiers. *ISPRS International Journal of Geo-Information*, 6(2):57, 2017.

[Yao *et al.*, 2022] Di Yao, Haonan Hu, Lun Du, Gao Cong, Shi Han, and Jingping Bi. Trajgat: A graph-based long-term dependency modeling approach for trajectory similarity computation. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pages 2275–2285, 2022.

[Yin *et al.*, 2016] Hongzhi Yin, Xiaofang Zhou, Bin Cui, Hao Wang, Kai Zheng, and Quoc Viet Hung Nguyen. Adapting to user interest drift for poi recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 28(10):2566–2581, 2016.

[Zhao *et al.*, 2020] Pengpeng Zhao, Anjing Luo, Yanchi Liu, Jiajie Xu, Zhixu Li, Fuzhen Zhuang, Victor S Sheng, and Xiaofang Zhou. Where to go next: A spatio-temporal gated network for next poi recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 34(5):2512–2524, 2020.

[Zheng *et al.*, 2008] Yu Zheng, Like Liu, Longhao Wang, and Xing Xie. Learning transportation mode from raw gps data for geographic applications on the web. In *Proceedings of the 17th international conference on World Wide Web*, pages 247–256, 2008.

[Zheng *et al.*, 2009] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th international conference on World wide web*, pages 791–800, 2009.

[Zheng, 2015] Yu Zheng. Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3):1–41, 2015.

[Zhu *et al.*, 2021] Yuanshao Zhu, Yi Liu, James J. Q. Yu, and Xingliang Yuan. Semi-supervised federated learning for travel mode identification from gps trajectories. *IEEE Transactions on Intelligent Transportation Systems*, 23(3):2380–2391, 2021.