Region Profile Enhanced Urban Spatio-Temporal Prediction via Adaptive Meta-Learning

Jie Chen
Shanghai University
Shanghai, China
chenjie_cs@shu.edu.cn

Tong Liu*
Shanghai University
Shanghai, China
tong_liu@shu.edu.cn

Ruiyuan Li
Chongqing University
Chongqing, China
ruiyuan.li@cqu.edu.cn

ABSTRACT

Urban spatio-temporal (ST) prediction plays a crucial role in smart city construction. Due to the high cost of ST data collection, improving ST prediction in a lack of data is significant. For this purpose, existing meta-learning methods have been demonstrated powerful by learning an initial network from training tasks and adjusting to target tasks with limited data. However, such shared knowledge from a set of tasks may contain irrelevant noise due to the gap of region-varying ST dynamics, resulting in the negative transfer issue. As a revelation of regional functional patterns, region profiles give rise to the diversity of ST dynamics. Thus, we design a novel adaptive meta-optimized model MetaRSTP, which conducts the initial prediction model in a finer-granularity of region level with region profiles as semantic evidence. To enhance the expressiveness of profiles, we firstly build a semantic alignment space to explore the inter-view co-semantics. Fusing it with view-specific uniqueness, the multi-view region profiles can be better applied in urban tasks. Then, a regional bias generator derives non-shared parameters in terms of profiles, which alleviates the divergence among regions. We set a new meta-learning strategy as initialize the network with fixed generalizable parameters and region-adaptive bias, thus enhancing the personalized prediction performance even in few-shot scenarios. Extensive experiments on real-world datasets illustrate the effectiveness of our MetaRSTP and our learned region profiles.

CCS CONCEPTS

• **Information systems** → **Spatial-temporal systems**.

KEYWORDS

Urban spatio-temporal prediction; Meta-learning; Deep learning

ACM Reference Format:
Jie Chen, Tong Liu, and Ruiyuan Li. 2023. Region Profile Enhanced Urban Spatio-Temporal Prediction via Adaptive Meta-Learning. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23), October 21–25, 2023, Birmingham, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3583780.3615027

*corresponding author

CIKM '23, October 21–25, 2023, Birmingham, United Kingdom

ACM ISBN 979-8-4007-0124-5/23/10…$15.00
https://doi.org/10.1145/3583780.3615027

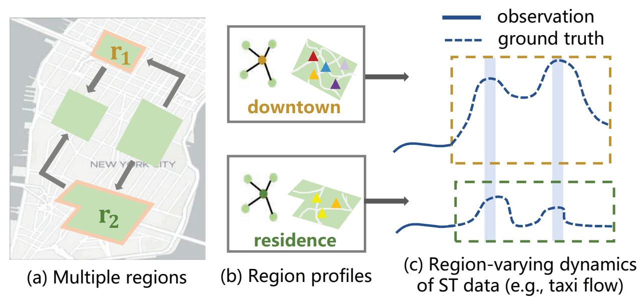


Figure 1: A motivation example : various region profiles lead to the diversity of ST dynamics but with shared characteristics of morning/evening peak.

1 INTRODUCTION

The fast advances in urban sensing technology and mobile computing have significantly facilitated the construction of intelligent cities. Urban ST prediction, one of the most prominent smart urban services, has enabled multi-domain applications such as trajectory prediction [24], route time estimation [4], and origin-destination demand prediction [31]. An accurate ST prediction system plays a crucial role in improving service efficiency, giving early warnings for emergency and ultimately providing insights for policymaking.

Existing hybrid deep learning models [3, 11, 12, 17] achieve encouraging results in ST prediction by mining non-linear ST correlations accurately. Unluckily, the superior performance of them is conditioned on large-scale training data which are probably inaccessible for sensor-scarce regions in real-world applications. Currently, advanced meta-learning methods [18, 26, 34] have been utilized to solve ST prediction tasks in few-shot scenarios, which learn a shared initial network from all tasks and fine-tune the network for a target task with limited data. To better transfer shared knowledge across similar tasks, task-clustering [4, 24, 28] is proposed to guide the initial network in a fine granularity of category level.

Nonetheless, a common pitfall is that existing meta-learning methods still fail to thoroughly avoid the risk of negative transfer due to the great differences among regions. As shown in Figure 1, for two different regions, it is possible that their observed dynamics is almost consistent in the beginning (e.g. before the dawn) but show totally different follow-up variation trends. Due to the similar observed data distribution, their ST dynamics are modeled in the same way even with the clustering meta strategy, which will result in inaccurate results for the future prediction.

To improve the prediction performance for overall regions, an intuitive idea is to build an effective model for each of them individually. Region profile, a unified representation uncovering the functionality and properties of an urban space, turn out to be effective in urban-related tasks [22, 29]. It is thus an excellent opportunity for few-shot ST prediction as well. In this paper, we aim to investigate adaptive prediction model in a finer-granularity of region level, simultaneously exploring model's generalization capability in few-shot scenarios. However, there exists two challenges:

- **How to attain comprehensive region profiles which can serve as reliable semantic evidence in ST prediction?** A practicable region profile requires multi-view representation learning for its various types of features. In a region, besides a simple combination of multi-views [7, 32], the possible inter-view interactions are complex to measure. For example, a downtown area shown in Figure 2(b) includes certain point-of-interest (POI) types such as malls and bars, as well as large-scale flows from certain regions, suggesting the latent semantic constraint within views. However, a residence zone with the same mobility features contains discriminate POIs such as residential buildings, which can better describe its function. The degree of semantic relevance among views affects the revelation of region profiles.
- **How to devise a meta-learning strategy which derives region-adaptive initial networks in few-shot scenarios?** Existing meta-learning methods [18, 24, 26, 28] learn a global or clustered shared initial network, which may cause uncertain noise for some regions. A fact is that varied region profiles shed light on the diversity of ST dynamics, as a downtown area r_1 presents greater fluctuation different from a residence zone r_2. To model such diversity, [15, 16] generate the weights of model based on regional geo-attributes, but may result in task-overfitting in few-shot situation. Actually, regions with their own specific ST dynamics may also obey the common trends such as the morning/evening peak [27]. Considering both generic regularity to fit few-shot scenarios and attached region-specific features for region-adaptive prediction remains unexplored.

To the end, we propose MetaRSTP (**Meta**-optimized Model for **R**egion-adaptive **S**patio-**T**emporal **P**rediction), an adaptive meta-learning framework which integrates regional multi-view profiles to enhance model's performance in few-shot scenarios. The model contains two modules for two challenges. For **challenge (1)**, we adopt a joint representation learning module to combine and align multi-view embeddings in a deep cooperation manner. With features from mobility view and geospatial view, a semantic alignment space is designed to preserve the co-semantic embedding under the effect of semantic constraints. With three types of embeddings including uniqueness and co-semantics, a cross-view attention fusion further enables the robustness of final multi-view region representation. For **challenge (2)**, we initialize the region-adaptive network via a combination of a shared initialization, and a weighted region-specific bias from a regional bias generator with the region profiles as input. The former is to keep the network's generalization ability for few-shot task, and the latter is to avoid being affected by the divergence of other regions. Through the bi-level optimization of meta-learning, the shared initialization will contain the general ST characteristics among all the regions, while the regional bias generator will bridge the inherent correlations between region profiles and ST dynamics. Finally, the prediction network for each region can perform in region-level granularity with reliable region profiles to offer more interpretability. Finally, the main contributions of our work are concluded as follows:

- We propose a novel deep meta-learned model MetaRSTP for region-adaptive ST prediction in a finer granularity, which is the first to integrate multi-view region profiles to alleviate the negative transfer issue in few-shot scenarios.
- We study the multi-view urban region profiling problem by building the semantic alignment space and cross-view fusion among multiple views, which can function as semantic support for multi-domain urban tasks.
- Extensive experiments and case study validate the effectiveness of our MetaRSTP and learned region profiles.

2 RELATED WORK

2.1 Region Profile Encoding

Region profile encoding aims to acquire versatile region representations, which reveals regions' functional property so that can be applied to multiple urban tasks. Early works focus on single-view study in terms of mobility data [6]. A popular scheme is to extract embeddings from flow connectivity graphs [21–23]. Recently, multi-view data [14] of mobility and geospatial data has been further exploited due to their more comprehensive expressiveness for region profiling. Along this line, some works adopt multi-view fusion methods, which learn view-specific embeddings separately and then fuse them jointly by concatenation [7, 32] or weighted attention summation [30]. Another works [29] attempt to propagate semantic information across multi-views based on the assumption that there exist strong constraints among distinct views.

However, there are no works explicitly modeling the inter-view cooperation, which makes the complex correlations among multi-views not fully exploited. Instead, we build a semantic alignment space to get co-semantic embeddings, and further fuse it with single-view features to ensure the effects of view-specific uniqueness.

2.2 Meta-Learning for ST Prediction

Meta-learning [5], a task-level learning paradigm, aims to learn the general knowledge from a set of training tasks to rapidly adapt to target tasks. As a classic optimization-based meta-learning algorithm, model-agnostic meta-learning (MAML) [5] has been widely used in various urban tasks. For example, MetaST [26] utilizes data from multiple source regions to enable the stable transfer in target regions for traffic prediction. However, the learned same initial network from MAML may cause sub-optimal results for some of tasks [10]. To alleviate the possible negative noise from irrelevant tasks, existing works [4, 24, 28] turn to adopt a task-clustering strategy. For instance, to conduct efficient POI recommendation, MetaPTP [24] group training users with similar trajectories in a category of certain preferences, thus offering cluster-aware initial networks for new users who belong to a certain category.

Inspired by the above works, we proceed to design a region-adaptive prediction network in a finer granularity so as to overcome the negative transfer issue. Besides a fixed sharing parameter with generalization ability, we propose a regional bias generator which contributes to bridging the inherent correlations between region profiles and corresponding ST dynamics with better interpretability.

3 PRELIMINARY

3.1 Problem Formulation

Following previous works [22, 30], a research city is divided into N_R disjoint regions, denoted as $R = \{r_1, r_2, \dots, r_{N_R}\}$. Two views are characterized for region profile learning. Afterwards, we present the formulation of region profile enhanced ST prediction problem.

Definition 1 (Inter-Region Mobility View). The mobility view is featured as a set of origin-destination (OD) region pairs $M = \{(r_o, r_d)\}$, which reveals the inter-region interactions. We compute the accessibility weight between $r_i, r'_j \in R$ as $w(r_i, r_j) = |(r_o, r_d) \in M | r_o = r_i, r_d = r_j|$, where $|\cdot|$ counts the co-occurance of a trip. Thus for region r_i, its OD contexts are described by the source distribution $p_o(r_k|r_i)$ and destination distribution $p_d(r_k|r_i)$:

$$p_o(r_k|r_i) = \frac{w(r_k, r_i)}{\sum_{r_j} w(r_j, r_i)},\ p_d(r_k|r_i) = \frac{w(r_i, r_k)}{\sum_{r_j} w(r_i, r_j)}. \tag{1}$$

Definition 2 (Intra-Region Geospatial View). The geospatial feature of a region can be depicted by its POI distribution $\mathbf{p}$ and road network density $\mathbf{d}$, i.e., $\mathbf{f} = \{\mathbf{p}, \mathbf{d}\} \in \mathbb{R}^{D_F}$, which reveals the regional function and attribute. Each dimension in $\mathbf{p} \in \mathbb{R}^{D_P}$ denotes the ratio of POIs of corresponding category, and $\mathbf{d} \in \mathbb{R}$ is computed as the area size divided by road length.

Definition 3 (Region Profile). For research regions, given their mobility patterns and geospatial features, a set of learned region representations $\mathbf{E} = \{\mathbf{e}_{r_i}\} \in \mathbb{R}^{N_R \times D_R}$ is expected to function as comprehensive region profiles with the uniform dimension D_R.

Definition 4 (Urban Dynamic State). Urban ST data includes taxi demand, POI check-in and so on, which varies across space and evolving over time [33]. We divide the time range into disjoint timeslots of equal length. For a region r, we denote the dynamic state (e.g. number of taxi pick-ups) at timeslot t as $x_{r_i}^{(t)}$.

Definition 5 (Urban ST Sequence). According to the First Law of Geography [19], adjacent regions naturally show explicit spatial dependency. At timeslot t, a region $r_i \in R$ is associated with a vector $\mathbf{x}_{r_i}^{(t)} = [x_{r_k}^{(t)}] \in \mathbb{R}^{N_S}$ where $r_k \in \mathcal{N}_{\mathcal{E}^s[r_i]}$, describing the gathered urban states of its spatial neighbors including itself and other $N_S - 1$ regions. Collecting such spatial features in a continuous timeslots T, we denote an urban ST sequence as $\mathrm{S}_{r_i} = [\mathbf{x}_{r_i}^{(t)}]_{t=t_c-T}^{t_c} \in \mathbb{R}^{T \times N_S}$ where t_c is the last timeslot.

Problem Formalization. For a region r_i, given a regional ST sequence with T historical dynamic features and corresponding region profile $\mathbf{e}_{r_i} \in \mathbf{E}$, the prediction task is formulated as learning a function $f_{\phi_{r_i}}(\cdot)$ to predict next urban state:

$$[\mathrm{S}_{r_i};\ \mathbf{e}_{r_i}] \xrightarrow{f_{\phi_{r_i}}(\cdot)} y_{r_i}^{(t_c+1)} \tag{2}$$

where $y_{r_i}^{(t_c+1)} = x_{r_i}^{(t_c+1)}$ is the ground-truth value, and $f_{\phi_{r_i}}(\cdot)$ is a specific prediction neural network parameterized by ϕ_{r_i}.

3.2 Meta-Learning Settings

In our work, we consider a region-level ST prediction as a learning task $\mathcal{T}_{r_i} = \{(S_{r_i,1}, y_{r_i,1}), \dots, (S_{r_i,N_L}, y_{r_i,N_L})\}$, where N_L is the number of all observed samples in region r. Following MAML-based training settings, each task is divided into a support set and a query set, i.e. $\mathcal{T}_{r_i} = \{\mathcal{D}_{r_i}^{sp}, \mathcal{D}_{r_i}^{qr}\}$. The former is used to fine-tune the initial network during local update, and the latter is used to learn a generalized initial network during global update, respectively. By adjusting the number of samples $\mathcal{T}_{r_i}$ to mimick few-shot scenarios and common scenarios, batches of training tasks are sampled to attain a prediction model with strong adaptability.

4 METHOD

4.1 Overview of MetaRSTP

The key challenges for region-adaptive ST prediction are two-fold. First, we need to extract comprehensive region features with strong expressiveness in functional properties. Second, based on the extracted region profiles, we need to devise effective meta-learning strategy containing extra region semantics to rule out negative noise. Therefore, our MetaRSTP consists of the following modules:

In **multi-view region profile learner** detailed in Figure 2(a), we consider mobility data and geospatial data for multi-view region representation learning. Two single-view encoders are implemented separately to derive discriminate feature extraction. Additionally, we innovatively construct a semantic alignment space which deals with the underlying semantic constraint between two views and uncovers co-semantic embeddings. Afterwards, a cross-view attention mechanism further enhances the profiling expressiveness by fusing the inter-view cooperation and view-specific uniqueness.

In **region-adaptive ST prediction model** detailed in Figure 2(b), considering region-varying ST patterns, we innovatively propose to customize a biased initialization for a specific region. Specifically, with the region profiles as input, a regional bias generator aims to offer personalized bias in the prediction model. Along with a fixed shared parameter denoting the common ST regularity through local and global update, the final generated initial network can better fit few-shot tasks in a finer granularity of region level. It also improves the confidence of prediction in an interpretable manner.

4.2 Multi-View Region Profile Learner

4.2.1 Single-View Feature Encoder. Multiple correlations among regions [29, 30] give insights for the revelation of region profiles. As they determine the proximity of regions in the embedding space, thus dominating the region profiling. In our work, we define correlations uncovered by human mobility and geospatial data as inter-region interactions and intra-region attributes. To capture single-view regional features, we firstly construct mobility view encoder and geospatial view encoder individually.

In the mobility view, regions with the same origin distribution or destination distribution are considered close to each other, giving hints about analogous urban functions. For each region $r_i \in R$, we connect it with regions with top N_M cosine similarity in terms of similar OD contexts:

$$c_o(r_i, r_j) = \mathrm{CosSim}(p_o(r_k|r_i), p_o(r_k|r_j)) \tag{3}$$

$$c_d(r_i, r_j) = \mathrm{CosSim}(p_d(r_k|r_i), p_d(r_k|r_j)) \tag{4}$$

With these OD contexts’ correlations, we construct a inter-region mobility graph $\mathcal{G}^m = (\mathcal{V}_R, \mathcal{E}^m)$, where $\mathcal{V}_R$ denotes N_R region vertices, and $\mathcal{E}^m$ is the set of predefined N_M correlated mobility neighbors for each region. In order to embed the mobility correlations into regions’ embeddings, We employ Graph Attention

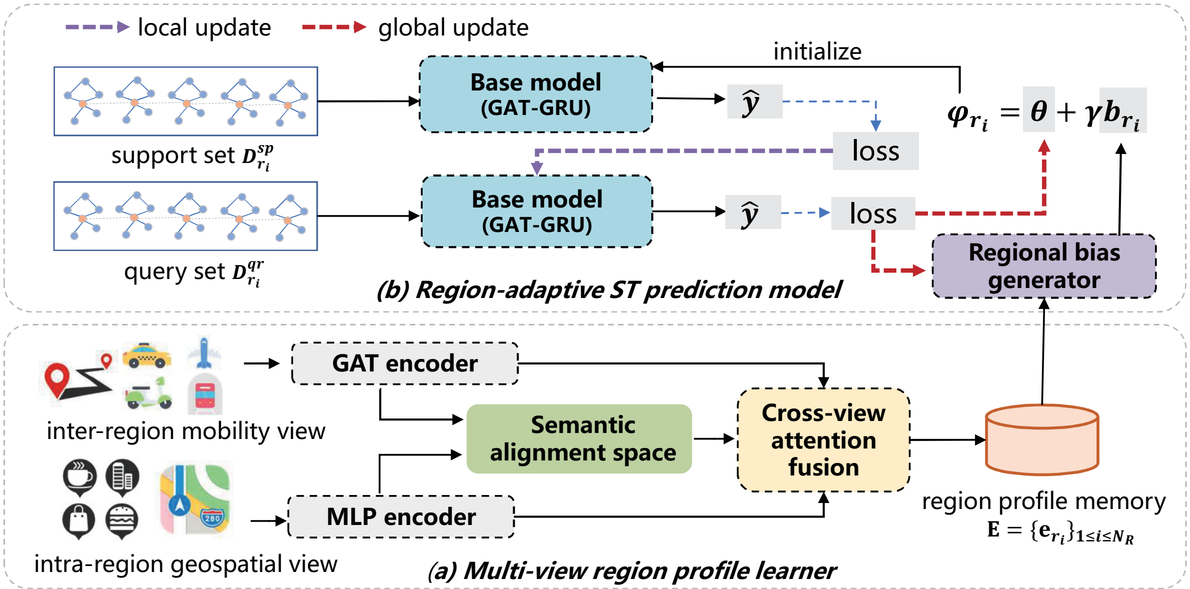


Figure 2: The framework of MetaRSTP. The first module is to extract the multi-view region profiles, including single-view encoders, semantic alignment and cross-view attention fusion. The second module is to generate regional personalized prediction model, including the base model, regional bias generator and meta optimization (local update and global update).

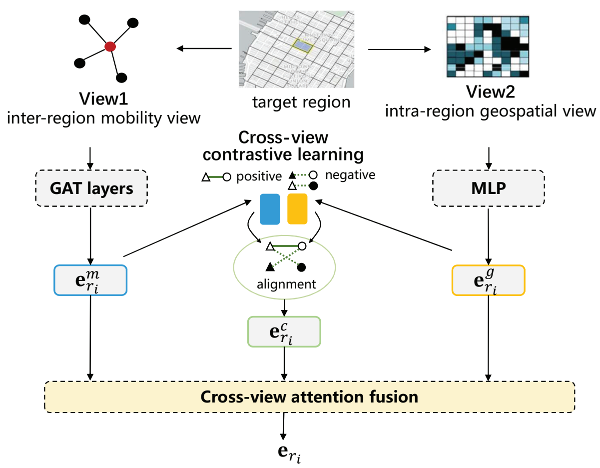


Figure 3: Multi-view region profile learner.

Network (GAT) [20] to update the representation of a region vertex by propagating information to its neighbors with attention scores. Formally, given the input OD features $\mathbf{v} = [v_1, v_2, \ldots, v_{N_R}]$ where $v_i \in \mathbb{R}^{N_R \times 2}$, region r_i's region embedding $\mathbf{e}^m_{r_i} \in \mathbb{R}^{D_M}$ in the mobility view is updated by the following steps:

$$\alpha^m_{ij} = \frac{\exp(\mathrm{LeakyReLU}(\mathbf{a}_m[\mathbf{W}_m v_i || \mathbf{W}_m v_j]))}{\sum_{k \in \mathcal{N}_{\mathcal{E}^m}[r_i]} \exp(\mathrm{LeakyReLU}(\mathbf{a}_m[\mathbf{W}_m v_i || \mathbf{W}_m v_k]))} \tag{5}$$

$$\mathbf{e}^m_{r_i} = \sum_{j \in \mathcal{N}_{\mathcal{E}^m}[r_i]} \alpha^m_{ij} \Phi_m v_j \tag{6}$$

where $\mathbf{W}_m, \Phi_m, \mathbf{a}_m$ are learnable parameters, $||$ is the concatenation, α_{ij} is the attention scores, and $\mathcal{N}_{\mathcal{E}^m}[r_i]$ is the set of region r_i's neighbors which share similar OD contexts.

In the geospatial view, there is a need to transform the coarse POI and road density features in r_i to a dense region embedding $\mathbf{e}^g_{r_i} \in \mathbb{R}^{D_G}$. Following [29], we employ the most commonly used structure Multi-Layer Perceptions (MLPs) to implement the geospatial encoder, which is formulated as $\mathbf{e}^g_{r_i} = \mathrm{MLP}(\mathbf{f}_i)$.

4.2.2 Semantic Alignment Space. To enhance the expressiveness and reliability of acquired region profiles, we proceed to explore the correlated inter-view cooperation. Though existing methods attempt to conduct cross-view attention [14] or establish latent constraints [29], the rich co-semantic representation is not well modeled in an explicit way, leading to incomplete semantics in region profiles. In our proposed semantic alignment space, an intuitive idea is that two view-specific features of a region can match each other as they reflect the same underlying regional patterns. For example, in a center region, mobility correlations with a wide range of regions matches high road density and multi-type POI distribution. Thus, exploring the coexistence of distinct view-specific features helps carry out highly effective urban region profiling.

Motivated by contrastive learning [2], we encode the co-semantic representation through comparing the positive pairs with negative ones by the Noise Contrastive Estimation (NCE) objective. The key is to construct view-to-view mapping alignment. Concretely, for region r_i, its region embedding in mobility view and geospatial view serve as the ground-truth for each other, acting as a positive pair $(\mathbf{e}^m_{r_i}, \mathbf{e}^g_{r_i})$. By reserving the embedding of r_i in one view and sampling the embeddings in another view from other regions, we can obtain two groups of representations $\hat{\mathbf{E}}^m = \{\mathbf{e}^m_{r_n} | n \neq i\}$ and $\hat{\mathbf{E}}^g = \{\mathbf{e}^g_{r_n} | n \neq i\}$, where $|\hat{\mathbf{E}}^g| = |\hat{\mathbf{E}}^m| = N_O$ is the predefined negative size. Thus negative pairs are extracted as $(\mathbf{e}^m_{r_n}, \mathbf{e}^g_{r_i})$ and $(\mathbf{e}^m_{r_i}, \mathbf{e}^g_{r_n})$. For any pair $(\mathbf{e}^m, \mathbf{e}^g)$, we take the following alignment step to capture the possible existing co-semantics $\hat{\mathbf{e}}^c \in \mathbb{R}^{D_C}$ within two views:

$$\hat{\mathbf{e}}^c = \mathrm{Align}(\mathbf{e}^m, \mathbf{e}^g) = \mathrm{ReLU}(\mathbf{W}_c([\mathbf{e}^m \| \mathbf{e}^g] + \mathbf{b}_c) \quad (7)$$

where $\mathrm{Align}(\cdot) : \mathbb{R}^{D_M+D_G} \rightarrow \mathbb{R}^{D_C}$ is a feature transformations function. To ensure the positive co-semantic representations take effect, we adopt the InfoNCE loss function [9] to model the inter-view matching rule in different regions:

$$\mathcal{L}_{inter} = \sum_{r_i \in R} \Big[-log\mathcal{M}(\mathbf{e}^m_{r_i}, \mathbf{e}^g_{r_i}) + \Big(log(\mathcal{M}(\mathbf{e}^m_{r_i}, \mathbf{e}^g_{r_i}) + \sum_{\mathbf{e}^m_{r_n} \in \hat{\mathbf{E}}^m} \mathcal{M}(\mathbf{e}^m_{r_n}, \mathbf{e}^g_{r_i}) + \sum_{\mathbf{e}^g_{r_n} \in \hat{\mathbf{E}}^g} \mathcal{M}(\mathbf{e}^m_{r_i}, \mathbf{e}^g_{r_n})\Big)\Big] \quad (8)$$

where $\mathcal{M}(\cdot,\cdot) = \exp(\mathrm{Align}(\cdot,\cdot))$ functions as an inter-view discriminator to evaluate the matching scores between view-specific region embeddings. After the model training, r_i's real positive co-semantics $\mathbf{e}^c_{r_i} = \mathrm{Align}(\mathbf{e}^m_{r_i}, \mathbf{e}^g_{r_i})$ is well learned and exploited .

4.2.3 Cross-View Attention Fusion. With the obtained sets of mobility view embeddings $\mathbf{E}^m = \{\mathbf{e}^m\}$, geospatial view embeddings $\mathbf{E}^g = \{\mathbf{e}^g\}$ and co-semantic embeddings $\mathbf{E}^c = \{\mathbf{e}^c\}$, a cross-view attention mechanism is used to better propagate knowledge across multiple aspects. Among them, co-semantic embeddings are viewed as keys while all the embeddings are taken as queries. We associate a key matrix $\mathbf{K}$ and a query matrix $\mathbf{Q}^*(* \in \{m, c, g\})$ as follows:

$$\mathbf{K} = \mathbf{E}^c\mathbf{W}_k,\ \mathbf{Q}^* = \mathbf{E}^*\mathbf{W}_q. \quad (9)$$

where $\mathbf{W}_k$ and $\mathbf{W}_q$ are learnable parameters. We then fuse single views with co-semantic view to achieve final region representations:

$$A^* = softmax(\frac{\mathbf{Q}^*\mathbf{K}^T}{\sqrt{D_R}}),\ \mathbf{E} = A^m\mathbf{E}^m + A^c\mathbf{E}^c + A^g\mathbf{E}^g \quad (10)$$

Thus, the attentively obtained region profile set $\mathbf{E} = \{\mathbf{e}\}$ contains comprehensive features of co-semantics and view-specific uniqueness, making it possible to assist urban tasks as semantic evidence.

4.2.4 Training Objective. To ensure that the region profiles possess strong applicability in fitting urban multi-tasks, we define a joint learning objective function from single-view and inter-view:

$$\mathcal{L}_{emb} = \beta_1\mathcal{L}_{mob} + \beta_2\mathcal{L}_{inter} + \beta_3\mathcal{L}_{geo} \quad (11)$$

where β_1, β_2 and β_3 are adjustable weights to balance mobility view loss, inter-view matching loss and geospatial view loss, respectively. The inter-view loss $\mathcal{L}_{inter}$ has been explained in Equation (8). In single-view, we expect the learned profiles are able to estimate the ground-truth distribution of mobility and geospatial data as accurate as possible. In the mobility view, with region profiles $\mathbf{e}_i$, $\mathbf{e}_j$ of region r_i, r_j, we can estimate the transition probability:

$$\hat{p}_o(r_j|r_i) = \frac{\exp(\mathbf{e}^T_{r_i}\mathbf{e}_{r_j})}{\sum_{r_k \in R}\exp(\mathbf{e}^T_{r_i}\mathbf{e}_{r_k})} \quad (12)$$

Likewise, $\hat{p}_d(r_j|r_i)$ is obtained. Then we define the learning objective $\mathcal{L}_{mob}$ to minimize the divergence.

$$\mathcal{L}_{mob} = \sum_{(r_i,r_j) \in R} -p_o(r_j|r_i)log\hat{p}_o(r_j|r_i) - p_d(r_j|r_i)log\hat{p}_d(r_j|r_i) \quad (13)$$

Similarly, to let the learned region profiles reserve the region similarity in terms of region attributes, we design a task to reconstruct region correlations based on region profiles. Taking $c_{geo}(r_i, r_j) = \mathrm{CosSim}(\mathbf{f}_{r_i}, \mathbf{f}_{r_j})$ as label, the learning objective $\mathcal{L}_{geo}$ in geospatial view is defined as follows:

$$\mathcal{L}_{geo} = \sum_{(r_i,r_j) \in R} \left(c_{geo}(r_i, r_j) - \mathbf{e}^T_{r_i}\mathbf{e}_{r_j}\right)^2. \quad (14)$$

4.3 Region-Adaptive ST Prediction Model

4.3.1 Base Model. The base model aims to predict next urban state based on historical ST sequences. There exists complex ST dependency in the dynamics of urban states, as adjacent regions are mutually affected and a region's previous states affect subsequent state. Similar to ST-MetaNet [15], we utilize a GAT to model the spatial correlations among regions, and a Gated Recurrent Unit (GRU) to capture the temporal evolvement of each region, which is illustrated as a GAT-GRU hybrid model in Figure 2(b).

For a region r_i, we regard its surrounding neighbors as $N_S - 1$ connected nodes in a distance-based spatial graph $\mathcal{G}^s_{r_i}$. At timeslot t, taking $\mathcal{G}^s_{r_i}$ and regional spatial features $\mathbf{x}^{(t)}_{r_i} \in \mathbb{R}^{N_S}$ as input, a local GAT computes the attention score and gets the output by weighted aggregation with regions in the neighbor set $\mathcal{N}_{\mathcal{E}^s}$:

$$\alpha^s_{ij} = \frac{\exp(\mathrm{LeakyReLU}(\mathbf{a}_s[\mathbf{W}_s x^{(t)}_{r_i} || \mathbf{W}_s x^{(t)}_{r_j}]))}{\sum_{k \in \mathcal{N}_{\mathcal{E}^s}[r_i]}\exp(\mathrm{LeakyReLU}(\mathbf{a}_s[\mathbf{W}_s x^{(t)}_{r_i} || \mathbf{W}_s x^{(t)}_{r_k}]))} \quad (15)$$

$$\mathbf{z}^{(t)}_{r_i} = \sum_{j \in \mathcal{N}_{\mathcal{E}^s}[r_i]} \alpha^s_{ij}\Phi_s x^{(t)}_{r_k} \quad (16)$$

where $\mathbf{W}_s, \Phi_s, \mathbf{a}_s$ are learnable parameters, and $\mathbf{z}^{(t)}_{r_i}$ is the representation containing spatial dependency. Afterwards, in order to forecast $y^{(t_c+1)}_{r_i}$, we model the temporal evolution of region r_i by passing all the spatial representations along the time span $\{t_c - T, \ldots, t_c\}$ through a GRU, which is formulated as:

$$\mathbf{u} = \sigma(\mathbf{W}_u\mathbf{z}^{(t)}_{r_i} + \mathbf{U}_u\mathbf{h}^{(t-1)}_{r_i} + \mathbf{b}_u) \quad (17)$$

$$\mathbf{r} = \sigma(\mathbf{W}_r\mathbf{z}^{(t)}_{r_i} + \mathbf{U}_r\mathbf{h}^{(t-1)}_{r_i} + \mathbf{b}_r) \quad (18)$$

$$\mathbf{h}'_{r_i} = \phi(\mathbf{W}_h\mathbf{z}^{(t)}_{r_i} + \mathbf{U}_h(\mathbf{r} \circ \mathbf{h}^{(t-1)}_{r_i}) + \mathbf{b}_h) \quad (19)$$

$$\mathbf{h}^{(t)}_{r_i} = \mathbf{u} \circ \mathbf{h}^{(t-1)}_{r_i} + (1 - \mathbf{u}) \circ \mathbf{h}'_{r_i} \quad (20)$$

where $\mathbf{z}^{(t)}_{r_i}, \mathbf{h}^{(t)}_{r_i}$ are the input vector and the hidden state at timeslot t, respectively. $\mathbf{u}, \mathbf{r}$ are update gate vector and reset gate vector. $\mathbf{W}_u, \mathbf{W}_r, \mathbf{W}_h$ and $\mathbf{U}_u, \mathbf{W}_r, \mathbf{U}_h$ are weight matrices. $\mathbf{b}_u, \mathbf{b}_r, \mathbf{b}_h$ are biases. $\circ$ is the element-wise multiplication, $\sigma(\cdot)$ is sigmoid function, and $\phi(\cdot)$ is tanh function. We take $\hat{y}^{(t_c+1)}_{r_i} = \mathbf{h}^{(t_c+1)}_{r_i}$ as the output prediction value.

4.3.2 Regional Bias Generator. In urban planning, ST dynamics closely interrelates with region profiles. For example, traffic patterns are totally different between business centers and industrial parks. In existing meta-learning methods, shared knowledge extracted from a set of training tasks is likely to cause negative transfer when fitting target tasks. To address this issue, a promising approach is to introduce distinctive region features as indicative weighted bias, so as to address the gap resulting from other regions.

Thus, we innovatively devise a regional bias generator to conduct bias assignment for each region. The key insight is to view region profiles as the important metadata of ST prediction neural networks for capturing specific ST correlations. For a region

r_i, we retrieve the corresponding region embedding $\mathbf{e}_{r_i}$ from the regional profile memory $\mathbf{E}$. Furthermore, the bias generator takes meta knowledge of region profile as input and outputs the non-shared bias parameter b_{r_i}. Formally, we denote the trainable weights of the whole bias generator network as η, and N_P is the number of parameters in the base model. In specific, our designed bias generator is a function $\mathcal{F}_\eta(\cdot)$ formed by two two transformation $\mathcal{F}_1 : \mathbb{R}^{D_R} \to \mathbb{R}^{D_O}$, $\mathcal{F}_2 : \mathbb{R}^{D_O} \to \mathbb{R}^{N_P}$. Similar to the parameter generator method in [13], we implement the region-adaptive bias generator by two fully connected layers. Compared with the task-clustering scheme, differentiability and interpretability can be better guaranteed in this way. Also, besides the consideration for diverse ST dynamics, when a region profile is analogous to another region, the initial prediction networks for them still tend to be similar via parameter adjustment, which ensures that the inherent correlations between region profile and ST dynamics are well bridged and learned.

4.3.3 Meta Optimization. Shared regularity among all tasks contribute to fitting regions with few-shot samples, and non-shared region-specific ST dynamics further improve the prediction model's confidence and alleviate the negative transfer issue. For these reasons, in the process of meta optimization, we present a biased initialization meta-learning strategy with a shared generalizable parameter and a variable bias to conduct region-adaptive prediction in a finer way. Algorithm 1 outlines the end-to-end meta training process. Firstly, for each training task $\mathcal{T}_{r_i} \in \mathcal{T}^{tr}$, its specific base model is initialized by shared parameters θ and characteristic bias b_{r_i} in terms of the meaningful region profile $\mathbf{e}_{r_i}$.

$$\phi_{r_i} \leftarrow \theta + \gamma b_{r_i},\ b_{r_i} = \mathcal{F}_\eta(\mathbf{e}_{r_i}) \tag{21}$$

where γ is the personalized update rate for controlling how much region-specific knowledge is added to the shared knowledge. Afterwards, keeping with MAML-based episode training process, there are following two stages: local update and global update.

During the **local update** stage, the local network for specific region is expected to converge to a good local optimum, which is similar to traditional neural networks' training. Thus, we update the local network by minimizing the prediction loss $\mathcal{L}_{\mathcal{D}^{sp}_{r_i}}$ of each $\mathcal{T}_{r_i} \in \mathcal{T}^{tr}$ according to its support set $\mathcal{D}^{sp}_{r_i}$ with a local learning rate α .

$$\phi^*_{r_i} \leftarrow \phi_{r_i} - \alpha \nabla_{\phi_{r_i}} \mathcal{L}_{\mathcal{D}^{sp}_{r_i}}(\phi_{r_i}) \tag{22}$$

During the **global update** stage, our goal is to obtain a shared generalizable prediction model θ^* which can fitting few-shot tasks, and an effective bias generator η^* which bridge the profiles and ST dynamics. After local update, one-step gradient decent is taken to update the two parameters based on the sum of the losses on query set $\mathcal{D}^{qr}_{r_i}$ of each training task $\mathcal{T}_{r_i} \in \mathcal{T}^{tr}$.

$$\theta^* \leftarrow \theta - \lambda \sum_{\mathcal{T}_{r_i} \in \mathcal{T}^{tr}} \nabla_\theta \mathcal{L}_{\mathcal{D}^{qr}_{r_i}}(\phi^*_{r_i}) \tag{23}$$

$$\eta^* \leftarrow \eta - \lambda \sum_{\mathcal{T}_{r_i} \in \mathcal{T}^{tr}} \nabla_\eta \mathcal{L}_{\mathcal{D}^{qr}_{r_i}}(\phi^*_{r_i}) \tag{24}$$

where λ is the learning rate of global update, and $\mathcal{L}_{\mathcal{T}^{qr}_{r_i}}$ is the prediction loss of query set $\mathcal{D}^{qr}_{r_i}$ for prediction task $\mathcal{T}_{r_i}$.

Algorithm 1: MetaRSTP Training Algorithm

input : ST prediction tasks $\{\mathcal{T}_r\}$, region embedding set $\mathbf{E}$, learning rate α, λ, and bias rate γ.
output: Trained base model parameter θ^* and bias generator parameterη^*

$\theta, \eta \leftarrow$ random initialization;
repeat
  randomly select a batch of training tasks $\mathcal{T}^{tr}$ from $\{\mathcal{T}_r\}$;
  for $\mathcal{T}_{r_i} \in \mathcal{T}^{tr}$ **do**
    $\phi_{r_i} \leftarrow \theta + \gamma b_{r_i}$; // $b_{r_i} \leftarrow \eta, \mathbf{e}_{r_i}$
    $\mathcal{D}^{sp}_{r_i}$, $\mathcal{D}^{qr}_{r_i} \leftarrow$ support set, query set from $\mathcal{T}_{r_i}$;
    evaluate $\mathcal{L}_{\mathcal{D}^{sp}_{r_i}}(\phi)$ with $\mathcal{D}^{sp}_i$ and update locally:
      $\phi^*_{r_i} \leftarrow \phi_{r_i} - \alpha \nabla_{\phi_{r_i}} \mathcal{L}_{\mathcal{D}^{sp}_i}(\phi_{r_i})$;
    evaluate $\mathcal{L}_{\mathcal{D}^{qr}_{r_i}}(\phi^*_{r_i})$ with $\mathcal{D}^{qr}_{r_i}$;
  update globally with $\sum_{\mathcal{T}_{r_i} \in \mathcal{T}^{tr}} \mathcal{L}_{\mathcal{D}^{qr}_{r_i}}(\phi^*_{r_i})$:
    $\theta^* \leftarrow \theta - \lambda \sum_{\mathcal{T}_{r_i} \in \mathcal{T}^{tr}} \nabla_\theta \mathcal{L}_{\mathcal{D}^{qr}_{r_i}}(\phi^*_{r_i})$;
    $\eta^* \leftarrow \eta - \lambda \sum_{\mathcal{T}_{r_i} \in \mathcal{T}^{tr}} \nabla_\eta \mathcal{L}_{\mathcal{D}^{qr}_{r_i}}(\phi^*_{r_i})$;
until *stopping criteria is met*;

5 EXPERIMENTS

In this section, we conduct extensive experiments to answer the following research questions through our evaluation:
RQ1: Does the proposed MetaRSTP outperform existing methods in few-shot scenarios and common scenarios?
RQ2: Do the proposed components (e.g., alignment space, bias generator, and so on) take effect?
RQ3: How the important parameters of the region profile dimension and the bias rate influence the experiment results?
RQ4: Can the generated region profiles possess practicability for other urban tasks?

5.1 Settings

5.1.1 Datasets. From NYC Open Data[1], we collect several real-world datasets including 180 regions split by streets in Manhattan. To carry out multi-view region profile learning, we apply taxi trip data as mobility view, and take POI data [25], road network data[2] as geospatial view. Additionally, the district division is provided by the community Boards from [1]. Further practicability assessment on region profiles involves the use of check-in data, crime data, and land usage data. We list the description of each dataset in Table 1.

Following [15], we set our tasks as predicting regional future pick-up and drop-off volumes and conduct experiments on taxi trip datasets, which contains time and geographical coordinates of pick-ups and drop-offs. For each region, we count the volumes per 30 minutes based on the number of taxis entering or exiting the region. Furthermore, a total number of observed data in roughly 5760 timeslots are divided into 80% for meta-training and 20% for meta-testing and evaluating.

[1]http://opendata.cityofnewyork.us/
[2]https://www.openstreetmap.org/

Table 1: Datasets

Dataset	Description
Census blocks	Boundaries of 180 regions split by streets in Manhattan.
Taxi trips	Around 30 million taxi trip records during four months.
Check-in & POIs	Over 200 thousand check-ins and 4 thousand POIs of 246 categories.
Road network	Road distribution including road length and area size in Manhattan.
Crime data	Around 40 thousand crime records during one year.
Land usage	12 types of districts classified by urban functions.

5.1.2 Baselines. We compare MetaRSTP with the following representative ST prediction methods. To be fair, all the meta-learning methods use GAT-GRU as base model.

- **DCRNN** [11], a classic model in traffic prediction, uses a diffusion convolution and Seq2Seq to mine ST dynamics.
- **MAML** [5] learns a globally-shared initialization from all tasks and generalize to an unseen task.
- **ST-MetaNet** [15] derives region-aware models based on learned representations from regional geo-graph graphs.
- **MetaPTP** [24] devises a meta strategy that learns clustered initial networks from ST tasks with similar samples.

Besides the evaluation on the whole model and our designed meta-learning strategy, we also compare with four embedding methods by replacing the multi-view region profile learner in MetaRSTP.

- **node2vec** [8] uses biased random walks to learn node latent representations by skip-gram models.
- **MVURE** [30] employs weighted multi-view fusion with human mobility and inherent region attribute data including features of POIs and check-ins.
- **MGFN** [22] designs a multi-level attention mechanism to learn region profiles from multiple mobility patterns based on intra-pattern and inter-pattern correlations.
- **ReMVC** [29] features in the schemes of hierarchical intra-view and inter-view contrastive learning.

5.1.3 Experiment Setup. In our method, two modules are trained by Adaptive Moment Estimation optimizer separately. In region profile learning, we set the dimension of embeddings as $D_M = D_C = D_G = D_R = 96$. The number of mobility neighbors N_M is 20, and the negative size N_O for contrastive learning is 10. Moreover, we set $\beta_1 = 1, \beta_2 = 10, \beta_3 = 5$ to balance the scale of three types of losses. The learning rate is 1e-3 over 500 epochs. In ST prediction module, we use $T = 10$ to predict urban state in next timeslot. GAT-GRU is implemented by single-layer GAT and GRU with 128 hidden units, and the number of spatial neighbors N_S is 20. In meta-training period, we sample 4 training tasks in each iteration, and set the training epochs to be 100 with early stopping strategy. The learning rates of local update α, global update λ are 1e-2 and 1e-4, respectively. For the bias generator, we build two-layer FCNs with hidden units [16, 64, N_P], where N_P is the number of parameters in the base model, and set the bias rate $\gamma = 0.2$. All the methods are implemented using PyTorch on a single Nvidia RTX 4090.

5.1.4 Evaluation Metrics. We leverage two commonly-used metrics to evaluate one-step prediction performance: Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|, \ \text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\|\hat{y}_i - y_i\|} \tag{25}$$

where n is the number of instances, $\hat{y}_i$ is the prediction result and y_i is the ground truth.

5.2 Performance Comparison (RQ1)

Table 2 presents the performance of our proposed MetaRSTP in comparison to eight baselines in predicting pick-ups and drop-offs. Denoting the ratio of draw-out training samples in each task as K, we choose $K = \{5\%, 10\%, 20\%\}$ to simulate prediction in few-shot scenarios, and use "overall" to represent prediction with all the training samples as common scenarios. From the table, we can summarize the following conclusions from two aspects.

5.2.1 Embedding Model Comparison. Compared with embedding methods, MetaRSTP with our designed multi-view region profile learner achieves superior performance. Among them, node2vec performs worst, suggesting that a simple concatenation is insufficient to fully exploit the complex multi-view information of a region. Its generated region profiles may even act as noise in the "overall" pick-up prediction, since the prediction model with them perform worse than MAML without extra semantic knowledge. MVURE with multi-view fusion lacks explicit modeling for co-semantic information, causing decline of its performance. MGFN generates single-view region profiles and suffers from limited performance due to the lack of the geospatial view. ReMVC depends on strong assumption that each regions' view-specific embeddings follow strict matching principle without the explicit co-semantic representation. Though MGFN and ReMVC get competitive results, our region profiles perform best, as we consider the view-specific uniqueness fusion and cross-view co-semantic alignment among multiple views, thus attaining high-quality region profiles for further prediction.

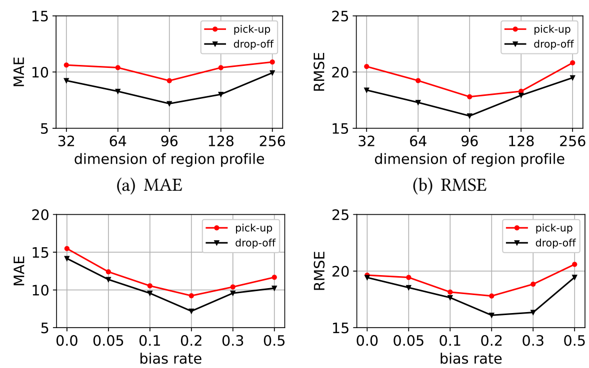
5.2.2 Meta-Learning Strategy Comparison. Among the ST prediction models, MetaRSTP outperforms all the baselines with varying numbers of training samples of each task and shows 12.28% and 7.74% average improvements on MAE and RMSE beyond the sub-optimal models, respectively. Particularly, the performance of DCRNN degrades significantly with limited data when $K = 5\%$ compared to other meta-learning methods. MAML achieves better results than DCRNN as it can allocate an effective initial network for tasks with limited samples. ST-MetaNet' s promising results verifies the importance of considering regional ST dynamics. Also, MetaPTP gives a finer initial model for regions in a same cluster with higher accuracy, improving the strategy of MAML which learns a global initialization and is more likely to introduce negative noise. MetaPTP's better results than ST-MetaNet in most cases can be explained as that ST-MetaNet focuses on region itself and may ignore common ST trends among regions. Finally, our proposed MetaRSTP further enhances the performance due to two reasons: (1) modeling the region-agnostic common ST dynamics as common initialization. (2) additionally considering the uniqueness of region profiles to offer finer-grained prediction in region level.

Table 2: Performance of MetaRSTP and baselines in terms of MAE and RMSE on taxi pick-up and drop-off prediction. The best and the second-best results are highlighted in boldface and underlined respectively.

Methods			node2vec	MVURE	MGFN	ReMVC	DCRNN	MAML	ST-MetaNet	MetaPTP	**MetaRSTP**
pick-up prediction	K=5%	MAE	21.949	17.949	18.071	17.602	28.116	24.974	22.889	19.230	**16.478**
		RMSE	28.833	26.803	26.136	25.564	33.367	29.327	28.915	26.128	**24.891**
	K=10%	MAE	16.445	15.802	15.526	14.255	24.974	16.250	16.013	15.247	**13.278**
		RMSE	26.274	24.484	23.888	23.694	31.209	27.795	27.128	25.109	**22.499**
	K=20%	MAE	15.629	14.407	13.120	13.379	19.729	15.911	15.119	14.239	**12.312**
		RMSE	25.711	23.815	22.436	23.068	29.327	26.963	26.046	22.291	**21.329**
	overall	MAE	13.950	11.471	10.517	11.532	15.569	12.773	13.329	10.329	**9.231**
		RMSE	23.228	20.647	19.117	20.007	25.573	23.118	23.831	20.548	**17.800**
drop-off prediction	K=5%	MAE	22.260	18.323	16.825	15.701	19.987	23.232	16.185	18.986	**15.120**
		RMSE	29.137	28.501	27.407	26.512	33.974	29.489	27.768	27.896	**25.978**
	K=10%	MAE	17.293	15.368	14.671	14.907	22.554	16.023	16.406	15.889	**13.424**
		RMSE	26.322	24.173	23.741	24.576	29.884	27.039	24.942	24.676	**22.762**
	K=20%	MAE	16.275	15.620	14.991	13.675	16.794	15.905	13.440	14.089	**12.969**
		RMSE	23.883	21.495	21.556	20.085	28.227	22.418	21.855	22.692	**20.856**
	overall	MAE	12.531	11.986	10.124	9.433	16.727	10.363	11.586	9.336	**7.889**
		RMSE	19.850	18.760	17.834	17.049	25.142	19.319	20.861	18.125	**16.097**

Table 3: Performance comparison on MetaRSTP's variants.

Methods	pick-up prediction		drop-off prediction	
	MAE	RMSE	MAE	RMSE
$w/o.mob$	11.239	19.870	10.382	18.392
$w/o.geo$	10.032	18.392	9.543	17.492
$w/o.align$	9.931	18.933	8.983	17.329
$w/o.fusion$	10.289	18.239	10.349	18.194
$w/o.meta$	16.219	24.493	15.403	23.506
$w/o.bias$	11.382	19.439	12.403	19.540
Full Version	**9.231**	**17.800**	**7.889**	**16.097**



(a) MAE (b) RMSE

(c) MAE (d) RMSE

Figure 4: Performance comparison with different dimensions of the region profile and different bias rate.

5.3 Ablation Study (RQ2)

To better understand the effect of each component in MetaRSTP, we investigate six MetaRSTP's variants: (1) $w/o.mob$ removes inter-region mobility view modeling; (2) $w/o.geo$ removes intra-region geospatial view modeling; (3) $w/o.align$ removes semantic alignment space within two views; (4) $w/o.fusion$ removes different views' combination and only use co-semantic representation; (5) $w/o.meta$ removes meta-learning framework and only use base model; (6) $w/o.bias$ removes the regional bias generator. The results of ablation tests on pick-up prediction and drop-off prediction are detailed in Table. 3. Through the table, we can observe that:

MetaRSTP which considers multi-view information generally has better performance than those considering only single-view data, i.e., $w/o.mob$ and $w/o.geo$. The result demonstrates that each view has its own merit to enhance the region representation. Meanwhile, the model $w/o.geo$ outperforms $w/o.mob$ on both two tasks. One possible reson is that POIs often suffer from data sparsity problem. The second observation is that the combination operation and the alignment operation are both integral. Since $w/o.fusion$ and $w/o.align$ both suffer from worse performance than the full version. Thus, it is proved that inter-view contrastive learning can fully exploit the importance of information propagation between views, and fusion operation ensures the unique expressiveness of each single view. In common scenarios, $w/o.meta$ degenerates into base model, suffering from failing to adapt to each region with fine-tune. Also, $w/o.bias$ without region-specific knowledge seems to degenerate into the MAML model, and performs worse than MetaRSTP due to the same initialization for all tasks, proving that the bias generation method helps to derive a reliable initial network for each region and transfer knowledge across various regions.

5.4 Parameter Sensitivity Analysis (RQ3)

In this section, we investigate how the performance of MetaRSTP varies with different dimensions of region embedding and different bias rates, and find the optimal settings.

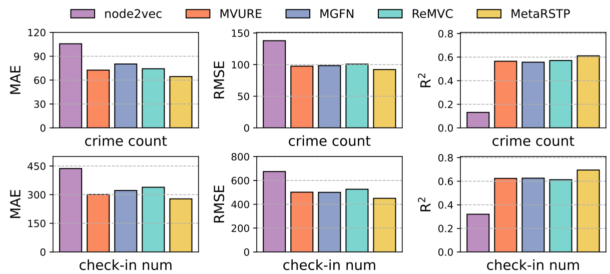


Figure 5: Performance comparison of different embedding models with Lasso regression model on downstream prediction tasks.

5.4.1 Effect of the Dimension of Region Profile. The dimension of region profile directly impacts the expressiveness of regions and the efficiency of bias generation. As shown in Figure 4(a) and Figure 4(b), increasing the value of region profile from 32 to 96 leads to visible improvements in performance for both tasks. This fact also illustrates that the bias learned from multi-view region profile essentially takes effect. When the dimension value exceeds 96, the prediction model seems to easily overfit and perform worse results. As a reasonable dimension can reflect the corresponding knowledge and keeps good performance, we select $D_R = 96$ as the region profile dimension.

5.4.2 Effect of the Bias Rate. Bias rate is an importance hyperparameter in adjusting the region-specific knowledge ratio. Figure 4(c) and Figure 4(d) depict the influences on MAE and RMSE with changing bias rate. A appropriate rate value can introduce region's personalized knowledge to make better prediction and don's influence the generation capability of the sharing initialization. It is observed that the performance has a slight drop when the rate is more than 0.2 on both tasks. We infer that this is because sampling too many unique region's knowledge will become harmful to preserve the common ST dynamics reflected by the sharing initialization. Thus we choose well-performing bias rate $\gamma = 0.2$.

5.5 Case Study: Inspecting the Region Profiles (RQ4)

In order to further explore the practical applicability of our learned region profiles, we present the results of two types of urban related tasks: prediction task and clustering task.

5.5.1 Prediction Tasks. Crime count prediction and check-in num prediction are two important urban downstream problems. To evaluate the effectiveness of different region profiles from different methods, we use a simple Lasso regression model for these downstream tasks and consider region profiles as features. Prediction performance is measured by MAE, RMSE and Coefficient of Determination (R^2). As shown in Figure 5, our learned region profiles outperforms all the embedding methods.

Thus, it is verified that the comprehensive region profiles can be applied to estimate multiple urban dynamic states with simple prediction models. As if two regions are close in an embedding space, their similar profiles will drive their similar states in multi-domains. In other words, region profiles function as a kind of fixed semantic features to solve multiple urban issues, alleviating the difficulty of acquiring data.

Table 4: Performance comparison on land usage classification

Method	NMI	ARI	F-measure
node2vec	0.676	0.427	0.481
MVURE	0.761	0.571	0.605
MGFN	0.749	0.566	0.581
ReMVC	0.730	0.536	0.589
MetaRSTP	**0.788**	**0.657**	**0.677**

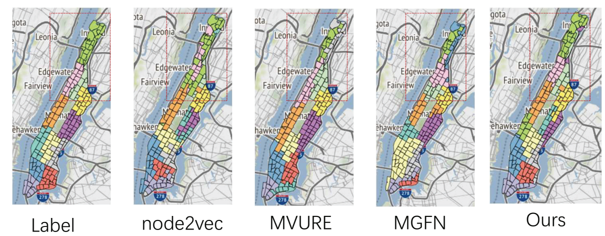


Figure 6: Clustering visualization results in Manhattan.

5.5.2 Clustering Tasks. We further turn to explore the effect of region profiles in land usage clustering task. In specific, we conduct K-means ($k = 12$) clustering with the region profiles as inputs, and regions that belong to the same cluster are categorized as the same type. As shown in Table 4, we evaluate the performance by three metrics: Normalized Mutual Information (NMI), Adjusted Rand Index (ARI) and F-measure. The table illustrates that our learned region profiles well fit clustering tasks and perform best.

We also visualize the clustering results in Figure 6, where regions in the same color are viewed as belong to the same type. It is observed that our results effectively delineate the boundaries of district division and are the closest to the label. It thus shows the interpretation of our designed region profile learner, as a good embedding network should derive the target embeddings which can reflect similarities. Especially in the top part marked with a red rectangle, some clusters are mixed in other embedding methods. Since our method gives more satisfying identification with clearer boundaries, the region functionalities are well preserved. These results further demonstrate the potential applicability of region profiles as powerful semantics in a variety of urban issues.

6 CONCLUSION

In this paper, we propose MetaRSTP, an adaptive meta-learning framework with robust region profiles to improve personalized ST prediction performance in few-shot scenarios. We first learn comprehensive multi-view region profiles via semantic alignment and fusion to uncover the functions of regions. A bias generator further models the inherent correlations between region profiles and ST dynamics. Then, the biased initialization meta-learning strategy results in better region-adaptive prediction in few-shot scenarios due to the combination of region-agnostic regularity and region-specific semantics for fine-tuning. Finally, extensive experiments on real-world datasets demonstrate the superiority of MetaRSTP and the availability of region profiles.

REFERENCES

[1] Bruce F Berg. 2007. New York City Politics: Governing Gotham. *New Republic* (2007), 352.

[2] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. 2020. A Simple Framework for Contrastive Learning of Visual Representations. In *ICML*, Vol. 119. 1597–1607.

[3] Jeongwhan Choi, Hwangyong Choi, Jeehyun Hwang, and Noseong Park. 2022. Graph Neural Controlled Differential Equations for Traffic Forecasting. In *AAAI*. 6367–6374.

[4] Yu Fan, Jiajie Xu, Rui Zhou, Jianxin Li, Kai Zheng, Lu Chen, and Chengfei Liu. 2022. MetaER-TTE: An Adaptive Meta-learning Model for En Route Travel Time Estimation. In *IJCAI*. 2023–2029.

[5] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *ICML*. 1126–1135.

[6] 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 Trans. Knowl. Data Eng.* 31, 3 (2019), 535–548.

[7] Yanjie Fu, Pengyang Wang, Jiadi Du, Le Wu, and Xiaolin Li. 2019. Efficient Region Embedding with Multi-View Spatial Networks: A Perspective of Locality-Constrained Spatial Autocorrelations. In *AAAI*. 906–913.

[8] Aditya Grover and Jure Leskovec. 2016. Node2vec: Scalable Feature Learning for Networks. In *KDD*. 855–864.

[9] Tengda Han, Weidi Xie, and Andrew Zisserman. 2020. Self-supervised Co-training for Video Representation Learning. *CoRR* abs/2010.09709 (2020).

[10] Yilun Jin, Kai Chen, and Qiang Yang. 2022. Selective Cross-City Transfer Learning for Traffic Prediction via Source City Region Re-Weighting. In *KDD*. 731–741.

[11] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. 2018. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In *ICLR*.

[12] Hao Liu, Yongxin Tong, Jindong Han, Panpan Zhang, Xinjiang Lu, and Hui Xiong. 2022. Incorporating Multi-Source Urban Data for Personalized and Context-Aware Multi-Modal Transportation Recommendation. *IEEE Trans. Knowl. Data Eng.* 34, 2 (2022), 723–735.

[13] Bin Lu, Xiaoying Gan, Weinan Zhang, Huaxiu Yao, Luoyi Fu, and Xinbing Wang. 2022. Spatio-Temporal Graph Few-Shot Learning with Cross-City Knowledge Transfer. In *KDD*. 1162–1172.

[14] Yan Luo, Fu-Lai Chung, and Kai Chen. 2022. Urban Region Profiling via Multi-Graph Representation Learning. In *CIKM*. 4294–4298.

[15] Zheyi Pan, Yuxuan Liang, Weifeng Wang, Yong Yu, Yu Zheng, and Junbo Zhang. 2019. Urban Traffic Prediction from Spatio-Temporal Data Using Deep Meta Learning. In *KDD*. 1720–1730.

[16] Zheyi Pan, Wentao Zhang, Yuxuan Liang, Weinan Zhang, Yong Yu, Junbo Zhang, and Yu Zheng. 2022. Spatio-Temporal Meta Learning for Urban Traffic Prediction. *IEEE Trans. Knowl. Data Eng.* 34, 3 (2022), 1462–1476.

[17] Junho Song, Jiwon Son, Dong-hyuk Seo, Kyungsik Han, Namhyuk Kim, and Sang-Wook Kim. 2022. ST-GAT: A Spatio-Temporal Graph Attention Network for Accurate Traffic Speed Prediction. In *CIKM*. 4500–4504.

[18] Huimin Sun, Jiajie Xu, Kai Zheng, Pengpeng Zhao, Pingfu Chao, and Xiaofang Zhou. 2021. MFNP: A Meta-optimized Model for Few-shot Next POI Recommendation. In *IJCAI*. 3017–3023.

[19] W. R. Tobler. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* 46 (1970), 234–240.

[20] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *ICLR*.

[21] Hongjian Wang and Zhenhui Li. 2017. Region Representation Learning via Mobility Flow. In *CIKM*. 237–246.

[22] Shangbin Wu, Xu Yan, Xiaoliang Fan, Shirui Pan, Shichao Zhu, Chuanpan Zheng, Ming Cheng, and Cheng Wang. 2022. Multi-Graph Fusion Networks for Urban Region Embedding. In *IJCAI*. 2312–2318.

[23] Congxi Xiao, Jingbo Zhou, Jizhou Huang, Hengshu Zhu, Tong Xu, Dejing Dou, and Hui Xiong. 2023. A Contextual Master-Slave Framework on Urban Region Graph for Urban Village Detection. In *ICDE*. 736–748.

[24] Yuan Xu, Jiajie Xu, Jing Zhao, Kai Zheng, An Liu, Lei Zhao, and Xiaofang Zhou. 2022. MetaPTP: An Adaptive Meta-optimized Model for Personalized Spatial Trajectory Prediction. In *KDD*. 2151–2159.

[25] Dingqi Yang, Daqing Zhang, Vincent W. Zheng, and Zhiyong Yu. 2015. Modeling User Activity Preference by Leveraging User Spatial Temporal Characteristics in LBSNs. *IEEE Trans. Syst. Man Cybern. Syst.* 45, 1 (2015), 129–142.

[26] Huaxiu Yao, Yiding Liu, Ying Wei, Xianfeng Tang, and Zhenhui Li. 2019. Learning from Multiple Cities: A Meta-Learning Approach for Spatial-Temporal Prediction. In *WWW*. 2181–2191.

[27] Huaxiu Yao, Xianfeng Tang, Hua Wei, Guanjie Zheng, and Zhenhui Li. 2019. Revisiting Spatial-Temporal Similarity: A Deep Learning Framework for Traffic Prediction. In *AAAI*. 5668–5675.

[28] Runsheng Yu, Yu Gong, Xu He, Yu Zhu, Qingwen Liu, Wenwu Ou, and Bo An. 2021. Personalized Adaptive Meta Learning for Cold-start User Preference Prediction. In *AAAI*. 10772–10780.

[29] Liang Zhang, Cheng Long, and Gao Cong. 2022. Region Embedding With Intra and Inter-View Contrastive Learning. *TKDE* (2022), 1–6.

[30] Mingyang Zhang, Tong Li, Yong Li, and Pan Hui. 2020. Multi-View Joint Graph Representation Learning for Urban Region Embedding. In *IJCAI*. 4431–4437.

[31] Ruixing Zhang, Liangzhe Han, Boyi Liu, Jiayuan Zeng, and Leilei Sun. 2022. Dynamic Graph Learning Based on Hierarchical Memory for Origin-Destination Demand Prediction. In *IJCAI*. 2383–2389.

[32] Yunchao Zhang, Yanjie Fu, Pengyang Wang, Xiaolin Li, and Yu Zheng. 2019. Unifying Inter-region Autocorrelation and Intra-region Structures for Spatial Embedding via Collective Adversarial Learning. In *KDD*. 1700–1708.

[33] Yingxue Zhang, Yanhua Li, Xun Zhou, and Jun Luo. 2020. cST-ML: Continuous Spatial-Temporal Meta-Learning for Traffic Dynamics Prediction. In *ICDM*. 1418–1423.

[34] Weida Zhong, Qiuling Suo, Abhishek Gupta, Xiaowei Jia, Chunming Qiao, and Lu Su. 2021. MetaTP: Traffic Prediction with Unevenly-Distributed Road Sensing Data via Fast Adaptation. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3 (2021), 141:1–141:28.