Journal of Computational and Graphical Statistics

ISSN: 1061-8600 (Print) 1537-2715 (Online) Journal homepage: www.tandfonline.com/journals/ucgs20

Spatio-temporal Prediction of Fine-Grained Origin-Destination Matrices with Applications in Ridesharing

Run Yang, Runpeng Dai, Siran Gao, Xiaocheng Tang, Fan Zhou & Hongtu Zhu

To cite this article: Run Yang, Runpeng Dai, Siran Gao, Xiaocheng Tang, Fan Zhou & Hongtu Zhu (21 Jan 2026): Spatio-temporal Prediction of Fine-Grained Origin-Destination Matrices with Applications in Ridesharing, Journal of Computational and Graphical Statistics, DOI: 10.1080/10618600.2026.2618112

To link to this article: https://doi.org/10.1080/10618600.2026.2618112

View supplementary material

Accepted author version posted online: 21 Jan 2026.

Submit your article to this journal

Article views: 45

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ucgs20

Check for updates

Spatio-temporal Prediction of Fine-Grained Origin-Destination Matrices with Applications in Ridesharing

Run Yang[1], Runpeng Dai[2], Siran Gao[1], Xiaocheng Tang[3], Fan Zhou[1,*], and Hongtu Zhu[2,*]

[1]Department of Statistics and Management, Shanghai University of Finance and Economics

[2]Department of Biostatistics, University of North Carolina at Chapel Hill

[3]Personal Researcher, California, USA

*Correspondence to: Fan Zhou zhoufan@mail.shufe.edu.cn, Shanghai University of Finance and Economics, and Hongtu Zhu htzhu@email.unc.edu , University of North Carolina at Chapel Hill

Accepted Manuscript

Abstract

Accurate spatial-temporal prediction of network-based travelers' requests is crucial for the effective policy design of ridesharing platforms. Having knowledge of the total demand between various locations in the upcoming time slots enables platforms to proactively prepare adequate supplies, thereby increasing the likelihood of fulfilling travelers' requests and redistributing idle drivers to areas with high potential demand to optimize the global supply-demand equilibrium. This paper delves into the prediction of Origin-Destination (OD) demands at a fine-grained spatial level, especially when confronted with an expansive set of local regions. While this task holds immense practical value, it remains relatively unexplored within the research community. To fill this gap, we introduce a novel prediction model called OD-CED, which comprises an unsupervised space coarsening technique to alleviate data sparsity and an encoder-decoder architecture to capture both semantic and geographic dependencies. Through practical experimentation, OD-CED has demonstrated remarkable results. It achieved an impressive reduction of up to 45% reduction in root-mean-square error and 60% in weighted mean absolute percentage error over traditional statistical methods when dealing with OD matrices exhibiting a sparsity exceeding 90%.

Keywords: Encoder-Decoder, Fine-grained Spatial Level, Origin-Destination Prediction, Ride-sharing Platforms, Space Coarsening

1 Introduction

Spatial-temporal processes have gained prominence in contemporary statistics and data science, driven by its extensive applications across fields like climate modeling, traffic management, public health surveillance (Wikle & Zammit-Mangion 2023). For instance, dynamic Origin-Destination (OD) matrices are pivotal in quantifying the flow between various spatial entities over time. However, analyzing and predicting OD matrices at a fine-grained spatial scale involving numerous locations present significant challenges within the research community (Zhang et al. 2021, Wang et al. 2019, Ke et al. 2021). Ride-sourcing platforms such as Uber exemplify this complexity. When a city area is divided into N locations, the OD demand flows during each time interval across these locations generate an $N \times N$ matrix. This matrix illustrates the travel requests between any pair of spatial cells, offering a fine-grained view of traffic movements. Such detailed OD matrices are crucial for developing more effective transportation strategies, as they provide comprehensive micro-level traffic insights, enabling precise traffic condition forecasts. However, the task of predicting large-scale, fine-grained OD matrices introduces three key challenges:

Scalability. The OD matrix grows dramatically with the number of spatial divisions N. For instance, analyzing traffic data across hundreds of locations can yield an OD matrix with hundreds of thousands of flows. Addressing these large matrices requires scalable prediction models, a challenge that many advanced methods struggle with. As spatial complexity increases, scalability issues arise, including rising memory requirements and decreased computational efficiency.

Data Sparsity. Adopting a fine-grained spatial approach often leads to a decrease in average demands at both individual locations and across flows, thereby heightening the issue of data sparsity. In the ridesharing industry, segmented urban areas frequently exhibit near-zero customer requests between distant or unrelated locations for the majority of the time, resulting in skewed OD demand distributions. This accentuates the challenges posed by irregular traffic patterns and lessens the effectiveness of identifying and leveraging patterns or group structures within OD data.

Semantic and Geographical Dependencies. Both semantic and geographical dependencies play crucial roles in OD prediction. Higher semantic dependence often leads to larger travel demands. For instance, two locations such as a residential area and a business district may be semantically correlated, even if they are geographically distant. Conversely, geographically adjacent cells are more likely to share similar traffic patterns due to similarities in their points of interest (POIs), even if the semantic dependency is weak in this scenario. However many current prediction methods overlook these two dependencies, which can make their forecasts less accurate.

Accepted Manuscript

1.1 Related work

OD data has been extensively studied by statisticians (Medina et al. 2002, Hazelton 2008, Ma & Qian 2018), focusing on estimating OD matrices or understanding commuting and migration patterns. We categorize these methods based on the data used.

One category employs data aggregation methods, which combines diverse sources like Call Detail Records (CDR) (Calabrese et al. 2011), mobile cellular signaling data (Janzen et al. 2016), and check-in data (Fekih et al. 2021). Another category uses models like the gravity (Pourebrahim et al. 2018) and radiation (Liu et al. 2020) models to link OD matrices with factors like population, socioeconomic variables, land types, and distances.

Much research focuses on OD estimation, often using Bayesian modeling Vardi (1996). Other methods include Bayesian inference (Tebaldi & West 1998), information-theoretic approaches (Zhang et al. 2005), and feature reduction models (Lakhina et al. 2004), applied in fields like internet anomaly detection (Papagiannaki et al. 2003).

This paper addresses OD forecasting, predicting future OD matrices based on historical data. Traditional methods focus on OD estimation or matrix completion and are not suited for forecasting. Models like ARIMA and Support Vector Regression (SVR) are limited due to inefficiency and lack of spatial correlation capture. Advanced methodologies are needed for efficient predictions capturing spatial dependencies.

Existing computational OD forecasting methods generally fall into two distinct categories. The first category conceptualizes the OD matrix as a single image and utilizes Convolutional Neural Networks (CNNs) for capturing spatial-temporal closeness, as seen in references (He et al. 2016, Zhang et al. 2016, 2017, Ma et al. 2017, Liu et al. 2019). A notable advancement in this category is the integration of Recurrent Neural Networks (RNNs) with CNNs, enabling more effective capture of long-term spatial-temporal correlations. ConvLSTM (Shi et al. 2015), initially developed for weather forecasting, has notably achieved success in traffic flow predictions. Further improvements include the integration of non-local modules with ConvLSTM (Liu et al. 2019) and the introduction of a look-up convolution layer, termed LC, for enhanced learning of time-series traffic patterns (Lv et al. 2018).

The second category, exemplified by (Wang et al. 2019, Zhang et al. 2021), approaches the OD matrix as a fully connected weighted bidirectional graph, employing Graph Neural Networks (GNNs) to discern dynamic traffic patterns. Innovations in this category include combining Graph Convolutional Networks (GCNs) and CNNs to capture spatio-temporal interactions, as in STGCN (Lu et al. 2020), and the development of multi-GCN architectures (Chai et al. 2018, Geng et al. 2019, Lu et al. 2020). Techniques such as ASTGCN (Guo et al. 2019) and STGNN (Wang, Ma, Wang, Jin, Wang, Tang, Jia & Yu 2020) incorporate spatio-temporal attention for improved GCN embedding, while SFGNN (Song et al. 2020) forms a spatio-temporal fusion graph. Other notable approaches include the use of self-adaptive adjacency matrices (Wu et al. 2019), Adaptive Graph Convolutional Recurrent Networks (Bai et al. 2020), and the STSGCN (Li & Zhu 2021) for learning dynamic spatial-temporal correlations. However, a significant limitation of these methods, across both categories, is the dramatic increase in spatiotemporal complexity as N grows, posing scalability challenges in effectively handling large-scale OD matrices.

1.2 Contributions

To tackle the aforementioned challenges, we introduce a novel lightweight OD prediction model called the **Coarseing-E ncoder-D ecoder** network for fine-grained **O rigin-D estination** data (**OD-CED**). As shown in Fig 2, the OD-CED framework comprises two primary stages:

Accepted Manuscript

• **Preprocess Stage.** We transform *N* fine-grained cells into *M* coarse-grained super-cells to significantly reduce computational demands ($M \ll N$) while preserving the essential spatial-temporal dependencies. Through a strategic two-step Label Propagation (LP) process, cells exhibiting similar OD patterns are merged into larger clusters, thereby enhancing the framework's robustness against irregular traffic demands.

• **Learning Stage.** We utilize a hierarchical embedding architecture to capture semantic and geographical dependencies. An encoder learns the representations of super-cells through a multi-head self-attention network and the decoder predicts fine-grained cells by estimating the similarity between each fine-grained cell and its corresponding super-cells.

In particular, the new method tries to address three critical questions: (Q1) How can we reduce prediction errors for OD flows with typically non-zero values, given that the model often yields sparse predictions due to the prevalence of zero values across most spatio-temporal locations? (Q2) How can we accurately predict exceptional or extreme non-zero demands, which may arise when fine-grained spatial divisions result in less evident periodic patterns? (Q3) How can we utilize both semantic and geographic information to better understand traffic patterns of OD demands? These questions drive the methodological development outlined in Section 3.

The main contributions of this paper are as follows: (i) We are the first to address large-scale fine-grained OD prediction with extremely sparse data. (ii) We propose a novel method to merge fine-grained cells into super-cells, reducing OD matrix size while maintaining dependencies. (iii) We design an encoder-decoder model to capture global dependencies effectively. (iv) We develop a permutation-invariant module to learn super-cell level representations.

2 Fine-grained origin-destination data in ridesharing

In this section, we analyze two real-world industrial datasets collected from a ridesharing company to highlight the challenges posed by fine-grained OD flows. In both datasets, each day is divided into 24 one-hour, non-overlapping time intervals, and the city area is partitioned into *N* local regions. The datasets record the number of customer travel requests between each origin-destination (OD) pair within each time interval. Here, each OD flow represents the total demand between two spatial locations per hour. Below is a brief overview of the two datasets.

City-C: This anonymized dataset, released by the ridesharing company, contains historical ride-hailing orders from November 1 to November 30, 2016, in City C. The urban area is divided into 632 hexagonal cells, each covering approximately 1.2 square kilometers. The dataset records hourly ride-hailing orders between any two cells. The sparsity ratio of City-C is 99.15%.

City-S:. Collected internally from June 1 to November 30, 2021, this dataset spans a longer duration (180 days) compared to City-C (30 days). In City-S, each local region covers $2.1km^2$, with 638 segments in total. The dataset exhibits a slightly lower sparsity ratio of 98.02%, indicating higher overall travel demand than City-C.

Accepted Manuscript

These datasets feature fine-grained spatial resolution, unlike traditional OD data that aggregate flows over large, homogeneous city zones (e.g., residential, commercial, or industrial areas). Figure 1 (b) and (d) illustrate the differences between coarse- and fine-grained spatial divisions in City-S, with heatmaps of daily traffic outflow. Figure 1 (a) and (c) present the corresponding OD matrices, highlighting the increased sparsity at finer spatial scales.

Fine-grained partitioning significantly increases OD data sparsity by subdividing large areas into smaller cells. For example, Figure 1 (e) shows high traffic volumes between a residential area and industrial zones at a coarse level. However, when divided into finer cells (Figure 1 (f) and (g)), demand becomes fragmented, disrupting temporal periodicity and leading to sparse traffic patterns. This poses significant challenges for traditional OD prediction methods, which often rely on stable periodic trends.

The irregular traffic patterns in these datasets violate the assumptions of many conventional statistical and deep-learning approaches, making it difficult to apply standard prediction models effectively. Addressing the complexities of fine-grained OD flows requires novel methodologies capable of robust learning and prediction under high sparsity and irregular demand distributions.

3 Spatio-temporal prediction model

In this section, we present the detailed architecture of the OD-CED model. As shown in Figure 2, OD-CED comprises two main stages: the *preprocessing stage* and the *learning stage*. In the preprocessing stage, we propose a novel non-parametric down-sampling method to perform space coarsening, merging the original *N* fine-grained cells into *M* coarse-grained super-cells to mitigate sparsity issues. In the learning stage, we introduce a novel OD embedding module to capture both the in-flow and out-flow information of super-cells. We design a specialized cross-shaped receptive field to precisely quantify the semantic relationships among super-cells. The encoder takes the concatenation of the learned OD embedding and some external POI embedding as input to refine the representations of super-cells, which are fed into the decoder along with a trainable embedding table of cells.

3.1 Problem statement

We first introduce some important notations and formally define the prediction problem.

• ***Cells.*** We divide the entire urban area into *N* non-overlapping hexagonal local regions, referred to as "cells", denoted by $G = \{g_1, g_2, \ldots, g_N\}$. The coverage area of each cell is defined by its center's longitude, latitude, and cell radius.

• ***Super-cells.*** Super-cells are sub-regions formed by merging multiple cells together, where the resulting regions have arbitrary shapes and share no common cells. The set of super-cells obtained from *G* is denoted as $S = \{s_1, s_2, \ldots, s_M\}$, where $M << N$. We propose a novel space coarsening method to perform the merging of cells to form super-cells.

• ***OD Matrix.*** The cell-level OD Matrix at time slot *t* is defined as $X_t = (x_t(i, j)) \in \mathcal{R}^{N \times N}$, where $x_t(i, j) \in \mathbb{R}$ represents the total traffic demand from cell *i* to cell *j*. Similarly, the OD

Accepted Manuscript

Matrix at the super-cell level is defined as $X_t^s \in \mathcal{R}^{M\times M}$, representing the traffic demand among the *M* super-cells at time slot *t*. For convenience, we define the series of OD Matrices between time slot t_1 and t_2 as an OD tensor $X_{t_1,t_2} = [X_{t_1}, X_{t_1+1}, \ldots, X_{t_2}] \in \mathbb{R}^{N\times N\times(t_2-t_1)}$, and similarly, we let $X^s_{t_1,t_2} = [X^s_{t_1}, X^s_{t_1+1}, \ldots, X^s_{t_2}] \in \mathbb{R}^{M\times M\times(t_2-t_1)}$ be the corresponding super-cell level OD tensor.

• ***POI Matrix.*** We categorize points of interest (POIs) into *p* different categories and represent the POI information at the super-cell level using a binary matrix $P^s = (p^s(i,j)) \in \mathcal{R}^{M\times p}$. The matrix element $p^s(i,j)$ is equal to 1 if a POI of category *j* (such as a restaurant or a cinema) is present in super-cell *i*, and 0 otherwise.

• ***Sparsity rate of OD Matrix.*** The sparsity rate of an origin-destination (OD) tensor $X_{t_1,t_2} \in \mathcal{R}^{N\times N\times(t_2-t_1)}$ is the proportion of zero elements in the tensor, given by:

$$S = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{t_2-t_1}\mathbb{I}(x_{i,j,t}=0)}{N^2(t_2-t_1)}, \quad (1)$$

where $\mathbb{I}(x_{i,j,t}=0)$ is an indicator function that returns 1 if $x_{i,j,t}$ is zero, and 0 otherwise.

The OD prediction problem involves forecasting the future OD demands $X_{t+1,t+\tau}$ for the next τ time slots time slots based on past OD data $X_{t-K+1,t}$.

3.2 Zero-Inflated Negative Binomial (ZINB) Distribution

In this study, we formulate the OD prediction problem in this paper as a non-parametric estimation problem of zero-inflated negative binomial (ZINB) distribution, as inspired by Zhuang et al. (2022). As discussed in Section 2, Fine-grained OD data are often characterized by significant sparsity, manifesting as an overabundance of zero counts. The ZINB distribution effectively addresses zero inflation by integrating a new parameter with an NB distribution. This integration allows the model to account for the excess zeros while capturing the variability in the count data. Formally, we assume that for each *t* and all $t+1 \le s \le t+\tau$, the conditional distributions of $x_s(i,j)$'s given historical OD tensor $X_{t-K+1,t}$ are independent and each $x_s(i,j)$ follows a ZINB distribution

$$f_{\mathrm{ZINB}}\left(x_s(i,j)\middle| X_{t-K+1,t}\right) = \begin{cases} \pi_s(i,j) + \left(1-\pi_s(i,j)\right) f_{\mathrm{NB}}\left(0; n_s(i,j), p_s(i,j)\right), & x_s(i,j)=0, \\ \left(1-\pi_s(i,j)\right) f_{\mathrm{NB}}\left(x_s(i,j); n_s(i,j), p_s(i,j)\right), & x_s(i,j)>0. \end{cases} \quad (2)$$

Here, f_{NB} denotes the negative binomial distribution characterized by shape parameters $n_s(i,j)$ and $p_s(i,j)$, while $\pi_s(i,j)$ represents the probability of an excess zero occurring. The parameters $\pi_s(i,j)$, $n_s(i,j)$, and $p_s(i,j)$ are modeled as functions of the historical OD tensor $X_{t-K+1,t}$ and are estimated using neural networks. This parameterization enables the

Accepted Manuscript

model to learn complex, non-linear relationships within the data, thereby enhancing its capacity to handle the sparsity inherent in OD datasets.

3.3 Prepossessing

The preprocessing module is build upon the Label *propagation method* (Zhou et al. 2003) and comprises four main steps: labeling, construction of transition matrices, propagation, and cell merging. The motivation behind this module is rooted in the observation that only a few cells exhibit high traffic volumes for most of the time, acting as "dense cells". Meanwhile, the majority of cells have extremely sparse traffic demands, interacting with only a small fraction of these dense cells. Thus, each dense cell can be considered the center of a community with the "sparse cells" in the same community being the most related ones. After space coarsening, we group each dense cell with nearby sparse cells that share close traffic relationships, forming the upper-level super-cell.

• ***Labeling.*** Given the historical OD data from time $t=1$ to time $\mathcal{T}_0$, we calculate a flow statistics $f_i = \mathcal{T}_0^{-1}\sum_{t=1}^{\mathcal{T}_0}\left[\sum_{j=1}^{N} x_{i,j,t} + \sum_{k=1}^{N} x_{k,i,t}\right]$ for each cell g_i, which is the mean value of OD flows associated with g_i over the $\mathcal{T}_0$ time slots. We can obtain $F' = [f_{(1)}, f_{(2)}, \ldots, f_{(N)}]$ satisfying $f_{(1)} \le f_{(2)} \le \ldots \le f_{(N)}$ by re-ordering the *N* elements in $F = [f_1, f_2, \ldots, f_N]$. Let $g_{(i)}$ be the cell corresponding to $f_{(i)}$ for $i = 1, \ldots, N$. Based on F', the *N* cells can be divided into two parts: the first $N-M$ cells of F', which have near-zero volumes for all the in- and out-flows, form a sparse set $G_s = \{g_{(N-M)}, \ldots, g_{(N)}\}$, and the remaining *M* cells form a dense set $G_d = \{g_{(1)}, ..., g_{(M)}\}$, which includes all the dense cells.

We first assign a unique label to each of the *M* dense cells in G_d such that any two of them would not fall into the same community. The corresponding initialized labeling matrix of G_d is denoted by $Y_l = [y_1, y_2, \ldots, y_M] = \mathbf{I}_{M\times M} \in \mathcal{R}^{M\times M}$, where each y_i is an one-hot labelling vector for g_i with only the *i*-th element being 1. Since all the $N-M$ "sparse cells" in G_s are unassigned to any community at the beginning, we initialize their labeling matrix as a zero matrix $Y_u = \mathbf{0}_{(N-M)\times M} \in \mathcal{R}^{(N-M)\times M}$. Thus, the initial labeling matrix Y^0 for all the *N* cells is the concatenation of Y_l and Y_u, denoted by:

$$Y^0 = [Y_l^T \;\; Y_u^T]^T = [\mathbf{I}^T \;\; \mathbf{0}^T]^T \in \mathcal{R}^{N\times M}. \quad (3)$$

• ***Transition Matrices Construction.*** We introduce two transition matrices that represent the probability of both cells *j* and *i* belonging to the same community from two different perspectives. One is a semantic transition matrix $T^{sem} \in \mathcal{R}^{N\times N}$ for measuring the traffic connections among cells. Cells connected by large OD flows usually tend to share similar traffic patterns, and are more likely to be categorized into the same community.

Accepted Manuscript

The other is a geographic transition matrix T^{geo} for capturing the geographic adjacency. Based on the first law that everything is related to everything else but nearby things are more related than distant things (Tobler 1970, Zhu et al. 2018), geographically adjacent cells usually share certain kinds of common traffic patterns and it is reasonable to merge them together. Thus, from the geographic perspective, the probability of g_j belonging to the same community with g_i is defined as $T_{i,j}^{geo} = \mathbb{I}(dis(g_i, g_j) < l) / \{\sum_{j=1}^{m} \mathbb{I}(dis(g_i, g_j) < l)\}$, where *dis* denotes the geographic distance between the centers of two cells and *l* is a pre-defined threshold. In practice, we let *l* be slightly larger than twice the cell radius to make $T_{i,j}^{geo}$ between g_i and each of its six neighboring cells be 1/6 and others be 0.

This design ensures that each community contains only one dense cell. Dense cells, representing urban residential, work zones, and entertainment venues in practice, exhibit distinct traffic patterns, not only in volume but also in their phases, with observed periodicity. By avoiding amalgamating these cells into larger communities, which would consist of multiple dense cells, the design preserves the model's ability to discern nuanced traffic patterns within the OD matrix. This configuration is essential for accurately capturing the intricate dynamics of traffic flow within and between these communities.

• ***Propagation.*** Given the initial labeling matrix Y^0 and T^{sem} and T^{geo} defined above, we infer the labels of sparse cells through a five-step labeling propagation procedure as follows:

Step 1: Set $Y^{0,sem} = Y^0$ and $Y^{0,geo} = Y^0$. Step 2: Labelling Update: For the *n*-th iteration, $Y^{n,sem} = T^{sem} Y^{n-1}$ and $Y^{n,geo} = T^{geo} Y^{n-1}$. Step 3: Label Fusion: $Y^n = \alpha Y^{n,sem} + \beta Y^{n,geo}$ with $\alpha + \beta = 1$. Step 4: Reset the first *M* rows of Y^n to be an identity matrix. Step 5: Repeat Steps 2 to 4 until convergence to obtain the labelling matrix $\tilde{Y} = [\tilde{Y}_l^T \tilde{Y}_u^T]^T$. Step 6: Apply row-wise Argmax operator to $\tilde{Y}$ to get the final clustering results $\hat{Y}$.

Note that in Step 3, we set $\alpha = \beta = 0.5$ to ensure that both semantic and geographical transitions equally contribute to the labeling propagation, facilitating the convergence of the algorithm.

In Step 4, to prevent merging dense cells, we reset the first *M* rows of Y^n to an identity matrix at the end of each iteration. This ensures that each resulting community contains only one dense cell. In Step 6, we apply row-wise Argmax to $\tilde{Y}$ to obtain $\hat{Y} = (\hat{Y}_{i,j})$, where $\hat{Y}_{i,j} = I(\tilde{Y}_{i,j} = \max_{k \in \{1,\ldots,M\}} \tilde{Y}_{i,k})$. This assigns each sparse cell to the community that is most related to from both semantic and geographical perspectives.

• ***Cell Merging.*** In this step, all the *N* cells are divided into *M* communities and each community represents a super-cell. Based on $\hat{Y}$, we can finally build the coarsened OD tensor $X_{0,\mathcal{T}_0}^s = X_{0,\mathcal{T}_0} \times_1 \hat{Y} \times_2 \hat{Y} \in \mathcal{R}^{M \times M \times \mathcal{T}_0}$ of length $\mathcal{T}_0$, where $\times_i$ denote *i*-mode product of tensor $X_{0,\mathcal{T}_0}$ with matrix $\hat{Y}$ for $i \in \{1,2\}$ [1]. Thus, each flow in $X_{0,\mathcal{T}_0}^s$ is the summation of a

Accepted Manuscript

group of OD flows in $X_{0,\mathcal{T}_0}$, making $X^s_{0,\mathcal{T}_0}$ much smaller in terms of the size of tensor, yet much denser than $X_{0,\mathcal{T}_0}$.

Our space coarsening module offers two advantages over traditional downsampling methods. First, it retains the original traffic pattern by reducing dimensions at the cell level, unlike methods like mean or max pooling, which can merge unrelated OD flows and alter the data structure. Second, our method accounts for both geographic and semantic dependencies, whereas traditional approaches often ignore semantic relationships among OD flows.

For example, in Figure 3 (b), g_1 and g_5 are geographically distant cells, yet they share similar traffic patterns due to a common OD neighbor g_7. Conversely, g_1 and g_3 are close to each other, but their traffic patterns can differ significantly when there are few traffic demands between them. Figure 3 (a) illustrates that some conventional spatial clustering methods based solely on geographic relationships fail to merge g_1 with g_5, whereas our method successfully groups them together by considering their high semantic dependence.

3.4 OD Embedding

We present a novel OD embedding module illustrated in Figure 4(b) to learn feature vectors for each super-cell, capturing both traffic flow and Point of Interest (POI) information. For each super-cell s_i, we define one origin matrix $O_i = X^s_{t-K+1,t}(i,\cdot,\cdot) \in \mathcal{R}^{M\times K}$ and one destination matrix $D_i = X^s_{t-K+1,t}(\cdot,i,\cdot) \in \mathcal{R}^{M\times K}$, where O_i includes all the coarsened OD flows starting from s_i and D_i includes the ones ending at s_i. We multiply O_i and D_i by two weight matrices $W^o \in \mathcal{R}^{K\times d}$ and $W^d \in \mathcal{R}^{K\times d}$, respectively, to learn temporal representations, leading to two embedded matrices $\hat{O}_i = O_i W^o \in \mathcal{R}^{M\times d}$ and $\hat{D}_i = D_i W^d \in \mathcal{R}^{M\times d}$. Each row of $\hat{O}_i$ (or $\hat{D}_i$) is a *d*-dimensional feature vector, encoding the temporal information of the corresponding coarsened OD flows related to s_i.

Let $\hat{o}^i_r \in \mathcal{R}^d$ be the *r*-th row of $\hat{O}_i$, $\hat{d}^i_r \in \mathcal{R}^d$ be the *r*-th row of $\hat{D}_i$, and $Q = [q_1,\ldots,q_{N_q}]^\intercal \in \mathcal{R}^{N_q\times d}$ be a set of N_q randomly initialized learnable queries. We define the similarity score between $\hat{o}^i_r$ and q_j and that between $\hat{d}^i_r$ and q_j as

$$\alpha^{\hat{o},i}_{r,j} = \frac{\exp(\hat{o}^{i\intercal}_r \cdot q_j)}{\sum_{k=1}^{N_q}\exp(\hat{o}^{i\intercal}_r \cdot q_k)} \text{ and } \alpha^{\hat{d},i}_{r,j} = \frac{\exp(\hat{d}^{i\intercal}_r \cdot q_j)}{\sum_{k=1}^{N_q}\exp(\hat{d}^{i\intercal}_r \cdot q_k)}, \quad (4)$$

respectively. A larger similarity score represents higher importance of $\hat{o}^i_r$ or $\hat{d}^i_r$ to the super-cell s_i. The final traffic flow representation of s_i is defined as

$$e^s_{od,i} = \sum_{r=1}^{M}\sum_{j=1}^{N_q}\alpha^{\hat{o},i}_{r,j}\cdot\hat{o}^i_r + \sum_{r=1}^{M}\sum_{j=1}^{N_q}\alpha^{\hat{d},i}_{r,j}\cdot\hat{d}^i_r, \quad (5)$$

Accepted Manuscript

which takes all the coarsened OD flows related to s_i into consideration. By applying this aggregation process to all the *M* super-cells, we can finally obtain a traffic feature matrix $E_{od}^s = [e_{od,1}^s, e_{od,2}^s, \ldots, e_{od,M}^s]^T \in \mathcal{R}^{M \times d}$.

This design offers two main advantages. Firstly, the architecture entails a fixed number of learning parameters. However, employing a simple aggregation method like a weighted sum of all the related OD flows, such as $e_{od,i}^s = \sum_{r=1}^{M} w_r^o \hat{o}_r^i + \sum_{r=1}^{M} w_r^d \hat{d}_r^i$ requires 2*M* parameters. This leads to a linear increase in parameters with the number of cells, resulting in computational inefficiency as the dataset grows.

Secondly, the aggregation process is permutation invariant such that the learned representation of each super-cell depends solely on the traffic demands of all related OD flows and remains unchanged regardless of their order in the input matrices $\hat{d}_r$ and $\hat{o}_r$. Thus, re-ordering the *M* rows of the coarsened OD matrix does not change the value of the learned feature vector when w_r^o and w_r^d are provided. However, for the naive weighted sum, if we exchange any two rows of $\hat{O}_i$ and $\hat{d}_i$, then the output $e_{od,i}^s$ changes accordingly, which is unreasonable.

To incorporate POI information, we multiply the super-cell level POI matrix $P^s \in R^{M \times p}$ by a weight matrix $W_{poi} \in R^{p \times d}$ to get the POI representation $E_{poi}^s = P^s W_{poi} \in R^{M \times d}$, and the final OD embedding for each super-cell is computed as $E^s = E_{od}^s + E_{poi}^s \in R^{M \times d}$.

3.5 OD Encoder-Decoder

The OD Encoder-Decoder, depicted in Figure 4(a), serves as the core learning module of the proposed method. Utilizing the OD embedding E^s, the encoder discerns geographical dependencies among super-cells, and its output is subsequently fed into the decoder to derive representations of fine-grained cells.

OD Encoder. The core of the OD Encoder utilizes a multi-head self-attention (MHSA) architecture (Vaswani et al. 2017). Each head assesses spatial-temporal similarities among super-cells and refines their representations accordingly. By integrating information from diverse representation subspaces, the multi-head design enhances expressive capabilities, leveraging the low-rank and sparse nature of individual heads (Wang, Li, Khabsa, Fang & Ma 2020).

We assume that all the *h* heads share the same architecture, but allow them to have different parameters. For each head $i \in \{1, \ldots, h\}$, the input E^s is projected by three weight matrices $W_{Q,i}^{enc} \in \mathcal{R}^{d \times d_Q}$, $W_{K,i}^{enc} \in \mathcal{R}^{d \times d_K}$, and $W_{V,i}^{enc} \in \mathcal{R}^{d \times d_V}$, respectively, to get $Q_{i,E} = E^s W_{Q,i}^{enc}$, $K_{i,E} = E^s W_{K,i}^{enc}$, and $V_{i,E} = E^s W_{Q,i}^{enc}$, where $d_Q = d_V = d_K = \left\lfloor \frac{d}{h} \right\rfloor$. Then, the output of the *i*-th head H_i is given by $H_i^{enc} = softmax\left(\frac{Q_{i,E} K_{i,E}^T}{\sqrt{d}} \right) V_{i,E}$.

Accepted Manuscript

In this case, the *j*-th row of H_i^{enc} represents a weighted sum of the rows in $V_{i,E}$, where the weights correspond to the similarities between the *j*-th row of $Q_{i,E}$ and the rows of $K_{i,E}$. Subsequently, we multiply the concatenation of the *h* heads by a weight matrix $W_O^{enc} \in \mathcal{R}^{h\left\lfloor \frac{d}{h} \right\rfloor \times d}$ to obtain the output $MHSA(E^s) = [H_1^{enc}, \ldots, H_h^{enc}]W^O$ of this MHSA module.

Specifically, we apply layer normalization (LN) to the input E^s and utilize a residual block design to enhance optimization efficiency (Narang et al. 2021). Therefore, the final architecture of the MHSA module is as $E^{s'} = MHSA(LN(E^s)) + E^s$ and $E^{s'}$ is then fed into a feed-forward network (FFN) to get the final output $\hat{E}^s = FFN(LN(E^{s'})) + E^{s'}$. The FFN here is simply a two-layer fully connected network with ReLU. $\hat{E}^s = \left\{\hat{e}_1^s, \hat{e}_2^s, \ldots, \hat{e}_M^s\right\}$ here encodes pairwise semantic relationship between super-girds.

OD Decoder. The decoder utilizes a randomly initialized trainable embedding matrix $E^g \in \mathcal{R}^{N \times d}$ along with the output of the OD encoder $\hat{E}^s$ to learn the representations of fine-grained cells. Incorporating E^g enables the OD-CED model to capture the long-term characteristics of fine-grained cells through end-to-end training with the entire dataset. The decoder employs a Multi-Head-Cross-Attention (MHCA) module, which measures spatial-temporal similarities between super-cells and fine-grained cells, aggregating the embedding of related super-cells to obtain the representation of the target fine-grained cell.

For each head $i \in \{1, \ldots, h\}$, we project E^g with $W_{Q,i}^{dec} \in \mathcal{R}^{d \times d_Q}$ to obtain the embedding $Q_{i,D} = E^g W_{Q,i}^{dec}$ of fine-grained cells. Similarly, we project $\hat{E}^s$ with $W_{V,i}^{dec} \in \mathcal{R}^{d \times d_V}$ and $W_{K,i}^{dec} \in \mathcal{R}^{d \times d_K}$ to obtain the embeddings $V_{i,D} = \hat{E}^s W_{V,i}^{dec}$ and $K_{i,D} = \hat{E}^s W_{K,i}^{dec}$ of super-cells, where $d_Q = d_V = d_K = \left\lfloor \frac{d}{h} \right\rfloor$. We then re-weight $V_{i,D}$ using the similarity between $K_{i,D}$ and $Q_{i,D}$ to derive the output of the *i*-th head $H_i^{dec} = \left[\hat{Y} \otimes softmax\left(\frac{Q_{i,D}K_{i,D}^T}{\sqrt{d}}\right)\right]V_{i,D}$.

$softmax(Q_{i,D}K_{i,D}^T / \sqrt{d}) \in \mathcal{R}^{N \times M}$ computes the similarity between the features of the *N* fine-grained cells and those of the *M* super-cells. The $\hat{Y} \in \{0,1\}^{N \times M}$ from the Propagation step indicates whether a fine-grained cell belongs to the same community as a super-cell, and its dot product with the softmax term helps to downplay weak connections among cells. Consequently, the representations of each fine-grained cell are influenced by its neighbors within the same community.

We then multiply the concatenation of the *h* heads with a weight matrix $W_O^{dec} \in \mathcal{R}^{h\left\lfloor \frac{d}{h} \right\rfloor \times d}$ to get the output $MHCA(E^g, \hat{E}^s) = [H_1^{dec}, \ldots, H_h^{dec}]W_O^{dec}$.

The residual mechanism, along with LN is applied, and the final representation $\hat{E}^g$ of each fine-grained cell *g* is obtained as follows,

Accepted Manuscript

$$E^{g'} = MHCA(LN(E^g), LN(\hat{E}^s)) + E^g \quad \text{and} \quad \hat{E}^g = FFN(LN(E^{g'})) + LN(E^{g'}). \quad (6)$$

The complexity of the OD Decoder is $\Omega(N \times M) \approx \Omega(N)$, as M is much smaller than N in practice and can be ignored. This makes our OD-CED model more scalable and computationally efficient compared to most existing OD prediction methods. Moreover, by introducing the cell embedding table E^g, we eliminate direct computations at the original fine-grained level, further reducing the computational complexity.

3.6 Optimization and Prediction

Using the spatial-temporal representations of fine-grained cells, denoted as $\hat{E}^g \in \mathcal{R}^{N \times d}$, and super-cells, represented as $\hat{E}^s \in \mathcal{R}^{M \times d}$ as input, we employ a neural network with linear transformations and 1×1 convolutions to predict $\{n_s(i,j), p_s(i,j), \pi_s(i,j); t+1 \le s \le t+\tau, 1 \le i, j \le N\}$ of the ZINB distribution as defined in Section 3.2. The final optimization of the full OD-CED model parameterized with θ is to minimize the negative log-likelihood $-\sum_{t=1}^{T} \log(L(X_{t+1,t+\tau} \mid X_{t-K+1,t}; \theta))$, where $L(X_{t+1,t+\tau} \mid X_{t-K+1,t})$ is the conditional likelihood function of $X_{t+1,t+\tau}$ given $X_{t-K+1,t}$ defined as follows,

$$L(X_{t+1,t+\tau} \mid X_{t-K+1,t}) = \prod_{s=1}^{\tau} \prod_{i=1}^{N} \prod_{j=1}^{N} f_{ZINB}(x_s(i,j); n_s(i,j), p_s(i,j), \pi_s(i,j)). \quad (7)$$

Upon obtaining $\hat{\theta}$, we naturally obtain estimators for $\hat{n}_s(i,j)$, $\hat{p}_s(i,j)$, and $\hat{\pi}_s(i,j)$. Subsequently, we utilize these estimators to predict the future OD tensor $X_{t'+1,t'+\tau} \in \mathcal{R}^{N \times N \times \tau}$ based on $X_{t'-K,t'}$, leveraging the expectation of the conditional mean as follows,

$$\hat{x}_s(i,j) = \left(\frac{1}{\hat{p}_s(i,j)} - 1\right)\hat{n}_s(i,j) \text{ for } t'+1 \le s \le t'+\tau \text{ and } 1 \le i, j \le N. \quad (8)$$

4 Real Analysis

In this section, we conduct comprehensive experiments on two fine-grained OD datasets City-C and City-S, as described in Section 2, to evaluate the proposed method. To better illustrate the advantages of our approach in addressing the challenges mentioned in the introduction, we compare OD-CED with a diverse set of baseline methods, including two traditional statistical approaches—Historical Average (HA) and Linear Regression—four convolution-based deep learning models—CSTN (Liu et al. 2019) and MRSTN (Noursalehi et al. 2021)—and several representative GCN-based spatio-temporal graph models, such as GEML (Wang et al. 2019), STGCN (Song et al. 2020), ASTGCN (Guo et al. 2019), STSGCN (Li & Zhu 2021), STZINB-GNN, and STTD.

All models use the previous $k = 6$ time intervals to predict the next one. For deep learning methods, we employ the Adam optimizer with an initial learning rate of 0.004 (halved every

Accepted Manuscript

50 epochs), a batch size of 32, and a maximum of 100 epochs. The embedding dimension and number of aggregation queries are set to $d = 64$ and $Q_n = 32$, respectively. We partition each city into $M = 60$ super-cells—roughly one-tenth of the total number of fine-grained cells. This proportion was not the result of extensive tuning; rather, it is an empirical, balanced choice that preserves dense flows while keeping computation and memory within budget. The embedding dimension in Section 3.4 was set to $d = 64$, and the number of aggregation queries to $Q_n = 32$. The initial learning rate was 0.004, halved every 50 epochs, with a fixed batch size of 32.

Model performance is evaluated using three standard metrics: Root Mean Square Error (RMSE), Weighted Mean Absolute Percentage Error (wMAPE), and Common Part of Commuters (CPC). Following previous works (Liu et al. 2019, Yao et al. 2018), we compute these metrics only over non-zero OD flows, as zero-demand entries carry no practical meaning in real mobility systems. We adhere to the approach outlined in prior studies (Liu et al. 2019, Yao et al. 2018) for computing the three metrics using non-zero OD demands. This choice is motivated by the fact that flows with no traveling records hold no significance in the ridesharing industry.

Because GCN-based methods explicitly construct localized spatio-temporal graphs where each node corresponds to an O–D pair, their memory complexity grows quadratically with the number of cells. For our fine-grained datasets, this results in an adjacency matrix of size $679^2 \times 679^2$, containing approximately 1.76×10^{11} entries—requiring over 350 GB of GPU memory even under fp16 precision. Therefore, these GCN-based models can only be evaluated on sampled subsets of the data, while all other methods, including OD-CED, are trained and tested on the full datasets. The results on the full datasets are summarized in Table 1, and those for GCN-based approaches on sampled subsets are presented in Table 2.

Table 1 presents the prediction results of each compared method. OD-CED outperforms all other methods in both datasets, with more significant improvement observed in City-S due to its larger data size. The superior performance of OD-CED can be attributed to two key factors. First, the space coarsening module effectively reduces computation costs and addresses data sparsity by grouping fine-grained cells into super-cells. Second, the OD Embedding module and the Encoder-Decoder capture both pairwise semantic relationships between regions and global geographic patterns. HA consistently performs poorly as it ignores data variations. Although OLSR and LASSO show slight improvements over HA, they fail to effectively capture the spatial characteristics of OD data, which are crucial for accurate OD prediction (Wang et al. 2019). CSTN and MRSTN show a significant increase in both wMAPEs and RMSEs, leveraging CNNs to extract spatial features. However, the improvement in the sparser dataset City-C is much less significant compared to City-S. This suggests that traditional CNNs are not well-suited for OD data, as there are no meaningful relationships between spatially nearby OD flows in an OD matrix.

Table 2 further compares OD-CED with representative GCN-based models on the data subsets where those methods can be executed feasibly. OD-CED consistently outperforms all graph-based baselines across both datasets. Compared with STSGCN and ASTGCN, OD-CED achieves substantial improvements in all three metrics, confirming its ability to capture temporal dependencies and O–D structural interactions without explicit graph construction. While GEML attains the second-best performance, its advantage arises mainly from its hybrid architecture combining GNNs and skip-LSTM layers. Despite using over 30 times

Accepted Manuscript

more parameters, GEML's wMAPE remains 29% higher than that of OD-CED, highlighting the superior efficiency of our model. Recent variants such as STZINB-GNN and STTD, which attempt to model zero-inflated distributions or temporal dynamics, also lag behind OD-CED due to limited scalability and incomplete modeling of spatial dependencies. Overall, these findings demonstrate that OD-CED's unified encoder–decoder design effectively integrates spatial semantics and temporal flow dynamics, achieving both higher accuracy and stronger generalization across cities of different scales.

To further illustrate why OD-CED outperforms GEML in practice, we conduct a minor case study comparing the prediction results of OD-CED and GEML for two randomly selected days in City-C and City-S. Figures 5(a) and 5(b) depict the ground truths in the first column, followed by the predictions of OD-CED and GEML in the subsequent columns. It's observed that OD-CED exhibits greater sensitivity to extremely large OD demands, leading to an overall more accurate prediction outcome compared to GEML.

To highlight the training efficiency of OD-CED, we compare the training time of each method against its model size. Table 3 presents the computation time of each method for one training epoch on a V100 GPU. OD-CED is significantly more efficient than the other four methods due to its lightweight design. Even the second-best approach, GEML, requires roughly twice as much training time. Given the necessity for real-time model updates in dynamic platforms, the proposed OD-CED stands out as a practical choice.

To further validate the robustness of OD-CED, we perform a sensitivity analysis on the number of super-cells M, which controls the granularity of spatial coarsening. The corresponding results and discussions are presented in the supplementary material (see *"Sensitivity Analysis on the Number of Super-cells M"*).

5 Conclusion

In this paper, we propose a novel prediction model OD-CED for large-scale fine-grained OD data, while addressing three main challenges. We propose a novel space coarsening module to group small fine-grained cells together to highly increase the computation efficiency and preserve the data structure. The OD Embedding module captures semantic and geographical dependencies in a global way based on a well-designed permutation invariant operator. At last, a novel Encoder-Decoder is introduced to predict the fine-grained OD matrices by utilizing the spatial-temporal information obtained in the coarsened space. Some empirical results show that OD-CED can exhibit great improvement over existing methods, especially when the OD data is extremely sparse.

Accepted Manuscript

SUPPLEMENTARY MATERIAL

The online supplementary materials contain the following components:

• **Supplementary.pdf (PDF File) :** Additional definitions and extended experimental results.

• **Code (ZIP files):** python scripts for model training, prediction, evalutation.

• **Data (folder):** Simulated origin–destination matrices generated for model validation and reproducibility.

• **README (TXT):** Documentation describing the folder structure, environment, and usage instructions.

All materials are available at the JCGS supplementary materials webpage accompanying this article.

Notes

[1] The $n-$mode (matrix) product of a tensor $\mathcal{X} \in \mathcal{R}^{I_1 \times I_2 \times \ldots \times I_N}$ with a matrix $U \in \mathcal{R}^{J \times I_n}$ is denoted by $\mathcal{X} \times_n U$ and is of the size $I_1 \times \ldots \times I_{n-1} \times J \times I_{n+1} \times \ldots \times I_N$. Elementwisely, we have $(\mathcal{X} \times_n U)_{i_1 \ldots i_{n-1} j i_{n+1} \ldots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \ldots i_N} u_{i_n j}$.

Accepted Manuscript

References

Bai, L., Yao, L., Li, C., Wang, X. & Wang, C. (2020), 'Adaptive graph convolutional recurrent network for traffic forecasting', *Advances in neural information processing systems* **33**, 17804–17815.

Calabrese, F., Di Lorenzo, G., Liu, L. & Ratti, C. (2011), 'Estimating origin-destination flows using mobile phone location data', *IEEE Pervasive Computing* **4**(10), 36–44.

Chai, D., Wang, L. & Yang, Q. (2018), Bike flow prediction with multi-graph convolutional networks, *in* 'Proceedings of the 26th ACM SIGSPATIAL international conference on advances in geographic information systems', pp. 397–400.

Fekih, M., Bellemans, T., Smoreda, Z., Bonnel, P., Furno, A. & Galland, S. (2021), 'A data-driven approach for origin–destination matrix construction from cellular network signalling data: a case study of lyon region (france)', *Transportation* **48**, 1671–1702.

Geng, X., Li, Y., Wang, L., Zhang, L., Yang, Q., Ye, J. & Liu, Y. (2019), Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting, *in* 'Proceedings of the AAAI conference on artificial intelligence', Vol. 33, pp. 3656–3663.

Guo, S., Lin, Y., Feng, N., Song, C. & Wan, H. (2019), Attention based spatial-temporal graph convolutional networks for traffic flow forecasting, *in* 'Proceedings of the AAAI conference on artificial intelligence', Vol. 33, pp. 922–929.

Hazelton, M. L. (2008), 'Statistical inference for time varying origin–destination matrices', *Transportation Research Part B: Methodological* **42**(6), 542–552.

He, K., Zhang, X., Ren, S. & Sun, J. (2016), Deep residual learning for image recognition, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 770–778.

Janzen, M., Vanhoof, M., Axhausen, K. W. & Smoreda, Z. (2016), Estimating long-distance travel demand with mobile phone billing data, *in* '16th Swiss Transport Research Conference (STRC 2016)', Swiss Transport Research Conference (STRC).

Ke, J., Qin, X., Yang, H., Zheng, Z., Zhu, Z. & Ye, J. (2021), 'Predicting origin-destination ride-sourcing demand with a spatio-temporal encoder-decoder residual multi-graph convolutional network', *Transportation Research Part C: Emerging Technologies* **122**, 102858.

Lakhina, A., Papagiannaki, K., Crovella, M., Diot, C., Kolaczyk, E. D. & Taft, N. (2004), Structural analysis of network traffic flows, *in* 'Proceedings of the joint international conference on Measurement and modeling of computer systems', pp. 61–72.

Li, M. & Zhu, Z. (2021), Spatial-temporal fusion graph neural networks for traffic flow forecasting, *in* 'Proceedings of the AAAI conference on artificial intelligence', Vol. 35, pp. 4189–4196.

Accepted Manuscript

Liu, L., Qiu, Z., Li, G., Wang, Q., Ouyang, W. & Lin, L. (2019), 'Contextualized spatial–temporal network for taxi origin-destination demand prediction', *IEEE Transactions on Intelligent Transportation Systems* **20**(10), 3875–3887.

Liu, Y., Fang, F. & Jing, Y. (2020), 'How urban land use influences commuting flows in wuhan, central china: A mobile phone signaling data perspective', *Sustainable Cities and Society* **53**, 101914.

Lu, B., Gan, X., Jin, H., Fu, L. & Zhang, H. (2020), Spatiotemporal adaptive gated graph convolution network for urban traffic flow forecasting, *in* 'Proceedings of the 29th ACM International conference on information & knowledge management', pp. 1025–1034.

Lv, Z., Xu, J., Zheng, K., Yin, H., Zhao, P. & Zhou, X. (2018), Lc-rnn: a deep learning model for traffic speed prediction, *in* 'Proceedings of the 27th International Joint Conference on Artificial Intelligence', pp. 3470–3476.

Ma, W. & Qian, Z. S. (2018), 'Statistical inference of probabilistic origin-destination demand using day-to-day traffic data', *Transportation Research Part C: Emerging Technologies* **88**, 227–256.

Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y. & Wang, Y. (2017), 'Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction', *Sensors* **17**(4), 818.

Medina, A., Taft, N., Salamatian, K., Bhattacharyya, S. & Diot, C. (2002), 'Traffic matrix estimation: Existing techniques and new directions', *ACM SIGCOMM Computer Communication Review* **32**(4), 161–174.

Narang, S., Chung, H. W., Tay, Y., Fedus, L., Févry, T., Matena, M., Malkan, K., Fiedel, N., Shazeer, N., Lan, Z. et al. (2021), Do transformer modifications transfer across implementations and applications?, *in* 'Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing', pp. 5758–5773.

Noursalehi, P., Koutsopoulos, H. N. & Zhao, J. (2021), 'Dynamic origin-destination prediction in urban rail systems: A multi-resolution spatio-temporal deep learning approach', *IEEE Transactions on Intelligent Transportation Systems* **23**(6), 5106–5115.

Papagiannaki, K., Taft, N., Zhang, Z.-L. & Diot, C. (2003), Long-term forecasting of internet backbone traffic: Observations and initial models, *in* 'IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428)', Vol. 2, IEEE, pp. 1178–1188.

Pourebrahim, N., Sultana, S., Thill, J.-C. & Mohanty, S. (2018), Enhancing trip distribution prediction with twitter data: comparison of neural network and gravity models, *in* 'Proceedings of the 2nd acm sigspatial international workshop on ai for geographic knowledge discovery', pp. 5–8.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k. & Woo, W.-c. (2015), Convolutional lstm network: a machine learning approach for precipitation nowcasting, *in*

Accepted Manuscript

‘Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 1’, pp. 802–810.

Song, C., Lin, Y., Guo, S. & Wan, H. (2020), Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting, *in* ‘Proceedings of the AAAI conference on artificial intelligence’, Vol. 34, pp. 914–921.

Tebaldi, C. & West, M. (1998), ‘Bayesian inference on network traffic using link count data’, *Journal of the American Statistical Association* **93**(442), 557–573.

Tobler, W. R. (1970), ‘A computer movie simulating urban growth in the detroit region’, *Economic geography* **46**(sup1), 234–240.

Vardi, Y. (1996), ‘Network tomography: Estimating source-destination traffic intensities from link data’, *Journal of the American statistical association* **91**(433), 365–377.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017), ‘Attention is all you need’, *Advances in neural information processing systems* **30**, 6000–6010.

Wang, S., Li, B. Z., Khabsa, M., Fang, H. & Ma, H. (2020), ‘Linformer: Self-attention with linear complexity’, *arXiv preprint arXiv:2006.04768* .

Wang, X., Ma, Y., Wang, Y., Jin, W., Wang, X., Tang, J., Jia, C. & Yu, J. (2020), Traffic flow prediction via spatial temporal graph neural network, *in* ‘Proceedings of the web conference 2020’, pp. 1082–1092.

Wang, Y., Yin, H., Chen, H., Wo, T., Xu, J. & Zheng, K. (2019), Origin-destination matrix prediction via graph convolution: a new perspective of passenger demand modeling, *in* ‘Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining’, pp. 1227–1235.

Wikle, C. K. & Zammit-Mangion, A. (2023), ‘Statistical deep learning for spatial and spatiotemporal data’, *Annual Review of Statistics and Its Application* **10**, 247–270.

Wu, Z., Pan, S., Long, G., Jiang, J. & Zhang, C. (2019), ‘Graph wavenet for deep spatial-temporal graph modeling’, *arXiv preprint arXiv:1906.00121* .

Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., Gong, P., Ye, J. & Li, Z. (2018), Deep multi-view spatial-temporal network for taxi demand prediction, *in* ‘Proceedings of the AAAI conference on artificial intelligence’, Vol. 32, pp. 2588–2595.

Zhang, D., Xiao, F., Shen, M. & Zhong, S. (2021), ‘Dneat: A novel dynamic node-edge attention network for origin-destination demand prediction’, *Transportation Research Part C: Emerging Technologies* **122**, 102851.

Zhang, J., Zheng, Y. & Qi, D. (2017), Deep spatio-temporal residual networks for citywide crowd flows prediction, *in* ‘Proceedings of the AAAI conference on artificial intelligence’, Vol. 31, pp. 1655–1661.

Accepted Manuscript

Zhang, J., Zheng, Y., Qi, D., Li, R. & Yi, X. (2016), Dnn-based prediction model for spatio-temporal data, *in* 'Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems', pp. 1–4.

Zhang, Y., Roughan, M., Lund, C. & Donoho, D. L. (2005), 'Estimating point-to-point and point-to-multipoint traffic matrices: An information-theoretic approach', *IEEE/ACM Transactions on networking* **13**(5), 947–960.

Zhou, D., Bousquet, O., Lal, T., Weston, J. & Schölkopf, B. (2003), 'Learning with local and global consistency', *Advances in neural information processing systems* **16**.

Zhu, A.-X., Lu, G., Liu, J., Qin, C.-Z. & Zhou, C. (2018), 'Spatial prediction based on third law of geography', *Annals of GIS* **24**(4), 225–240.

Zhuang, D., Wang, S., Koutsopoulos, H. & Zhao, J. (2022), Uncertainty quantification of sparse travel demand prediction with spatial-temporal graph neural networks, *in* 'Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining', pp. 4639–4647.

Accepted Manuscript

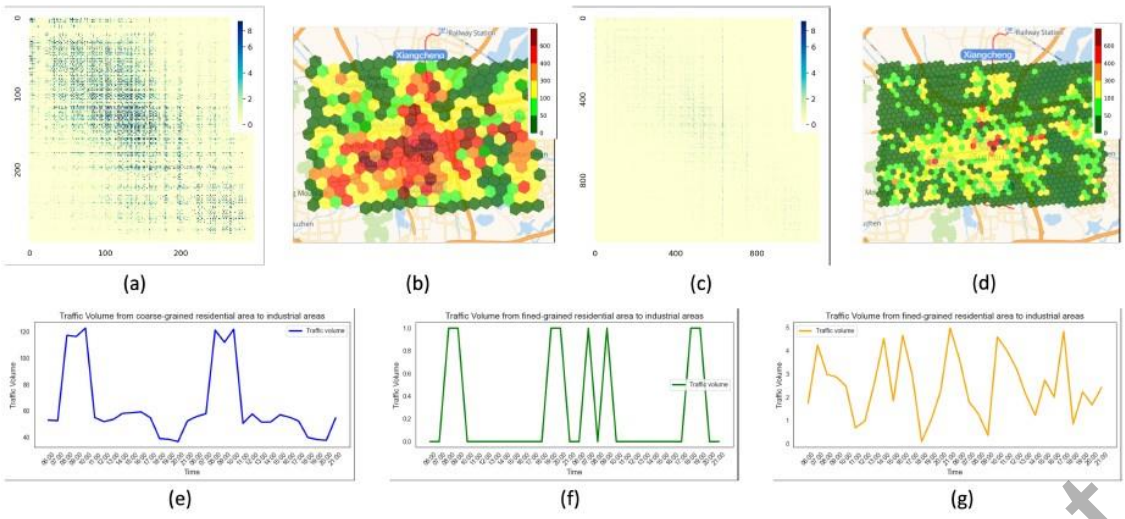

Figure 1: (a) OD matrix for 297 selected cells in City-S; (b) outflow counts from these cells in (a) over a single day; (c) OD matrix for 1111 selected cells in City-S; (d) outflow counts from these cells in (c) over a single day; (e) Temporal trends of an OD flow between residential and industrial areas at a coarse-grained level, compared to fine-grained spatial divisions in (f) and (g).

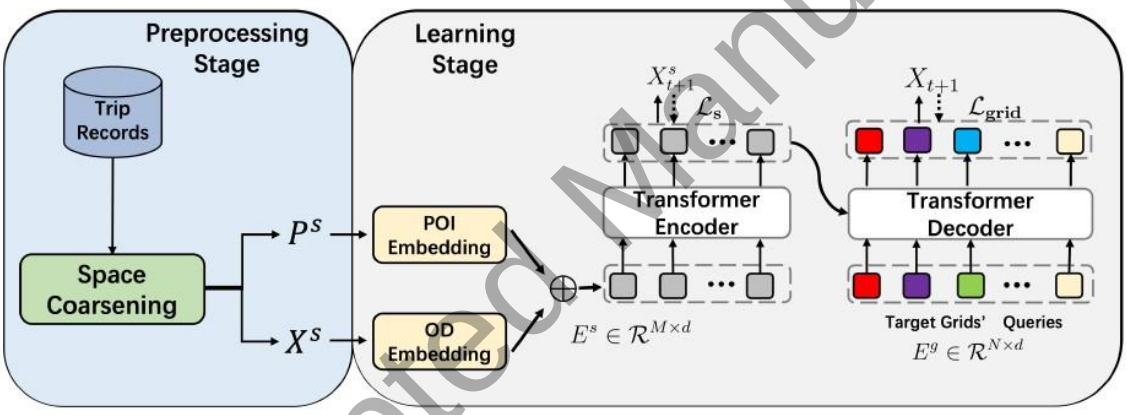

Figure 2: The architecture of OD-CED, consists of a *Preprocessing Stage* and a *Learning Stage*. In the Preprocessing Stage, a space coarsening procedure is applied to adaptively merge fine-grained cells into coarse-grained super-cells. In the Learning Stage, an encoder-decoder network is employed for coarse-grained encoding and fine-grained decoding.

Accepted Manuscript

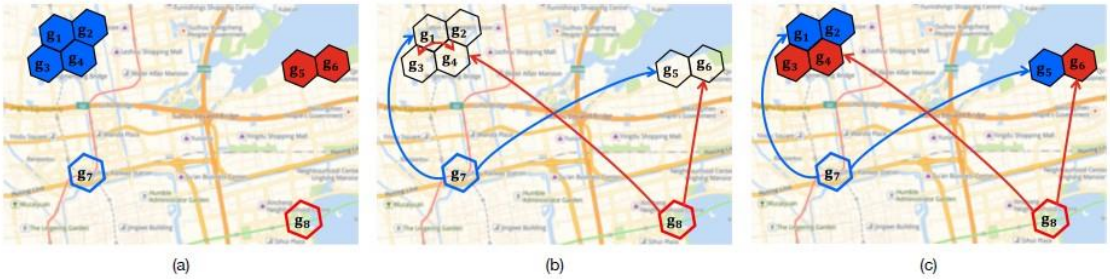

Figure 3: (a) Cell merging based on geographical closeness. Cells clustered together are assigned the same color; (b) Directed OD flows between cells; and (c) Clustering result by our method which jointly considers geographic and semantic dependence.

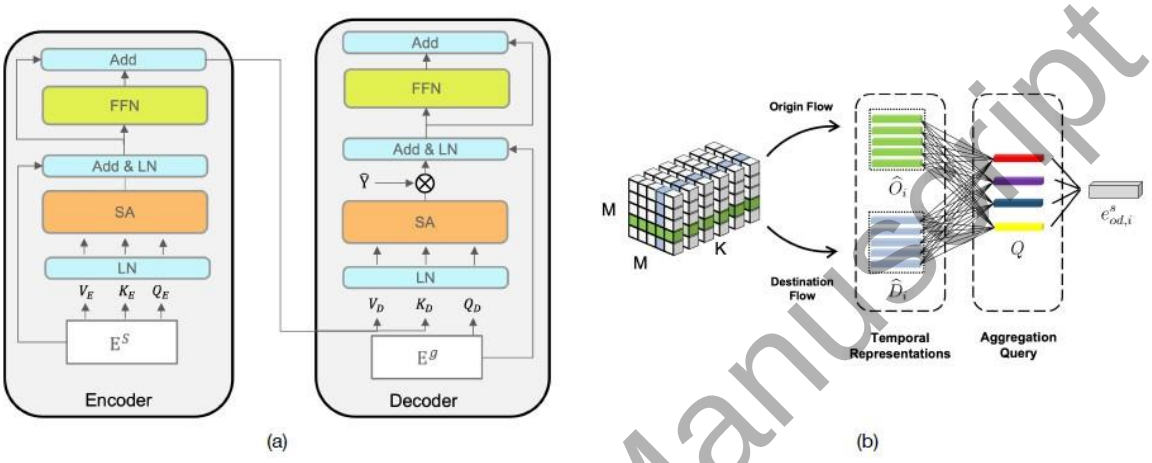

Figure 4: (a) The main architecture of the OD Encoder-Decoder module and (b) The architecture of the OD Embedding module.

Accepted Manuscript

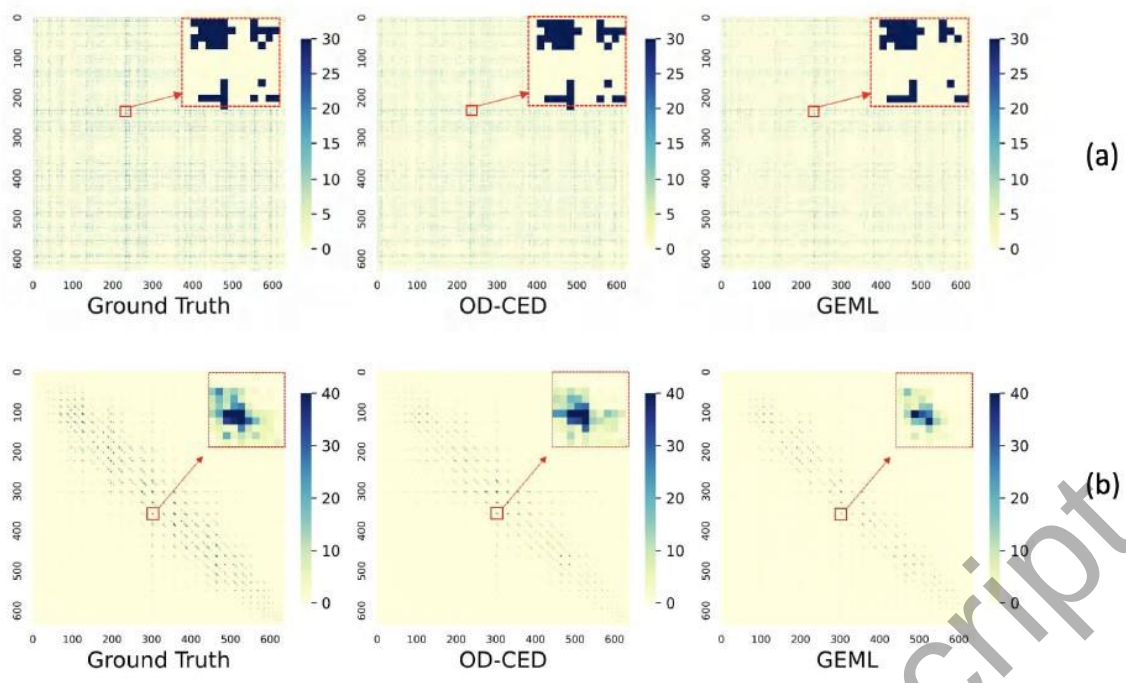

Figure 5: (a) The prediction results of the OD-CED and GEML for City-C (Nov. 14); (b) The prediction results of the OD-CED and GEML for City-S (Dec. 13).

Accepted Manuscript

Table 1: Prediction performance of all the compared methods on the test data across City-C and City-S, evaluated by Root Mean Square Error (RMSE), Weighted Mean Absolute Percentage Error (wMAPE), and Common Part of Commuters (CPC).

Method	**City-C**			**City-S**		
	wMAPE	**RMSE**	**CPC**	**wMAPE**	**RMSE**	**CPC**
HA	0.813	1.442	0.348	0.821	1.435	0.355
OLSR	0.822	1.419	0.324	0.816	1.351	0.333
LASSO	0.807	1.424	0.359	0.813	1.349	0.337
CSTN	0.782	1.370	0.354	0.721	1.217	0.451
MRSTN	0.788	1.380	0.351	0.766	1.253	0.464
OD-CED	**0.411**	**0.905**	**0.776**	**0.323**	**0.740**	**0.889**

Table 2: Quantitative comparison of OD-CED and representative GCN-based baselines

Method	**City-C**			**City-S**		
	wMAPE	**RMSE**	**CPC**	**wMAPE**	**RMSE**	**CPC**
GEML	0.645	1.210	0.665	0.590	1.095	0.715
STGCN	0.681	1.337	0.488	0.596	1.210	0.674
ASTGCN	0.682	1.368	0.647	0.601	1.052	0.688
STSGCN	0.663	1.348	0.619	0.583	0.972	0.712
STZINB-GNN	0.701	1.486	0.556	0.623	1.157	0.605
STTD	0.797	1.420	0.337	0.797	1.440	0.359
OD-CED	**0.432**	**0.958**	**0.791**	**0.341**	**0.782**	**0.918**

Table 3: Number of Parameters and Training time of one epoch for each compared method

	CSTN	**MRSTN**	**GEML**	**STGCN**	**OD-CED**
# of Params (in millions)	0.54M	0.67M	2.9M	1.6M	**0.1M**
Training Time (in seconds)	1222.13s	1602.14s	39.63s	28.81s	**22.12s**

Accepted Manuscript