Vehicle Trajectory Data Processing, Analytics, and Applications: A Survey

CHENXI LIU, College of Computer Science and Electronic Engineering, Hunan University, Changsha, China
ZHU XIAO, College of Computer Science and Electronic Engineering, Hunan University, Changsha, China
WANGCHEN LONG, The College of Artificial Intelligence, Zhuhai City Polytechnic, Zhuhai, China
TONG LI, Tsinghua University, Beijing, China
HONGBO JIANG, College of Computer Science and Electronic Engineering, Hunan University, Changsha, China
KEQIN LI, Department of Computer Science, State University of New York, New York, United States

Vehicles traveling through cities generate extensive vehicle trajectory collected by scalable sensors, providing excellent opportunities to address urban challenges such as traffic congestion and public safety. In this survey, we systematically review vehicle trajectory collection, preprocessing, analytics, and applications. First, we focus on the standard techniques for vehicle trajectory collection and corresponding datasets. Next, we introduce representative approaches for the latest advances in vehicle trajectory processing. We further discuss individual travel behavior and collective mobility analytics using vehicle trajectory. Since private cars constitute the majority of urban vehicles and form the basis for many recent research findings, we emphasize analytics based on private car trajectory data. We then compile vehicle trajectory-boosted applications from the perspective of computing vehicle trajectory. Finally, we go through unresolved problems with vehicle trajectory and outline potential future research directions.

CCS Concepts: • **Applied computing**;

Additional Key Words and Phrases: Vehicle trajectory, trajectory processing, spatio-temporal data, travel behavior, mobility pattern

This work was supported in part by the National Natural Science Foundation of China under Grants U24A20247 and 62272152, in part by the National Key Research and Development Program of China under Grant 2023YFC3321601, in part by the Characteristic Innovation Project of Universities in Guangdong Province (No. 2024KTSCX228) and the Engineering Technology Research Center Project of Universities in Guangdong Province (No. 2024GCZX013).
Authors' Contact Information: Chenxi Liu, College of Computer Science and Electronic Engineering, Hunan University, Changsha, Hunan, China; e-mail: cxliu@hnu.edu.cn; Zhu Xiao (Corresponding author), College of Computer Science and Electronic Engineering, Hunan University, Changsha, Hunan, China; e-mail: zhxiao@hnu.edu.cn; Wangchen Long, The College of Artificial Intelligence, Zhuhai City Polytechnic, Zhuhai, China; e-mail: longwangchen@zhcpt.edu.cn; Tong Li, Tsinghua University, Beijing, Beijing, China; e-mail: tongli@mail.tsinghua.edu.cn; Hongbo Jiang, College of Computer Science and Electronic Engineering, Hunan University, Changsha, China, China; e-mail: hongbojiang2004@gmail.com; Keqin Li, Department of Computer Science, State University of New York, New York, Washington, United States; e-mail: lik@newpaltz.edu.

ACM 0360-0300/2025/05-ART231
https://doi.org/10.1145/3715902

ACM Reference Format:
Chenxi Liu, Zhu Xiao, Wangchen Long, Tong Li, Hongbo Jiang, and Keqin Li. 2025. Vehicle Trajectory Data Processing, Analytics, and Applications: A Survey. *ACM Comput. Surv.* 57, 9, Article 231 (May 2025), 36 pages.
https://doi.org/10.1145/3715902

1 Introduction

Urban vehicles have been on the rise in many cities during the past few years while the growing number of vehicles supports people's daily travel needs, such as running errands, commuting, shopping, and traveling [77, 81, 116, 173]. The rapid increase of urban vehicles provides considerable convenience for people's work and life, it also brings several challenges to cities, such as traffic congestion [91], environmental pollution [66], transportation safety [78], and parking difficulties [240]. Meanwhile, vehicles traveling in cities have led to a substantial increase in the volume and variety of vehicle trajectory collected from scalable sensors [120, 131], such as cameras, **radio frequency identification** (**RFID**), **global positioning system** (**GPS**) devices, and **on-board diagnostics** (**OBD**) systems. The vehicle trajectory contains valuable knowledge on human mobility patterns and travel preferences, offering excellent opportunities to solve the aforementioned problems [35, 152]. Consequently, vehicle trajectory processing and analytics can provide effective services for a series of applications, including smart city management [126], intelligent transportation systems [117], and location-based social networks [36].

Vehicle trajectory has garnered broad interest from academia and industry. Scholars have provided several surveys related to trajectory data from different perspectives [1, 3, 6, 158, 164, 180, 181]. For instance, Sousa et al. [164] surveyed vehicle trajectory similarity, including vehicle trajectory representations, similarity computation methods, and analysis applications. The authors in [181] categorized the spatio-temporal and introduced the deep learning methods for various trajectory data mining. However, these works did not discuss real-world vehicle trajectory collection. Alternatively, some existing surveys focused on vehicle trajectory analytics and applications [6]. For example, Rettore et al. [158] sorted out vehicular data sources and how they may be used and leveraged to deliver urban applications and services. Additionally, the authors in [1, 3, 180] discussed the recent developments in trajectory data management research, including trajectory preprocessing, storage, tools for trajectory analytics, and applications. Alsahfi et al. [3]. reviewed trajectory data models and indexing techniques, emphasizing the need for efficient storage and retrieval mechanisms. Alam et al. [1] discussed the trajectory data in spatio-temporal database systems, highlighting its role in applications like traffic management and urban planning. Wang et al. [180] focused on the advancements in trajectory query processing and privacy-preserving techniques. Nevertheless, existing surveys do not take travel behavior and mobility analysis into account, resulting in a lack of a comprehensive analysis of vehicle mobility. A systematic understanding of vehicle trajectory is necessary to depict the whole ecosystem, from data collection and processing to analytics and applications.

In this article, we aim at providing a systematic review of vehicle trajectory collection, preprocessing, mobility analytics, and applications to bridge the existing gap. Specifically, our survey seeks to answer the following questions: *In which ways and with what devices can collect the vehicle trajectory? How to process these trajectory data and mine the knowledge contained therein? What are the typical application scenarios of vehicle trajectory?* To that end, this work presents a survey of the ecosystem of vehicle trajectory. We first discuss the mainstream methods for vehicle trajectory collection and present various vehicle trajectory datasets. Next, we explore recent advances in vehicle trajectory processing by introducing representative methods. Following this, we elaborate on individual travel behavior and collective mobility analytics. Notably, we pay

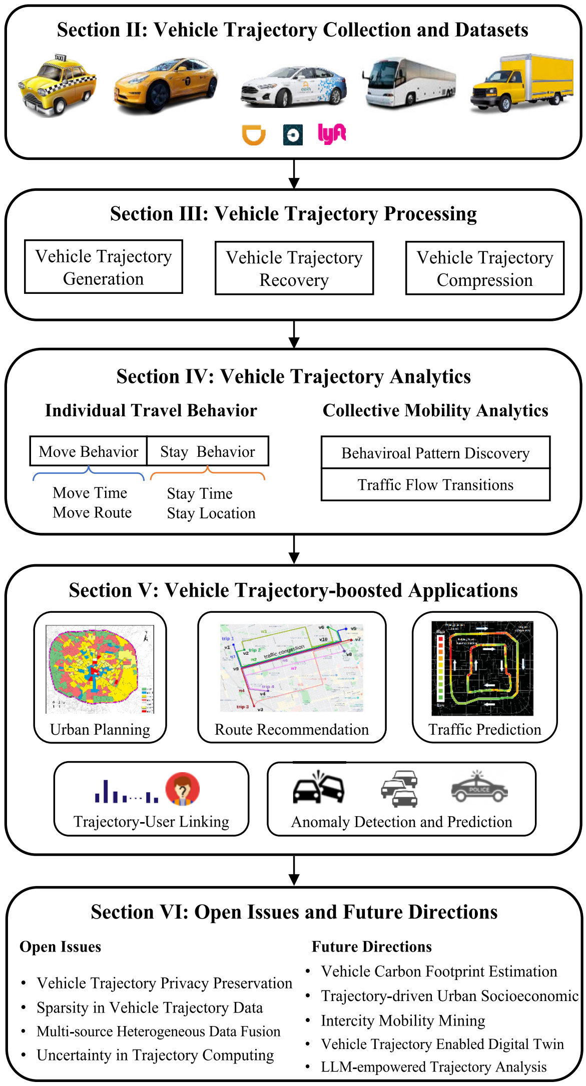
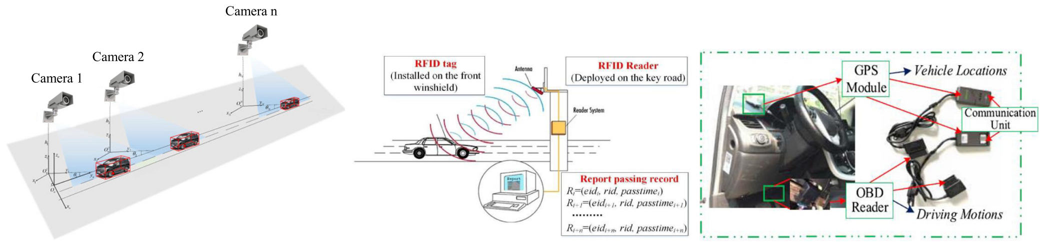


Fig. 1. Overall framework.

particular attention to analytics based on private car trajectory data, as private cars constitute the majority of urban vehicles and have been the focus of emerging research findings in recent years. Subsequently, we summarize the vehicle trajectory-boosted applications from the perspective of vehicle trajectory computing. Finally, we discuss the open issues related to vehicle trajectory and foresee future research directions.

The framework of this article is shown in Figure 1. We focus on taxis, ride-hailing services, private cars, buses, trucks, and **connected autonomous vehicles** (**CAVs**), as depicted in Table 2. In Section 2, we present vehicle trajectory collection methods and datasets. Next, we discuss recent progress in vehicle trajectory processing techniques in Section 3. After that, in Section 4, we introduce vehicle trajectory analytics. Additionally, in Section 5, we summarize vehicle-boosted applications. In Section 6, we outline the current open issues and foresee future research directions. Finally, we conclude the survey in Section 7.



(a) Cameras in [169] (b) RFID in [19] (c) GPS and OBD in [202]

Fig. 2. Sensors for vehicle trajectory collection.

Table 1. Comparison of Vehicle Trajectory Collection Sensors

Sensors	Type	Deployer	Scale	Cost	Collected information
Cameras	Static	Government	Citywide	High	Vehicle flow and speed
RFID	Static	Government, enterprises	Regional	High	Vehicle identification and flow
GPS	Mobile	Enterprises, individuals	Million	Low	Longitude, latitude, altitude, and speed
OBD	Mobile	Enterprises, individuals	Million	Low	Speed, acceleration, and steering direction

2 Vehicle Trajectory Collection and Datasets

In this section, we explore the methods for collecting vehicle trajectory, as well as various datasets that provide a foundation for subsequent analysis. Understanding these collection methods is crucial as they directly influence the quality of the vehicle trajectory processing and analysis.

2.1 Vehicle Trajectory Collection

Vehicle trajectory collection mainly relies on fixed sensors, mobile sensors, or a combination of these two [62, 63]. As illustrated in Figure 2, fixed sensors, which are pre-installed on the road or infrastructure such as loop detectors [194] and surveillance cameras [220, 230], can be utilized to extract vehicle trajectory. Mobile sensors refer to **Global Navigation Satellite System** (**GNSS**) device and OBD [16, 202, 203], which are equipped in vehicles to record the trajectory data. As a combination of fixed sensors and mobile sensors, the authors in [19, 247] proposed to use **Electronic Registration Identification** (**ERI**) to conduct vehicle trajectory acquisition. Table 1 shows the comparison of various vehicle trajectory collection methods.

Cameras. Modern smart cities and intelligent transportation systems depend heavily on real-time traffic monitoring [101]. Due to their value in managing and regulating traffic, security camera use has expanded quickly in recent years. As shown in Figure 2(a), a property pre-deployed surveillance camera can recognize individual vehicles. For instance, the authors in [230] proposed a queue length estimation model from cameras to extract vehicle trajectory. Yu et al. [220] collected the vehicle trajectory from a surveillance camera, as being called the camera-based trajectory. However, they could only determine traffic densities from the images by counting the number of pixels in the camera's images. While not covering the road segments without cameras, millions of camera-based trajectories can be utilized to predict the transitions between urban regions. To resolve this issue, the authors in [171] suggested using the installed traffic cameras as a sensor network to follow vehicle travel and try to reconstruct large-scale vehicle trajectories or estimate citywide traffic flow based on the insufficient camera-based trajectories.

RFID. Using RFID technology, vehicles can be accurately identified. As shown in Figure 2(b), every vehicle has an RFID tag attached, and critical roadways and crossroads have RFID scanners as acquisition points built into the infrastructure. Passive RFID tags on vehicles are activated, and passing records are created when the vehicle approaches an RFID reader (an acquisition point). Similar to video surveillance, this is collecting traffic data at a fixed site. Video surveillance and RFID both have advantages over one another. For instance, they assist in precisely identifying a specific vehicle and record information about all vehicles operating on roadways. Notably, the RFID-based trajectory collection technique offers benefits over video surveillance, such as a long recognition distance, high identification accuracy, quick reading speed, low production cost, and more, making it suited for collecting urban traffic data [19].

GPS. GPS is a frequently utilized worldwide approach for vehicle location in the trajectory gathering process since it can deliver precise and dependable vehicle localization and navigation performance in urban situations [203]. As shown in Figure 2(c), GPS devices, often integrated with OBD systems, are installed in vehicles to collect trajectory data [199]. These devices record essential information such as longitude, latitude, altitude, and speed over time. The deployment of GPS devices is highly scalable, capable of covering millions of vehicles, and is relatively low-cost compared to fixed sensors. The accuracy and reliability of GPS data make it an essential tool in intelligent transportation systems, enabling precise vehicle tracking and enhancing traffic management efficiency.

OBD. Using the OBD, we can collect vehicle trajectory from on-board sensors to monitor traffic events in urban regions. The authors in [203] proposed a new paradigm of **GPS and OBD Integration** (**GOI**), as shown in Figure 2(c), which offers a feasible way for large-scale trajectory collection especially suitable for private cars. TrajData, a systematic approach based solely on plug-and-play OBD devices, was developed by Xiao et al. [202]. They aggregated vehicle data from multiple sources using low-cost **commercial-off-the-shelf** (**COTS**) GPS and OBD modules.

2.2 Vehicle Trajectory Datasets

Vehicle trajectory represents vehicles' spatial and temporal locations as they travel [9], offering vital information for facilitating various applications (e.g., smart city, intelligent transportation, and location-based social networks) [102, 164]. Table 2 provides the overview of vehicle trajectory datasets collected from different sources. We divide the vehicle trajectory sources into the following categories: taxis, ride-hailing vehicles, private cars, buses, trucks, and CAV.

Taxi, typically equipped with GPS devices and operating under regulatory supervision, naturally generate large-scale and high-frequency trajectory data [33, 132, 189]. The scale of taxi trajectory datasets is generally large due to both the large number of taxis and their extended operation time [34, 84]. Representative taxi trajectory datasets include **T-Drive** [224], **TaxiNYC** [184], **TaxiUrbComp** [176], **TaxiPorto** [160], and the more recent **SynMob** [228]. **T-Drive** contains 4.96 million trajectories collected from over 33,000 taxis in Beijing, China, spanning three months. The example of taxi trajectories for T-Drive dataset is illustrated in Table 3. **TaxiNYC** focuses on New York, U.S., and includes more than 160 million trajectories from both yellow and green taxis, continuously updated over 12 months. **TaxiUrbComp** involves 12.8 million trajectories generated by 35,300 taxis in Chengdu and Beijing over two months. **TaxiPorto** comprises 1.7 million trajectories recorded by 442 taxis in Porto, Portugal, over an entire year, with GPS traces logged every 15 seconds, enabling analyses of recurring urban mobility patterns such as daily commuting flows. Meanwhile, **SynMob** is a synthetic dataset of 2 million trajectories across one month, offering a complementary source of data for model training and validation without the privacy concerns inherent in real-world data. Due to their richness, taxi trajectory datasets are widely employed in numerous transportation and urban planning tasks, including but not limited to traffic flow prediction [121, 241], traffic management [84], route optimization [18], and urban function classification [71, 257]. They provide insights into spatiotemporal travel patterns, informing decision-making to enhance mobility, reduce congestion, and improve overall transportation efficiency.

Table 2. Overview of Vehicle Trajectory Datasets

Source	Dataset	Public	Sensors	Statistics	Time Span	Collection Site
Taxi	T-Drive[224]	✓	GPS	33,000 taxis, 4.96 million trajectories	3 Months	Beijing, China
	TaxiNYC[184]	✓	GPS	160 million trajectories	12 Months	New York, U.S.
	TaxiUrbComp[176]	✓	GPS	35,300 taxis, 12.8 million trajectories	2 Months	Chengdu and Beijing, China
	TaxiPorto[160]	✓	GPS	442 taxis, 1.7 million trajectories	1 Year	Porto, Portugal
	SynMob[228]	✓	Synthetic GPS	2 million trajectories	1 Month	-
Ride-hailing	DiDiChuXing[216]	-	GPS	5,674,266 million trajectories	12 Months	Chengdu and Xi'an, China
	UberNYC[88]	✓	GPS	18.8 million trajectories	12 Months	New York, U.S.
	Grab-Posisi[88]	✓	GPS	84,000 trajectories	18 Months	Southeast Asia
Private Car	PriSH[174]	-	GPS	1,275 cars, 114 million trajectories	10 Months	Shanghai, China
	PriFX[32]	-	GPS, OBD	74 cars, 4,859 trajectories	7 Months	Shanghai, China
	PriIta[51]	-	GPS	779,000 cars, 128 million trajectories	1 Month	The whole Italy
	PriTraj[198]	✓	GPS, OBD	60,000 cars, 10 million trajectories	12 Months	China
Bus	SCD[150]	-	RFID, GPS	40 million trajectories	14 Days	Beijing, China
	BusGPS[147]	-	GPS	1,809 buse, 67,709 trajectories	28 Days	Beijing, China
	BusTra[64]	-	GPS	40 buse, 31,500 trajectories	2 Months	Singapore
	BusXA[133]	-	GPS	5,000 buses, 75 million trajectories	1 Month	Xi'an, China
Truck	TruckSC[125]	-	GPS	63 trucks, 679,849 trajectories	1 Month	Sichuan, China
	TruckTJ[25]	✓	GPS	940,000 trajectories	1 Month	Tianjin, China
	Greek[25]	✓	GPS	50 trucks, 276 trajectories	33 Days	Athens, Greece
CAV	Argoverse[12]	✓	GPS, Cameras	324,557 trajectories	41 Days	Miami and Pittsburgh, U.S.
	ApolloScape[134]	✓	GPS, Cameras	100,000 trajectories	155 Minutes	BeiJing, China
	LyftCA[69]	✓	GPS	20 vehicles and 170,000 trajectories	4 Months	Palo Alto, California, U.S.

Table 3. An Example of Taxi Trajectories in T-Drive Datasets

ID	Timestamp	Longitude	Latitude
10009	2008/2/2 13:39	116.26991	39.95215
10009	2008/2/2 13:49	116.30055	39.96021
10009	2008/2/2 13:59	116.31418	39.96519
10010	2008/2/2 13:44	116.56454	40.07186
10010	2008/2/2 13:54	116.56753	40.06372
10010	2008/2/2 14:04	116.5829	40.06109

Ride-hailing. Ride-hailing services, also known as carpool services, refer to the process in which a customer requests a customized ride through an online third-party platform, such as the well-known Uber and Didi. Representative ride-hailing trajectory datasets include **UberNYC**,[1] **DiDiChuXing**,[2] and **Grab-Posisi** [88]. **UberNYC**, a service offered by the NYC **Taxi & Limousine Commission** (**TLC**) in collaboration with Uber, encompasses nearly 4.5 million Uber pickups in New York City from April to September 2014, plus 14.3 million trips from January to June 2015. **DiDiChuXing** provides the original trajectories and order information of special-vehicle fleets operating primarily in Chengdu and Xi'an, China. Meanwhile, **Grab-Posisi** contains around 84,000 trajectories collected over 18 months across Southeast Asia. Drawing from these datasets, extensive research has emerged on topics such as traffic demand prediction [159, 216] and travel time estimation [111, 176].

After appropriate handling of users' privacy concerns, the **PriTraj**[3] dataset publicly provides both GPS and OBD trajectories of more than 60,000 private cars from over 20 cities in China. Table 4 presents sample trips from this dataset. Other private car trajectory datasets include **PriSH** [174], **PriFX** [32], and **PriIta** [51], which are either non-public or partially public. **PriSH** contains 114,030,503 records from 1,275 private cars in Shanghai between July 2014 and April

[1]https://github.com/fivethirtyeight/uber-tlc-foil-response
[2]https://gaia.didichuxing.com
[3]https://github.com/HunanUniversityZhuXiao/PrivateCarTrajectoryData

Table 4. An Example of Private Car Trajectory Trips

ID	Stay time	Move time	Stay Longtitude	Stay Latitude	Stay Duration
97811	2018/09/01 09:19:22	2018/09/01 09:38:03	113.779126	22.728238	00:18:41
97811	2018/09/01 09:52:45	2018/09/01 16:36:00	113.783398	22.698229	06:42:15
97811	2018/09/01 16:48:39	2018/09/01 17:54:23	113.784248	22.698069	01:05:54
104783	2018/09/03 14:30:44	2018/09/03 14:39:39	113.876633	22.518154	00:08:55
104783	2018/09/03 14:57:39	2018/09/03 17:29:11	113.895668	22.533185	02:31:42
104783	2018/09/03 17:32:23	2018/09/03 19:01:05	113.898268	22.543391	01:28:42

2015. **PriFX** data, collected from July 2014 to January 2015 in Shanghai's Fengxian district, integrate OBD and GPS information. **PriIta** was mainly gathered for insurance purposes and covers 128,363,000 journeys from 779,000 private cars in Italy. Leveraging these diverse datasets, researchers have pursued a wide range of studies, including semantic travel pattern mining [174], human mobility discovery [51], vehicle flow prediction [119], road network construction [76], and stay duration prediction [17].

Bus plays a crucial role in urban transportation systems [142]. Representative datasets include **SCD** [150] and **BusGPS** [147], both collected in Beijing, China. **SCD** comprises over 40 million bus trajectories, while **BusGPS** includes 67,709 trajectories from 1,809 buses. **BusTra** records the trajectories of five bus routes in Singapore between 6 May and 7 July, 2017, whereas **BusXA** compiles more than 75 million trajectories from over 5,000 buses in Xi'an, China. For other open-source bus datasets, see the data platform.[4] Existing studies utilizing bus trajectory data focus on mobility pattern prediction [150], travel time estimation [64, 133, 147], and mobility event detection [5].

Truck trajectories, typically recorded via GPS devices installed on heavy-duty diesel vehicles, serve as key data sources for analyzing emission patterns and other transportation-related impacts. They enable policymakers to develop targeted measures that alleviate negative externalities such as air pollution and congestion. Representative datasets include **TruckSC** [125], **TruckTJ** [25], and Greek [25], each highlighting different operational contexts for heavy trucks. Specifically, **TruckSC** contains 679,849 trajectories collected over one month in Sichuan, China, while **TruckTJ** includes 940,000 trajectories in Tianjin, China, also within a one-month span. **Greek** presents a smaller-scale dataset, 276 trajectories from 50 trucks, recorded over 33 days in Athens, Greece. These truck trajectory datasets provide critical insights for advancing environmental sustainability [151], improving logistics efficiency [96], and enhancing congestion management [52, 99].

Connected autonomous vehicles (CAV). Two notable CAV datasets include **Argoverse**[5] [12] and **ApolloScape**[6] [134], alongside **LyftCA**[7] [69]. **Argoverse** stands as the first large-scale autonomous driving dataset offering HD maps with geometric and semantic annotations. **ApolloScape** provides manually annotated trajectories, LiDAR point clouds, and camera-based images collected under various lighting and traffic conditions, featuring cyclists, pedestrians, and automobiles in complex urban environments. Meanwhile, **LyftCA** comprises approximately 170,000 scenarios acquired from 20 Lyft self-driving vehicles over four months in Palo Alto, California; each scenario lasts 25 seconds and includes perceptual outputs of the self-driving system. These datasets underpin the development of algorithms for perception [10, 206], navigation [48], control

[4] https://data.world/datasets/bus
[5] https://github.com/argoai/argoverse-api
[6] http://apolloscape.auto/trajectory.html
[7] https://level-5.global/level5/data/

Table 5. Overview of Vehicle Trajectory Generation Data Methods

Reference	Methods	Evaluation Metrics		
		Similarity	Privacy	Diversity
He [65]	Transfer Learning	✓	-	-
Ouyang [146]	Generative Adversarial Network, Convolutional Neural Network	✓	-	-
SVAE [74]	Variational Autoencoder, Long Short-Term Memory network	✓	-	✓
MoveSim [45]	Generative Adversarial Network, Convolutional Neural Network	✓	✓	-
TrajGAIL [28]	Generative Adversarial Imitation Learning, Recurrent Neural Network	✓	-	-
TrajGen [11]	Generative Adversarial Network, Map Matching	✓	-	-
STULIG [251]	Variational Autoencoder, Convolutional Neural Network	✓	-	-
Wang [192]	Gated Recurrent Unit, Mixture Density Network	✓	-	-
Benarous [8]	Long Short-Term Memory network, Variable-order Markov model	✓	✓	✓
ActSTD [226]	Generative Adversarial Imitation Learning, Neural Differential Equations	✓	-	-
Jiang [85]	Generative Adversarial Network, A* Search Algorithm	-	-	✓
TrajGDM [29]	Diffusion Model, Transformer, Long Short-Term Memory network	✓	✓	✓
DiffTraj [259]	Diffusion Model, Convolutional Neural Network	✓	✓	✓

[166], and simulation [233], ultimately advancing the deployment of safe autonomous driving systems [205].

3 Vehicle Trajectory Processing

Building on the data collection methods discussed in Section 2, this Section delves into the recent advancements in vehicle trajectory processing techniques to enhance the usability and capability of the vehicle trajectory, including vehicle trajectory generation (Section 3.1), vehicle trajectory recovery (Section 3.2) and vehicle trajectory compression (Section 3.3).

3.1 Vehicle Trajectory Generation

Vehicle trajectory generation aims at producing synthetic vehicle trajectories similar to real-world trajectories, serving a dual purpose. (i) It provides an effective solution to address data deficiencies. Data collection can be costly or challenging in practical scenarios, leading to insufficient data to meet application requirements. Trajectory data generation technology resolves this issue by providing sufficient synthetic yet realistic data, fostering diverse applications. (ii) It addresses the privacy leakage problem. During trajectory data collection, private information such as vehicle ownership details, locations, and personal data may be included in real-world vehicle trajectories. To safeguard privacy while maintaining the qualities of real data, trajectory data generation can replace or remove privacy-sensitive data while ensuring a tolerable level of resemblance. Table 5 presents the overview of vehicle trajectory generation methods.

Evaluation metrics. Generating synthetic vehicle trajectories requires realistically reproducing the mobility patterns, keeping the diversity of trajectories, and protecting privacy. Determining the quality of synthetic data can be difficult as it is not always evident how to do so. As shown in Table 5, generating synthetic vehicle trajectories can be evaluated in terms of three aspects: *similarity*, *privacy*, and *diversity*. The similarity is a widely-used evaluation index, while the similarity evaluation methods in various application scenarios are usually different. Privacy is another important evaluation metric used to measure whether the model can effectively protect users' sensitive information and is not easy to reidentify. Diversity evaluates if a dataset consists of a single repeating diary by determining whether the produced diaries have the same diversity of diaries as the original diaries. Benarous et al. [8] investigated the usage of several synthetic data production models for extended location sequences and provided a road map for establishing synthetic data creation strategies. They evaluated the effectiveness of the various models using various criteria, including privacy, statistical similarity, per-instance similarity, and diversity.

Trajectory generation methods. Trajectory generation has been tackled using various techniques, including **generative adversarial networks** (**GANs**), **variational autoencoders** (**VAEs**), **sequence-to-sequence** (**Seq2Seq**) models, **partially observable Markov decision processes** (**POMDPs**), and transfer learning. These methods address the spatial and temporal complexities of human mobility, with each offering unique contributions and perspectives. For examples, Ouyang et al. [146] converted spatial locations into a 2D matrix and employed GANs to generate data points, which were later reconstructed into sequential trajectories. This approach highlighted the potential of GANs in modeling discrete spatial data. Building on this, Huang et al. [74] combined VAEs to learn latent trajectory features with Seq2Seq models to ensure coherent temporal patterns, improving trajectory realism. Extending this line of work, Feng et al. [45] introduced a self-attention-based network to capture temporal dynamics and pre-trained GANs to generate trajectories that respect mobility regularities. He et al. [65] shifted the focus to transfer learning by leveraging data from multiple source cities to predict travel paths for target cities, showcasing the utility of cross-city mobility knowledge transfer. Choi et al. [28] approached trajectory generation through POMDPs, framing it as an imitation learning problem. Cao et al. [11] further diversified the methods by separating spatial and temporal components: Seq2Seq models handled temporal information, while GANs generated spatial features, ensuring modular and flexible trajectory generation. Recently, Jiang et al. [85] proposed TS-TrajGen, a two-stage GAN framework that integrates domain-specific knowledge to generate continuous trajectories over road networks, emphasizing seamless spatial continuity. Yuan et al. [226] added to this by modeling the dynamics of activity-based mobility, combining continuous flow transitions with updates at activity locations. Similarly, Zhu et al. [259] introduced DiffTraj, which utilizes diffusion models to reconstruct trajectories from noise, effectively capturing both spatial and temporal features through reverse denoising.

In summary, these methods collectively advance trajectory generation by addressing diverse aspects of human mobility, from spatial discretization to activity-based modeling, cross-city transfer, and spatio-temporal regularities. Each builds on previous insights, creating a comprehensive framework for realistic and adaptable synthetic trajectory data generation.

3.2 Vehicle Trajectory Recovery

A high sampling rate of trajectories is essential for urban applications based on vehicle trajectory. However, due to communication loss and storage limitations, massive trajectories are collected at a low sampling rate in realistic conditions. To address such data sparsity issues, trajectory recovery/interpolation/reconstruction/completion/imputation and data cleaning methods are proposed. These methods infer missing data or impute unobserved data, thereby improving the completeness of the trajectory data. Table 6 provides an overview of vehicle trajectory recovery methods.

Statistical methods. Vehicle trajectory recovery is to impute missing or unobserved data for low-sampling-rate trajectories and rebuild high-quality trajectories. Provided that vehicles are moving with uniform speeds [68] and the traffic follows a probability distribution, statistical methods are designed to utilize the statistical features of the probability distribution to model multi-dimensional spatio-temporal correlations of trajectory data. In this line, various statistical methods have been proposed to achieve the goal of trajectory data imputation. For instance, Chen et al. [21] addressed missing data recovery as a tensor completion issue and suggested a three-procedure architecture. In so doing, it allows for discovering traffic patterns to solve the

Table 6. Overview of Vehicle Trajectory Recovery Methods

Category	Reference	Methods	Evaluation metrics
Statistic	STD [21]	Tensor Decomposition	RMSE, MAE, MRE
	nGPF-VSE [208]	Generalized Error Distribution, Particle Filter	RMSE
	BGCP [20]	Bayesian Probabilistic Matrix Factorization	MAPE, RMSE
	Kaur [90]	Boosting Regressor, Bias Correction	MSB, MSPE, MAPE
	ST-TRPCA [47]	Principal Component Analysis, Attention Mechanism	MAE, RMSE, MAPE
Deep Learning	DHTR [178]	Kalman Filter, Attention Mechanism	nDTW, LCSS, EDR
	Bi-GTPPP [195]	Long Short-Term Memory network	Recall, F1, MAP
	MTrajRec [157]	Multi-task Learning, Gated Recurrent Unit	MAE, RMSE, Recall
	AttnMove [197]	Attention Mechanism	Recall, MAP, Distance
	R2C-TrA [198]	Transfer Learning	RMSE
	STCPA [212]	Self-attention	RMSE
	GraphMM [127]	Conditional Model, Graph Neural Network,	Accuracy, R-LCS
	RNTrajRec [24]	Graph Attention Network, Transformer	MAE, RMSE, Recall, Precision, F1
	TERI [22]	Transformer, Contrastive Learning	Precision, Recall
	LightTR [128]	Federated Learning, Gated Recurrent Unit	Precision, Recall, MAE, RMSE

RMSE=Root Mean Squared Error, MAE=Mean Absolute Error, MRE=Mean Relative Error, MAPE=Mean Absolute Percentage Error, MSB=Mean Squared Bias, MSPE=Mean Squared Prediction Error, nDTW=normalized Dynamic Time Warping, LCSS=Longest Common Subsequence, EDR=Endpoint Detection and Response, R-LCS=Ratio of Longest Common Subsequence.

data recovery problem. To compute the non-Gaussian probability density during the vehicle state estimation, Xiao et al. [208] regarded the noise in the trajectory to be non-Gaussian and took advantage of the generalized error distribution. In [20], the authors used a Bayesian probabilistic matrix factorization model to determine the spatio-temporal correlations for trajectory data imputation. Regarding realizing a computationally efficient data imputation, Kaur et al. [90] provided a dual-stage error-corrected boosting regressor-based imputation strategy that also evaluated the uncertainty around imputed values. Tensor robust principal component analysis was used by Feng et al. [47] to capture the spatial and temporal correlations in trajectories for data recovery.

Deep learning methods. Existing methods explore trajectory data's spatial, temporal, and contextual relationships to improve data imputation processes. Initially, Wang et al. [178] utilize both spatial and temporal attention to capture spatio-temporal correlations, in which the improved Kalman filter is integrated to calibrate noise for trajectory recovery. Subsequently, Ren et al. [157] attempt to leverage multi-task learning to learn sequential dependencies from low sampling-rate trajectories and solve the map-constrained trajectory recovery problem. Xia et al. [197] introduced an attention mechanism to exploit intra-trajectory and inter-trajectory correlations from long-term history for trajectory recovery. Rempe et al. [156] utilized deep convolutional neural networks to capture spatio-temporal correlations between urban grids for trajectory recovery. Yuan et al. [227] tried to learn the distribution of incomplete data and utilized a generative adversarial network to generate the corresponding imputed data. Xiao et al. [198] integrate ensemble learning and transfer learning to construct a fine-grained prediction model to implement vehicle trajectory recovery. Xu et al. [212] provide the STCPA to the trajectory speed imputation. Through the attention mechanism, STCPA captures intricate traffic connections between the spatial and temporal dimensions, assisting in mitigating the data sparsity problem. Liu et al. [127] propose GraphMM, which is a graph-based approach that leverages graph neural networks and conditional models to align road segments and trajectories in latent space, effectively capturing correlations in road and trajectory graph topologies with efficient training and inference algorithms for scalability.

Table 7. Overview of Vehicle Trajectory Compression Methods

Category	Reference	Methods	Mode	Evaluation Metrics
Line Simplification	GRTS [98]	Segment Heuristic	Online	RE, CC
	SQUISH [144]	Priority Queue	Online	SED, Speed
	Long [129]	Binary Search	Offline	AD
	BQS [124]	Bounded Quadrant System	Online	CR, PP
	REST [244]	Greedy Algorithm, Douglas-Peucker	Offline	SRS, CR
	Lin [110]	Synchronous Euclidean Distance	Offline	CR, AE, RT
	RLTS [190]	Reinforcement Learning, Markov Decision Process	Online	SED
	MARL4TS [191]	Multi-Agent Reinforcement Learning	Online	CR, RT
Map Matching	Sandu [161]	Douglas-Peucker	Offline	DC, ASE
	COMPRESS [61]	Spatio-temporal Decomposer	Offline	CR, RT
	Yang [215]	Trajectories Representation	Offline	CR, RT
	CiNCT [95]	Suffix Arrays	Offline	CR, RT
	TrajCompressor [14]	Heading Change Compression	Online	CR, RT
	UTCQ [104]	Referential Representation	Offline	CR, RT, MC
	TRACE [103]	Referential Representation	Online	CR, RT, MC

RE=Reduction Efficiency, CC=Computational Costs, SED=Synchronous Euclidean Distance, AD=Angular Difference, CR=Compression Ratios, PP=Pruning Power, SRS=Size of Reference Set, AE=Average Errors, RT=Running Time, MC=Memory Cost, DC=Degree of Compression, ASE=Average Synchronous Error.

Chen et al. [24] propose a RNTrajRec framework to combine road network representation with GPS trajectory representation for trajectory recovery. To address trajectory recovery without previous knowledge, Chen et al. [22] present a TERI framework, a two-stage technique that first detects recovery positions and then imputes the missing data points. Most recently, Liu et al. [128] propose LightTR, a framework for federated trajectory recovery based on a client-server architecture, to trajectory recovery while maintaining the privacy and decentralization of the data in each client.

To sum up, statistical methods aim at discovering specific patterns or distribution characteristics inherent in the trajectory data to achieve trajectory recovery. Deep learning methods attempt to implement trajectory recovery by exploring the spatio-temporal correlations in an end-to-end manner, which relies on a training model and has higher data volume and quality requirements.

3.3 Vehicle Trajectory Compression

During trajectory collecting, several Internet-of-Vehicle devices send massive amounts of vehicle trajectory to the data center [246], which poses various sustainable problems concerning storage, transmission, and processing [14, 16]. Many trajectory data compression methodologies are proposed to resolve these problems [137]. They are mainly categorized into line simplification methods and map-matching methods. Table 7 presents the overview of vehicle trajectory compression methods.

Line simplification methods. Line simplification techniques aim at reducing trajectory data by eliminating unnecessary or irrelevant points, typically using distance-based measurements. These methods play a crucial role in optimizing storage, transmission, and analysis of trajectory data. For instance, Lange et al. [98] proposed an online approach using dead reckoning to maximize message delivery over wireless networks while reducing data points. Muckell et al. [144] employed a priority queue to compress GPS trajectory streams, removing low-information points based on synchronized Euclidean distance estimation. Similarly, Long et al. [129] focused on direction-preserving trajectory simplification, using dynamic programming and binary search to minimize direction-based errors in the simplified trajectory. Liu et al. [124] introduced the Bounded Quadrant System algorithm, designed for resource-constrained environments, to produce error-bounded simplified trajectories. Zhao et al. [244] represented raw trajectories as a series of sub-trajectories, leveraging greedy and dynamic programming algorithms for reference trajectory construction and compression. In related work, Lin et al. [110] also employed synchronous Euclidean distance as an error metric for simplification.

Later, Wang et al. [190, 191] treated trajectory simplification as a sequential decision process, applying reinforcement learning to determine which points to drop, eliminating the need for human-defined rules. In summary, line simplification methods have evolved from heuristic-based approaches to advanced algorithms like dynamic programming and reinforcement learning. These methods effectively balance data reduction with the preservation of essential trajectory features, catering to diverse application scenarios and resource constraints.

Map-matching methods. Map-matching algorithms project trajectories onto road networks, representing them as sequences of road segments to reduce storage requirements while preserving spatial and temporal relationships [23, 252]. Sandu et al. [161] laid the foundation for trajectory compression in road networks by integrating the Douglas-Peucker technique, enabling compression with predictable error bounds. Building on this, Han et al. [61] extended the approach by separating trajectories into spatial routes and temporal sequences, allowing for parallel compression tailored to each component. Yang et al. [215] took a different direction by introducing a distance-time pair representation for trajectories in road networks. Koide et al. [95] further refined these ideas by treating trajectories as sequences of road edges, employing pattern matching and sub-path extraction to enhance compression while maintaining trajectory integrity. Expanding the focus to mobile applications, Chen et al. [14] developed a lightweight online map-matching system. Their approach utilized vehicle heading data to align noisy GPS points efficiently and introduced a compressor leveraging heading changes at junctions for compact trajectory representation. Building on this practical application, Li et al. [104] addressed the challenge of compressing uncertain trajectories, proposing a referential representation. They later extended this work to an online setting [103], incorporating speed-based and multi-reference techniques to balance compression ratio and processing speed.

In summary, map-matching methods have evolved from foundational compression techniques to sophisticated algorithms leveraging spatial, temporal, and contextual road network data. These methods effectively reduce storage requirements and enhance trajectory utility, although they rely on updated road network data and can have high computational costs.

4 Vehicle Trajectory Analytics

With a foundation of robust data collection and advanced processing techniques established in Sections 2 and 3, we now move on to the analytics of vehicle trajectory. This Section analyses individual travel behavior and collective mobility patterns derived from the processed trajectory data. Understanding these analytics is crucial for deriving meaningful insights that can inform various applications, from traffic management to urban planning. Notably, we pay particular attention to analytics based on private car trajectory data, as private cars constitute the majority of urban vehicles and have been the focus of emerging research findings in recent years.

4.1 Individual Travel Behavior

Trajectories consist of multiple spatio-temporal points, reflecting individual travel behaviors. Figure 3 illustrates this, where each location represents a stay behavior, and the paths between them denote move behaviors. This subsection reviews studies on both move and stay behaviors.

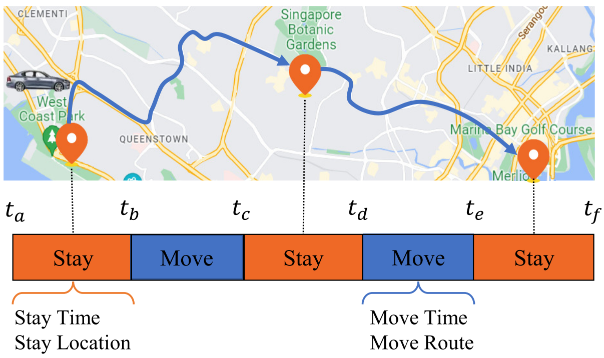


Fig. 3. Individual travel behavior analytics.

4.1.1 *Move Behavior.* Move behavior can be quantified by mobility metrics [163], such as move time and route, as presented in Table 8. Move time, also known as travel time, is the primary feature of a vehicle when traveling from the origin to the destination. The **estimated time of arrival** (**ETA**) or travel time estimation is one of the crucial areas in vehicle trajectory analytics [67].

In the following, we elaborate on the four latest models of ETA. To specify, the DeepOD model was developed by Yuan et al. [223] to completely utilize past trajectories, road networks, weather, and traffic conditions data for ETA. Hong et al. [67] leveraged heterogeneous information graphs in the ETA task. They proposed a network based on vehicle trajectory to analyze traffic behavior patterns simultaneously. Jin et al. [89]proposed a spatiotemporal graph search framework to estimate travel time. The combined spatio-temporal correlations of intersections and road segments are gathered using a hierarchical neural architecture search approach with internal and exterior search space. Fu et al. [49] developed CompactETA, a real-time ETA learning system that offered an accurate online travel time inference within 100 microseconds. From the upper part of Table 8, we can conclude that the main techniques for ETA have been deep graph learning over the past two years.

Apart from move time of the move behavior, many scholars have analyzed mobility metrics during the routes of the vehicle's travel, such as move route, mile, and speed. In [218], the authors stated that discovering common movement paths would be essential for comprehending the principles behind travel behavior. They used extended label propagation clustering to find frequent paths and created association rules to extract frequent move paths over time. Su et al. [165] developed a personalized route description system that uses previous trajectory data to produce higher-quality customized route descriptions for urban commuters. Todi et al. proposed a path-planning algorithm [170]. They created random trees with long short-term memory to forecast the course of obstacles in a real-world setting and integrated it with a probabilistic and dynamic planning framework. The uncertainty problem, particularly, has received special attention in move path planning. Zyner et al. [260] developed a method for predicting multi-modal routes with uncertainty. The multi-modal data is processed using a mixed-density network. In [26], the authors examined the problem of determining a priory shortest paths in a bus network under trip time uncertainty, where residents who use public transportation as their routing options are less flexible than those of private car drivers. Besides, move mileage and speed are two critical features of the move behavior. **Vehicle miles traveled** (**VMT**), which refers to the total number of miles driven by vehicles, is one of the most significant performance indicators for traffic planning. In [39], the authors calculated statewide VMT by functional class of routes through Maryland using extensive vehicle GPS trajectory data. For the moving speed, the authors in [167] predicted bus speed by identifying critical intrinsic and extrinsic factors influencing bus speed and their importance in various conditions. They investigate spatial, temporal, and contextual connections to extract features for training the model to achieve accurate bus speed prediction.

4.1.2 *Stay Behavior.* Stay behavior can be observed in people's daily travel. Specifically, many people travel from different parts of the city, arrive, and *"stay"* in several regions to participate in

Table 8. Studies on Move Behavior Analytics

Metric	Reference	Data Source	Methods	Remarks
Move Time	DeepOD [223]	Ride-hailing	TG	Road network OD ETA with external data.
	HetETA [67]	Ride-hailing	HIN	ETA considering traffic behavior pattern.
	CompactETA [49]	Ride-hailing	GAT	A fast inference system for real-time ETA.
	Auto-STDGCN [89]	Taxi	NAS, GCN	Hierarchical neural search for path ETA.
Move Route	Zheng [26]	Bus	PTA	Route Planning under travel time uncertainty.
	Yu [218]	Taxi	LPC, AR	Frequent route discovery for urban travel flow.
	PerRD [165]	Taxi, bus	KT	A system to generate customized route descriptions.
	Fan [39]	Truck	SPAH	Miles travelled estimation based on Spark.
	Zyner [260]	CAV	MDN	Multi-modal route prediction with uncertainty.

TG=Temporal Graph, HIN=Heterogeneous Information Network, GAT=Graph Attention Network, MDN=Mixture Density Network, NAS=Neural Architecture Search, PTA=Polynomial-Time Algorithm, LPC=Label Propagation Clustering, KT=Knowledge Trees, AR=Association Rule, SPAH=Shortest Path Algorithm Heuristic.

their activities [201]. Such stay behaviors, in essence, indicate the visit locations of vehicles and represent people's travel demand and preferences [115]. Existing works on stay behavior inference have focused on two aspects, i.e., stay location [40] and stay time [7].

On the location level, many scholars analyzed the user's stay behavior for location prediction [42]. In [15], the authors devised a semantic-aware technique for predicting user stay locations. They retrieved user interests to reflect the user's preference for various visit places. The user interests were then input into a semantic-aware recurrent neural network to capture users' travel patterns. Hu et al. [72] developed a platform for globally applicable stay point-based location analytics that obtains stay points with customizable parameters, stay points with locations, extracts location profiles, and visualizes analysis findings. They demonstrated three applications based on the combination of these capabilities: illegal location finding and popular location ranking. Wei et al. modeled the stay behavior with generative adversarial imitation learning and integrated the stochastic constraints from system dynamics in the learning process [193]. They were the first to learn to model the state transition of moving agents with system dynamics. In [130], by concurrently collecting the regularity and preference based on the stay time, location, frequency, and duration features, the authors proposed a deep neural network to forecast the stay location of private cars. In [13], the authors proposed a topic-based station-free data mining approach to understand shared bike riders' stay behavior better. Fan et al. [40] proposed a location prediction approach that uses stay feature extraction. The approach built historical trajectory connections using trajectory data. They applied location discovery methods to convert the historical trajectory links into stay point links, then used density-based spatial clustering of applications with noise to cluster the stay points to generate clustering linkages. Huang et al. [75] explored how different private car users traveled. They obtained stay features from private car trajectories using DBSCAN, which is most likely to match stay locations, in other words, frequently-visit places. The key findings in [75] indicate that private car users with fewer FVPs are more likely to exhibit travel regularity, and their travel behavior is relatively fixed.

Regarding the time level, we define the stay time as two categories: stay duration and stay time, the former is exact time when the stay starts or ends, while the later is duration of the stay. As depicted in Table 9, most studies of stay time focus on private cars. Private cars tend to stay in specific regions during daily driving [75, 174]. Therefore, most studies on the analysis of private car trajectories focus on the stay behavior. Zhang [238] et al. proposed an RNN-based encoder model to address the difficulties in stay time prediction, namely the associated randomness and uncertainty. The prediction model uses neural arithmetic logic units to increase the neural network's capability

Table 9. Studies on Stay Behavior Analytics

Metric	Reference	Data Source	Methods	Remarks
Stay Location	Sem-LSTM [15]	Private car	DBSCAN	Semantic stay location prediction based on AOI.
	SALON [72]	Truck, taxi	DBSACN	A stay point-based location analysis platform.
	MoveSD [193]	Cameras	GAIL	Location prediction based on system dynamics.
	DeepRP [130]	Private car	AM	Location prediction with regularity and preference.
Stay Time	Li [107]	Taxi	VD	The predictability analysis of stay time.
	SOI [17]	Private car	KDE	Stay duration prediction.
	MSFD [115]	Private car	SSE	Regular travel behavior analysis.
	Zhang [238]	Private car	RNN, NALU	Stay duration prediction in various time scales.

DBSCAN=Density-Based Spatial Clustering of Applications with Noise, AOI = area of interest, AM=Attention Mechanism, KPCA=Kernel Principal Component Analysis, GAIL=Generative Adversarial Imitation Learning, VD=Voronoi Diagram, KDE=Kernal Density Estimation, SSE=the Sum of Squares due to Error, NALU=Neural Arithmetic Logic Units.

to handle linear relationships and, more crucially, lessen the effect of sparsity and unpredictability on staying events while forecasting stay time. In [17], the authors proposed the stay-of-interest model for forecasting the stay duration. The goal is to predict how long a private car would stay in a specific location, which is essential information for car services like semantic travel analysis and innovative recommendation services. The authors in [107] investigated the prediction of stay duration using real-world taxi trajectories. They computed the entropy, provided the bounds of predictability in vehicle stay time, and observed that the average stay duration followed the same pattern across days. Considering the substantial correlation between stay location and stay time, many researchers investigate how to integrate these two metrics. For instance, Liu et al. [115] explored the stay behavior in light of three stay features, i.e., stay location frequency, stay time, and stay duration. Based on the sum of the squared errors, they specifically got the fitting distribution of stay characteristics. Baumann et al. [7] first drew the normal probability of arrival time and duration predictability vectors, then predicted the time and duration of users at their stay locations.

Most of the existing studies of stay behavior are focused on private cars. The reason is that one can extract fine-grained stay behavior from private car trajectory data, while other trajectories, such as taxis and buses, do not contain meaningful information on stay behavior. The reasons behind this can be explained as follows: (i) The private car trajectory data offers easy-to-extract attributes characterizing people's stay behavior as it is straightforward to retrieve the stay location and stay duration from the time series trajectory trips. Recall Table 4, the private car dataset records the trajectory trip, which indicates the stay time, location, and duration. (ii) Public transportation such as buses travel along with predefined routes on a preset schedule. Their stay locations are preset, and the stay times are often unchanged. Their departure/routing choices are less flexible than private cars [26]. Those trajectories of buses help design but fail to provide essential information for investigating stay behaviors. (iii) Taxis and ride-hailing cars offer transportation for a single passenger or a small group of passengers; these vehicles' stay behaviors are primarily focused on picking up and dropping off passengers. As such, trajectory data of taxis and ride-hailing cars contain incomplete data on stay behavior due to a lack of information related to stay behavior, especially stay time.

To summarize, understanding stay behavior is an intriguing problem as it provides not only essential information about people's daily travel but also benefits broad applications such as POI recommendations, parking lot settings, and urban planning. On this point, the stay behavior deserves further exploration, and private car trajectory offers the best dataset that properly reflects people's stay behavior when looking into the current variety of vehicle trajectory.

Table 10. Overview of Collective Mobility Analytics

Category	Main focus	Reference	Data source	Methods	Remarks
Behavioral Pattern Discovery	Movement	Naveh [145]	Bus	TF	Weekly periodicity of movement pattern.
		IERP [210]	Private car	KPCA	Regular travel behavior during workdays.
		Wang [179]	Taxi	PR	Multiscale time-span traffic patterns.
	Aggregation	Wang [175]	Private car	KDE	Daily aggregation pattern prediction.
		VBGMM [200]	Private car	BGM	Evolution of urban region attractiveness.
		Yu [217]	Taxi	Cluster	Daily hotspot patterns discovery.
	Emission	Li[105]	Taxi	Statistics	Emission patterns of traffic analysis zones.
		Yu [221]	Camera	Statistics	Uneven distribution of vehicle emissions
		Cheng [25]	Truck	Statistics	Summarize 16 pollutant emission patterns.
Traffic Flow Transitions	Traffic flow	JMDI [168]	Taxi	RL	Citywide traffic flow estimation.
		Zhao [245]	Probe car	PBNN, GAN	Estimate queue length and traffic flow.
		MBA-STNet [138]	Taxi	BNN, GAN	Multi-task learning for traffic flow prediction.
	O-D Flow	Riascos [149]	Taxi	Statistics	Dynamical analysis on spatial activity of taxis.
		CAS–CNN [234]	Rail metro	CNN	Short-term O-D demand prediction.
		MG-GAN [118]	Private car	GAN	Regional transfer flow prediction
	Travel Demand	PGCM [100]	Bus	GCN, BNN	Confidence interval bus demand prediction.
		DTCNN [37]	Taxi	CNN	Citywide traffic demand prediction.
		MLRNN [232]	Taxi	RNN	Region-level taxi demand prediction.

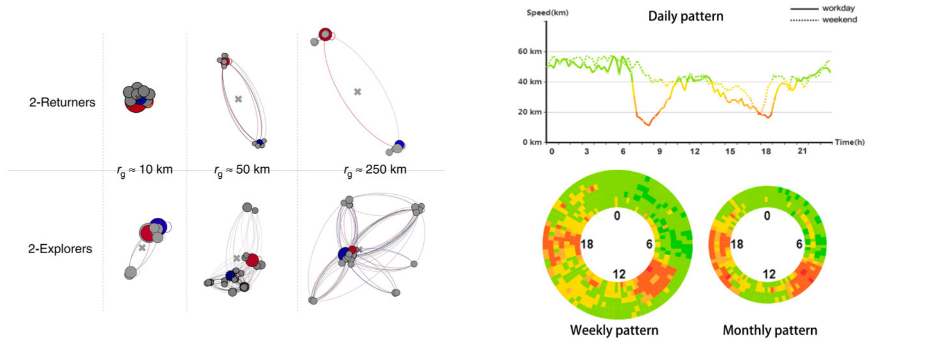
TF=Tensor Factorization, KPCA=Kernel PCA, PR=Pattern Recognition, KDE=Kernel Density Estimation, BGM=Bayesian Gaussian Mixture, RL=Reinforcement Learning, PT=Probability Theory, O-D=Origin-Destination, CNN=Convolutional Neural Network, BNN=Bayesian Neural Network, GAN=Generative Adversarial Network.

4.2 Collective Mobility Analytics

The collective mobility analysis can be divided into two categories. One is behavioral pattern discovery, which focuses on the spatio-temporal evolution regularity of vehicle travel behavior, such as movement patterns, aggregation patterns, and emission patterns. The other is traffic flow transitions, which indicate the dynamic migration characteristics of vehicle groups traveling across urban regions, including traffic flow prediction, **origin-destination** (**O-D**) flow analysis, and travel demand estimation. Table 10 presents the overview of collective mobility analytics.

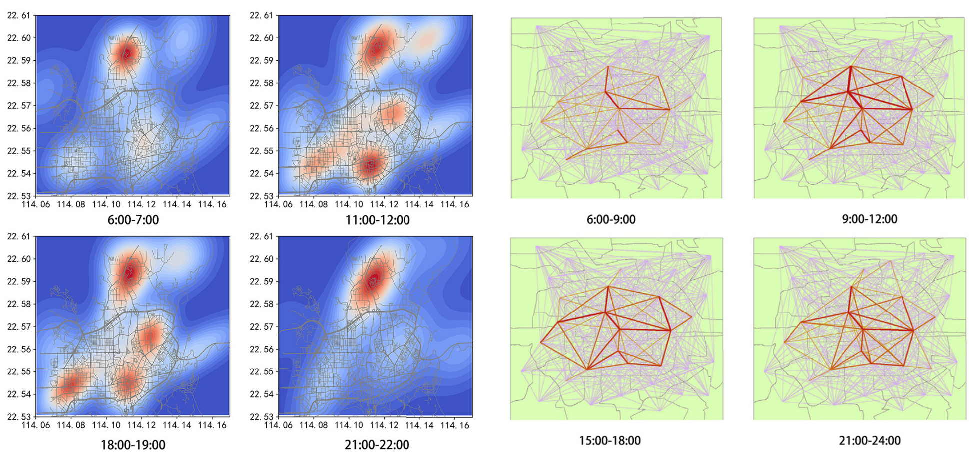
4.2.1 Behavioral Pattern Discovery. The vehicle trajectory contains a wealth of user behavior parameters that allow the discovery of vehicle behavioral patterns that support a wide range of applications, from transportation management to urban computing.

Movement pattern discovery aims at finding out spatio-temporal regularity from the vehicle trajectory dataset. In [148], Barabási et al. unveiled two distinct mobility patterns, i.e., returners and explorers. As shown in Figure 4(a), the distance traveled by returners and explorers is estimated by the total radius of gyration r_g. For instance, the traveled distance of a two-returner is mainly determined by the two frequently-visited places, namely the FVPs (see the circles in Figure 4(a)), which typically respond to the residence and working place. A two-explorer, on the other hand, frequently goes between many different locations, e.g., occasionally-visited places. Naveh et al. [145] studied spatio-temporal movement patterns derived from large-scale urban public transportation trajectory data. They discovered that public transportation mobility patterns were geographically more limited, with identifiable peaks and valleys in their temporal profiles. Xiao et al. [210] investigated the regular travel patterns based on large-scale private car trajectory data. Their findings are two-folds, (i) showing that private car movement in urban regions has a degree of regularity. One such viewpoint is consistent with people's everyday travel patterns, as proved by the results in [148, 209]; (ii) revealing the spatio-temporal coupling correlation of



(a) The returners and explorers patterns [148].

(b) Time-scale interactive movement patterns [179].



(c) Vehicle aggregations in Shenzhen [200].

(d) Emission pattern discovered in Beijing [105].

Fig. 4. Behavioral patterns. (a) Returners limit much of their mobility to a few locations, typically corresponding to their FVPs, while the mobility of explorers cannot be reduced to a few locations and spread to many OVPs. (b) The real-time interactive traffic patterns based on taxi trajectories present various time-scale mobility. (c) Private car trajectory data gives enough information to illustrate the spatiotemporal development of the urban aggregation effect. (d) Vehicle emissions denoted by different colors. Widths of lines between urban regions show various patterns over time.

urban hot regions by exploring private car users' **arrive-stay-leave** (**ASL**) behaviors [119]. Wang et al. [179] exploited the real-time interactive urban traffic patterns based on massive taxi GPS trajectories and discovered various time-scale traffic patterns, which are shown in Figure 4(b).

Aggregation patterns stem from the fact that vehicles driving in cities are frequently concentrated in certain regions such as residential, workplaces, and hotspots, producing an aggregation effect, displaying the spatiotemporal pattern of vehicle collective movement in various urban regions. Wang et al. [175] first utilized private car trajectory data to study the spatio-temporal evolution of aggregation patterns in urban environments. On this basis, Xiao et al. [200, 201] clarified the distinction between the weekend and weekday aggregation patterns, and they continued to investigate the weekend private vehicle aggregation impact by using a deep learning technique. Figure 4(c) illustrates the shifting strength of the aggregation effect in Shenzhen City based on private car trajectory data, with deep red and deep blue indicating the aggregation core and boundary of the aggregation region, respectively. Following the First Law of Geography, the deep red eventually fades into deep blue. The above works [175, 200, 201] demonstrate that compared with the trajectory data of taxi and public transportation, the private car trajectory data is more suitable to reflect the spatio-temporal aggregation pattern during the evolution of urban traffic since it exhibits stay behavior (recall Section 4.1.2). In this line, large numbers of vehicles, the majority of which being private cars, drive to and stay for certain periods in FVPs (see in Figure 4(c)), and their time-varying aggregation patterns are tightly related to the formation and disappearance of urban hot zones. The authors in [217] studied the spatio-temporal distribution of the hotspots based on the taxi trajectory dataset. Yu et al. [217] studied the hotspot detection problem using an enhanced quality threshold clustering method based on a neighborhood association, which could generate effective and reasonable urban hotspots from taxi trajectory data and offer helpful information to traffic control systems.

Emission pattern discovery concentrates on exploring the emission periodicity for low-carbon urban planning. Cheng et al. [25] tried to estimate the heavy-duty diesel truck emission inventory and examine the peculiarities of their spatio-temporal evolution based on truck trajectories. They also did the hotspot and local-outlier analysis to determine the spatio-temporal shifting trend of pollutant emission intensity and cluster and outlier patterns. Yu et al. [221] evaluated and studied regional traffic emission characteristics based on the trajectory gathered by automatic license plate recognition detectors. Li et al. [105] established traffic analysis zones and used them as an analytical unit to investigate Beijing's spatial and temporal dynamic emission patterns using taxi GPS data. The representative results are shown in Figure 4(d). They discovered that traffic analysis zones featuring commercial regions, entertainment venues, and transit hubs have more significant emissions than other zones. Additionally, they demonstrated that the three types of taxi emission patterns, corresponding to the hours of 0:00–3:00, 3:00–6:00, and 6:00–24:00, could be distinguished.

4.2.2 Traffic Flow Transitions. Many scholars investigate traffic flow transitions based on the collective vehicle trajectory dataset, including traffic flow prediction, O-D flow analysis, and travel demand estimation.

Traffic flow prediction is a research highlight in mining vehicle trajectory datasets since it helps a variety of transportation operations and urban applications [83]. Zhan et al. [231] presented a hybrid system to predict citywide traffic flow based on taxi trajectory that combines machine learning methods and traffic flow theory. Tang et al. [168] focused on real-time traffic flow inference based on dense taxi trajectories and incomplete trajectories captured by camera surveillance systems. Xing et al. [211] attempted to estimate the traffic flows using taxi trajectories and small-scale license plate recognition data based on ensemble support vector regression. Using probability theory, Zhao et al. [245] calculated the traffic flow distribution based on the trajectory data from ride-sharing cars. Moreover, Liu et al. [119] studied the private car flows in irregular regions and proposed a multigraph-dense convolutional network to predict private car flows by exploiting spatio-temporal semantic information.

Origin-Destination (O-D) flow analysis is an essential technique to reveal traffic dynamics, which concentrates on the types and intensity of traffic in the city change along with the time [82]. Riascos et al. [149] studied the vehicle flow transition probabilities based on taxi trips. They generated origin-destination matrices that described the overall activity of the taxi flow, identified high-demand zones, and then examined the likelihood of transition between high-demand zones. Zhang et al. [234] investigated the challenge of predicting short-term origin-destination flow based on train trajectory and suggested a channel-wise attentive split convolutional neural network model to estimate the origin-destination flow. Liu et al. [118] investigated private car transitions between urban regions based on multi-source data and proposed multiple graph-based generative adversarial networks to predict private car transitions.

Travel demand estimation can help the government and businesses make better management and operational choices by delivering data-driven insights [30, 37, 216]. Yao et al. [216] argued that travel demand is influenced by the spatial, temporal, and semantic views, and then they leveraged CNN, LSTM, and graph embedding to model correlations in the three views. Du et al. [37] hold the same opinion that traffic demand is affected by various factors and proposed a dynamic transition CNN approach to estimate travel demand. Chu et al. [30] mainly analyzed travel demands' complex spatial distributions and temporal dynamics and proposed a multi-scale convolutional long short-term memory network to model and forecast the travel demand.

5 Vehicle Trajectory-boosted Applications

Building upon the insights gained from vehicle trajectory analytics in Section 4, this Section explores the applications that are enhanced by vehicle trajectory. We summarize various applications where trajectory data plays a critical role, providing a deep understanding of how to mine and utilize knowledge from vehicle trajectories. These applications range from traffic prediction and route recommendation to anomaly detection [43, 93, 243].

5.1 Traffic Prediction

Traffic-related prediction tasks mainly include predicting flow [135], demand [37], speed [59], and condition [109]. With vehicle trajectory-driven insights, precise traffic prediction could aid in improved management and decision-making on the part of the government [143, 182, 183, 185, 186]. Existing related works usually focus on traffic prediction in three entities: stations, road segments, and regions [113, 140].

The station level is based on road sensors, intersections, or actual/virtual stations for traffic prediction. In [58], the authors proposed an online learning collaborative method to predict traffic flow in road intersections. Du et al. [37] discovered virtual stations in cities by density-peak clustering. They proposed a dynamic transition convolutional neural network to forecast traffic demand in virtual stations. Gong et al. [55] used the online latent space technique for the metro systems to handle the difficulty of predicting the distribution of population flows over the whole network. Zhang et al. [234] presented a channel-wise attentive split–convolutional network for forecasting short-term origin-destination flows in the metro system.

Urban road segments have regular traffic congestion near the city center during rush hours, making it critical to estimate traffic flow on road segments [114]. Anwar et al. [4] proposed a complete framework for capturing flow evolution by progressively updating partitions in an effective two-layer technique. In [153], using partially observed traffic data, the authors created a reliable network-wide traffic state imputation framework. Mallah et al. [135] used a deep neural network trained to anticipate multi-tasks with data from connected vehicles to predict short-term traffic flow on a particular road section. Miao et al. [139] proposed the first unified replay-based continuous learning framework for spatio-temporal prediction on streadming data, such as traffic flow.

Traffic flow-based vehicle trajectory can be modeled as tensors. Some authors divided the spatial range into grids and sliced the temporal sections equally to form tensors of next time intervals [57]. Each grid has a number that represents the number of trips to or from that zone. For example, Guo et al. [57] divided a city into a grid map and represented crowd flows data in each grid. Graph-based modeling can effectively handle non-Euclidean data. In this regard, Geng et al. [54] encoded non-Euclidean correlations between zones into multiple graphs, including neighborhood, functional similarity, and transportation connectivity graphs. They proposed a spatio-temporal convolution network based on multiple graphs for traffic demand prediction.

5.2 Route Recommendation

Route recommendations are designed to recommend cruise routes to taxis or private cars so they can rapidly discover and pick up passengers or find the best route to destinations. Most of the existing work on route recommendation focuses on taxis in cities [18, 56, 79, 154, 172].

Tu et al. [172] developed a system for electric taxi drivers that suggests routes that consider both cruising and station recharging. In their work, taxi trip information, such as the likelihood of picking up passengers and the distribution of destinations, is learned from the GPS trajectories. Guo et al. [56] employed a force-directed strategy to address the issue of ride-on-demand services' requesting route recommendations. Ji et al. [79] proposed a deep reinforcement learning approach to make dynamic route recommendations for available taxis. They were the first to consider real-time internal and exterior spatio-temporal characteristics to recommend a dynamic taxi route. Qu et al. [154] put out a strategy for recommending taxi routes termed adaptive shortest anticipated cruising route. To determine the probable cruising distance of taxis, they consider the load distribution between passengers and taxis and introduce the shortest anticipated cruising distance. Yuen et al. [229] developed a route recommendation system to forecast the path that has the best likelihood of discovering appropriate consumers while remaining within the permitted detour distances. Chen et al. [18] proposed a parallel split-and-combine method to allow taxis to find routes based on their locations. The resultant capability is aimed toward several applications, such as ridesharing, location-based services, and route planning and suggestion.

5.3 Trajectory-User Linking

Trajectory-User Linking (**TUL**) is a newly-introduced mining task based on trajectory data[41, 239], which aims at linking unknown trajectories to users who generate them, enabling broad applications ranging from personalized location-based recommendation [53] to potential criminal identification [44, 87]. TUL is introduced in [53], which correlates unlabeled trajectories to their potential users and gradually steps into the hot topic of spatio-temporal data mining. TULER [53] is proposed by utilizing RNN-based models to learn the trajectory sequence for capturing the dependencies and linking them to users. TULVEA [250] incorporates variational autoencoder into the TUL task, leveraging a semi-supervised framework to learn the hierarchical semantics of sequence trajectories. However, the data sparsity issue remains in dense trajectories, and most works usually ignore it. In the follow-up, DeepTUL [136] uses the attentive recurrent network to learn the multi-periodic properties of human mobility for more accurate user-trajectories matching. Still, the RNN-based model used in DeepTUL has limited performance in capturing long-term dependencies.

A few works focus on identifying users from different mobility datasets. Feng et al. [46] then developed a co-attention mechanism and a multi-modal embedding network to address the low-quality issue with mobility data. Dealing with the massive number of users, TULSN [219] proposed a Siamese network to capture semantic information in the trajectory and only requires small-scale labeled trajectory data to complete training.

5.4 Anomaly Detection and Prediction

Urban anomalies, if not properly addressed, can lead to significant risks, including loss of life and property [92]. To help prevent such adverse outcomes, a growing number of data-driven frameworks now leverage big data and machine learning techniques to automatically detect and predict these anomalies [237]. Accurate anomaly detection is critical for governments and local communities, as it underpins the development of smart city applications–from intelligent transportation systems to public safety management [73].

Many studies focus on vehicle trajectories [236]; for instance, [112] frames anomaly detection as a hypothesis-testing problem and introduces a novel fault localization indicator using specialized mathematical methods. Beyond vehicle trajectories, Lam et al. [97] devised a clustering-based detection strategy that harnesses publicly available bike-sharing data to identify spatio-temporal events deviating from normal daily patterns. When a cluster is determined to be out of the ordinary, an anomalous event is flagged for that time and location. Alfeo et al. [2] focused on densely populated urban regions, examining their behavior over time to uncover anomalies. They evaluated the method's effectiveness in handling smaller anomalies by assessing the relationship between an anomaly index and observed urban incidents. Chiang et al. [27] introduced a technique that identifies congestion based on bus trajectories, spotlighting anomalous traffic health conditions and the structural traits of congestion cascades. Meanwhile, Zhu et al. [258] presented a method integrating location-specific time series decomposition and outlier detection for uncovering urban events. These events are defined as anomalies that substantially diverge from the forecasts derived from established trend and periodicity patterns.

In summary, researchers have developed a wide spectrum of approaches for identifying and understanding urban anomalies. Such efforts are pivotal for building resilient and responsive smart city systems, guiding policymakers and communities in proactively managing urban challenges.

5.5 Urban Planning

In recent years, urban planning applications based on vehicle trajectories have mainly focused on the regional level, such as urban functional region identification and region attractiveness discovery. Different functional regions in a city, such as residential, business, and educational regions, are fostered by urbanization and contemporary civilization [225]. The urban functional region reflects the city's spatial structures, which is critical to urban planning [253]. In recent years, researchers have combined vehicle trajectories and other multi-source heterogeneous data to identify urban functional regions. Zhang et al. [235] focused on learning an embedding space from urban data for urban regions, including taxi trajectory, street block, and POI. They presented a multi-view joint learning technique to learn complete and representative urban area embeddings. Insightful information on the architecture and dynamics of cities can be gained from the quantitative representations of urban regions that aid in better examining the linkages between urban features. Fu et al. [50] defined urban form as a synthesis of urban functions and related community portfolios. They suggested a collective learning strategy to model individual-level heterogeneous human mobility data to identify and quantify the urban forms of residential communities. Liu et al. [123] proposed a bi-clustering method to partition the bike stations into urban functional regions. For the construction of the bike system, it is intended to group bike stations with comparable POI features and near geographic distances.

Numerous people go by vehicle to designated functional regions, where they then stop and stay for a while, causing the attractiveness of such functional regions to change over time [200]. Utilizing latent activity taxi trajectories, Yuan et al. [225] identified urban functional regions that served as a calibration for urban development. In [200], the authors uncovered the appeal of metropolitan regions by analyzing private automobile trajectory statistics. This is the first study to use private car trajectory data to examine urban region attractiveness, offering a fresh viewpoint on understanding the evolution of urban mobility. Besides, exploring the urban aggregation effect contributes to urban planning. Xiao et al. [201] developed a spatio-temporal attention network to understand the dynamic aggregation effect of private cars. Wang et al. [175] investigated the stay behavior and urban aggregation effect based on large-scale private car trajectory data. Those works on regional aggregation discovery provide essential information for understanding urban dynamics and facilitate various applications in urban planning.

6 Open Issues and Future Directions

Despite significant advancements in vehicle trajectory collection, processing, analytics, and applications, several open issues and future directions remain to be addressed. These include challenges related to privacy preservation, data sparsity, multi-source heterogeneous data fusion, and uncertainty in trajectory computing. Additionally, emerging areas such as vehicle carbon footprint estimation, urban socioeconomics, intercity mobility mining, vehicle trajectory with digital twin technology and large language models present promising avenues for future research. This section outlines these open issues and explores potential directions for advancing the field.

6.1 Open Issues

Vehicle trajectory analysis while offering remarkable insights and practical applications still faces several outstanding challenges that limit its effectiveness and scalability. These challenges encompass data privacy, sparsity, multi-source heterogeneous data fusion, and uncertainty in trajectory computing. This subsection delves into these pressing issues, highlighting their impact on current methodologies and identifying research gaps that would be addressed.

Vehicle Trajectory Privacy Preservation. Researchers have turned their attention in recent years from protecting location privacy in the trajectory to protecting vehicle owners' trajectory privacy [80, 122]. In this line, the emerging **Federated Learning** (**FL**) is proposed to solve data security exchange and privacy protection in distributed environments, and several studies have applied federal learning to vehicle trajectory privacy preservation [141, 249]. Federated learning requires that the training data owned by different consortium parties must share the same feature space, which limits the practicability of FL [60]. This problem is addressed by federated transfer learning, which instead of using differential privacy, applies homomorphic encryption and polynomial approximation to provide a safer and more dependable solution for vehicle trajectory [108]. Federated transfer learning participants may have their own feature space based on the properties of transfer learning, making it appropriate for more trajectory data privacy protection varieties. However, how to carefully design federated transfer learning methods for trajectory privacy preservation is an open issue.

Sparsity in Vehicle Trajectory. According to the sparsity origins of the vehicle trajectory, the authors in [255] categorize the sparsity into two scenarios in spatio-temporal data, intrinsic sparsity, and fake sparsity. Vehicle trajectory is spontaneously dense, which is caused by the sparse dispersion of sensor equipment. Talking about vehicle sparse trajectory data usually means fake-sparsity vehicle trajectory. Existing studies address sparsity problem by applying generative adversarial networks [118], deep factorization machines [254], and natural language processing [162] methods. So far, sparsity remains a challenge in vehicle trajectory due to the uncertainty of information in the process of trajectory acquisition.

Multi-source Heterogeneous Data Fusion. Multi-source heterogeneous data fusion combines many dataset types into a standardized format, improving performance for applications that use vehicle trajectory [187]. Combining many sources of heterogeneous data is a workable solution to the problem of limited sensor coverage and the lack of precise information for validation. There are mainly three types of multi-source heterogeneous data in the cities, including spatio-temporal static data (AOI, POI and road network, etc.), spatial static time dynamic data (weather, events, holidays, etc.) and spatio-temporal dynamic data (vehicle trajectory) [248]. However, there is no unified architecture for integrating the above multi-source heterogeneous data in real-time and efficiently. Such a common architecture is needed to serve a wide variety of applications.

Uncertainty in Trajectory Computing. Uncertainty quantification is a vital problem that has been brought up by the high dynamics and diverse interactions in trajectory computing [256]. Research interest in assessing prediction uncertainty has increased recently [38]. However, few works have made efforts to quantify the spatio-temporal uncertainty for vehicle trajectory computing. Epistemic uncertainty and aleatoric uncertainty are two types of uncertainty. Epistemic uncertainty results from a lack of understanding of the training data. Inherent unpredictability in data observations is captured by aleatoric uncertainty. It would be interesting to investigate how to quantify the above two categories of uncertainties in different vehicle trajectory-based applications for a better decision.

6.2 Future Directions

Looking ahead, vehicle trajectory research holds immense promise in advancing intelligent transportation systems and informing urban policy. Recent technological strides, ranging from vehicle carbon footprint estimation to emerging paradigms such as digital twins and large language models, offer exciting opportunities to push the boundaries of what can be achieved with trajectory data. This subsection explores these prospective directions, illuminating how novel techniques and interdisciplinary approaches could catalyze the next wave of breakthroughs in vehicle trajectory analytics and applications.

Vehicle Carbon Footprint Estimation. With the proposal of the "Dual Carbon", i.e., Peak Carbon Dioxide Emissions and Carbon Neutrality [117, 242], increasing attention has been given to carbon footprint and carbon neutrality of road traffic based on urban vehicle trajectory [204]. In this line, the followings provide two main open issues that can be regarded as future directions.

(1) Carbon footprint calculation. The reasonable measurement method is the premise of slowing down road traffic's carbon footprint rate effectively. The carbon footprint of road traffic is calculated based on urban vehicle trajectory. Compared with the traditional top-down measurement method, the heterogeneity and spatial and temporal differences of urban vehicle carbon emissions can be clearly and standardized to control carbon emission reduction in the key link. However, due to the limitation of technical background, fuel type and other parameters of vehicle trajectory cannot be obtained, which poses a new challenge to carbon footprint measurement. In addition, the electrification of public transport has made road traffic carbon emission-reduction focus on private cars, which are still dominated by traditional fuels. It has become an essential goal of low-carbon emission reduction to develop urban vehicle emission models, compile vehicle emission inventory of urban road traffic, and measure individual traffic carbon footprint based on urban private car trajectory data.
(2) Individual carbon trading. As carbon emission reduction at the individual level, personal carbon trading is significant to achieving carbon neutrality efficiently. Urban vehicle trajectory can be combined with blockchain, cloud computing, artificial intelligence, and other high-tech network information technologies to build a personal carbon trading market and achieve "bottom-up" carbon reduction and carbon control. However, establishing an individual carbon trading market depends on real-time trajectory data acquisition.

Vehicle Trajectory-driven Urban Socioeconomics. Vehicle trajectory can provide fine-grained information for urban socioeconomic in a multi-level structure. Urban socioeconomics has a series of indicators, such as neighborhood-level indicators (economic activeness, resident consumption, etc) and street-level indicators (POIs, commercial activeness, etc) [106, 196]. In the neighborhood region-level urban structure, the trajectories generated by vehicles moving between the neighborhood regions of different categories can reflect the dynamic correlations between these regions. In the street-level urban structure, combing POI indicators and vehicle trajectory provides support for mining semantic interactions between streets.

Intercity Mobility Mining from Vehicle Trajectory. The majority of the current research on urban mobility based on trajectory data is focused on analyzing urban mobility patterns [207]. Human intercity or intercity travel behavior has received very little consideration. In addition to moving within a city, individuals often travel between cities in their vehicles [209]. This clearly illustrates how urban social functional regions are evolving. Due to the accelerated economic expansion and urbanization, neighboring cities now perform a variety of tasks within big, multi-city economic areas. Since individuals must travel between these nearby cities to work in various functional regions, this logically improves intercity mobility. Analyzing intercity mobility using vehicle trajectory offers a chance to investigate intercity traffic and its connection to the growth of new urban agglomerations.

Vehicle Trajectory Enabled Digital Twin for ITS. The Digital Twin, as an emerging technology, has attracted the interest of the traffic community, and it is considered one of the most effective solutions for **intelligent transportation systems** (**ITSs**). The vehicle trajectory datasets are considered the cornerstone of the digital twin for the intelligent transportation system. How to utilize vehicle trajectory datasets to build the digital twin of complex traffic scenes, and achieve the perfect integration and accurate mutual feedback between physical space and digital space, is a frontier scientific problem that needs to be solved urgently. Some studies have been conducted to investigate how to build a virtual digital space mapping the real traffic system [70, 188]. However, it is still challenging to twin the real-world traffic system based on the vehicle trajectory datasets since it is difficult to express the dynamic traffic elements and the operation, evolution, and interaction laws between them only through models.

Large Language Model-empowered Vehicle Trajectory Analysis. Large Language Models (**LLMs**) have shown tremendous potential in various domains, including natural language processing [155, 214], computer vision [94, 222], time series analytics [121, 213], and urban planing [86]. In the context of vehicle trajectory analysis, LLMs could offer new ways to understand complex trajectory data [177]. By leveraging the capabilities of LLMs, researchers could enhance the analysis of trajectory data, enabling more accurate predictions, anomaly detection, and improved decision-making [31]. LLMs assist in extracting semantic information from trajectory data and identifying patterns and trends that may not be evident through traditional methods. Furthermore, they facilitate the integration of multi-modal data [120], helping to create comprehensive vehicle travel behavior analysis models. As LLMs continue to evolve, their application in vehicle trajectory analysis holds the promise of unlocking new insights and driving innovations in transportation research and urban planning. However, effectively incorporating LLMs into trajectory analysis frameworks presents challenges, such as managing computational resources and maintaining the interpretability of the models. Addressing these challenges will be crucial for realizing the full potential of LLMs in this field.

7 Conclusion

This study provides a comprehensive overview of vehicle trajectory collection, processing, analytics, and applications. Specifically, we reviewed various data collection technologies and offered an overview of different vehicle trajectory datasets to give readers a thorough background. Next, we elaborated on vehicle trajectory processing techniques by introducing their representative methods and analyzing their pros and cons. We then categorized vehicle trajectory analysis technologies into individual travel behavior and collective mobility analytics based on the scope of analysis objectives. Furthermore, we systematically analyzed existing vehicle trajectory analysis technologies to help researchers quickly understand the current landscape. Following this, we summarized vehicle trajectory-boosted applications to offer a deeper understanding of how to mine and utilize knowledge from vehicle trajectory. Finally, we discussed the remaining challenges and future research directions related to vehicle trajectory.

References

[1] Md Mahbub Alam, Luis Torgo, and Albert Bifet. 2021. A survey on spatio-temporal data analytics systems. *ACM Computing Surveys* 54, 10 (2021), 1–38.

[2] Antonio Luca Alfeo, Mario G. C. A. Cimino, Sara Egidi, Bruno Lepri, and Gigliola Vaglini. 2018. A stigmergy-based analysis of city hotspots to discover trends and anomalies in urban transportation usage. *IEEE Transactions on Intelligent Transportation Systems* 19, 7 (2018), 2258–2267.

[3] Tariq Alsahfi, Mousa Almotairi, and Ramez Elmasri. 2020. A survey on trajectory data warehouse. *Spatial Information Research* 28, 1 (2020), 53–66.

[4] Tarique Anwar, Chengfei Liu, Hai Le Vu, Md. Saiful Islam, and Timos Sellis. 2018. Capturing the spatiotemporal evolution in road traffic networks. *IEEE Transactions on Knowledge and Data Engineering* 30, 8 (2018), 1426–1439.

[5] Shunsuke Aoki, Kaoru Sezaki, Nicholas Jing Yuan, and Xing Xie. 2021. BusBeat: Early event detection with real-time bus GPS trajectories. *IEEE Transactions on Big Data* 7, 2 (2021), 371–382.

[6] Gowtham Atluri, Anuj Karpatne, and Vipin Kumar. 2018. Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys* 51, 4 (2018), 1–41.

[7] Paul Baumann, Wilhelm Kleiminger, and Silvia Santini. 2013. How long are you staying?: Predicting residence time from human mobility traces. In *Proceedings of the MobiCom*. 231–234.

[8] Benarous, Eran Toch, and Irad Ben-Gal. 2022. Synthesis of longitudinal human location sequences: Balancing utility and privacy. *ACM Transactions on Knowledge Discovery from Data* 16, 6 (2022), 1–27.

[9] Jiawei Cai, Dong Wang, Hongyang Chen, Chenxi Liu, and Zhu Xiao. 2024. Modeling dynamic spatiotemporal user preference for location prediction: A mutually enhanced method. *World Wide Web* 27, 2 (2024), 1–17.

[10] Zekun Cai and Aiping Xiong. 2023. Understand users' privacy perception and decision of V2X communication in connected autonomous vehicles. In *Proceedings of the 32nd USENIX Security Symposium*. 2975–2992.

[11] Chu Cao and Mo Li. 2021. Generating mobility trajectories with retained data utility. In *Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2610–2620.

[12] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, et al. 2019. Argoverse: 3D tracking and forecasting with rich maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8748–8757.

[13] Ximing Chang, Jianjun Wu, Zhengbing He, Daqing Li, Huijun Sun, and Weiping Wang. 2020. Understanding user's travel behavior and city region functions from station-free shared bike usage data. *Transportation Research Part F: Traffic Psychology and Behaviour* 72 (2020), 81–95.

[14] Chao Chen, Yan Ding, Xuefeng Xie, Shu Zhang, Zhu Wang, and Liang Feng. 2020. TrajCompressor: An online map-matching-based trajectory compression framework leveraging vehicle heading direction and change. *IEEE Transactions on Intelligent Transportation Systems* 21, 5 (2020), 2012–2028.

[15] Huiling Chen, Dong Wang, and Chenxi Liu. 2020. Towards semantic travel behavior prediction for private car users. In *2020 IEEE 22nd International Conference on High Performance Computing and Communications; IEEE 18th International Conference on Smart City; IEEE 6th International Conference on Data Science and Systems*. 950–957.

[16] Jie Chen, Zhu Xiao, Dong Wang, Daiwu Chen, Vincent Havyarimana, Jing Bai, and Hongyang Chen. 2018. Toward opportunistic compression and transmission for private car trajectory data collection. *IEEE Sensors Journal* 19, 5 (2018), 1925–1935.

[17] Jie Chen, Zhu Xiao, Dong Wang, Wangchen Long, and Vincent Havyarimana. 2019. Stay of interest: A dynamic spatiotemporal stay behavior perception method for private car users. In *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems*. 1526–1532.

[18] Lisi Chen, Shuo Shang, Christian S. Jensen, Bin Yao, Zhiwei Zhang, and Ling Shao. 2019. Effective and efficient reuse of past travel behavior for route recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 488–498.

[19] Li Chen, Linjiang Zheng, Dong Xia, Xiaolin Cai, Dihua Sun, and Weining Liu. 2022. Recognizing and analyzing private car commuters using big data of electronic registration identification of vehicles. *IEEE Transactions on Intelligent Transportation Systems* 23, 9 (2022), 15629–15643.

[20] Xinyu Chen, Zhaocheng He, and Lijun Sun. 2019. A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation. *Transportation Research Part C: Emerging Technologies* 98 (2019), 73–84.

[21] Xinyu Chen, Zhaocheng He, and Jiawei Wang. 2018. Spatial-temporal traffic speed patterns discovery and incomplete data recovery via SVD-combined tensor decomposition. *Transportation Research Part C: Emerging Technologies* 86 (2018), 59–77.

[22] Yile Chen, Gao Cong, and Cuauhtemoc Anda. 2023. TERI: An effective framework for trajectory recovery with irregular time intervals. *PVLDB* 17, 3 (2023), 414–426.

[23] Yile Chen, Xiucheng Li, Gao Cong, Zhifeng Bao, Cheng Long, Yiding Liu, Arun Kumar Chandran, and Richard Ellison. 2021. Robust road network representation learning: When traffic patterns meet traveling semantics. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*. 211–220.

[24] Yuqi Chen, Hanyuan Zhang, Weiwei Sun, and Baihua Zheng. 2023. Rntrajrec: Road network enhanced trajectory recovery with spatial-temporal transformer. In *Proceedings of the 2023 IEEE 39th International Conference on Data Engineering*. 829–842.

[25] Shifen Cheng, Feng Lu, and Peng Peng. 2020. A high-resolution emissions inventory and its spatiotemporal pattern variations for heavy-duty diesel trucks in Beijing, China. *Journal of Cleaner Production* 250 (2020), 119445.

[26] Yi cheng Zheng, Yi Zhang, and Li Li. 2016. Reliable path planning for bus networks considering travel time uncertainty. *IEEE Intelligent Transportation Systems Magazine* 8, 1 (2016), 35–50.

[27] Meng-Fen Chiang, Ee-Peng Lim, Wang-Chien Lee, and Agus Trisnajaya Kwee. 2017. BTCI: A new framework for identifying congestion cascades using bus trajectory data. In *Proceedings of the IEEE BigData*. 1133–1142.

[28] Seongjin Choi, Jiwon Kim, and Hwasoo Yeo. 2021. Trajgail: Generating urban vehicle trajectories using generative adversarial imitation learning. *Transportation Research Part C: Emerging Technologies* 128 (2021), 103091.

[29] Chen Chu, Hengcai Zhang, Peixiao Wang, and Feng Lu. 2024. Simulating human mobility with a trajectory generation framework based on diffusion model. *International Journal of Geographical Information Science* 38, 5 (2024), 847–878.

[30] Kai-Fung Chu, Albert Y. S. Lam, and Victor OK Li. 2019. Deep multi-scale convolutional LSTM network for travel demand and origin-destination predictions. *IEEE Transactions on Intelligent Transportation Systems* 21, 8 (2019), 3219–3232.

[31] Can Cui, Yunsheng Ma, Xu Cao, Wenqian Ye, and Ziran Wang. 2024. Receive, reason, and react: Drive as you say, with large language models in autonomous vehicles. *IEEE Intelligent Transportation Systems Magazine* 16, 4 (2024), 81–94.

[32] Jingze Cui, Xian Zhou, Yanmin Zhu, and Yanyan Shen. 2018. A road-aware neural network for multi-step vehicle trajectory prediction. In *Database Systems for Advanced Applications: 23rd International Conference, DASFAA 2018, Gold Coast, QLD, Australia, May 21-24, 2018, Proceedings, Part I 23*. 701–716.

[33] Jiewen Deng, Jinliang Deng, Renhe Jiang, and Xuan Song. 2023. Learning gaussian mixture representations for tensor time series forecasting. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*. 2077–2085.

[34] Jiewen Deng, Jinliang Deng, Du Yin, Renhe Jiang, and Xuan Song. 2024. TTS-Norm: Forecasting tensor time series via multi-way normalization. *ACM Transactions on Knowledge Discovery from Data* 18, 1 (2024), 3:1–3:25.

[35] Liwei Deng, Yan Zhao, Hao Sun, Changjie Yang, Jiandong Xie, and Kai Zheng. 2023. Fusing local and global mobility patterns for trajectory recovery. In *Proceedings of the International Conference on Database Systems for Advanced Applications*. 448–463.

[36] Zhijun Ding, Xiaolun Li, Changjun Jiang, and Mengchu Zhou. 2018. Objectives and state-of-the-art of location-based social network recommender systems. *ACM Computing Surveys* 51, 1 (2018), 1–28.

[37] Bowen Du, Xiao Hu, Leilei Sun, Junming Liu, Yanan Qiao, and Weifeng Lv. 2021. Traffic demand prediction based on dynamic transition convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems* 22, 2 (2021), 1237–1247.

[38] Chao Du, Zhifeng Gao, Shuo Yuan, Lining Gao, Ziyan Li, Yifan Zeng, Xiaoqiang Zhu, Jian Xu, Kun Gai, and Kuang-Chih Lee. 2021. Exploration in online advertising systems with deep uncertainty-aware learning. In *Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2792–2801.

[39] Junchuan Fan, Cheng Fu, Kathleen Stewart, and Lei Zhang. 2019. Using big GPS trajectory data analytics for vehicle miles traveled estimation. *Transportation Research Part C: Emerging Technologies* 103 (2019), 298–307.

[40] Li Fan, Xia Jizhe, Huang Zhao, Li Xiaoming, and Li Qingquan. 2020. Predicting personal next location based on stay point feature extraction. *Geomatics and Information Science of Wuhan University* 45, 12 (2020), 1970–1980.

[41] Zipei Fan, Quanjun Chen, Renhe Jiang, Ryosuke Shibasaki, Xuan Song, and Kota Tsubouchi. 2019. Deep multiple instance learning for human trajectory identification. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 512–515.

[42] Zipei Fan, Xiaojie Yang, Wei Yuan, Renhe Jiang, Quanjun Chen, Xuan Song, and Ryosuke Shibasaki. 2022. Online trajectory prediction for metropolitan scale mobility digital twin. In *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*. 1–12.

[43] Yuchen Fang, Jiandong Xie, Yan Zhao, Lu Chen, Yunjun Gao, and Kai Zheng. 2024. Temporal-frequency masked autoencoders for time series anomaly detection. In *Proceedings of the 2024 IEEE 40th International Conference on Data Engineering*. 1228–1241.

[44] Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, and Depeng Jin. 2018. DeepMove: Predicting human mobility with attentional recurrent networks. In *Proceedings of the 2018 World Wide Web Conference*. 1459–1468.

[45] Jie Feng, Zeyu Yang, Fengli Xu, Haisu Yu, Mudan Wang, and Yong Li. 2020. Learning to simulate human mobility. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 3426–3433.

[46] Jie Feng, Mingyang Zhang, Huandong Wang, Zeyu Yang, Chao Zhang, Yong Li, and Depeng Jin. 2019. DPLink: User identity linkage via deep neural network from heterogeneous mobility data. In *Proceedings of the World Wide Web*. 459–469.

[47] Xinxin Feng, Haitao Zhang, Can Wang, and Haifeng Zheng. 2022. Traffic data recovery from corrupted and incomplete observations via spatial-temporal TRPCA. *IEEE Transactions on Intelligent Transportation Systems* 23, 10 (2022), 17835–17848.

[48] Nicolette Formosa, Mohammed Quddus, Mohit Kumar Singh, Cheuk Ki Man, Craig Morton, and Cansu Bahar Masera. 2023. An experiment and simulation study on developing algorithms for CAVs to navigate through roadworks. *IEEE Intelligent Transportation Systems Magazine* 25, 1 (2023), 120–132.

[49] Kun Fu, Jieping Ye, and Zheng Wang. 2020. CompactETA: A fast inference system for travel time prediction. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 3337–3345.

[50] Yanjie Fu, Guannan Liu, Yong Ge, Pengyang Wang, Hengshu Zhu, Chunxiao Li, and Hui Xiong. 2019. Representing urban forms: A collective learning model with heterogeneous human mobility data. *IEEE Transactions on Knowledge and Data Engineering* 31, 3 (2019), 535–548.

[51] Riccardo Gallotti, Armando Bazzani, Sandro Rambaldi, and Marc Barthelemy. 2016. A stochastic model of randomly accelerated walkers for human mobility. *Nature Communications* 7, 1 (2016), 1–7.

[52] Jiajing Gao and Lu Zhen. 2024. Intelligent distribution transportation route planning considering traffic congestion. *IEEE Intelligent Transportation Systems Magazine* 16, 3 (2024), 96–110.

[53] Qiang Gao, Fan Zhou, Kunpeng Zhang, Goce Trajcevski, Xucheng Luo, and Fengli Zhang. 2017. Identifying human mobility via trajectory embeddings. In *Proceedings of the IJCAI*. 1689–1695.

[54] Xu Geng, Yaguang Li, Leye Wang, Lingyu Zhang, Qiang Yang, Jieping Ye, and Yan Liu. 2019. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 3656–3663.

[55] Yongshun Gong, Zhibin Li, Jian Zhang, Wei Liu, and Yu Zheng. 2022. Online spatio-temporal crowd flow distribution prediction for complex metro system. *IEEE Transactions on Knowledge and Data Engineering* 34, 2 (2022), 865–880.

[56] Suiming Guo, Chao Chen, Jingyuan Wang, Yan Ding, Yaxiao Liu, Xu Ke, Zhiwen Yu, and Daqing Zhang. 2020. A force-directed approach to seeking route recommendation in ride-on-demand service using multi-source urban data. *IEEE Transactions on Mobile Computing* 21, 6 (2020), 1909–1926.

[57] Shengnan Guo, Youfang Lin, Shijie Li, Zhaoming Chen, and Huaiyu Wan. 2019. Deep spatial-temporal 3D convolutional neural networks for traffic data forecasting. *IEEE Transactions on Intelligent Transportation Systems* 20, 10 (2019), 3913–3926.

[58] Zhengang Guo, Yingfeng Zhang, Jingxiang Lv, Yang Liu, and Ying Liu. 2021. An online learning collaborative method for traffic forecasting and routing optimization. *IEEE Transactions on Intelligent Transportation Systems* 22, 10 (2021), 6634–6645.

[59] Liangzhe Han, Bowen Du, Leilei Sun, Yanjie Fu, Yisheng Lv, and Hui Xiong. 2021. Dynamic and multi-faceted spatio-temporal deep learning for traffic speed forecasting. In *Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 547–555.

[60] Mu Han, Kai Xu, Shidian Ma, Aoxue Li, and Haobin Jiang. 2022. Federated learning-based trajectory prediction model with privacy preserving for intelligent vehicle. *International Journal of Intelligent Systems* 37, 12 (2022), 10861–10879.

[61] Yunheng Han, Weiwei Sun, and Baihua Zheng. 2017. COMPRESS: A comprehensive framework of trajectory compression in road networks. *ACM Transactions on Database Systems* 42, 2 (2017), 1–49.

[62] Vincent Havyarimana, Zhu Xiao, Thabo Semong, Jing Bai, Hongyang Chen, and Licheng Jiao. 2021. Achieving reliable intervehicle positioning based on redheffer weighted least squares model under multi-GNSS outages. *IEEE Transactions on Cybernetics* 53, 2 (2021), 1039–1050.

[63] Vincent Havyarimana, Zhu Xiao, Alexis Sibomana, Di Wu, and Jing Bai. 2019. A fusion framework based on sparse Gaussian–Wigner prediction for vehicle localization using GDOP of GPS satellites. *IEEE Transactions on Intelligent Transportation Systems* 21, 2 (2019), 680–689.

[64] Peilan He, Guiyuan Jiang, Siew-Kei Lam, and Dehua Tang. 2019. Travel-time prediction of bus journey with multiple bus trips. *IEEE Transactions on Intelligent Transportation Systems* 20, 11 (2019), 4192–4205.

[65] Tianfu He, Jie Bao, Ruiyuan Li, Sijie Ruan, Yanhua Li, Li Song, Hui He, and Yu Zheng. 2020. What is the human mobility in a new city: Transfer mobility knowledge across cities. In *Proceedings of the World Wide Web*. 1355–1365.

[66] Kethmi Hirushini Hettige, Jiahao Ji, Shili Xiang, Cheng Long, Gao Cong, and Jingyuan Wang. 2024. AirPhyNet: Harnessing physics-guided neural networks for air quality prediction. In *Proceedings of the Twelfth International Conference on Learning Representations*. 1–17.

[67] Huiting Hong, Yucheng Lin, Zang Li, Kun Fu, Zheng Wang, Xiaohu Qie, and Jieping Ye. 2020. HetETA: Heterogeneous information network embedding for estimating time of arrival. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2444–2454.

[68] Sahar Hoteit, Stefano Secci, Stanislav Sobolevsky, Carlo Ratti, and Guy Pujolle. 2014. Estimating human trajectories and hotspots through mobile phone data. *Comput. Networks* 64 (2014), 296–307.

[69] John Houston, Guido Zuidhof, Luca Bergamini, Yawei Ye, Long Chen, Ashesh Jain, Sammy Omari, Vladimir Iglovikov, and Peter Ondruska. 2020. One thousand and one hours: Self-driving motion prediction dataset. In *Proceedings of the Conference on Robot Learning*. 409–418.

[70] Chunhua Hu, Weicun Fan, Elan Zeng, Zhi Hang, Fan Wang, Lianyong Qi, and Md Zakirul Alam Bhuiyan. 2022. Digital twin-assisted real-time traffic data prediction method for 5G-Enabled internet of vehicles. *IEEE Transactions on Industrial Informatics* 18, 4 (2022), 2811–2819.

[71] Sheng Hu, Song Gao, Liang Wu, Yongyang Xu, Ziwei Zhang, Haifu Cui, and Xi Gong. 2021. Urban function classification at road segment level using taxi trajectory data: A graph convolutional neural network approach. *Computers, Environment and Urban Systems* 87 (2021), 101619.

[72] Yue Hu, Sijie Ruan, Yuting Ni, Huajun He, Jie Bao, Ruiyuan Li, and Yu Zheng. 2021. SALON: A universal stay point-based location analysis platform. In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*. 407–410.

[73] Chao Huang, Chuxu Zhang, Peng Dai, and Liefeng Bo. 2020. Cross-interaction hierarchical attention networks for urban anomaly prediction. In *Proceedings of the 29th International Conference on International Joint Conferences on Artificial Intelligence*. 4359–4365.

[74] Dou Huang, Xuan Song, Zipei Fan, Renhe Jiang, Ryosuke Shibasaki, Yu Zhang, Haizhong Wang, and Yugo Kato. 2019. A variational autoencoder based generative model of urban human mobility. In *Proceedings of the 2019 IEEE Conference on Multimedia Information Processing and Retrieval*. 425–430.

[75] Yourong Huang, Zhu Xiao, Dong Wang, Hongbo Jiang, and Di Wu. 2020. Exploring individual travel patterns across private car trajectory data. *IEEE Transactions on Intelligent Transportation Systems* 21, 12 (2020), 5036–5050.

[76] Yourong Huang, Zhu Xiao, Xiaoyou Yu, Dong Wang, Vincent Havyarimana, and Jing Bai. 2019. Road network construction with complex intersections based on sparsely sampled private car trajectory data. *ACM Transactions on Knowledge Discovery from Data* 13, 3 (2019), 35:1–35:28.

[77] Christian S. Jensen. 2022. Digitalization in the service of society: The case of big vehicle trajectory data. In *Proceedings of the 34th International Conference on Scientific and Statistical Database Management*. 1–1.

[78] Hwanseok Harrison Jeong, Yiwen Chris Shen, Jaehoon Paul Jeong, and Tae Tom Oh. 2021. A comprehensive survey on vehicular networking for safe and efficient driving in smart transportation: A focus on systems, protocols, and applications. *Vehicular Communications* 31 (2021), 100349–100371.

[79] Shenggong Ji, Zhaoyuan Wang, Tianrui Li, and Yu Zheng. 2020. Spatio-temporal feature fusion for dynamic taxi route recommendation via deep reinforcement learning. *Knowledge-based System* 205 (2020), 106302.

[80] Hongbo Jiang, Jie Li, Ping Zhao, Fanzi Zeng, Zhu Xiao, and Arun Iyengar. 2021. Location privacy-preserving mechanisms in location-based services: A comprehensive survey. *ACM Computing Surveys* 54, 1 (2021), 1–36.

[81] Renhe Jiang, Xuan Song, Dou Huang, Xiaoya Song, Tianqi Xia, Zekun Cai, Zhaonan Wang, Kyoung-Sook Kim, and Ryosuke Shibasaki. 2019. Deepurbanevent: A system for predicting citywide crowd dynamics at big events. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2114–2122.

[82] Renhe Jiang, Zhaonan Wang, Zekun Cai, Chuang Yang, Zipei Fan, Tianqi Xia, Go Matsubara, Hiroto Mizuseki, Xuan Song, and Ryosuke Shibasaki. 2021. Countrywide origin-destination matrix prediction and its application for COVID-19. In *Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 1317, 2021, Proceedings, Part IV 21*. 319–334.

[83] Renhe Jiang, Zhaonan Wang, Jiawei Yong, Puneet Jeph, Quanjun Chen, Yasumasa Kobayashi, Xuan Song, Shintaro Fukushima, and Toyotaro Suzumura. 2023. Spatio-temporal meta-graph learning for traffic forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 8078–8086.

[84] Renhe Jiang, Du Yin, Zhaonan Wang, Yizhuo Wang, Jiewen Deng, Hangchen Liu, Zekun Cai, Jinliang Deng, Xuan Song, and Ryosuke Shibasaki. 2021. DL-Traff: Survey and benchmark of deep learning models for urban traffic prediction. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*. 4515–4525.

[85] Wenjun Jiang, Wayne Xin Zhao, Jingyuan Wang, and Jiawei Jiang. 2023. Continuous trajectory generation based on two-stage GAN. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 4374–4382.

[86] Yue Jiang, Qin Chao, Yile Chen, Xiucheng Li, Shuai Liu, and Gao Cong. 2024. UrbanLLM: Autonomous urban activity planning and management with large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*. 1810–1825.
[87] Guangyin Jin, Chenxi Liu, Zhexu Xi, Hengyu Sha, Yanyun Liu, and Jincai Huang. 2022. Adaptive dual-view WaveNet for urban spatial-temporal event prediction. *Information Sciences* 588 (2022), 315–330.
[88] Guangyin Jin, Zhexu Xi, Hengyu Sha, Yanghe Feng, and Jincai Huang. 2020. Deep multi-view spatiotemporal virtual graph neural network for significant citywide ride-hailing demand prediction. *Neurocomputing* 510 (2022), 79–94.
[89] Guangyin Jin, Huan Yan, Fuxian Li, Yong Li, and Jincai Huang. 2021. Hierarchical neural architecture search for travel time estimation. In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*. 91–94.
[90] Mankirat Kaur, Sarbjeet Singh, and Naveen Aggarwal. 2022. Missing traffic data imputation using a dual-stage error-corrected boosting regressor with uncertainty estimation. *Information Sciences* 586 (2022), 344–373.
[91] Jintao Ke, Hai Yang, and Zhengfei Zheng. 2020. On ride-pooling and traffic congestion. *Transportation Research Part B: Methodological* 142 (2020), 213–231.
[92] Tung Kieu, Bin Yang, Chenjuan Guo, Razvan-Gabriel Cirstea, Yan Zhao, Yale Song, and Christian S. Jensen. 2022. Anomaly detection in time series with robust variational quasi-recurrent autoencoders. In *Proceedings of the 2022 IEEE 38th International Conference on Data Engineering*. 1342–1354.
[93] Tung Kieu, Bin Yang, Chenjuan Guo, Christian S. Jensen, Yan Zhao, Feiteng Huang, and Kai Zheng. 2022. Robust and explainable autoencoders for time series outlier detection. In *Proceedings of the 2022 IEEE 38th International Conference on Data Engineering*. 3038–3050.
[94] Dohwan Ko, Joonmyung Choi, Hyeong Kyu Choi, Kyoung-Woon On, Byungseok Roh, and Hyunwoo J. Kim. 2023. MELTR: Meta loss transformer for learning to fine-tune video foundation models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20105–20115.
[95] Satoshi Koide, Yukihiro Tadokoro, Chuan Xiao, and Yoshiharu Ishikawa. 2018. CiNCT: Compression and retrieval for massive vehicular trajectories via relative movement labeling. In *Proceedings of the 2018 IEEE 34th International Conference on Data Engineering*. 1097–1108.
[96] Fanhui Kong, Jianqiang Li, Bin Jiang, Huihui Wang, and Houbing Song. 2022. Trajectory optimization for drone logistics delivery via attention-based pointer network. *IEEE Transactions on Intelligent Transportation Systems* 24, 4 (2022), 4519–4531.
[97] Alex Lam, Matthew Schofield, and Shen-Shyang Ho. 2019. Detecting (unusual) events in urban areas using bike-sharing data. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Analytics for Local Events and News*. 5:1–5:7.
[98] Ralph Lange, Frank Dürr, and Kurt Rothermel. 2011. Efficient real-time trajectory tracking. *VLDBJ* 20, 5 (2011), 671–694.
[99] Jorge A. Laval. 2023. Self-organized criticality of traffic flow: Implications for congestion management technologies. *Transportation Research Part C: Emerging Technologies* 149 (2023), 104056.
[100] Can Li, Lei Bai, Wei Liu, Lina Yao, and S. Travis Waller. 2020. Graph neural network for robust public transit demand prediction. *IEEE Transactions on Intelligent Transportation Systems* 23, 5 (2020), 4086–4098.
[101] Rui Li, Fan Zhang, Tong Li, Ning Zhang, and Tingting Zhang. 2022. DMGAN: Dynamic multi-hop graph attention network for traffic forecasting. *IEEE Transactions on Knowledge and Data Engineering* 35, 9 (2022), 9088–9101.
[102] Tong Li, Ahmad Alhilal, Anlan Zhang, Mohammad A. Hoque, Dimitris Chatzopoulos, Zhu Xiao, Yong Li, and Pan Hui. 2019. Driving big data: A first look at driving behavior via a large-scale private car dataset. In *Proceedings of the 2019 IEEE 35th International Conference on Data Engineering*. 61–68.
[103] Tianyi Li, Lu Chen, Christian S. Jensen, and Torben Bach Pedersen. 2021. TRACE: Real-time compression of streaming trajectories in road networks. *PVLDB* 14, 7 (2021), 1175–1187.
[104] Tianyi Li, Ruikai Huang, Lu Chen, Christian S. Jensen, and Torben Bach Pedersen. 2020. Compression of uncertain trajectories in road networks. *PVLDB* 13, 7 (2020), 1050–1063.
[105] Tingting Li, Jianping Wu, Anrong Dang, Lyuchao Liao, and Ming Xu. 2019. Emission pattern mining based on taxi trajectory data in Beijing. *Journal of Cleaner Production* 206 (2019), 688–700.
[106] Tong Li, Shiduo Xin, Yanxin Xi, Sasu Tarkoma, Pan Hui, and Yong Li. 2022. Predicting multi-level socioeconomic indicators from structural urban imagery. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*. ACM, 3282–3291.
[107] Yong Li, Wenyu Ren, Depeng Jin, Pan Hui, Lieguang Zeng, and Dapeng Wu. 2014. Potential predictability of vehicular staying time for large-scale urban environment. *TVT* 63, 1 (2014), 322–333.
[108] Xinle Liang, Yang Liu, Tianjian Chen, Ming Liu, and Qiang Yang. 2023. Federated transfer reinforcement learning for autonomous driving. In *Federated and Transfer Learning* 27 (2023). 357–371.

[109] Binbing Liao, Jingqing Zhang, Chao Wu, Douglas McIlwraith, Tong Chen, Shengwen Yang, Yike Guo, and Fei Wu. 2018. Deep sequence learning with auxiliary information for traffic prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 537–546.

[110] Xuelian Lin, Jiahao Jiang, Shuai Ma, Yimeng Zuo, and Chunming Hu. 2019. One-pass trajectory simplification using the synchronous Euclidean distance. *VLDBJ* 28, 6 (2019), 897–921.

[111] Yan Lin, Huaiyu Wan, Jilin Hu, Shengnan Guo, Bin Yang, Youfang Lin, and Christian S. Jensen. 2023. Origin-destination travel time oracle for map-based services. *SIGMOD* 1, 3 (2023), 217:1–217:27.

[112] Zenan Ling, Robert C. Qiu, Xing He, and Lei Chu. 2021. A new approach of exploiting self-adjoint matrix polynomials of large random matrices for anomaly detection and fault location. *IEEE Transactions on Big Data* 7, 3 (2021), 548–558.

[113] Aoyu Liu and Yaying Zhang. 2024. An efficient spatial-temporal transformer with temporal aggregation and spatial memory for traffic forecasting. *Expert Syst. Appl.* 250 (2024), 123884.

[114] Aoyu Liu and Yaying Zhang. 2024. Spatial-temporal dynamic graph convolutional network with interactive learning for traffic forecasting. *IEEE Transactions on Intelligent Transportation Systems* 25, 7 (2024), 7645–7660.

[115] Chenxi Liu, Jiawei Cai, Dong Wang, Jiaxin Tang, Lei Wang, Huiling Chen, and Zhu Xiao. 2021. Understanding the regular travel behavior of private vehicles: An empirical evaluation and a semi-supervised model. *IEEE Sensors Journal* 21, 17 (2021), 19078–19090.

[116] Chenxi Liu, Dong Wang, Huiling Chen, and Renfa Li. 2021. Study of forecasting urban private car volumes based on multi-source heterogeneous data fusion. *Journal on Communication* 42, 3 (2021), 54–64.

[117] Chenxi Liu, Zhu Xiao, Cheng Long, Dong Wang, Tao Li, and Hongbo Jiang. 2024. MVCAR: Multi-view collaborative graph network for private car carbon emission prediction. *IEEE Transactions on Intelligent Transportation Systems* 26, 1 (2024), 472–483.

[118] Chenxi Liu, Zhu Xiao, Dong Wang, Minhao Cheng, Hongyang Chen, and Jiawei Cai. 2022. Foreseeing private car transfer between urban regions with multiple graph-based generative adversarial networks. *World Wide Web* 25, 6 (2022), 2515–2534.

[119] Chenxi Liu, Zhu Xiao, Dong Wang, lei Wang, Hongbo Jiang, Hongyang Chen, and Jiangxia Yu. 2021. Exploiting spatiotemporal correlations of arrive-stay-leave behaviors for private car flow prediction. *IEEE Transactions on Network Science and Engineering* 9, 2 (2021), 834–847.

[120] Chenxi Liu, Qianxiong Xu, Hao Miao, Sun Yang, Lingzheng Zhang, Cheng Long, Ziyue Li, and Rui Zhao. 2025. TimeCMA: Towards LLM-empowered multivariate time series forecasting via cross-modality alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

[121] Chenxi Liu, Sun Yang, Qianxiong Xu, Zhishuai Li, Cheng Long, Ziyue Li, and Rui Zhao. 2024. Spatial-temporal large language model for traffic prediction. In *Proceedings of the 2024 25th IEEE International Conference on Mobile Data Management*.

[122] Jiaxin Liu, Liwei Deng, Hao Miao, Yan Zhao, and Kai Zheng. 2022. Task assignment with federated preference learning in spatial crowdsourcing. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*. 1279–1288.

[123] Junming Liu, Leilei Sun, Qiao Li, Jingci Ming, Yanchi Liu, and Hui Xiong. 2017. Functional zone based hierarchical demand prediction for bike system expansion. In *Proceedings of the 23th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 957–966.

[124] Jiajun Liu, Kun Zhao, Philipp Sommer, Shuo Shang, Brano Kusy, Jae-Gil Lee, and Raja Jurdak. 2016. A novel framework for online amnesic trajectory compression in resource-constrained environments. *IEEE Transactions on Knowledge and Data Engineering* 28, 11 (2016), 2827–2841.

[125] Sijing Liu, Gang Chen, Long Wei, and Guoqi Li. 2021. A novel compression approach for truck GPS trajectory data. *IET Intelligent Transport Systems* 15, 1 (2021), 74–83.

[126] Wei Liu, Yoshito Watanabe, and Yozo Shoji. 2020. Vehicle-assisted data delivery in smart city: A deep learning approach. *TVT* 69, 11 (2020), 13849–13860.

[127] Yu Liu, Qian Ge, Wei Luo, Qiang Huang, Lei Zou, Haixu Wang, Xin Li, and Chang Liu. 2023. Graphmm: Graph-based vehicular map matching by leveraging trajectory and road correlations. *IEEE Transactions on Knowledge and Data Engineering* 36, 1 (2023), 184–198.

[128] Ziqiao Liu, Hao Miao, Yan Zhao, Chenxi Liu, Kai Zheng, and Huan Li. 2024. LightTR: A lightweight framework for federated trajectory recovery. In *Proceedings of the 2024 IEEE 40th International Conference on Data Engineering*.

[129] Cheng Long, Raymond Chi-Wing Wong, and HV Jagadish. 2014. Trajectory simplification: On minimizing the direction-based error. *PVLDB* 8, 1 (2014), 49–60.

[130] Wangchen Long, Tao Li, Zhu Xiao, Dong Wang, Rui Zhang, Amelia C. Regan, Hongyang Chen, and Yongdong Zhu. 2022. Location prediction for individual vehicles via exploiting travel regularity and preference. *TVT* 71, 5 (2022), 4718–4732.

[131] Yandi Lun, Hao Miao, Jiaxing Shen, Renzhi Wang, Xiang Wang, and Senzhang Wang. 2023. Resisting TUL attack: Balancing data privacy and utility on trajectory via collaborative adversarial learning. *GeoInformatica* (2023), 1–21.

[132] Haoyuan Ma, Mintao Zhou, Xiaodong Ouyang, Du Yin, Renhe Jiang, and Xuan Song. 2022. Forecasting regional multimodal transportation demand with graph neural networks: An open dataset. In *ITSC*. 3263–3268.

[133] Jiaman Ma, Jeffrey Chan, Goce Ristanoski, Sutharshan Rajasegarar, and Christopher Leckie. 2019. Bus travel time prediction with real-time traffic information. *Transportation Research Part C: Emerging Technologies* 105 (2019), 536–549.

[134] Yuexin Ma, Xinge Zhu, Sibo Zhang, Ruigang Yang, Wenping Wang, and Dinesh Manocha. 2019. TrafficPredict: Trajectory prediction for heterogeneous traffic-agents. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 6120–6127.

[135] Ranwa Al Mallah, Alejandro Quintero, and Bilal Farooq. 2022. Prediction of traffic flow via connected vehicles. *IEEE Transactions on Mobile Computing* 21, 1 (2022), 264–277.

[136] Congcong Miao, Jilong Wang, Heng Yu, Weichen Zhang, and Yinyao Qi. 2020. Trajectory-user linking with attentive recurrent network. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems*. 878–886.

[137] Hao Miao, Ziqiao Liu, Yan Zhao, Chenjuan Guo, Bin Yang, Kai Zheng, and Christian S. Jensen. 2024. Less is more: Efficient time series dataset condensation via two-fold modal matching. In *Proceedings of the VLDB Endowment*.

[138] Hao Miao, Jiaxing Shen, Jiannong Cao, Jiangnan Xia, and Senzhang Wang. 2022. MBA-STNet: Bayes-enhanced discriminative multi-task learning for flow prediction. *IEEE Transactions on Knowledge and Data Engineering* 35, 7 (2022), 7164–7177.

[139] Hao Miao, Yan Zhao, Chenjuan Guo, Bin Yang, Zheng Kai, Feiteng Huang, Jiandong Xie, and Christian S. Jensen. 2024. A unified replay-based continuous learning framework for spatio-temporal prediction on streaming data. In *Proceedings of the 2024 IEEE 40th International Conference on Data Engineering*.

[140] Hao Miao, Yan Zhao, Chenjuan Guo, Bin Yang, Kai Zheng, and Christian S Jensen. 2025. Spatio-temporal prediction on streaming data: A unified federated continuous learning framework. *IEEE Transactions on Knowledge and Data Engineering* 37, 4 (2025), 2126–2140.

[141] Hao Miao, Xiaolong Zhong, Jiaxin Liu, Yan Zhao, Xiangyu Zhao, Weizhu Qian, Kai Zheng, and Christian S. Jensen. 2024. Task assignment with efficient federated preference learning in spatial crowdsourcing. *IEEE Transactions on Knowledge and Data Engineering* 36, 4 (2024), 1800–1814.

[142] Songsong Mo, Zhifeng Bao, Baihua Zheng, and Zhiyong Peng. 2022. Towards an optimal bus frequency scheduling: When the waiting time matters. *IEEE Transactions on Knowledge and Data Engineering* 34, 9 (2022), 4484–4498.

[143] Songsong Mo, Yile Chen, Hao Wang, Gao Cong, and Zhifeng Bao. 2023. Lemo: A cache-enhanced learned optimizer for concurrent queries. *SIGMOD* 1, 4 (2023), 1–26.

[144] Jonathan Muckell, Jeong-Hyon Hwang, Vikram Patil, Catherine T. Lawson, Fan Ping, and S. S. Ravi. 2011. SQUISH: An online approach for GPS trajectory compression. In *Proceedings of the 2nd International Conference on Computing for Geospatial Research and Applications*. 1–8.

[145] Kianoosh Soltani Naveh and Jiwon Kim. 2019. Urban trajectory analytics: Day-of-week movement pattern mining using tensor factorization. *IEEE Transactions on Intelligent Transportation Systems* 20, 7 (2019), 2540–2549.

[146] Kun Ouyang, Reza Shokri, David S. Rosenblum, and Wenzhuo Yang. 2018. A non-parametric generative model for human trajectories. In *Proceedings of the IJCAI*. 3812–3817.

[147] Junbiao Pang, Jing Huang, Yong Du, Haitao Yu, Qingming Huang, and Baocai Yin. 2019. Learning to predict bus arrival time from heterogeneous measurements via recurrent neural network. *IEEE Transactions on Intelligent Transportation Systems* 20, 9 (2019), 3283–3293.

[148] Luca Pappalardo, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-László Barabási. 2015. Returners and explorers dichotomy in human mobility. *Nature Communications* 6, 1 (2015), 1–8.

[149] Alejandro Pérez Riascos and José Mateos. 2020. Networks and long-range mobility in cities: A study of more than one billion taxi trips in New York City. *Scientific Reports* 10 (03 2020), 4022–4036.

[150] Geqi Qi, Ailing Huang, Wei Guan, and Lingling Fan. 2019. Analysis and prediction of regional mobility patterns of bus travellers using smart card data and points of interest data. *IEEE Transactions on Intelligent Transportation Systems* 20, 4 (2019), 1197–1214.

[151] Qiujun Qian, Mi Gan, Lifei Wei, Zhu Yao, and Yuxin He. 2024. Behavior-driven planning of electric truck charging infrastructure for intercity operations. *IEEE Transactions on Intelligent Transportation Systems* 25, 9 (2024), 12049–12065.

[152] Tangwen Qian, Yongjun Xu, Zhao Zhang, and Fei Wang. 2022. Trajectory prediction from hierarchical perspective. In *Proceedings of the 30th ACM International Conference on Multimedia*. 6822–6830.

[153] Huiling Qin, Xianyuan Zhan, Yuanxun Li, Xiaodu Yang, and Yu Zheng. 2021. Network-wide traffic states imputation using self-interested coalitional learning. In *Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1370–1378.

[154] Boting Qu, Wenxin Yang, Ge Cui, and Xin Wang. 2020. Profitable taxi travel route recommendation based on big taxi trajectory data. *IEEE Transactions on Intelligent Transportation Systems* 21, 2 (2020), 653–668.

[155] Aida Ramezani and Yang Xu. 2023. Knowledge of cultural moral norms in large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*. 428–446.

[156] Felix Rempe, Philipp Franeck, and Klaus Bogenberger. 2022. On the estimation of traffic speeds with deep convolutional neural networks given probe data. *Transportation Research Part C: Emerging Technologies* 134 (2022), 103448.

[157] Huimin Ren, Sijie Ruan, Yanhua Li, Jie Bao, Chuishi Meng, Ruiyuan Li, and Yu Zheng. 2021. MTrajRec: Map-constrained trajectory recovery via Seq2Seq multi-task learning. In *Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1410–1419.

[158] Paulo H. Rettore, Guilherme Maia, Leandro A. Villas, and Antonio A. F. Loureiro. 2019. Vehicular data space: The data point of view. *IEEE Communications Surveys and Tutorials* 21, 3 (2019), 2392–2418.

[159] Can Rong, Tong Li, Jie Feng, and Yong Li. 2023. Inferring origin-destination flows from population distribution. *IEEE Transactions on Knowledge and Data Engineering* 35, 1 (2023), 603–613.

[160] Alberto Rossi, Gianni Barlacchi, Monica Bianchini, and Bruno Lepri. 2020. Modelling taxi drivers' behaviour for the next destination prediction. *IEEE Transactions on Intelligent Transportation Systems* 21, 7 (2020), 2980–2989.

[161] Iulian Sandu Popa, Karine Zeitouni, Vincent Oria, and Ahmed Kharrat. 2015. Spatio-temporal compression of trajectories in road networks. *GeoInformatica* 19, 1 (2015), 117–145.

[162] Kangjia Shao, Yang Wang, Zhengyang Zhou, Xike Xie, and Guang Wang. 2021. TrajForesee: How limited detailed trajectories enhance large-scale sparse information to predict vehicle trajectories?. In *Proceedings of the 2021 IEEE 37th International Conference on Data Engineering*. 2189–2194.

[163] Chaoyang Shi, Qingquan Li, Shiwei Lu, and Xiping Yang. 2021. Exploring temporal intra-urban travel patterns: An online car-hailing trajectory data perspective. *Remote Sensing* 13, 9 (2021), 1825.

[164] Roniel S. De Sousa, Azzedine Boukerche, and Antonio A. F. Loureiro. 2020. Vehicle trajectory similarity: Models, methods, and applications. *ACM Computing Surveys* 53, 5 (2020), 1–32.

[165] Han Su, Guanglin Cong, Wei Chen, Bolong Zheng, and Kai Zheng. 2019. Personalized route description based on historical trajectories. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 79–88.

[166] Wenbo Sun, Fangni Zhang, Wei Liu, and Qingying He. 2023. Optimal control of connected autonomous vehicles in a mixed traffic corridor. *IEEE Intelligent Transportation Systems Magazine* 25, 5 (2023), 4206–4218.

[167] Yidan Sun, Guiyuan Jiang, Siew-Kei Lam, Shicheng Chen, and Peilan He. 2019. Bus travel speed prediction using attention network of heterogeneous correlation features. In *Proceedings of the 2019 SIAM International Conference on Data Mining*. 73–81.

[168] Xianfeng Tang, Boqing Gong, Yanwei Yu, Huaxiu Yao, Yandong Li, Haiyong Xie, and Xiaoyu Wang. 2019. Joint modeling of dense and incomplete trajectories for citywide traffic volume inference. In *Proceedings of the World Wide Web*. 1806–1817.

[169] Xinyao Tang, Huansheng Song, Wei Wang, and Yanni Yang. 2020. Vehicle spatial distribution and 3D trajectory extraction algorithm in a cross-camera traffic scene. *Sensors* 20, 22 (2020), 6517.

[170] Vasudha Todi, Gunjan Sengupta, and Sourangshu Bhattacharya. 2019. Probabilistic path planning using obstacle trajectory prediction. In *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*. ACM, 36–43.

[171] Panrong Tong, Mingqian Li, Mo Li, Jianqiang Huang, and Xiansheng Hua. 2021. Large-scale vehicle trajectory reconstruction with camera sensing network. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 188–200.

[172] Wei Tu, Ke Mai, Yatao Zhang, Yang Xu, Jincai Huang, Min Deng, Long Chen, and Qingquan Li. 2021. Real-time route recommendations for E-Taxies leveraging GPS trajectories. *TII* 17, 5 (2021), 3133–3142.

[173] Jarrett Walker. 2024. *Human Transit, Revised Edition: How Clearer Thinking about Public Transit Can Enrich Our Communities and Our Lives*. Island Press.

[174] Chengcheng Wan, Yanmin Zhu, Jiadi Yu, and Yanyan Shen. 2018. SMOPAT: Mining semantic mobility patterns from trajectories of private vehicles. *Information Sciences* 429 (2018), 12–25.

[175] Dong Wang, Jiaojiao Fan, Zhu Xiao, Hongbo Jiang, Hongyang Chen, Fanzi Zeng, and Keqin Li. 2019. Stop-and-wait: Discover aggregation effect based on private car trajectory data. *IEEE Transactions on Intelligent Transportation Systems* 20, 10 (2019), 3623–3633.

[176] Dong Wang, Junbo Zhang, Wei Cao, Jian Li, and Yu Zheng. 2018. When will you arrive? Estimating travel time based on deep neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 2500–2507.

[177] Yue Jiang, Qin Chao, Yile Chen, Xiucheng Li, Shuai Liu, and Gao Cong. 2024. UrbanLLM: Autonomous urban activity planning and management with large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*. 1810–1825.

[178] Jingyuan Wang, Ning Wu, Xinxi Lu, Xin Zhao, and Kai Feng. 2021. Deep trajectory recovery with fine-grained calibration using kalman filter. *IEEE Transactions on Knowledge and Data Engineering* 33, 3 (2021), 921–934.
[179] Qi Wang, Min Lu, and Qingquan Li. 2020. Interactive, multiscale urban-traffic pattern exploration leveraging massive GPS trajectories. *Sensors* 20, 4 (2020), 1084.
[180] Sheng Wang, Zhifeng Bao, J Shane Culpepper, and Gao Cong. 2021. A survey on trajectory data management, analytics, and learning. *ACM Computing Surveys* 54, 2 (2021), 1–36.
[181] Senzhang Wang, Jiannong Cao, and Philip Yu. 2020. Deep learning for spatio-temporal data mining: A survey. *IEEE Transactions on Knowledge and Data Engineering* 34, 8 (2020), 3681–3700.
[182] Senzhang Wang, Hao Miao, Hao Chen, and Zhiqiu Huang. 2020. Multi-task adversarial spatial-temporal networks for crowd flow prediction. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*. 1555–1564.
[183] Senzhang Wang, Hao Miao, Jiyue Li, and Jiannong Cao. 2021. Spatio-temporal knowledge transfer for urban crowd flow prediction via deep attentive adaptation networks. *IEEE Transactions on Intelligent Transportation Systems* 23, 5 (2021), 4695–4705.
[184] Senzhang Wang, Hao Miao, Jiyue Li, and Jiannong Cao. 2022. Spatio-temporal knowledge transfer for urban crowd flow prediction via deep attentive adaptation networks. *IEEE Transactions on Intelligent Transportation Systems* 23, 5 (2022), 4695–4705.
[185] Senzhang Wang, Jiaqiang Zhang, Jiyue Li, Hao Miao, and Jiannong Cao. 2021. Traffic accident risk prediction via multi-view multi-task spatio-temporal networks. *IEEE Transactions on Knowledge and Data Engineering* 35, 12 (2021), 12323–12336.
[186] Senzhang Wang, Meiyue Zhang, Hao Miao, Zhaohui Peng, and Philip S. Yu. 2022. Multivariate correlation-aware spatio-temporal graph convolutional networks for multi-scale traffic prediction. *TIST* 13, 3 (2022), 1–22.
[187] Yang Wang, Yong Zhang, Zhen Sean Qian, Shaofan Wang, Yongli Hu, and Baocai Yin. 2019. Multi-source traffic data reconstruction using joint low-rank and fundamental diagram constraints. *IEEE Intelligent Transportation Systems Magazine* 11, 3 (2019), 221–234.
[188] Ziran Wang, Rohit Gupta, Kyungtae Han, Haoxin Wang, Akila Ganlath, Nejib Ammar, and Prashant Tiwari. 2022. Mobility digital twin: Concept, architecture, case study, and future challenges. *IOTJ* 9, 18 (2022), 17452–17467.
[189] Zhaonan Wang, Renhe Jiang, Hao Xue, Flora D. Salim, Xuan Song, and Ryosuke Shibasaki. 2022. Event-aware multimodal mobility nowcasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 4228–4236.
[190] Zheng Wang, Cheng Long, and Gao Cong. 2021. Trajectory simplification with reinforcement learning. In *Proceedings of the 2021 IEEE 37th International Conference on Data Engineering*. 684–695.
[191] Zheng Wang, Cheng Long, Gao Cong, and Qianru Zhang. 2021. Error-bounded online trajectory simplification with multi-agent reinforcement learning. In *Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1758–1768.
[192] Zhichang Wang, Binjie Zhang, Nu Xia, and Jiahui Jin. 2022. Generating dynamic urban traffic based on stochastic origin-destination matrix. In *Proceedings of the 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design*. 570–575.
[193] Hua Wei, Dongkuan Xu, Junjie Liang, and Zhenhui Li. 2021. How do we move: Modeling human movement with system dynamics. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 4445–4452.
[194] David Wilkie, Jason Sewall, and Ming C. Lin. 2013. Flow reconstruction for data-driven traffic animation. *ACM Transactions on Graphics* 32, 4 (2013), 89:1–89:10.
[195] Dongbo Xi, Fuzhen Zhuang, Yanchi Liu, Hengshu Zhu, Pengpeng Zhao, Chang Tan, and Qing He. 2020. Exploiting bi-directional global transition patterns and personal preferences for missing POI category identification. *Neural Networks* 132 (2020), 75–83.
[196] Yanxin Xi, Tong Li, Huandong Wang, Yong Li, Sasu Tarkoma, and Pan Hui. 2022. Beyond the first law of geography: Learning representations of satellite imagery by leveraging point-of-interests. In *Proceedings of the World Wide Web*. 3308–3316.
[197] Tong Xia, Yunhan Qi, Jie Feng, Fengli Xu, Funing Sun, Diansheng Guo, and Yong Li. 2021. AttnMove: History enhanced trajectory recovery via attentional network. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 4494–4502.
[198] Jianhua Xiao, Zhu Xiao, Dong Wang, Vincent Havyarimana, Chenxi Liu, Chengming Zou, and Di Wu. 2022. Vehicle trajectory interpolation based on ensemble transfer regression. *IEEE Transactions on Intelligent Transportation Systems* 23, 7 (2022), 7680–7691.
[199] Zhu Xiao, Yanxun Chen, Mamoun Alazab, and Hongyang Chen. 2022. Trajectory data acquisition via private car positioning based on tightly-coupled GPS/OBD integration in urban environments. *IEEE Transactions on Intelligent Transportation Systems* 23, 7 (2022), 9680–9691.

[200] Zhu Xiao, Hui Fang, Hongbo Jiang, Jing Bai, Vincent Havyarimana, and Hongyang Chen. 2021. Understanding urban area attractiveness based on private car trajectory data using a deep learning approach. *IEEE Transactions on Intelligent Transportation Systems* 23, 8 (2021), 12343–12352.
[201] Zhu Xiao, Hui Fang, Hongbo Jiang, Jing Bai, Vincent Havyarimana, Hongyang Chen, and Licheng Jiao. 2021. Understanding private car aggregation effect via spatio-temporal analysis of trajectory data. *IEEE Transactions on Cybernetics* 53, 4 (2021), 2346–2357.
[202] Zhu Xiao, Fancheng Li, Ronghui Wu, Hongbo Jiang, Yupeng Hu, Ju Ren, Chenglin Cai, and Arun Iyengar. 2020. TrajData: On vehicle trajectory collection with commodity plug-and-play OBU devices. *IOTJ* 7, 9 (2020), 9066–9079.
[203] Zhu Xiao, Pingting Li, Vincent Havyarimana, Georges Maigary Hassana, Dong Wang, and Keqin Li. 2018. GOI: A novel design for vehicle positioning and trajectory prediction under urban environments. *IEEE Sensors Journal* 18, 13 (2018), 5586–5594.
[204] Zhu Xiao, Bo Liu, Linshan Wu, Hongbo Jiang, Beihao Xia, Tao Li, and Cassandra C Wang. 2024. Exploring spatio-temporal carbon emission across passenger car trajectory data. *IEEE Transactions on Intelligent Transportation Systems* 26, 2 (2024), 1812–1825.
[205] Zhu Xiao, Jinmei Shu, Hongbo Jiang, Geyong Min, Hongyang Chen, and Zhu Han. 2023. Perception task offloading with collaborative computation for autonomous driving. *IEEE Journal on Selected Areas in Communications* 41, 2 (2023), 457–473.
[206] Zhu Xiao, Jinmei Shu, Hongbo Jiang, Geyong Min, Jinwen Liang, and Arun Iyengar. 2024. Toward collaborative occlusion-free perception in connected autonomous vehicles. *IEEE Transactions on Mobile Computing* 23, 5 (2024), 4918–4929.
[207] Zhu Xiao, Linshan Wu, Hongbo Jiang, Zheng Qin, Chengxi Gao, You Li, Hongyang Chen, and Jiangchuan Liu. 2023. Exploring intercity mobility in urban agglomeration: Evidence from private car trajectory data. *TCSS* 21, 12 (2023), 5036–5050.
[208] Zhu Xiao, Dapeng Xiao, Vincent Havyarimana, Hongbo Jiang, Daibo Liu, Dong Wang, and Fanzi Zeng. 2019. Toward accurate vehicle state estimation under non-Gaussian noises. *IOTJ* 6, 6 (2019), 10652–10664.
[209] Zhu Xiao, Hui Xiao, Wenjie Chen, Hongyang Chen, Amelia Regan, and Hongbo Jiang. 2021. Exploring human mobility patterns and travel behavior: A focus on private cars. *IEEE Intelligent Transportation Systems Magazine* 14, 5 (2021), 129–146.
[210] Zhu Xiao, Shenyuan Xu, Tao Li, Hongbo Jiang, Rui Zhang, Amelia C. Regan, and Hongyang Chen. 2020. On extracting regular travel behavior of private cars based on trajectory data analysis. *TVT* 69, 12 (2020), 14537–14549.
[211] Jiping Xing. 2022. Estimating traffic volumes in an urban network based on taxi GPS and limited LPR data using machine learning techniques. In *Proceedings of the 2022 IEEE 25th International Conference on Intelligent Transportation Systems.* 20–25.
[212] Qianxiong Xu, Sijie Ruan, Cheng Long, Liang Yu, and Chen Zhang. 2022. Traffic speed imputation with spatio-temporal attentions and cycle-perceptual training. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management.* 2280–2289.
[213] Ronghui Xu, Hao Miao, Senzhang Wang, Philip S Yu, and Jianxin Wang. 2024. PeFAD: A parameter-efficient federated framework for time series anomaly detection. In *Proceedings of the 30th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*
[214] Sun Yang, Qiong Su, Zhishuai Li, Ziyue Li, Hangyu Mao, Chenxi Liu, and Rui Zhao. 2024. SQL-to-schema enhances schema linking in text-to-SQL. In *Proceedings of the International Conference on Database and Expert Systems Applications.*
[215] Xiaochun Yang, Bin Wang, Kai Yang, Chengfei Liu, and Baihua Zheng. 2017. A novel representation and compression for queries on trajectories in road networks. *IEEE Transactions on Knowledge and Data Engineering* 30, 4 (2017), 613–629.
[216] Huaxiu Yao, Fei Wu, Jintao Ke, Xianfeng Tang, Yitian Jia, Siyu Lu, Pinghua Gong, Jieping Ye, and Zhenhui Li. 2018. Deep multi-view spatial-temporal network for taxi demand prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence.* 2588–2595.
[217] Qingying Yu, Chuanming Chen, Liping Sun, and Xiaoyao Zheng. 2021. Urban Hotspot area detection using nearest-neighborhood-related quality clustering on taxi trajectory data. *ISPRS International Journal of Geo-Information* 10, 7 (2021), 473.
[218] Wenhao Yu. 2019. Discovering frequent movement paths from taxi trajectory data using spatially embedded networks and association rules. *IEEE Transactions on Intelligent Transportation Systems* 20, 3 (2019), 855–866.
[219] Yong Yu, Haina Tang, Fei Wang, Lin Wu, Tangwen Qian, Tao Sun, and Yongjun Xu. 2020. TULSN: Siamese network for trajectory-user linking. In *Proceedings of the 2020 International Joint Conference on Neural Networks.* 1–8.
[220] Yanwei Yu, Xianfeng Tang, Huaxiu Yao, Xiuwen Yi, and Zhenhui Li. 2019. Citywide traffic volume inference with surveillance camera records. *IEEE Transactions on Big Data* 7, 6 (2019), 900–912.

[221] Zhi Yu, Weichi Li, Yonghong Liu, Xuelan Zeng, Yongming Zhao, Kaiying Chen, Bin Zou, and Jiajun He. 2021. Quantification and management of urban traffic emissions based on individual vehicle data. *Journal of Cleaner Production* 328 (2021), 129386.

[222] Zhongzhi Yu, Shang Wu, Yonggan Fu, Shunyao Zhang, and Yingyan Celine Lin. Hint-Aug: Drawing hints from foundation vision transformers towards boosted few-shot parameter-efficient tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11102–11112.

[223] Haitao Yuan, Guoliang Li, Zhifeng Bao, and Ling Feng. 2020. Effective travel time estimation: When historical trajectories over road networks matter. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*. 2135–2149.

[224] Jing Yuan, Yu Zheng, Chengyang Zhang, Wenlei Xie, Xing Xie, Guangzhong Sun, and Yan Huang. 2010. T-drive: Driving directions based on taxi trajectories. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 99–108.

[225] Nicholas Jing Yuan, Yu Zheng, Xing Xie, Yingzi Wang, Kai Zheng, and Hui Xiong. 2015. Discovering urban functional zones using latent activity trajectories. *IEEE Transactions on Knowledge and Data Engineering* 27, 3 (2015), 712–725.

[226] Yuan Yuan, Jingtao Ding, Huandong Wang, Depeng Jin, and Yong Li. 2022. Activity trajectory generation via modeling spatiotemporal dynamics. In *Proceedings of the 28th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 4752–4762.

[227] Ye Yuan, Yong Zhang, Boyue Wang, Yuan Peng, Yongli Hu, and Baocai Yin. 2022. STGAN: Spatio-temporal generative adversarial network for traffic data imputation. *IEEE Transactions on Big Data* 9, 1 (2022), 200–211.

[228] Ying Wu Xiangyu Zhao Yuanshao Zhu', Yongchao Ye' and J.Q. Yu James. 2023. SynMob: Creating high-fidelity synthetic GPS trajectory dataset for urban mobility analysis. In *Proceedings of the Advances in Neural Information Processing Systems*.

[229] Chak Fai Yuen, Abhishek Pratap Singh, Sagar Goyal, Sayan Ranu, and Amitabha Bagchi. 2019. Beyond shortest paths: Route recommendations for ride-sharing. In *Proceedings of the World Wide Web*.

[230] Xianyuan Zhan, Ruimin Li, and Satish V. Ukkusuri. 2015. Lane-based real-time queue length estimation using license plate recognition data. *Transportation Research Part C: Emerging Technologies* 57 (2015), 85–102.

[231] Xianyuan Zhan, Yu Zheng, Xiuwen Yi, and Satish V Ukkusuri. 2016. Citywide traffic volume estimation using trajectory data. *IEEE Transactions on Knowledge and Data Engineering* 29, 2 (2016), 272–285.

[232] Chizhan Zhang, Fenghua Zhu, Yisheng Lv, Peijun Ye, and Fei-Yue Wang. 2021. MLRNN: Taxi demand prediction based on multi-level deep learning and regional heterogeneity analysis. *IEEE Transactions on Intelligent Transportation Systems* 23, 7 (2021), 8412–8422.

[233] Jiawei Zhang, Cheng Chang, Zimin He, Wenqin Zhong, Danya Yao, Shen Li, and Li Li. 2023. Cavsim: A microscopic traffic simulator for evaluation of connected and automated vehicles. *IEEE Intelligent Transportation Systems Magazine* 24, 9 (2023), 10038–10054.

[234] Jinlei Zhang, Hongshu Che, Feng Chen, Wei Ma, and Zhengbing He. 2021. Short-term origin-destination demand prediction in urban rail transit systems: A channel-wise attentive split-convolutional neural network method. *Transportation Research Part C: Emerging Technologies* 124 (2021), 102928–102948.

[235] Mingyang Zhang, Tong Li, Yong Li, and Pan Hui. 2020. Multi-view joint graph representation learning for urban region embedding. In *Proceedings of the 29th International Conference on International Joint Conferences on Artificial Intelligence*. 4431–4437.

[236] Mingyang Zhang, Tong Li, Hongzhi Shi, Yong Li, and Pan Hui. 2019. A decomposition approach for urban anomaly detection across spatiotemporal data. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. 6043–6049.

[237] Mingyang Zhang, Tong Li, Yue Yu, Yong Li, Pan Hui, and Yu Zheng. 2020. Urban anomaly analytics: Description, detection and prediction. *IEEE Transactions on Big Data* 8, 3 (2020), 809–826.

[238] Qibo Zhang, Fanzi Zeng, Zhu Xiao, Hongbo Jiang, Amelia C. Regan, Kehua Yang, and Yongdong Zhu. 2022. Toward predicting stay time for private car users: A RNN-NALU approach. *TVT* 71, 6 (2022), 6007–6018.

[239] Sen Zhang, Senzhang Wang, Xiang Wang, Shigeng Zhang, Hao Miao, and Junxing Zhu. 2022. Multi-task adversarial learning for semi-supervised trajectory-user linking. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. 1–17.

[240] Weijia Zhang, Hao Liu, Yanchi Liu, Jingbo Zhou, and Hui Xiong. 2020. Semi-supervised hierarchical recurrent graph neural network for city-wide parking availability prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 1186–1193.

[241] Jie Zhao, Chao Chen, Wanyi Zhang, Ruiyuan Li, Fuqiang Gu, Songtao Guo, Jun Luo, and Yu Zheng. 2023. Coupling makes better: An intertwined neural network for taxi and ridesourcing demand co-prediction. *IEEE Transactions on Intelligent Transportation Systems* 25, 2 (2023), 1691—1705.

[242] Xin Zhao, Xiaowei Ma, Boyang Chen, Yuping Shang, and Malin Song. 2022. Challenges toward carbon neutrality in China: Strategies and countermeasures. *Resources, Conservation and Recycling* 176 (2022), 105959.

[243] Yan Zhao, Xuanhao Chen, Liwei Deng, Tung Kieu, Chenjuan Guo, Bin Yang, Kai Zheng, and Christian S. Jensen. 2022. Outlier detection for streaming task assignment in crowdsourcing. In *Proceedings of the World Wide Web*. 1933–1943.

[244] Yan Zhao, Shuo Shang, Yu Wang, Bolong Zheng, Quoc Viet Hung Nguyen, and Kai Zheng. 2018. REST: A reference-based framework for spatio-temporal trajectory compression. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2797–2806.

[245] Yan Zhao, Jianfeng Zheng, Wai Wong, Xingmin Wang, Yuan Meng, and Henry X. Liu. 2019. Various methods for queue length and traffic volume estimation using probe vehicle trajectories. *Transportation Research Part C: Emerging Technologies* 107 (2019), 70–91.

[246] Kai Zheng, Yan Zhao, Defu Lian, Bolong Zheng, Guanfeng Liu, and Xiaofang Zhou. 2019. Reference-based framework for spatio-temporal trajectory compression and query processing. *IEEE Transactions on Knowledge and Data Engineering* 32, 11 (2019), 2227–2240.

[247] Linjiang Zheng, Dong Xia, Li Chen, and Dihua Sun. 2020. Understanding citywide resident mobility using big data of electronic registration identification of vehicles. *IEEE Transactions on Intelligent Transportation Systems* 21, 10 (2020), 4363–4377.

[248] Yu Zheng. 2015. Trajectory data mining: An overview. *TIST* 6, 3 (2015), 29:1–29:41.

[249] Xiaolong Zhong, Hao Miao, Dazhuo Qiu, Yan Zhao, and Kai Zheng. 2023. Personalized location-preference learning for federated task assignment in spatial crowdsourcing. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 3534–3543.

[250] Fan Zhou, Qiang Gao, Goce Trajcevski, Kunpeng Zhang, Ting Zhong, and Fengli Zhang. 2018. Trajectory-user linking via variational autoencoder. In *Proceedings of the IJCAI*. 3212–3218.

[251] Fan Zhou, Xin Liu, Kunpeng Zhang, and Goce Trajcevski. 2020. Toward discriminating and synthesizing motion traces using deep probabilistic generative models. *TNNLS* 32, 6 (2020), 2401–2414.

[252] Haicang Zhou, Weiming Huang, Yile Chen, Tiantian He, Gao Cong, and Yew-Soon Ong. 2024. Road network representation learning with the third law of Geography. In *Proceedings of the Advances in Neural Information Processing Systems*.

[253] Wen Zhou, Dongping Ming, Xianwei Lv, Keqi Zhou, Hanqing Bao, and Zhaoli Hong. 2020. SO–CNN based urban functional zone fine division with VHR remote sensing image. *Remote Sensing of Environment* 236 (2020), 111458–111478.

[254] Zhengyang Zhou, Yang Wang, Xike Xie, Lianliang Chen, and Hengchang Liu. 2020. RiskOracle: A minute-level citywide traffic accident forecasting framework. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 1258–1265.

[255] Zhengyang Zhou, Yang Wang, Xike Xie, Lianliang Chen, and Chaochao Zhu. 2022. Foresee urban sparse traffic accidents: A spatiotemporal multi-granularity perspective. *IEEE Transactions on Knowledge and Data Engineering* 34, 8 (2022), 3786–3799.

[256] Zhengyang Zhou, Yang Wang, Xike Xie, Lei Qiao, and Yuantao Li. 2021. STUaNet: Understanding uncertainty in spatiotemporal collective human mobility. In *Proceedings of the World Wide Web*. 1868–1879.

[257] Di Zhu, Yu Liu, Xin Yao, and Manfred M. Fischer. 2023. Correction to: Spatial regression graph convolutional neural networks: A deep learning paradigm for spatial multivariate distributions. *GeoInformatica* 27, 3 (2023), 641–642.

[258] Xi Zhu and Diansheng Guo. 2017. Urban event detection with big data of taxi OD trips: A time series decomposition approach. *Transactions in GIS* 21, 3 (2017), 560–574.

[259] Yuanshao Zhu, Yongchao Ye, Shiyao Zhang, Xiangyu Zhao, and James Yu. 2023. DiffTraj: Generating GPS trajectory with diffusion probabilistic model. In *Proceedings of the Advances in Neural Information Processing Systems*.

[260] Alex Zyner, Stewart Worrall, and Eduardo M. Nebot. 2020. Naturalistic driver intention and path prediction using recurrent neural networks. *IEEE Transactions on Intelligent Transportation Systems* 21, 4 (2020), 1584–1594.

Received 7 February 2023; revised 14 January 2025; accepted 22 January 2025