International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 3
May-Jun 2026, PP 01-10

Joint Contrastive Representation Learning for Road Networks and Trajectory Data: A Review

Shweta Santosh Bhoye[1], Namrata D. Ghuse[2]
[1]Student, Department of Computer Engineering MET Bhujbal Knowledge City, Nashik, India
[2]Assistant Professor, Department of Computer Engineering MET Bhujbal Knowledge City, Nashik, India

Abstract. Road network structures and trajectory data are essential components of intelligent transportation systems (ITS), as they represent spatial infrastructure and temporal movement patterns, respectively. While road networks capture structural relationships and contextual information, trajectory data re- flects dynamic mobility behavior over time. Recent research has increasingly focused on integrating these two data sources using self-supervised and contrastive learning techniques to produce unified and meaningful representations. This paper presents a comprehensive review of joint contrastive representa- tion learning methods that model both intra-domain relationships (road–road and trajectory–trajectory) and inter-domain interac- tions (road–trajectory). Findings reported across multiple real- world mobility datasets indicate that these approaches achieve improved performance in tasks such as traffic prediction, route optimization, and trajectory similarity analysis compared to tra- ditional non-contrastive methods. In addition, this study examines commonly used evaluation strategies, highlights scalability and ethical challenges, and suggests future research directions for building adaptive and multimodal transportation systems.

Keywords: Contrastive Learning, Road Networks, Trajec- tory Data, Self-Supervised Learning, Graph Neural Networks, Transformers, Urban Mobility Analytics

I. Introduction

With the rapid growth of intelligent transportation systems (ITS) have increasingly relied on data-driven methods to analyze and optimize road traffic operations. The growing availability of high-fidelity GPS trajectories, digital maps, and road network data has accelerated research on spatial-temporal representation learning. Understanding these representations is vital for downstream applications such as traffic forecasting, route optimization, travel time estimation, and congestion detection.

Earlier deep learning methods generally handled road net- works and trajectory data as separate inputs, without capturing the relationship between them. Techniques such as Node2Vec and DeepWalk were commonly used to represent the structural aspects of road networks, while sequence-based models like RNNs and LSTMs focused on analyzing movement patterns in trajectory data. This independent treatment limits the model's ability to understand how spatial structure and temporal be- havior are interconnected. In practice, vehicle movement is in- fluenced by the road network, so separating these components reduces the effectiveness of representation learning [1], [11].

A significant challenge, therefore, is to establish a meaningful
link between the topology of road networks and the behavioral patterns observed in trajectories, particularly when labeled data is limited.

To address this, recent advances in self-supervised learning (SSL), particularly contrastive representation learning (CL), have emerged. CL enables models to learn discriminative embeddings from unlabeled data by maximizing the agree- ment (mutual information) between two augmented views of the same sample, thereby reducing dependence on human- annotated datasets [9], [10].

The objective of this review is threefold: 1) To systemati- cally analyze the core model architectures and learning strate- gies used in JCRL for road networks and trajectory data. 2) To provide a detailed breakdown of the contrastive objectives and data data augmentation methods required for achieving effective joint representations. 3) To highlight trends, limita- tions, and future directions, specifically focusing on how these models are assessed and the challenges in achieving real-world scalability.

II. Background Concepts

A. Graph Neural Networks (GNNs)

Road networks can be modeled as graph structures, where vertices represent intersections or road segments, while links indicate connections define their connections. Graph Neural Networks (GNNs), including Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs), extend convo- lution techniques to graph-based data. Well-known approaches such as the DCRNN and Spatio-Temporal Graph Convolu- tional Networks (STGCN) introduced effective ways to fuse spatial (GNN) and temporal (RNN/Conv) operations for traffic applications [11], [14]. GNNs are crucial for capturing the non-Euclidean, topological dependencies inherent in urban environments.

B. Transformers for Sequential Modeling

Trajectories are ordered sequences of locations or events over time. Transformer-based models are commonly applied in this domain due to their self-attention mechanism, which ef- fectively captures long-range dependencies within a sequence. Unlike traditional sequential models (RNNs/LSTMs), Trans- formers process entire sequences in parallel, enabling more efficient and comprehensive modeling of complex temporal movement patterns [10].

C. Contrastive Learning Principles

Contrastive learning aims to bring positive (similar) pairs closer in embedding space while separating negative (dissimi- lar) pairs apart. It is built on InfoNCE objective that maximizes mutual information between different augmented views of the same data [9], [13]. In graph and trajectory contexts, this helps align different spatial and temporal perspectives.

III. Literature Review

This review adopts a structured methodology to analyze existing research by classifying studies according to their representation focus (road networks, trajectories, or combined

approaches), learning strategies, and application areas. Such an organized evaluation enables the identification of major trends, key contributions, and existing research gaps in the field of road network and trajectory representation learning.I.

A. Foundational Spatio-Temporal Models

Wang et al. (2021) introduced ST-MetaNet,Yu et al. (2024) presented a self-supervised urban mobility learning framework using graph-based spatial reasoning. The model learns spa- tiotemporal embeddings without requiring annotations, cap- turing road structure, traffic flow, and movement patterns. It addresses complexity in large-scale mobility systems and per- forms strongly across multiple traffic forecasting bench- marks. integrating CNN, LSTM, and meta-learning for spatiotemporal prediction across diverse cities. CNN extracts spatial features, LSTM models temporal sequences, and meta-learning enables transferability across geographic regions. The model per- forms strongly in traffic forecasting, flow prediction, and other spatiotemporal tasks. ST-MetaNet is foundational for models requiring cross-city generalization.

Piaggesi and Panisson (2022) proposed HOSGNS, a dy- namic graph representation learning model for evolving spa- tiotemporal data. It extends skip-gram to higher-order se- quences and incorporates negative sampling for efficient train- ing. HOSGNS ef- fectively separates spatial and temporal de- pendencies in time-varying graphs. It out- performs predecessor models in dynamic network reconstruction and forecasting.

Li et al. (2022) introduced STGNN, a Spatio-Temporal Graph Neural Network that captures dynamic traffic conditions through graph-based modeling. It integrates spatial relation- ships from road graphs and temporal traffic patterns for mobil- ity prediction tasks. STGNN improves congestion estimation and traffic flow forecasting. It estab- lished benchmarks for dynamic traffic representation learning.

Yu et al. (2024) presented a self-supervised urban mobility learning framework using graph-based spatial reasoning. The model learns spatiotemporal embeddings without requiring an- notations, capturing road structure, traffic flow, and movement patterns. It addresses complexity in large-scale mobility sys- tems and performs strongly across multiple traffic forecasting benchmarks.

B. Road Network Focused Models

Ruan et al. (2020) introduced MapGAN, a GAN-based architecture that automatically generates accurate road network maps from noisy GPS trajectories. The model treats map generation as an image-to-image translation task. Its generated maps achieved high topological accuracy and strong F1 scores. MapGAN is widely used for automated map construction and road network reconstruction.

Sainju and Jiang (2020) developed a CNN–LSTM model to identify road safety features from street-view images. The CNN extracts visual features while the LSTM captures spatial continuity along road segments. The model outperformed existing methods in detecting barriers, signals, and road safety structures. It demonstrates the role of deep visual encoders in transportation safety analysis.

Chen et al. (2021) proposed Toast, which integrates a traffic- aware skip-gram with a Transformer module for robust road network embedding. Toast captures both structural

and seman- tic properties of road segments. The model achieved superior performance in transport planning and traffic prediction tasks. Its design makes it effective in handling incomplete or noisy road network data.

Chen et al. (2021b) extended Toast by introducing DyToast, which incorporates temporal dynamics using trigonometric functions and Transformer blocks. DyToast models periodic traffic variations and dynamic mobility conditions. It signifi- cantly improves road-speed inference and other time-sensitive tasks. DyToast is well known for its strong temporal modeling capability.

Pung et al. (2022) designed a graph simplification algo- rithm that reduces large-scale road networks while preserving topological integrity. The model removes redundant nodes and edges while maintaining shortest-path distributions. It achieved up to 90 percent reduction in network size, greatly improving computational efficiency for downstream GNN applications.

C. Trajectory Focused Models

Fu and Lee (2020) introduced Trembr, a model that inte- grates CNN-based graph modeling with hierarchical trajec- tory embedding through Traj2Vec and Road2Vec. The frame- work focuses on trajectory representation learning using road network constraints. Trembr improves trajectory similarity ranking and prediction accuracy compared to state-of-the-art models. Its ability to encode network-constrained mobility patterns makes it a foundational model for trajectory-centric learning.

Haydari et al. (2022) developed DPMM, a privacy- preserving map-matching framework for trajectory data. It applies Laplace and Exponential differential privacy mechanisms to protect origins, destinations, and complete paths. Despite privacy noise injection, the model maintains high map- matching accuracy. DPMM is essential for secure trajectory analytics where user identity protection is required.

D. Joint and Cross-Modal Contrastive Learning

Zhou et al. (2023) introduced Bidirectional Trajectory Con- trastive Learning (BTCL), a self-supervised model that learns trajectory embeddings using bidirectional contrastive objec- tives. The method generates multiple augmented trajectories and aligns them in the embedding space to preserve semantic

III. Literature Review

TABLE

Title	Architectu Core	Main Contribution
Yu et al. (2024): Self-Supervised Graph Representation Learning for Urban Mobility Forecasting	Graph Rep- resentation Learn- ing	Self-supervised graph learning for urban mo- bility forecasting, ad- dressing temporal dynamics.
Zhou et al. (BTCL) (2023): Bidirectional Trajectory Contrastive Learning for Unlabeled Trajectory Data	Transfor mer + GNN	Proposed Bidirectional Trajectory Contrastive Learning (**BTCL**).
Chang et al. (TrajCL) (2023): Contrastive Trajectory Similarity Learning with Dual-Feature Attention Encoder	Transform En- coder	erIntroduced **TrajCL** for trajectory similarity through contrastive objectives.
He et al. (2023): Hierarchi cal Contrastive Learning for Robust Trajectory	Hierarchic Con- trastive	aFl ocused on hierarchi- cal contrastive learn- ing for robust

Representation		trajectory representation.
Liu et al. (Geo-Graph-CL) (2023): Geo-Graph-CL: A Contrastive Learning Framework for Geo-Spatial and Graph Data	GNN + MLP	A Contrastive Learning Framework for Geo- Spatial and Graph Data.
Pung et al. (2022): A Road Network Simplification Algorithm Preserving Shortest Path Distribution	Algorithm /Graph Theory	Developed road network simplification algorithm preserving shortest paths.
Piaggesi & Panisson (HOSGNS) (2022): Time-Varying Graph Representation Learning with HOSGNS	Dynamic GNN	Learned time-varying graph embeddings for evolving traffic (**HOSGNS**).
Haydari et al. (2022): Differentially Private Map Matching Using Laplace and Exponential Mechanisms	GCN + Privacy Module	Introduced differentially private map-matching technique (**DPMM**).
Saki & Hagen (2022): A Practical Guide to an Open-Source Map-Matching Approach Using Valhalla	Algorithm /Map- Matching	Provided a practical guide to the open- source Valhalla map- matching approach.
Li et al. (2022): Graph-Based Dynamic Traffic Representation Learning via Spatio-Temporal GNN	Spatio-Temporal GNN	Graph-based dynamic traffic representation learning via Spatio- Temporal GNNs.
Chen et al. (Toast) (2021): Robust Road Network Representation Learning for Transport Planning	GCN + Trans- former	Robust representation learning under incom- plete road data (**Toast**).
Wang et al. (ST-MetaNet) (2021): ST-MetaNet: A Meta-Learning Approach for Spatio-Temporal Prediction	GNN + LSTM + Meta- Learning	Used meta-learning to improve generalization across different cities.
Chen et al. (DyToast) (2021): DyToast: Semantic-Enhanced Representation Learning for Temporal Road Networks	GCN + Recur- rent/Attent	Enhanced Toast for semantic-enhanced iolenarning for temporal road networks.
Fu & Lee (Trembr) (2020): Trembr: Exploring Road Networks with Trajectory-based Representation Learning	CNN + Graph Model	Introduced **Trembr** for exploring road networks using hierarchical embedding.
Ruan et al. (MapGAN) (2020): Learning to Generate Maps from Trajectories Using MapGAN	GCN + Seq2Seq	Generated digital maps from GPS trajectories (**MapGAN**).
Sainju & Jiang (2020): Mapping Road Safety Features from Streetview Imagery Using Deep Learning	CNN	Mapped road safety features from streetview imagery using Deep Learning.

similarity. BTCL does not rely on labeled data, making it highly effective for large-scale mobility datasets. The model achieved strong improvements in driving intention prediction and outperformed several state-of-the-art baselines. It also demonstrated high generalization across unseen geographic regions.

Chang et al. (2023) proposed TrajCL, a trajectory con- trastive learning framework that uses a dual-feature Trans- former encoder for embedding learning. The model captures both the structural and sequential properties of trajectories and applies contrastive loss to enhance representation robustness. TrajCL is computationally efficient due to the removal of recurrent structures like LSTMs. The framework achieved significant gains in trajectory similarity search, retrieval, and clustering. It is particularly suitable for large-scale applications requiring fast and accurate trajectory comparison.

Ile et al. (2023) developed a hierarchical contrastive learning framework for trajectory representation. The model orga- nizes features across multiple hierarchical levels and applies contrastive objectives at each level to improve robustness. This hierarchical

structuring reduces noise sensitivity and im- proves discrimination between various trajectory patterns. The method enhances prediction, clustering, and retrieval tasks. Its multi-level learning design supports complex mobility analysis where trajectories differ in scale and granularity.

Liu et al. (2024) introduced GeoGraphCL, a cross-modal contrastive learning framework combining GNNs and MLPs to jointly learn geospatial and graph structural features. The model aligns multimodal information using contrastive objec- tives, improving understanding of spatial–graph relationships. GeoGraphCL performs well in mobility prediction tasks across diverse geographic regions. The framework also improves generalization, making it suitable for large-scale spatial data analytics.

IV. Methodology

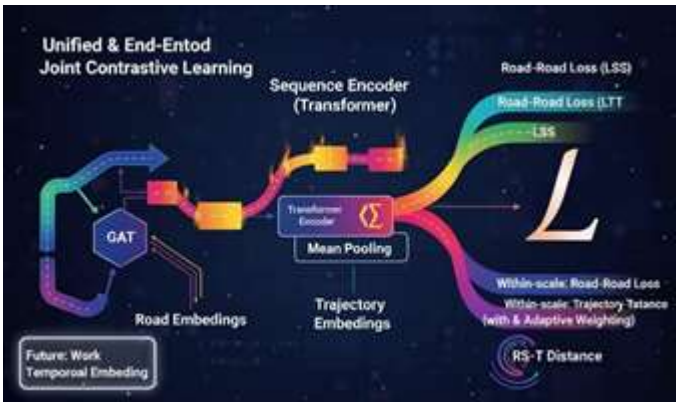

Fig. 1 . Architecture Diagram

A. Graph Encoder: Road Network Embedding

The spatial encoder, often a GAT, takes the initial road network feature matrix H(0) and adjacency matrix A as input. The l-th layer updates the node embeddings hˆ(l+1) as follows:

$$h_i^{(l+1)} = \sum_{j \in N(i)} \alpha_{ij}^{(l)} W^{(l)} h_j^{(l)}$$

similarity. BTCL does not rely on labeled data, making it highly effective for large-scale mobility datasets. The model

ˆ(l+1) (l) (l) (l)
i ij j
j∈N (i)

Where N(i) is the neighborhood of node i, αij is the attention coefficient determined by the relative importance of neighbor j, and W (l) are the layer-specific weights. The output is the final road embedding matrix Zr.

B. Sequence Encoder: Trajectory Embedding

The temporal encoder, typically a multi-head Transformer, processes the trajectory sequence T = p1, ..., pn . The core operation is the self-attention mechanism:

Attention(Q, K, V) = softmax QK⊤ V

Positional encodings are added to pi to maintain the temporal order. The output is a set of trajectory embeddings Zt.

C. Data Augmentation for Contrastive Learning

Effective contrastive learning heavily relies on generating meaningful positive pairs through data augmentation.

- **Road Augmentations (Spatial):** Techniques include node/edge dropping, feature masking, and subgraph sam- pling.
- **Trajectory Augmentations (Temporal):** Techniques in- clude subsequence cropping, subsampling, and time- based perturbation. Specialized hierarchical methods, such as those proposed by He et al. [?], ensure robustness across different levels of spatial and temporal resolution.

D. Joint Contrastive Objective

The framework maximizes mutual information across three distinct objective functions:
1) Within-Modal Losses:

- **Road-Road Loss (LSS):** Maximizes similarity between two modified versions of the same road segment or graph structure.
- **Trajectory-Trajectory Loss (LT T):** Aligns similar tra- jectory patterns.

2) Cross-Modal Loss: Road-Trajectory Loss (ST w): This is the critical inter-modal term, exemplified by works like Geo-Graph-CL [15]. It maximizes the similarity between a trajectory embedding zt and the road segment embedding zr that the trajectory traverses (positive pair), while contrasting it against irrelevant road segments (negative pairs).
The overall loss is the weighted sum of these objectives:
Ljoint = λ1LSS + λ2LTT + λ3STw

Where λ1, λ2, and λ3 are hyperparameters. All losses are based on the standard **InfoNCE loss**:

exp(sim(zi, zj)/τ)

Algorithm

1) Step 1: Graph Encoding Compute spatial embedding using GAT: h(l+1) = σ(Σ α(l)W (l)h(l)).

2) Step 2: Trajectory Encoding Encode sequential points via Transformer self-attention: Attn(Q, K, V) = softmax(√d)V . Obtain latent
embedding zk for trajectory tk.

3) Step 3: Positive and Negative Sampling Generate mul- tiple augmented views for both road and trajectory em- beddings to form intra- and inter-modal positive/negative pairs.
4) Step 4: Contrastive Loss Computation Compute LSS, LT T , and LST (or ST w). Total loss: Ltotal = λ1LSS + λ2LTT + λ3LST .
5) Step 5: Optimization Update model parameters θ: θ ← θ − η∇θLtotal, using Adam or similar optimizers.

6) Step 6: Joint Embedding Output Return final unified embeddings Zr and Zt for downstream tasks.

V. Conceptual Unification Framework

Many existing approaches handle road network represen- tation and trajectory modeling as separate tasks, or only combine them in a limited way. This fragmented treatment restricts the ability of models to capture the deep and complex interactions between the physical structure of transportation networks and the dynamic behavior of moving entities. In real- world scenarios, these two aspects are highly interdependent, as vehicle movement is strongly influenced by the layout and connectivity of the road system. Ignoring this relationship leads to incomplete or less accurate representations of urban mobility.

To address this limitation, it is important to develop a unified framework that can jointly learn from both road network data and trajectory information in a coordinated manner. Such a framework should enable seamless integration of spatial and temporal features, allowing the model to understand how infrastructure and movement patterns influence each other. By treating road structures and trajectories as complementary sources of information rather than independent components, the model can produce richer and more meaningful represen- tations. This integrated perspective is essential for accurately modeling real-world transportation systems and improving the performance of various downstream applications.

The unified framework consists of three major components. First, a graph encoder is applied to represent the spatial structure of road networks, including road segments, intersec- tions, and their connectivity patterns. This component captures topological relationships and structural semantics that shape how vehicles move across the city. Second, a sequence en- coder processes trajectories as ordered movement sequences, learning temporal dynamics, driving behaviors, and long-range movement patterns. Finally, a multi-scale contrastive learning mechanism aligns the representations learned from both encoders by comparing road–road, trajectory–trajectory, and road–trajectory relationships. This alignment ensures that the embeddings capture consistency across spatial and temporal dimensions.

Unlike traditional models that are trained for one specific task, this unified design emphasizes learning general-purpose mobility representations. These representations can be reused across various downstream applications, including travel time prediction, anomaly detection, congestion estimation, traffic state forecasting, and route recommendation. By combining spatial structure with temporal movement patterns, the frame- work promotes richer, more accurate, and more adaptable understanding of urban mobility.

VI. Challenges

Despite the significant progress enabled by contrastive learning in transportation analytics, several challenges still limit its widespread adoption and optimal performance. One major issue is data noise and privacy, as GPS trajectories often contain missing points, irregular sampling rates, and sensor inaccuracies that degrade model quality.

Additionally, trajectory data may expose sensitive personal information such as home–work locations, creating strong privacy concerns. Designing robust learning mechanisms that can handle noisy inputs while ensuring differential privacy remains a critical research need.

Another key challenge lies in scalability, particularly when dealing with large metropolitan datasets that may include millions of road segments and billions of trajectory points. Training contrastive learning models at such scales requires efficient sampling strategies, memory-optimized architectures, and distributed training pipelines. Without these improve- ments, computational costs become prohibitively high.

The selection of positive pairs and data augmentations also presents difficulty. In mobility data, defining “similar” or “related” samples is non-trivial; poor augmentation choices can lead to representation collapse or loss of important spa- tial–temporal semantics. Designing domain-aware augmenta- tions that preserve mobility characteristics is therefore an important direction for advancement.

Furthermore, the field currently lacks standardized evalu- ation metrics and benchmarks for self-supervised road and trajectory representation learning. This makes it challenging to compare models fairly, reproduce experiments, or assess generalization across different regions and datasets. Establish- ing unified datasets and evaluation protocols will significantly enhance research consistency.

In the future, research should concentrate on multimodal data integration by incorporating diverse inputs such as weather factors, major events, traffic disruptions, and social mobility patterns to better understand complex urban dy- namics. Online and continual learning approaches are also essential for real-time traffic systems, enabling models to adapt quickly to evolving mobility patterns. Lastly, the energy- efficient deployment of large models on edge devices and in smart vehicles will be crucial for enabling practical, low- latency intelligent transportation applications at scale.

VII. Conclusion

Joint Contrastive Representation Learning represents the cutting edge in urban mobility analytics by synergistically unifying spatial road network data and dynamic temporal trajectory data. The framework, leveraging dual encoders (GNNs and Transformers) and a multi-objective contrastive loss, successfully generates generalized and robust spatio- temporal embeddings. Our review highlights the critical role of data augmentation and the cross-modal loss term in enforcing consistency of meaning across both data streams. While JCRL models demonstrate superior performance across essential downstream tasks, future work must focus on addressing the challenges of dynamic environment modeling, large-scale computational efficiency, and ethical data handling to realize their full potential in real-time intelligent transportation systems.

References

1. H. Fu and K. Lee, "Trembr: Exploring Road Networks with Trajectory- based Representation Learning," IEEE ,2020.

2. X. Ruan, Y. Wu, and L. Zhao, "Learning to Generate Maps from Trajectories Using MapGAN," in IEEE ,2020.
3. P. Sainju and J. Jiang, "Mapping Road Safety Features from Streetview Imagery Using Deep Learning," IEEE ,2020.
4. T. Chen, Q. Yang, and Y. Li, "Robust Road Network Representation Learning for Transport Planning," ACM ,2021.
5. Y. Chen, G. Cong, and C. Long, “Robust Road Network Repre- sentation Learning: When Traffic Patterns Meet Traveling Semantics,” Proc.ACM 2021.
6. G. Piaggesi and A. Panisson, "Time-Varying Graph Representation Learning with HOSGNS," Applied Netw. Sci.,2022.
7. M. Haydari, S. Das, and A. Bhatnagar, "Differentially Private Map Matching Using Laplace and Exponential Mechanisms," IEEE ,2022.
8. R. Saki and T. Hagen, "A Practical Guide to an Open-Source Map- Matching Approach Using Valhalla," IEEE ,2022.
9. Z. Zhou, L. Yu, and J. Wang, "Bidirectional Trajectory Contrastive Learning for Unlabeled Trajectory Data," in IEEE ,2023.
10. S. Chang, X. Lu, and H. Liu, "Contrastive Trajectory Similarity Learning with Dual-Feature Attention Encoder," IEEE ,2023.
11. T. Li, Y. Gao, and F. Zhang, "Graph-Based Dynamic Traffic Repre- sentation Learning via Spatio-Temporal Graph Neural Networks," IEEE ,2022.
12. J. Yu, W. Zhang, and P. Liu, "Self-Supervised Graph Representation Learning for Urban Mobility Forecasting," Knowl.-Based Syst.,2024.
13. Z. Liu, H. Chen, J. Feng, Y. Liu, and Y. Zhang, "Geo-Graph-CL: A Contrastive Learning Framework for Geo-Spatial and Graph Data," IEEE, 2023.
14. S. Wang, H. Ma, Y. Li, and K. Chen, "ST-MetaNet: A Meta-Learning Approach for Spatio-Temporal Prediction," in Proc. AAAI Conf. Artif. Intell.,2021.
15. H. He, X. Zhang, and J. Chen, "Hierarchical Contrastive Learning for Robust Trajectory Representation," ACM Trans. Spatial Algorithms Syst., 2023.