Learning spatio-temporal dynamics on mobility networks for adaptation to open-world events

Zhaonan Wang [a], Renhe Jiang [b,*], Hao Xue [c], Flora D. Salim [c], Xuan Song [b], Ryosuke Shibasaki [b], Wei Hu [a], Shaowen Wang [a]

[a] *Department of Geography and Geographic Information Science, University of Illinois Urbana-Champaign, USA*
[b] *Center for Spatial Information Science, The University of Tokyo, Japan*
[c] *School of Computer Science and Engineering, University of New South Wales, Australia*

ARTICLE INFO

ABSTRACT

As a decisive part in the success of Mobility-as-a-Service (MaaS), spatio-temporal dynamics modeling on mobility networks is a challenging task particularly considering scenarios where open-world events drive mobility behavior deviated from the routines. While tremendous progress has been made to model high-level spatio-temporal regularities with deep learning, most, if not all of the existing methods are neither aware of the dynamic interactions among multiple transport modes on mobility networks, nor adaptive to unprecedented volatility brought by potential open-world events. In this paper, we are therefore motivated to improve the canonical spatio-temporal network (ST-Net) from two perspectives: (1) design a heterogeneous mobility information network (HMIN) to explicitly represent intermodality in multimodal mobility; (2) propose a memory-augmented dynamic filter generator (MDFG) to generate sequence-specific parameters in an on-the-fly fashion for various scenarios. The enhanced event-aware spatio-temporal network, namely **EAST-Net**, is evaluated on several real-world datasets with a wide variety and coverage of open-world events. Both quantitative and qualitative experimental results verify the superiority of our approach compared with the state-of-the-art baselines. What is more, experiments show generalization ability of EAST-Net to perform zero-shot inference over different open-world events that have not been seen.

1. Introduction

Mobility-as-a-Service (MaaS), as an emerging paradigm of transport service, seamlessly integrates multimodal mobility services (*e.g.* public transport, ride-hailing, bike-sharing), which streamlines trip planning, ticketing (for users), operating optimization, emergency response (for providers), and traffic management (for city managers). For a smooth operation of MaaS, spatio-temporal dynamics modeling on multimodal mobility networks is indispensable. However, the existing methods either implicitly handle the interactions between different modes of transportation or assume it to be time-invariant [1]. This task is even more challenging in scenarios where open-world events (*e.g.,* holiday, severe weather, epidemic) take place and deviate collective human mobility significantly from routine behaviors (*e.g.,* daily, weekday patterns). Moreover, as illustrated in Fig. 1, the impacts of different events
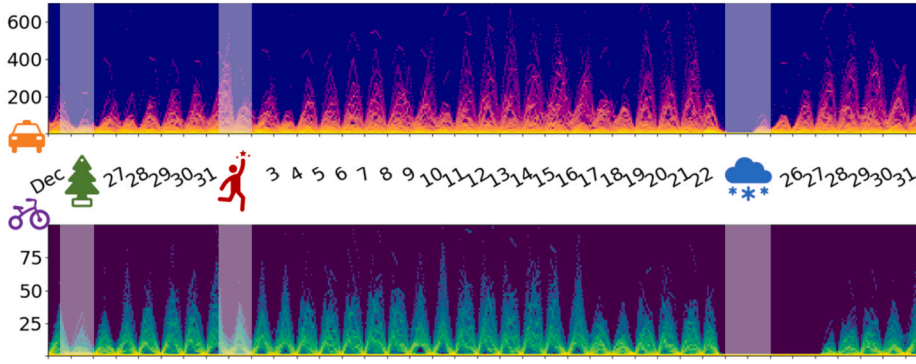

Fig. 1. Time series histograms of citywide *taxi* and *share bike* demands in Washington DC from 24 Dec. 2015 to 31 Jan. 2016, during which Christmas, New Year, and a major blizzard "*Jonas*" took place.

differ, *e.g.* taxi demand spiked on New Year's eve but vanished at Christmas and during the blizzard, and the volatility brought to each transport mode varied, *e.g.* recovery of share bike demand takes longer than the one of taxi after the blizzard.

To model such complex spatio-temporal dynamics, deep learning has been recently utilized to handle potential non-linearity underlying the data [2–6], which is beyond the classical autoregressive moving average (ARMA) based time series models. While the stationarity is not a prior assumption, most, if not all, of the deep learning-based spatio-temporal models are trained to fit general patterns, such as sequential or periodic orders (*e.g.,* daily, weekly). In the real-world context, these patterns could easily fail whenever an open-world event takes place and causes sudden changes. By far, the volatility brought by open-world events is downplayed and usually handled by simple rectifications, such as incorporating temporal covariates (*e.g.* time-of-day, day-of-week, whether-holiday) as auxiliary input [7,8], adding a memory bank to reuse similar patterns in history [9,10]. These manipulations to a certain degree bring time and holiday awareness, but they mainly help with the periodic and precedent parts and would still fail under more extreme scenarios like unprecedented events (*e.g.* historic blizzard, COVID-19 pandemic). There is another line of research [11–13] attempting to capture anomalous mobility tendency under events in an online fashion based on low-order Markov assumption and fine-grained time slot setting. These practices are arguably circumventing the inherent difficulty, namely non-stationarity, of the task instead of truly tackling it.

In this paper, we are motivated to tackle the identified twofold unawareness of the existing spatio-temporal networks, namely intermodality-unaware and event-unaware, correspondingly via: (1) explicitly representing the dynamic interactions among multiple modes on mobility networks; (2) intrinsically enhancing awareness and adaptivity of spatio-temporal models to various open-world events, even including unprecedented ones. Specifically, we design a heterogeneous information network to build the intermodal interactions into the widely adopted spatio-temporal modeling strategy; then leverage techniques of memory-augmented and dynamic filter networks that encourage the model to learn to distinguish and generalize to diverse open-world event scenarios without relying on external features. Based on the above two motivations, we propose an Event-Aware Spatio-Temporal Network (**EAST-Net**). Our contributions are summarized as follows:

- We design a new heterogeneous mobility information network (HMIN) to explicitly represent intermodal interactions (or inter-modality) for spatio-temporal dynamics modeling on mobility networks.
- We propose a novel memory-augmented dynamic filter generator (MDFG) to produce sequence-specific parameters on-the-fly, intrinsically improving both event-awareness and adaptivity of spatio-temporal predictive models.
- We conduct a series of experiments on five real-world mobility datasets with impactful open-world events, and the results quantitatively and qualitatively validate the adaptation and generalization capabilities of EAST-Net.

The remainder of this paper is organized as follows. Section 2 reviews related work within the scope of network-based spatio-temporal forecasting and adaptation. Section 3 describes the necessary definitions and problem formulation of this work. Section 4 presents the details of each main component as well as our proposed model, followed by Section 5 with a series of research questions for evaluating our method in open-world events. In Section 6, we discuss the implications, limitations and future directions of this research.

2. Related work

2.1. Network-based spatio-temporal forecasting

The first line of related studies stems from the recent advancement of Graph Neural Networks (GNNs) [14,15]. Leveraging the natural graph structure of spatial networks (e.g., road networks, mobility networks), network-based traffic forecasting was among the first major applications using GNNs [16,4]. Upon development, researchers adopted the intuitive measurement of physical distance between sensors to construct graph for GNN input. Then, it was found out that the pre-defined graph, based on empirical law (e.g., the

First Law of Geography), was not necessarily the optimal topology structure to facilitate forecasting. Thus, the idea to parameterize graph structure and jointly optimize with GNN was introduced into traffic forecasting [5,17,18]. This technique is known as graph structure learning (GSL).

A similar trend, namely transitioning from empirical network definition to GSL, can be observed in other spatio-temporal forecasting applications. For example, for transportation demand forecasting, region-wise connectivity or functional similarity [19] was substituted by learnable graph [20]; in infectious disease modeling, use of human mobility networks (e.g., origin-destination matrix) [21] evolved to causal graph learning [22]. In some cases, there is even not grounded understanding about the underlying graph structure in observations, such as pure time series [23,24], and spatio-temporal events [25], in which GSL plays an indispensable role. Similarly, there is little prior knowledge that can be utilized to guide modeling of multimodal dynamics on mobility networks. In this work, we take advantage of GSL to jointly represent intermodal and spatio-temporal dependencies on the proposed heterogeneous mobility information network (HMIN).

2.2. Spatio-temporal adaptation

2.2.1. Spatial adaptation

While facilitating end-to-end GNN optimization to a certain task (i.e., forecasting), GSL brings the potential challenge of generalization. To be specific, the optimized graph structure is only suitable for forecasting tasks on the trained location. For this problem, there is a recent line of research that aims to learn an adaptable graph structure from one or multiple source domains to a target domain [26–28]. This problem is known as spatial adaptation, namely the generalization of GSL to unseen spatial domains. While the work focus of this work is on the temporal adaptation, we also show the applicability of the proposed memory-augmented module to this task.

2.2.2. Temporal adaptation

From the temporal perspective, forecasting is by nature an extrapolation problem. In contrast to interpolation [29], in which the underlying data distribution is basically known, forecasting comes with more uncertainty that an unseen open-world event can cause sudden and unannounced changes in the environment. In order to regain acceptable performance, the prediction model must be able to detect, characterize, and adapt to such event with limited time and experience. This problem is well known as concept drift adaptation [30,31]. Here we briefly summarize the existing solutions by employed techniques:

(1) Normalization: Non-stationary Transformer [32] is a good example in this category by transforming input time series into a normal distribution.
(2) Re-training: DDG-DA [33] utilizes a drift detector to determine the time to retrain a new model with the most recent observations.
(3) Ensemble: some methods [34,35] pre-train a set of predictors or patterns and use a classifier to find the right match for prediction under drift.
(4) Online/incremental learning [36]. Similarly to MemDA [37], our proposed method is built upon the last two categories by learning multiple prototypical spatio-temporal dynamics on mobility networks with memory bank and further generating dynamic filter parameters on-the-fly to detect and adapt to open-world events respectively.

3. Preliminaries

In this section, we firstly formulate *multimodal mobility nowcasting* problem, then briefly revisit a standard solution, namely spatio-temporal network (ST-Net), for this task.

It is noteworthy that throughout this paper we use the term set {*modal, multimodal, intermodal, intermodality*} particularly as derivatives of *mode* of mobility, or of transportation, rather than *modality of data* as in common usage.

3.1. Problem definition

Given a specified spatio-temporal granularity, the time and space can be discretized into a set of equal-length time slots and regions (not necessarily equal-area), respectively, denoted by $\mathcal{T} = \{\tau_t | t \in (1, \cdots, T)\}$ and $\mathcal{R} = \{\eta_n | n \in (1, \cdots, N)\}$. Considering there are in total M modes of mobility, we can build a *multimodal mobility tensor* $\underline{\mathcal{M}} \in \mathbb{R}^{T \times N \times C}$, where $C = 2 \cdot M$ if modeling the supply and demand of multiple transport modes; $C = M$ if modeling the visit volume of multiple travel purposes. Accordingly, *multimodal mobility nowcasting* problem can be formulated as follows:

Given α-step consecutive observations in $\underline{\mathcal{M}}$, denoted by $(\mathbf{X}_{t-\alpha+1}, \cdots, \mathbf{X}_t)$, where $\mathbf{X} \in \mathbb{R}^{N \times C}$, return the immediate expectations for the next β-step, *i.e.* $(\hat{\mathbf{X}}_{t+1}, \cdots, \hat{\mathbf{X}}_{t+\beta})$. Note that auxiliary temporal covariates can be available from time slot $\tau_{t-\alpha+1}$ to $\tau_{t+\beta}$, denoted by $\mathbf{T}_{cov} \in \mathbb{R}^{(\alpha+\beta) \times \upsilon}$, where υ is total number of the covariates. Formally written as:

$$\hat{\mathbf{X}}_{t+1}, \cdots, \hat{\mathbf{X}}_{t+\beta} =$$

$$\underset{\mathbf{X}_{t+1}, \cdots, \mathbf{X}_{t+\beta}}{\text{argmax}} \ log \, P \left(\mathbf{X}_{t+1}, \cdots, \mathbf{X}_{t+\beta} | \mathbf{X}_{t-\alpha+1}, \cdots, \mathbf{X}_t; \mathbf{T}_{cov} \right) \tag{1}$$

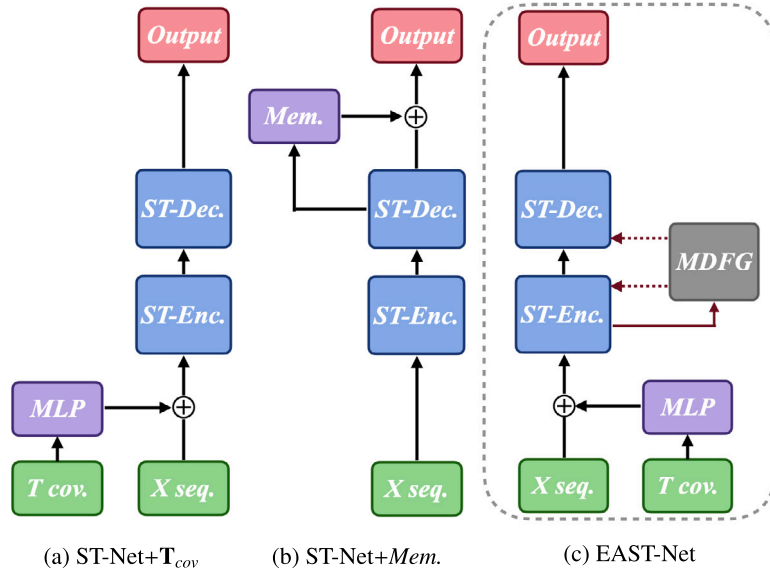(a) ST-Net+$\mathbf{T}_{cov}$ (b) ST-Net+*Mem.*   (c) EAST-Net

Fig. 2. Comparison of Abstract Structure between the Existing ST-Net with Rectifications (a) (b) and Proposed EAST-Net (c).

3.2. Spatio-temporal network

To solve the above problem, recent studies commonly exploit high-level spatio-temporal dependency in observations [3,4,1,23]. Particularly, convolutional and recurrent neural networks (*e.g.* CNN, GCN, TCN, RNN) are two typical submodules utilized to handle the underlying dependencies over the space $\mathscr{R}$ and time $\mathscr{T}$, respectively. This class of models arguably share a similar spatio-temporal view, which prioritizes the first and second dimensions in $(\mathbf{X}_{t-\alpha+1}, \cdots, \mathbf{X}_t) \in \mathbb{R}^{\alpha \times N \times C}$. We term this modeling strategy *Spatio-Temporal Graph* (STG), as demonstrated in Fig. 3, in which the third dimension of the observations is treated as features evolving on STG. We further term models built on top of STG *Spatio-Temporal Network* (ST-Net). Without loss of generality, we combine GCN and RNN to denote a ST-Net, which handles spatial dependency by a graph and temporal dependency in a recurrent form:

$$\mathbf{H} = \sigma(\mathbf{X} \star_{\mathcal{G}} \Theta) = \sigma(\sum_{k=0}^{K} \tilde{\mathcal{P}}^k \mathbf{X} \mathbf{W}_k) \tag{2}$$

$$\begin{cases} \mathbf{u}_t = sigmoid([\mathbf{X}_t^{(l)}, \mathbf{H}_{t-1}^{(l)}] \star_{\mathcal{G}} \Theta_{\mathbf{u}} + b_{\mathbf{u}}) \\ \mathbf{r}_t = sigmoid([\mathbf{X}_t^{(l)}, \mathbf{H}_{t-1}^{(l)}] \star_{\mathcal{G}} \Theta_{\mathbf{r}} + b_{\mathbf{r}}) \\ \mathbf{C}_t = tanh([\mathbf{X}_t^{(l)}, (\mathbf{r}_t \odot \mathbf{H}_{t-1}^{(l)})] \star_{\mathcal{G}} \Theta_{\mathbf{C}} + b_{\mathbf{C}}) \\ \mathbf{H}_t^{(l)} = \mathbf{u}_t \odot \mathbf{H}_{t-1}^{(l)} + (1 - \mathbf{u}_t) \odot \mathbf{C}_t \end{cases} \tag{3}$$

Equation (2) defines the basic graph convolution operation $\star_{\mathcal{G}}$, which takes input $\mathbf{X} \in \mathbb{R}^{N \times p}$ and returns $\mathbf{H} \in \mathbb{R}^{N \times q}$ given a graph topology matrix $\mathcal{P} \in \mathbb{R}^{N \times N}$ ($\tilde{\mathcal{P}}$ is its normalized form), approximation order K, and trainable parameters $\Theta \in \mathbb{R}^{(K+1) \times p \times q}$. Equation (3) defines an extended version of GRU (a form of RNN), namely GCRU, with matrix multiplications replaced by graph convolutions (Equation (2)). It is noteworthy that DCGRU [4] can be seen as a special form of GCRU by restricting $\tilde{\mathcal{P}}$ to be random walk normalized transition matrix and performing bidimensional graph diffusion. Then, stacking multiple layers (denoted by l) of GCRU forms encoder and decoder of a ST-Net, abbreviated as *ST-Enc/ST-Dec* in Fig. 2.

Besides, as illustrated in Fig. 2a, temporal covariates can be used as auxiliary input [7,8] to equip ST-Net with time and holiday awareness. In this case, $\mathbf{X}_t^{(0)} = [\mathbf{X}_t, \mathbf{T}_t']$, where [,] denotes a concatenation operation and $\mathbf{T}_t'$ is the linear projected representation of $\mathbf{T}_{cov}$ at τ_t. Another rectification for ST-Net (demonstrated in Fig. 2b) attaches an external memory bank [9,10] to the decoder such that some typical spatio-temporal patterns can be stored for reuse. This memory is implemented by a parameter matrix $\mathbf{M} \in \mathbb{R}^{m \times D}$, where m and d denote the total number of memory records and dimension of each one. Before making the final output, decoder makes a query to $\mathbf{M}$ for find similar representations, which is implemented by attention mechanism [38,39]. Formally,

$$\begin{cases} Q_t = \overline{\mathbf{H}}_t^{(l)} \mathbf{W}_Q + b_Q \\ \phi_j = \dfrac{e^{Q_t * \mathbf{M}[j]}}{\sum_{j=1}^{m} e^{Q_t * \mathbf{M}[j]}} \\ \tilde{\mathbf{V}} = (\sum_{j=1}^{m} \phi_j \cdot \mathbf{M}[j]) \mathbf{W}_V + b_V \end{cases} \tag{4}$$
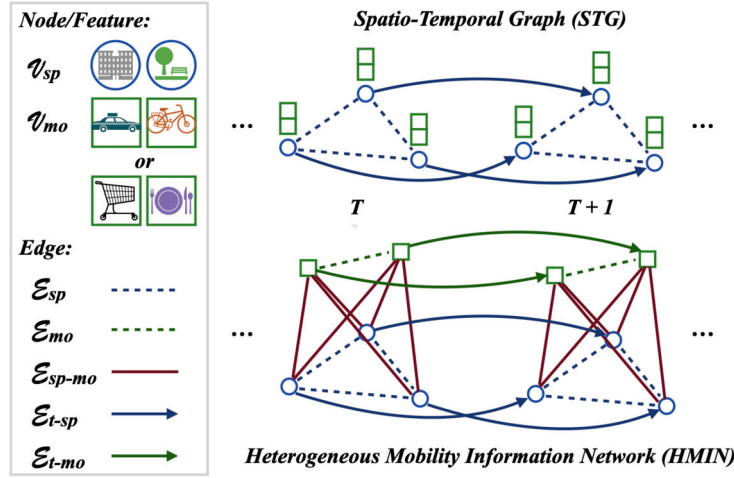
Fig. 3. Comparison between Spatio-Temporal Graph (STG) and Heterogeneous Mobility Information Network (HMIN) for Multimodal Mobility Modeling.

where $Q_t \in \mathbb{R}^D$ denotes the query vector projected from flattened $\mathbf{H}_t^{(l)}$; ϕ_j is the attention score corresponding to j-th memory record. The obtained vector $\tilde{\mathbf{V}}$ can be reshaped back and concatenated with $\mathbf{H}_t^{(l)}$ for output, $\mathbf{H}_t^{(out)} = [\mathbf{H}_t^{(l)}, \mathbf{V}]$.

4. Methodology

In this section, we elaborate the motivations and techniques for improving ST-Net, and present <u>E</u>vent-<u>A</u>ware <u>S</u>patio-<u>T</u>emporal Network (**EAST-Net**) as a more adaptive framework for multimodal mobility nowcasting.

4.1. Heterogeneous mobility information network

As presented in Fig. 3, STG, the fundamental of ST-Net, prioritizes spatio-temporal modeling while restricting all features (*i.e.* mobility modes) to evolve together on the fixed STG. We argue that this spatio-temporal view restricts the modeling of dynamic interactions among different modes of mobility, which is in fact the operating mechanism of MaaS. As demonstrated in Fig. 1, an open-world event may impact different transport modes variously, which confirms the necessity for intermodality modeling. Thus, we are motivated to design a new underlying structure, *i.e. Heterogeneous Mobility Information Network* (HMIN), to jointly represent intermodal interaction and spatio-temporal dependency.

As illustrated in Fig. 3, HMIN is defined as $\mathcal{G} = (\mathcal{V}_{sp} \cup \mathcal{V}_{mo}, \mathcal{E}_{sp} \cup \mathcal{E}_{mo} \cup \mathcal{E}_{sp\text{-}mo} \cup \mathcal{E}_{t\text{-}sp} \cup \mathcal{E}_{t\text{-}mo})$, where $\mathcal{V}_{sp} = \{\eta_1, \cdots, \eta_N\}$ and $\mathcal{V}_{mo} = \{\mu_1, \cdots, \mu_M\}$ denote node set of regions and mobility modes, respectively; $\mathcal{E}_{sp}$, $\mathcal{E}_{mo}$, $\mathcal{E}_{sp\text{-}mo}$, $\mathcal{E}_{t\text{-}sp}$, $\mathcal{E}_{t\text{-}mo}$ denote five edge sets for the relations in region-to-region, mode-to-mode, region-to-mode, time-to-region, time-to-mode. By this definition, the intermodal relationship and its dynamicity can be represented by $\mathcal{E}_{mo}$ and $\mathcal{E}_{t\text{-}mo}$; and the task of *multimodal mobility nowcasting* is reformulated as a *link prediction* task for edge set $\mathcal{E}_{sp\text{-}mo}$ ($|\mathcal{E}_{sp\text{-}mo}| = N \cdot C$) from τ_{t+1} to $\tau_{t+\beta}$.

Here we propose a simple yet generic framework to encode-decode HMIN by applying handy GCRU in a similar fashion to ST-Net. Denoting the framework of *ST-Enc/ST-Dec* (GCRU in multi-layer) as $\mathbf{H}_{t+1}, \cdots, \mathbf{H}_{t+\beta} = \text{GCRU}_{\text{Enc-Dec}}(\mathbf{X}_{t-\alpha+1}, \cdots, \mathbf{X}_t)$, processing of HMIN can be decomposed into two views (*i.e.* spatial and intermodal) followed by a stepwise fusion layer, formally denoted by:

$$
\begin{cases}
\mathbf{H}_{t+1}^{(sp)}, \cdots, \mathbf{H}_{t+\beta}^{(sp)} = \text{GCRU}_{\text{Enc-Dec}}^{(sp)}(\mathbf{X}_{t-\alpha+1}, \cdots, \mathbf{X}_t) \\
\mathbf{H}_{t+1}^{(mo)}, \cdots, \mathbf{H}_{t+\beta}^{(mo)} = \text{GCRU}_{\text{Enc-Dec}}^{(mo)}(\mathbf{X}_{t-\alpha+1}^{\top}, \cdots, \mathbf{X}_t^{\top}) \\
\hat{\mathbf{X}}_{t+\varepsilon} = \sigma(\mathbf{H}_{t+\varepsilon}^{(sp)} \mathbf{W}_{out} \mathbf{H}_{t+\varepsilon}^{(mo)\top})
\end{cases}
\tag{5}
$$

where $\varepsilon \in (1, \cdots, \beta)$ denotes the step index within horizon β; $\mathbf{H}_{t+\varepsilon}^{(sp)} \in \mathbb{R}^{N \times q}$ and $\mathbf{H}_{t+\varepsilon}^{(mo)} \in \mathbb{R}^{C \times q}$ denote spatial and modal embeddings on $\mathcal{V}_{sp}$ and $\mathcal{V}_{mo}$ at time slot $\tau_{t+\varepsilon}$, respectively; $\mathbf{W}_{out} \in \mathbb{R}^{q \times q}$ denotes a parameter matrix to fuse the node embeddings for link generation. For simplicity, we denote this framework by HMINet (Equation (5)) and consider it to be a general case of ST-Net, which only takes the spatial view and let $\mathbf{W}_{out} \in \mathbb{R}^{q \times C}$, $\mathbf{H}_{t+\varepsilon}^{(mo)} = \mathbf{I}_C$. Essentially, edge sets $\mathcal{E}_{sp}$ and $\mathcal{E}_{mo}$, representing spatial and intermodal dependencies, are handled by graph convolution in each domain; unidirectional temporal edges $\mathcal{E}_{t\text{-}sp}$ and $\mathcal{E}_{t\text{-}mo}$ are encoded by the recurrent structure; HMINet learns the mapping from α-step to β-step in edge set $\mathcal{E}_{sp\text{-}mo}$.

4.2. Memory-augmented dynamic filter generator

Although enhancing ST-Net in an intermodality-aware way, HMINet introduces extra parameters by approximately same amount that ST-Net has. To control the model size and, more importantly, empower it to be aware of and adaptive to various scenarios, we propose a novel *Memory-augmented Dynamic Filter Generator* (MDFG).
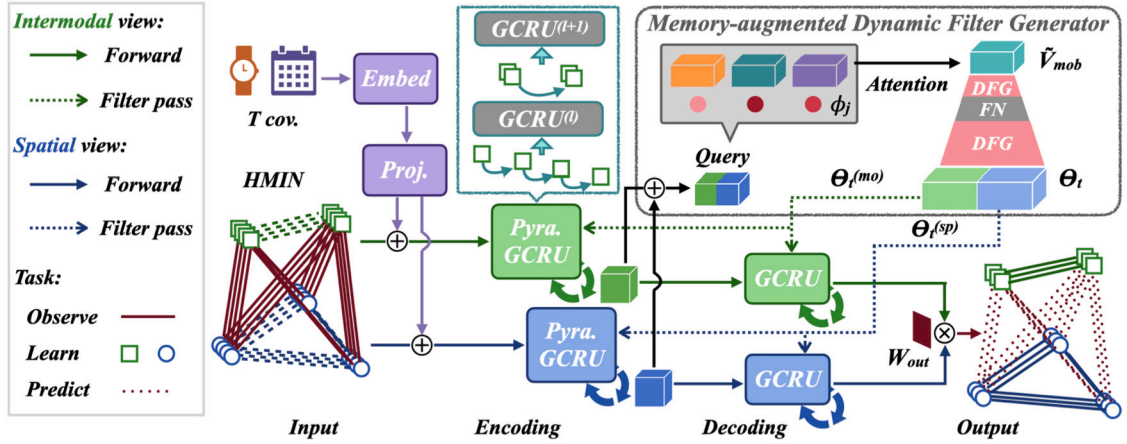
Fig. 4. Framework of EAST-Net: (1) Input Multimodal Mobility Tensor and Temporal Covariates; (2) Encode with Heterogeneous Mobility Information Network in Two Pyramidal GCRU Branches; (3) Query Memory-augmented Dynamic Filter Generator to Produce Sequence-specific Graph Convolution Kernels; (4) Decode with GCRU and Generate Links.

MDFG is motivated by a line of research on dynamic filter networks (DFN) [40–42], which have been mainly studied on convolutional kernels for image and video-related tasks. The core idea behind DFN is instead of sharing a same trainable filter for all samples in a dataset, dynamically generating filters conditioned on an input sample, which by nature increases the flexibility and adaptivity of model. In light of DFN, we argue that the indistinguishability between normal and event scenarios roots in the way that a same set of parameters (*e.g.* $\Theta' = [\Theta_{\mathbf{u}}, \Theta_{\mathbf{r}}, \Theta_{\mathbf{C}}]$ in Equation (3)) is shared for all observational sequences $(\mathbf{X}_{t-\alpha+1}, \cdots, \mathbf{X}_t)$ by vanilla ST-Net. In other words, parameters in ST-Net are sequence-agnostic. We thereby utilize the idea of DFN and further put parameters conditioned on a plugin memory bank $\mathbf{M}_{mob} \in \mathbb{R}^{m \times D}$ to encourage discovery of high-level mobility prototypes, which are representations incorporating spatial, temporal, and multimodal knowledge. To be specific, $\mathbf{M}_{mob}$ takes concatenated node embeddings $[\mathbf{H}_t^{(sp)}, \mathbf{H}_t^{(mo)}] \in \mathbb{R}^{(N+C) \times q}$ as a query and returns a reconstructed prototype vector $\tilde{\mathbf{V}}_{mob} \in \mathbb{R}^D$, which further passes through a DFN to produce momentary filters $[\Theta_t^{(sp)}, \Theta_t^{(mo)}]$ for $\text{GCRU}_{\text{Enc-Dec}}^{(sp)}$ and $\text{GCRU}_{\text{Enc-Dec}}^{(mo)}$. This interaction between HMINet and MDFG occurs in an on-the-fly manner, which generates sequence-specific parameters (denoted by Θ_t). Formally,

$$\begin{cases} \mathbf{Q}_t = [\overline{\mathbf{H}}_t^{(sp)}, \overline{\mathbf{H}}_t^{(mo)}]\mathbf{W}_Q + b_Q \\ \phi_j = \dfrac{e^{\mathbf{Q}_t * \mathbf{M}_{mob}[j]}}{\sum_{j=1}^{m} e^{\mathbf{Q}_t * \mathbf{M}_{mob}[j]}} \\ \tilde{\mathbf{V}}_{mob} = \sum_{j=1}^{m} \phi_j \cdot \mathbf{M}_{mob}[j] \\ \Theta_t = [\Theta_t^{(sp)}, \Theta_t^{(mo)}] = \int(\varphi(\int(\tilde{\mathbf{V}}_{mob}))) \end{cases} \tag{6}$$

where $\int$ denotes a dynamic filter generation (DFG) layer, which can be implemented in various ways. Without loss of generality, we utilize a linear projection in this case. φ denotes a filter normalization (FN) layer [42], used to normalize the generated parameters and avoid gradient vanishing and exploding.

4.3. Open-world spatio-temporal network

Based on HMINet and MDFG, we further make three refinements to the framework of *Event-Aware Spatio-Temporal Network* (EAST-Net), as illustrated in Fig. 4, which can be trained in an end-to-end fashion by minimizing a specified loss function using the standard backpropagation.

- *Temporal covariates* are fused stepwise for basic time and holiday awareness for both spatial and intermodal views, following the common practice [9].

$$\mathbf{X}_{t+\varepsilon}^{(0)} = \begin{cases} [\mathbf{X}_{t+\varepsilon}, \mathbf{T}'_{t+\varepsilon}] & , if \; \varepsilon \in (1-\alpha, \cdots, 0) \\ [\hat{\mathbf{X}}_{t+\varepsilon}, \mathbf{T}'_{t+\varepsilon}] & , if \; \varepsilon \in (1, \cdots, \beta) \end{cases} \tag{7}$$

Then, $\mathbf{X}_{t+\varepsilon}^{(0)}$ is fed into HMINet (Equation (5)) as input.

- *Pyramidal structure* [8] is leveraged in GCRU encoders to help accelerate the training of HMINet and discover multi-level temporal pattern for mobility prototype extraction. In a case by a factor of 2:

Algorithm 1 EAST-Net.

Input: $(\mathbf{X}_{t-\alpha+1}, \cdots, \mathbf{X}_t) \in \underline{\mathcal{M}}$, $\mathbf{T}_{cov} \in \mathbb{R}^{(\alpha+\beta) \times \upsilon}$
Parameter: $\mathbf{M}_{mob} \in \mathbb{R}^{m \times \bar{D}}$, $\int \in \mathbb{R}^{D \times |\Theta|}$, $\mathbf{W}_{out} \in \mathbb{R}^{q \times q}$
Output: $(\hat{\mathbf{X}}_{t+1}, \cdots, \hat{\mathbf{X}}_{t+\beta})$

1: Initialize $\mathbf{M}_{mob}$, $\int_{proj}$ randomly; initialize $\mathbf{H}_i^{(sp\text{-}0)}$, $\mathbf{H}_i^{(mo\text{-}0)}$ with zeros
2: Query MDFG for filter Θ_t' and graphs $\mathcal{P}_t^{(sp)}$, $\mathcal{P}_t^{(mo)}$ of Pyra-GCRU$_t^{(sp)}$, Pyra-GCRU$_t^{(mo)}$ # Equation (6)
3: Embed and project $\mathbf{T}_{cov}$ to $\mathbf{T}'^{(sp)}$, $\mathbf{T}'^{(mo)}$
4: $\mathbf{X}_{t-\varepsilon}^{(sp\text{-}0)} = [\mathbf{X}_{t-\varepsilon}, \mathbf{T}_{t-\varepsilon}'^{(sp)}]$, $\mathbf{X}_{t-\varepsilon}^{(mo\text{-}0)} = [\mathbf{X}_{t-\varepsilon}^T, \mathbf{T}_{t-\varepsilon}'^{(mo)}]$ for $\varepsilon \in (\alpha-1, \cdots, 0)$
5: **for** $l = 1, \cdots, L$ **do**
6:    **for** $i = t - \alpha + 1, \cdots, t$ **do**
7:       $\mathbf{H}_{i+1}^{(sp\text{-}l)} = $ Pyra-GCRU$_t^{(sp)}(\mathbf{H}_i^{(sp\text{-}l)}, \Theta_t'^{(sp)}, \mathcal{P}_t^{(sp)})$
8:       $\mathbf{H}_{i+1}^{(mo\text{-}l)} = $ Pyra-GCRU$_t^{(mo)}(\mathbf{H}_i^{(mo\text{-}l)}, \Theta_t'^{(mo)}, \mathcal{P}_t^{(mo)})$
9:    **end for**
10: **end for**
11: Update $[\mathbf{H}_t^{(sp\text{-}l)}, \mathbf{H}_t^{(mo\text{-}l)}]$ for MDFG query
12: Initialize a zero tensor $\mathbf{X}_{i-1} \in \mathbb{R}^{N \times C}$
13: **for** $i = t + 1, \cdots, t + \beta$ **do**
14:    $\mathbf{X}_i^{(sp\text{-}0)} = [\mathbf{X}_{i-1}, \mathbf{T}_i'^{(sp)}]$, $\mathbf{X}_{i-1}^{(mo\text{-}0)} = [\mathbf{X}_{i-1}^T, \mathbf{T}_i'^{(mo)}]$
15:    **for** $l = 1, \cdots, L$ **do**
16:       $\mathbf{H}_i^{(sp)\text{-}(l+1)} = $ GCRU$_t^{(sp)}(\mathbf{H}_i^{(sp\text{-}l)}, \Theta_t'^{(sp)}, \mathcal{P}_t^{(sp)})$
17:       $\mathbf{H}_i^{(mo)\text{-}(l+1)} = $ GCRU$_t^{(mo)}(\mathbf{H}_i^{(mo\text{-}l)}, \Theta_t'^{(mo)}, \mathcal{P}_t^{(mo)})$
18:    **end for**
19:    $\hat{\mathbf{X}}_i = \sigma(\mathbf{H}_i^{(sp)} \mathbf{W}_{out} \mathbf{H}_i^{(mo)\top})$
20: **end for**

$$\mathbf{H}_t^{(l+1)} = [\mathbf{H}_{2t}^{(l)}, \mathbf{H}_{2t+1}^{(l)}] \tag{8}$$

- *Adaptive edge sets* $\mathcal{E}_{sp}, \mathcal{E}_{mo}$ are learnt in HMINet without making any prior assumptions on either intermodal or spatial relationship [4,5]. Essentially, a pair of parameterized node embeddings are initialized for both GCRU$_{\text{Enc-Dec}}^{(sp)}$ and GCRU$_{\text{Enc-Dec}}^{(mo)}$ to derive corresponding topology for graph convolutions:

$$\begin{cases} \tilde{\mathcal{E}}_{(sp)} = softmax(relu(\mathbf{E}_{(sp)}\mathbf{F}_{(sp)}^\top)) \\ \tilde{\mathcal{E}}_{(mo)} = softmax(relu(\mathbf{E}_{(mo)}\mathbf{F}_{(mo)}^\top)) \end{cases} \tag{9}$$

where embeddings $\mathbf{E}_{(sp)}, \mathbf{F}_{(sp)} \in \mathbb{R}^{N \times \mu_{sp}}$ and $\mathbf{E}_{(mo)}, \mathbf{F}_{(mo)} \in \mathbb{R}^{C \times \mu_{mo}}$ are trained to learn the underlying region-to-region and mode-to-mode dependencies within node sets $\mathcal{V}_{sp}$ and $\mathcal{V}_{mo}$; the derived topology is normalized to $[0, 1]$ by softmax to simulate signal diffusion in each domain (replacing $\tilde{\mathcal{P}}$ in Equation (2)).

5. Experiment

5.1. Datasets

To evaluate the proposed method, we collect five real-world datasets with different spatio-temporal scales/coverage and represent *multimodal mobility* with *transport modes* on three city-level datasets (for New York City, Washington DC, Chicago in Table 1), and with *travel purpose* on one state and one country-level dataset (for Florida and the United States in Table 2). Fig. 5 illustrates the spatial granularity and temporal coverage of these datasets. Essentially, three city-level datasets (JONAS-{NYC, DC}, COVID-CHI) are grid-based, while the other two (DORIAN-FL, COVID-US) is graph-structured by treating each county/state as a node and the
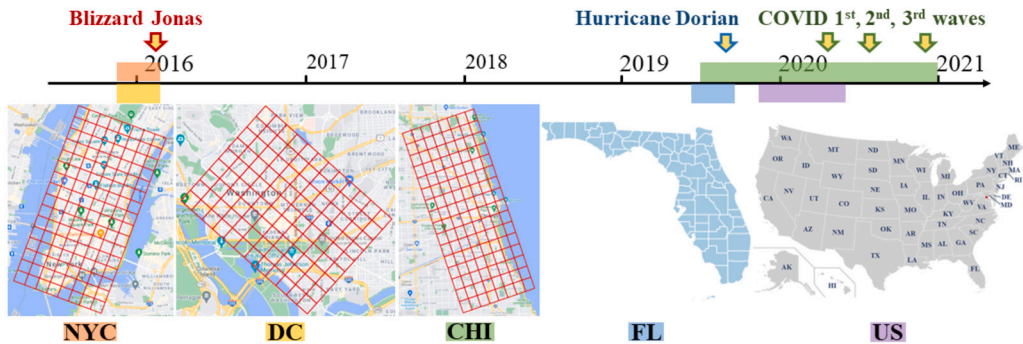

Fig. 5. Spatio-Temporal Coverage of Five Collected Mobility Datasets with Open-World Events.

Table 1

Summary of Three Open-World Datasets of *Transport Mode*.

Dataset	JONAS-NYC		JONAS-DC		COVID-CHI	
Span	$2015/10/24 \sim 2016/1/31$		$2015/10/24 \sim 2016/1/31$		$2019/7/1 \sim 2020/12/31$	
Temporal	100 days	by 30-minute	100 days	by 1-hour	550 days	by 2-hour
Spatial	16×8 grid	in $0.5 \times 0.5\ km$	9×12 grid	in $0.5 \times 0.5\ km$	14×8 grid	in $1.5 \times 1.2\ km$
Mobility	{Demand, Supply} of {Taxi, Share Bike, Scooter*}					
Event	Thanksgiving, Christmas, New Year, M.L.K. Day, **Blizzard Jonas** $(2016/1/22 \sim 24)$				Holidays, **COVID-19 Pandemic***	

* Scooter trip data is only available in **COVID-CHI** set.

Table 2

Summary of Two Open-World Datasets of *Travel Purpose*.

Dataset	DORIAN-FL	COVID-US
Span	$2019/6/11 \sim 2019/9/18$	$2019/11/14 \sim 2020/5/31$
Temporal	100 days by 1-hour	200 days by 1-hour
Spatial	67 counties of Florida	50 states + DC
Mobility	Point-of-Interest visits of 10 types: {grocery store, retailer, transportation, office, school, healthcare, entertainment, hotel, restaurant, service}[1]	
Event	Independence & Labor Day, **Hurricane Dorian** $(2019/9/1 \sim 3)$	Holidays, **COVID-19 Pandemic***

* COVID-19 pandemic outbroke in late March 2020; **COVID-US** set depicts **the early stage** (first wave in April), and **COVID-CHI** set depicts **the progression** (first to third waves till end of 2020) of the pandemic.

pairwise relationships within as edges. Similarly to the previous studies [3,1,13], trip records (*e.g.* taxi, share bike) or POI visits are processed as in/outflow (supply/demand) or visit volume to be further aggregated onto a given spatio-temporal granularity.

Particularly, each dataset is designed to cover a set of holidays and a historic event with big social impact, *i.e.* the winter storm Jonas, hurricane Dorian, and COVID-19 pandemic. Following the common practice [8,7], we encode temporal covariates of each time slot (*i.e. time-of-day*, *day-of-week*, *month-of-year*, *whether-holiday*) in an one-hot manner as temporal covariates input.

5.2. Settings

We chronologically split each dataset for training, validation, testing with a ratio of $7 : 1 : 2$, such that the lengths of test sets are roughly last 20 days for JONAS-{NYC, DC} and DORIAN-FL, 110 days for COVID-CHI, and 40 days for COVID-US. Lengths of observational and nowcasting sequences are set to $\alpha = 8$ and $\beta = 8$, respectively; number of GCRU layers $L = 2$ with approximation order $K = 3$ and hidden dimension $q = 32$; embedding dimensions for $\mathbf{T}_{cov}$ $v' = 2$, $\mu_{(sp)} = 20$ and $\mu_{(mo)} = 3$; mobility prototype memory $m = 8$ and $D = 16$. We provide a justification for the choices of some important hyper-parameters with Fig. 11.

For model training, batch size $= 32$; learning rate $= 5 \times 10^{-4}$; maximum epoch $= 100$ with an early stopper with a patience of 10; *MAE* is chosen to be optimized using *Adam*. We implement EAST-Net with *PyTorch* and carry out experiments on a GPU server with *NVIDIA GeForce GTX 3080 Ti* graphic cards. For evaluation, we adopt three commonly used metrics, namely Root Mean Square Error (*RMSE*), Mean Absolute Error (*MAE*) and Mean Absolute Percentage Error (*MAPE*).

5.3. Evaluations

In this section, to understand the performance of our approach, we develop a group of research questions and design a series of experiments correspondingly:

(1) How does EAST-Net perform compared with the existing methods?
(2) How does EAST-Net perform compared with its model variants?
(3) How does EAST-Net behave in different scenarios of open-world events?
(4) How does EAST-Net generalize over different scenarios of open-world events?

5.3.1. Quantitative evaluation 1

To quantitatively evaluate the overall prediction accuracy of EAST-Net on the multimodal mobility nowcasting problem, we implement eight baselines on mobility/traffic-related spatio-temporal prediction for comparison, including:

- **Historical Average (HA):** Average the values of same time slot in the training set for prediction.
- **Naive Forecast (NF):** Naively repeat the latest sequence observation for the next β time slots. This practice is proven to be rather effective under events [13].

---

[1] According to the NAICS industry codes (https://www.naics.com/search-naics-codes-by-industry/)

- **Vector AutoRegression (VAR):** A classic multivariate time series (MTS) model that handles linear dependency between variables. In practice, we firstly apply PCA (Principal Component Analysis) to reduce the tensor dimension of mobility networks for VAR, i.e., $\underline{\mathcal{M}} \in \mathbb{R}^{T \times N \cdot C} \to \mathbb{R}^{T \times 100}$.

  It is an advance time-series prediction model designed for high dimensional data
- **Transformer**[†] [39]: The enhanced version [43] with convolutional self-attention is implemented to capture local temporal pattern for univariate time series forecasting.
- **CoST-Net** [1]: A two-stage co-predictive model for multimodal transport demands. It models each mode individually with convolutional auto-encoder and uses a heterogeneous LSTM for collaborative modeling.
- **DCRNN** [4]: A special form of GCRU that requires a pre-defined transition matrix as auxiliary input to perform bidimensional graph diffusion convolutions.
- **Graph WaveNet (GW-Net)** [5]: The spatiotemporal forecasting model that firstly introduces parameterized node embedding for graph structure learning. GW-Net builds upon a fully convolutional structure, using graph and temporal convolutions to handle spatial and sequence dependencies.
- **MTGNN** [23]: A recent MTS forecasting model that features an efficient unidirectional graph constructor and multi-kernel temporal convolutions.
- **StemGNN** [44]: Another recent MTS forecasting model that jointly models spatial and temporal dependencies in the spectral domain.
- **MegaCRN** [18]: The state-of-the-art traffic forecasting model built upon GCRU, with a unique memory-augmented graph structure learning mechanism.

We present the performance comparison of EAST-Net and baselines in Table 3. It is noticeable that the error range on five datasets varies in magnitude: among three city-level sets, DC and CHI have relatively smaller transport volume than NYC; COVID-US is apparently the most tricky set which is state-level, of ten modes for travel purpose, and being tested at the very early stage (first wave) of the pandemic. Besides, acceptable results obtained by HA on JONAS-DC and DORIAN-FL, NF on JONAS-NYC and COVID-CHI indicate a rather strong short-term temporal dependency in JONAS-NYC and COVID-CHI, and a daily periodicity in JONAS-DC and DORIAN-FL. By treating the problem simply as time series, Transformer does not acquire satisfactory accuracy. Taking spatial locality into consideration, CoST-Net performs better than Transformer on JONAS-{NYC, DC}, but the pre-trained convolutional structure not only fails it on COVID-CHI but limits it from handling graph-based data like DORIAN-FL and COVID-US. Then, among four graph-based models, GW-Net prevails in terms of most metrics on all datasets. Lastly, speaking of EAST-Net, we can observe a consistent and dramatic improvement throughout JONAS-{NYC, DC}, DORIAN-FL and COVID-US, which undoubtedly confirms the efficacy of EAST-Net. The exception on COVID-CHI, we think, can be explained by: (1) A coarse time slot (2-hour) setting "smoothes" sudden changes, making the task easier for other models; (2) Along with the progression (first to third waves) of COVID-19 pandemic, other models gradually learn the pandemic pattern as a new normality.

Table 3

Performance of EAST-Net and Baselines on Five Open-World Mobility Datasets.

Model	JONAS-NYC			JONAS-DC			COVID-CHI		
	RMSE	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*
HA	48.95	39.22	75.3%	6.32	3.11	<u>38.9%</u>	37.16	9.94	190.5%
NF	29.93	28.37	59.3%	7.75	3.59	69.0%	12.91	5.66	79.2%
VAR	42.75	30.62	59.2%	5.97	3.22	41.2%	10.95	3.79	50.71%
Transformer[†]	35.05	23.43	47.1%	6.54	2.96	65.7%	12.67	5.11	80.5%
CoST-Net	33.72	22.49	41.0%	6.27	2.97	52.1%	15.26	6.88	83.7%
DCRNN	28.72	<u>18.72</u>	39.0%	5.47	3.07	50.4%	10.57	6.48	51.2%
GW-Net	<u>28.58</u>	19.37	37.0%	5.09	2.33	51.0%	<u>8.37</u>	3.72	**45.4%**
MTGNN	28.87	19.12	<u>36.4%</u>	5.16	2.69	48.0%	8.82	4.35	51.6%
StemGNN	30.71	21.49	40.8%	5.32	3.07	50.4%	8.40	4.50	<u>50.3%</u>
MegaCRN	29.09	19.80	44.84%	<u>4.82</u>	<u>2.29</u>	48.4%	**8.00**	**3.03**	52.1%
EAST-Net	**23.63**	**15.79**	**33.3%**	**4.10**	**2.00**	**35.0%**	9.38	<u>3.38</u>	61.5%
-Δ%	-17.3%	-15.7%	-8.5%	-14.9%	-12.7%	-10.0%	-	-	-

Model	DORIAN-FL			COVID-US		
	RMSE	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*
HA	132.44	45.38	49.5%	2822.12	1218.61	159.7%
NF	293.32	109.43	130.7%	2385.28	1258.12	185.1%
VAR	174.96	51.14	89.4%	1813.53	898.27	326.7%
Transformer[†]	195.58	68.49	69.4%	1767.71	862.82	180.1%
CoST-Net	-	-	-	-	-	-
DCRNN	144.94	47.85	53.3%	1194.38	722.34	155.9%
GW-Net	147.75	45.94	<u>47.8%</u>	<u>1022.82</u>	<u>490.97</u>	77.6%
MTGNN	142.76	<u>42.66</u>	**46.2%**	1083.00	535.61	<u>75.9%</u>
StemGNN	165.45	51.17	53.0%	1279.04	709.16	146.7%
MegaCRN	<u>123.72</u>	57.39	61.3%	1269.08	562.69	125.0%
EAST-Net	**107.12**	**38.13**	50.6%	**799.51**	**371.78**	**51.8%**
-Δ%	-13.4%	-13.0%	-	-21.8%	-24.3%	-31.8%

Table 4
Performance of EAST-Net and Its Variants on Five Open-World Mobility Datasets.

Variant	JONAS-NYC			JONAS-DC			COVID-CHI		
	RMSE	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*
ST-Net	31.38	20.22	44.0%	5.44	2.37	56.0%	12.17	5.06	80.0%
ST-Net+T_{cov}	<u>25.35</u>	<u>16.96</u>	<u>34.0%</u>	<u>4.45</u>	<u>2.04</u>	<u>43.1%</u>	**8.67**	**2.82**	59.0%
ST-Net+*Mem.*	30.73	20.16	40.4%	5.08	2.60	44.0%	9.92	3.02	**57.0%**
ST-Net+MDFG	28.00	18.38	38.0%	4.67	2.11	43.2%	<u>9.21</u>	3.83	61.8%
HMINet	28.71	18.21	38.0%	4.57	2.07	48.3%	11.44	4.48	78.5%
EAST-Net	**23.63**	**15.79**	**33.3%**	**4.10**	**2.00**	**35.0%**	9.38	<u>3.38</u>	61.5%

Variant	DORIAN-FL			COVID-US		
	RMSE	*MAE*	*MAPE*	*RMSE*	*MAE*	*MAPE*
ST-Net	144.21	47.58	59.3%	1123.91	519.07	62.2%
ST-Net+T_{cov}	128.59	43.88	55.8%	1434.33	720.08	82.4%
ST-Net+*Mem.*	136.91	46.02	58.6%	1058.52	528.29	63.4%
ST-Net+MDFG	<u>127.22</u>	<u>39.21</u>	<u>52.5%</u>	<u>874.11</u>	<u>398.02</u>	54.1%
HMINet	132.00	41.89	**46.5%**	906.85	399.35	**43.5%**
EAST-Net	**107.12**	**38.13**	50.6%	**799.51**	**371.78**	51.8%

5.3.2. Quantitative evaluation 2

To understand how EAST-Net improves from the canonical ST-Net, we implement ST-Net (in Equation (3)), ST-Net with two rectified forms (adding temporal covariates in Fig. 2a and memory bank 2b) and with Memory-augmented Dynamic Filter Generator (MDFG), as well as HMINet (in Equation (5)) for comparison. As presented in Table 4, within the ST-Net family, a regular memory bank improves ST-Net in most cases, but not as significantly as temporal covariates do. However, adding T_{cov} deteriorates the performance on COVID-US, which is actually reasonable because the reinforced awareness of periodicity backfires especially at the early stage of a historic epidemic when human mobility began to deviate (because of quarantine measures). In comparison, adopting MDFG or HMINet drops all metrics compared with regular ST-Net especially on COVID-US, which validates our motivation for explicit intermodality modeling and event-awareness design. Besides, adopting HMIN on COVID-CHI seems not as helpful as on other datasets. This issue, we think, may be caused by including the scooter data, which is in fact a pilot program in Chicago and thus has some months without any data. Lastly, comparing HMINet and EAST-Net side by side, we can observe a consistent performance improvement, which verifies the effectiveness of MDFG in various scenarios.

5.3.3. Qualitative evaluation

To understand how EAST-Net behaves in diverse scenarios including open-world events, we conduct two case studies on JONAS-NYC and COVID-US.
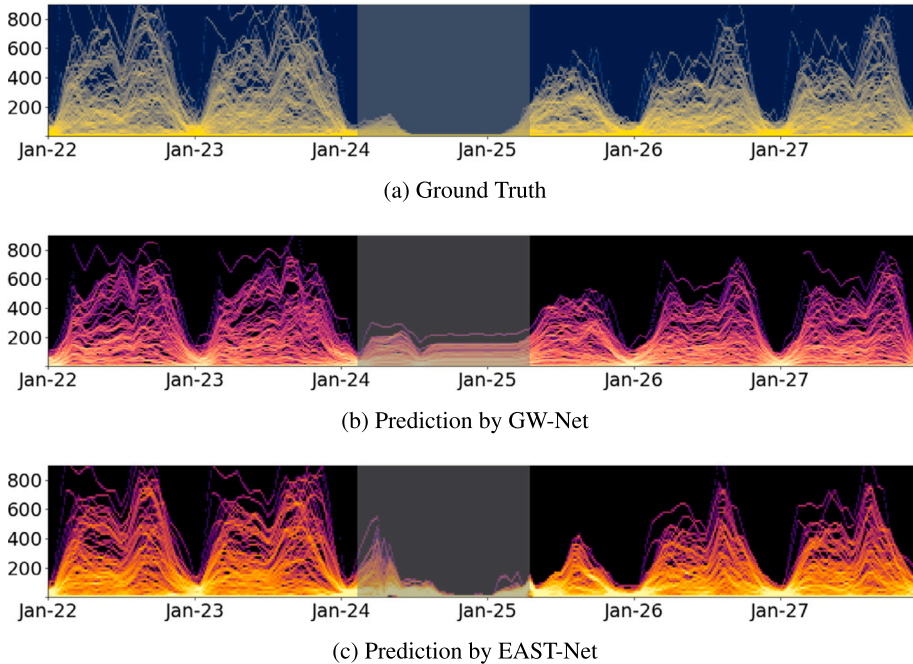

(a) Ground Truth

(b) Prediction by GW-Net

(c) Prediction by EAST-Net

Fig. 6. Time Series Histograms of Ground Truth and (2-hour ahead) Predictions of Citywide Taxi Demand in NYC from 22 Jan. 2016 to 27 Jan. 2016 (6 days).
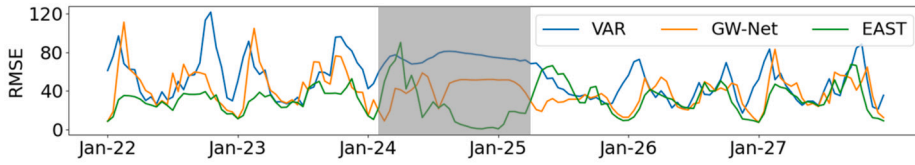
Fig. 7. City-averaged Prediction Accuracy (2-hour ahead) of Three Selected Models in NYC from 22 Jan. 2016 to 27 Jan. 2016 (6 days).
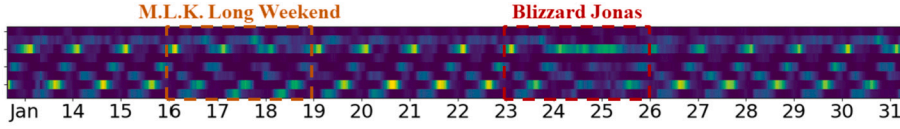

Fig. 8. On-the-fly Attention Score (ϕ_j) in MDFG at JONAS-NYC Test Set (from 12 Jan. to 31 Jan. 2016).

In Fig. 6a, a clear *no-mobility* period is expected under the impact of the historic blizzard "Jonas". GW-Net, a state-of-the-art model according to Table 3, simply makes native forecasting (repeating the latest observation) during this anomalous period (in Fig. 6b). In contrast, EAST-Net can quickly adapt to a *declining-to-zero* tendency (although causing under-estimations afterwards in Fig. 6c). In addition, as illustrated in Fig. 8, the composition of mobility prototypes in memory records for generating momentary filters is clearly differentiated between (1) normal workdays and weekend with a holiday; (2) a long weekend and the "Jonas" period. These observations demonstrate the event-awareness and adaptivity of EAST-Net under a short-term event causing sudden volatility.

Fig. 9a presents stream graphs [45] for state averaged POI visits in 10 categories during the first wave of COVID pandemic. Stream graph is a variation of stacked area graph by positioning layers to minimize weighted wiggle (sum of the squared slopes). In our case, an overall negative "tendency" is expected according to the ground truth. While an opposite positive "tendency" is produced by GW-Net, EAST-Net can capture the overall shape correctly (in Fig. 9b and 9c). In detail, EAST-Net also better catches a smaller volume of POI visits than GW-Net on the Memorial Day (25 May 2020). Besides, based on Fig. 10, EAST-Net becomes aware of a
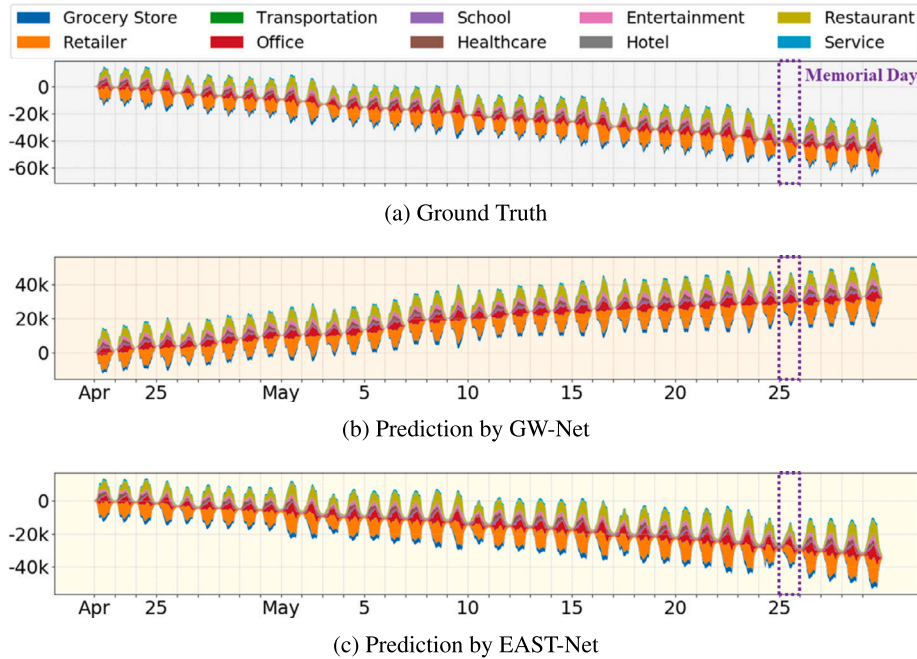

(a) Ground Truth

(b) Prediction by GW-Net

(c) Prediction by EAST-Net

Fig. 9. Stream Graphs of Ground Truth and (4-hour ahead) Predictions of State-averaged POI Visits in 10 Categories from 22 Apr. 2020 to 29 May. 2020 (38 days).
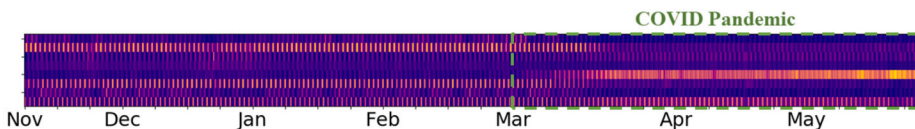

Fig. 10. On-the-fly Attention Score (ϕ_j) in MDFG at COVID-US Set (from 15 Nov. 2019 to 25 May. 2020).

Table 5

Spatial Knowledge Transfer: {NYC, DC} under JONAS and {CHI, US} under COVID.

Transfer	RMSE	MAE	MAPE
DC only	4.103	2.004	35.03%
NYC→DC zero-shot	4.813	2.263	40.29%
NYC→DC fine-tune	**3.936**	**1.841**	**32.40%**
NYC only	**23.632**	**15.790**	**33.33%**
DC→NYC zero-shot	36.372	25.537	48.96%
DC→NYC fine-tune	24.062	16.150	34.59%
CHI only	9.381	3.380	**61.50%**
US→CHI zero-shot	**9.125**	**3.274**	62.45%
US→CHI fine-tune	11.687	4.691	78.98%

Table 6

Spatio-Temporal Knowledge Transfer: JONAS-{NYC, DC} to COVID-CHI and JONAS-{NYC, DC} to COVID-US.

Transfer	RMSE	MAE	MAPE
COVID-CHI only	9.381	3.380	61.50%
JONAS-NYC→COVID-CHI zero-shot	10.145	3.368	65.43%
JONAS-NYC→COVID-CHI fine-tune	**8.863**	**3.125**	61.28%
JONAS-DC→COVID-CHI zero-shot	9.229	3.560	**60.51%**
JONAS-DC→COVID-CHI fine-tune	10.922	3.795	79.75%
COVID-US only	**799.51**	**371.78**	**51.84%**
JONAS-NYC→COVID-US zero-shot	3418.61	1476.85	461.48%
JONAS-NYC→COVID-US fine-tune	1049.54	482.44	60.83%
JONAS-DC→COVID-US zero-shot	2145.44	1046.92	319.29%
JONAS-DC→COVID-US fine-tune	1003.05	456.91	60.77%

new mobility pattern as early as March, the very beginning of the epidemic in US. These observations reconfirm the event-awareness and adaptivity of EAST-Net, particularly under a long-term event imposing lasting impact.

5.3.4. Additional evaluation

To understand how EAST-Net generalizes over different scenarios of open-world events, we conduct a group of knowledge transfer studies on the learnt representations in mobility prototype memory $\mathbf{M}_{mob}$, including *spatial transfer across regions* (under a same event), *spatio-temporal transfer across events*. In Table 5 and 6, **zero-shot** denotes the memory bank is loaded directly for testing; **fine-tune** denotes the memory values are loaded for initializing the model training.

In Table 5, it can be observed that learnt knowledge from a source city needs to be further adapted to the target city, under the blizzard case. However, the COVID case is interestingly the opposite, despite that US and CHI have different spatio-temporal scales and measure different nature of mobility (transport mode *v.s.* travel purpose). This phenomenon may be caused a shared low mobility status under the pandemic, so that the learnt patterns can be directly applied.

In Table 6, it can be observed that the nature of mobility matters when we are transferring over both spatial and temporal domains. Specifically, reusing historical lessons learnt from the blizzard Jonas can to some extent helps training or forecasting the multimodal transport demands under COVID pandemic, though the knowledge of DC is not as effective as the one of NYC for CHI. Not surprisingly, the knowledge learnt from different region and event for different nature of mobility is less mutually shareable, based on the results at COVID-US.

6. Concluding discussion

In this paper, we tackle the multimodal mobility nowcasting problem in response to various open-world event scenarios. By designing a heterogeneous mobility information network for explicitly representing intermodality and a memory-augmented dynamic filter generator for producing sequence-specific parameters on-the-fly, we propose an event-aware spatio-temporal network. Experiments on five real-world datasets verify the event-awareness and adaptivity of EAST-Net, which is even applicable to unprecedented events.

6.1. Implications

We discuss two major implications of our proposed model. (1) According to our experiments, demonstrated in Table 3 and 4, ST-Net+MDFG can already outperform the baseline models and consistently improve performance of the backbone. This observation indicates a promising applicability of MDFG as a plugin module to improve the model adaptivity to open-world events in the general multivariate time series forecasting task. It is also broadly related to out-of-distribution generalization towards self-adaptive AI. (2) We illustrate the promising generalizing capability of the core memory module in Table 5 and 6, and thereby envision it will play an essential role in a future spatio-temporal foundation model, which is pre-trained on unlabeled data and adaptable to unseen scenarios no matter over space (i.e., spatial adaptation) and time, against open-world events.
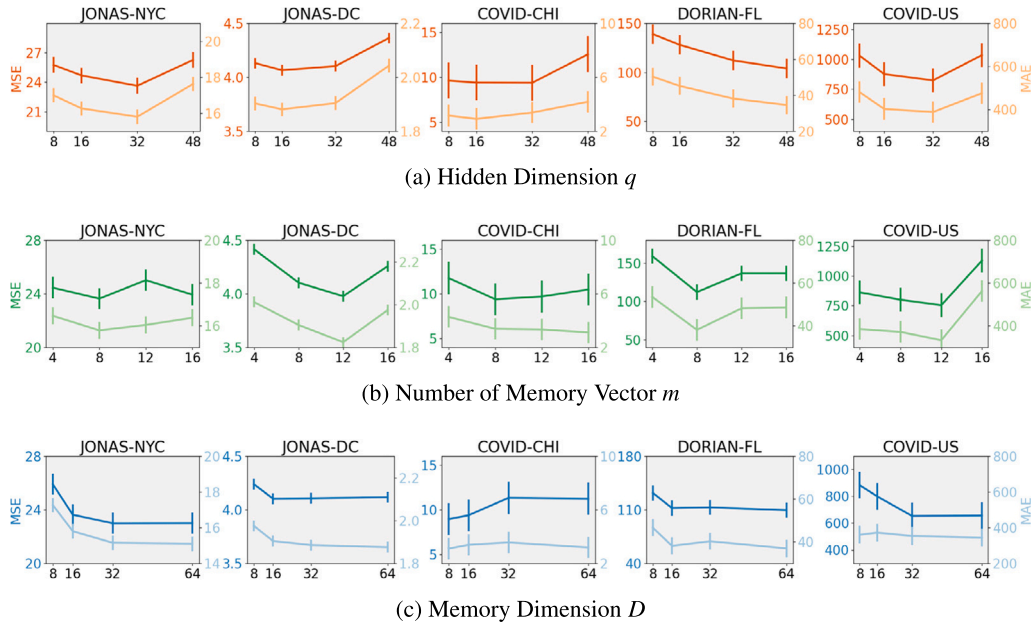
(a) Hidden Dimension q

(b) Number of Memory Vector m

(c) Memory Dimension D

Fig. 11. Sensitivity Analysis to Hyper-parameters.

6.2. Limitations

Here we identify two limitations of the proposed methods. (1) The proposed MDFG (i.e., memory-augmented dynamic filter generator) generates graph kernel parameters currently in fully parameter updating schema [31], which is a parameter-heavy solution [46] and leads to low efficiency in both training and inference stages. As a future direction, we would further improve the parameter generation efficiency (e.g., partially parameter updating) and find a balance with prediction accuracy. (2) According to Fig. 7, after an open-world event have just occurred or passed (i.e., January 24 and 25 in Blizzard JONAS case), a more-than-usual error level can be observed, which indicates our model still needs some time to adapt to and recover from the impacts of open-world events. In the future, we plan to investigate and enhance the speed of model adaptation.

CRediT authorship contribution statement

Zhaonan Wang: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Renhe Jiang:** Funding acquisition, Investigation, Methodology, Project administration, Resources. **Hao Xue:** Investigation, Project administration, Resources, Visualization, Writing – original draft, Writing – review & editing. **Flora D. Salim:** Conceptualization, Project administration, Supervision, Writing – original draft. **Xuan Song:** Funding acquisition, Supervision. **Ryosuke Shibasaki:** Funding acquisition, Supervision. **Wei Hu:** Writing – review & editing. **Shaowen Wang:** Funding acquisition, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I shall make data/code open upon acceptance.

Acknowledgements

References

[1] J. Ye, L. Sun, B. Du, Y. Fu, X. Tong, H. Xiong, Co-prediction of multiple transportation demands based on deep spatio-temporal neural network, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery &, Data Mining, 2019, pp. 305–313.

[2] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-c. Woo, Convolutional lstm network: a machine learning approach for precipitation nowcasting, in: Advances in Neural Information Processing Systems, 2015, pp. 802–810.

[3] J. Zhang, Y. Zheng, D. Qi, Deep spatio-temporal residual networks for citywide crowd flows prediction, in: Thirty-First AAAI Conference on Artificial Intelligence, 2017, pp. 1655–1661.

[4] Y. Li, R. Yu, C. Shahabi, Y. Liu, Diffusion convolutional recurrent neural network: data-driven traffic forecasting, in: International Conference on Learning Representations (ICLR'18), 2018.

[5] Z. Wu, S. Pan, G. Long, J. Jiang, C. Zhang, Graph wavenet for deep spatial-temporal graph modeling, in: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, 2019, pp. 1907–1913.

[6] C. Zheng, X. Fan, C. Wang, J. Qi, Gman: a graph multi-attention network for traffic prediction, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 2020, pp. 1234–1241.

[7] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, Z. Li, Deep multi-view spatial-temporal network for taxi demand prediction, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2018, pp. 2588–2595.

[8] A. Zonoozi, J.-j. Kim, X.-L. Li, G. Cong, Periodic-crn: a convolutional recurrent model for crowd density prediction with recurring periodic patterns, in: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-18, 2018, pp. 3732–3738.

[9] H. Yao, Y. Liu, Y. Wei, X. Tang, Z. Li, Learning from multiple cities: a meta-learning approach for spatial-temporal prediction, in: The World Wide Web Conference, 2019, pp. 2181–2191.

[10] X. Tang, H. Yao, Y. Sun, C. Aggarwal, P. Mitra, S. Wang, Joint modeling of local and global temporal dynamics for multivariate time series forecasting with missing values, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2020, pp. 5956–5963.

[11] Z. Fan, X. Song, R. Shibasaki, R. Adachi, Citymomentum: an online approach for crowd behavior prediction at a citywide level, in: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2015, pp. 559–569.

[12] R. Jiang, X. Song, Z. Fan, T. Xia, Q. Chen, S. Miyazawa, R. Shibasaki, Deepurbanmomentum: an online deep-learning system for short-term urban mobility prediction, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018, pp. 784–791.

[13] R. Jiang, X. Song, D. Huang, X. Song, T. Xia, Z. Cai, Z. Wang, K.-S. Kim, R. Shibasaki, Deepurbanevent: a system for predicting citywide crowd dynamics at big events, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery &, Data Mining, 2019, pp. 2114–2122.

[14] M. Defferrard, X. Bresson, P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, Adv. Neural Inf. Process. Syst. 29 (2016).

[15] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, arXiv preprint, arXiv:1609.02907, 2016.

[16] B. Yu, H. Yin, Z. Zhu, Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting, arXiv preprint, arXiv:1709.04875, 2017.

[17] L. Bai, L. Yao, C. Li, X. Wang, C. Wang, Adaptive graph convolutional recurrent network for traffic forecasting, Adv. Neural Inf. Process. Syst. 33 (2020) 17804–17815.

[18] R. Jiang, Z. Wang, J. Yong, P. Jeph, Q. Chen, Y. Kobayashi, X. Song, S. Fukushima, T. Suzumura, Spatio-temporal meta-graph learning for traffic forecasting, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, 2023, pp. 8078–8086.

[19] X. Geng, Y. Li, L. Wang, L. Zhang, Q. Yang, J. Ye, Y. Liu, Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 3656–3663.

[20] J. Ye, L. Sun, B. Du, Y. Fu, H. Xiong, Coupled layer-wise graph convolution for transportation demand prediction, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, 2021, pp. 4617–4625.

[21] R. Jiang, Z. Wang, Z. Cai, C. Yang, Z. Fan, T. Xia, G. Matsubara, H. Mizuseki, X. Song, R. Shibasaki, Countrywide origin-destination matrix prediction and its application for covid-19, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2021, pp. 319–334.

[22] L. Wang, A. Adiga, J. Chen, A. Sadilek, S. Venkatramanan, M. Marathe, Causalgnn: Causal-based graph neural networks for spatio-temporal epidemic forecasting, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, 2022, pp. 12191–12199.

[23] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, C. Zhang, Connecting the dots: multivariate time series forecasting with graph neural networks, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery &, Data Mining, 2020, pp. 753–763.

[24] C. Shang, J. Chen, J. Bi, Discrete graph structure learning for forecasting multiple time series, arXiv preprint, arXiv:2101.06861, 2021.

[25] Z. Wang, R. Jiang, Z. Cai, Z. Fan, X. Liu, K.-S. Kim, X. Song, R. Shibasaki, Spatio-temporal-categorical graph neural networks for fine-grained multi-incident co-prediction, in: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, 2021, pp. 2060–2069.

[26] B. Lu, X. Gan, W. Zhang, H. Yao, L. Fu, X. Wang, Spatio-temporal graph few-shot learning with cross-city knowledge transfer, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022, pp. 1162–1172.

[27] Y. Tang, A. Qu, A.H. Chow, W.H. Lam, S. Wong, W. Ma, Domain adversarial spatial-temporal network: a transferable framework for short-term traffic forecasting across cities, in: Proceedings of the 31st ACM International Conference on Information &, Knowledge Management, 2022, pp. 1905–1915.

[28] Y. Jin, K. Chen, Q. Yang, Transferable graph structure learning for graph-based traffic forecasting across cities, in: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023, pp. 1032–1043.

[29] Y. Wu, D. Zhuang, A. Labbe, L. Sun, Inductive graph neural networks for spatiotemporal Kriging, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, 2021, pp. 4478–4485.

[30] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, G. Zhang, Learning under concept drift: a review, IEEE Trans. Knowl. Data Eng. 31 (12) (2018) 2346–2363.

[31] L. Yuan, H. Li, B. Xia, C. Gao, M. Liu, W. Yuan, X. You, Recent advances in concept drift adaptation methods for deep learning, in: IJCAI, 2022, pp. 5654–5661.

[32] Y. Liu, H. Wu, J. Wang, M. Long, Non-stationary transformers: exploring the stationarity in time series forecasting, Adv. Neural Inf. Process. Syst. 35 (2022) 9881–9893.

[33] W. Li, X. Yang, W. Liu, Y. Xia, J. Bian, Ddg-da: data distribution generation for predictable concept drift adaptation, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, 2022, pp. 4092–4100.

[34] Y. Du, J. Wang, W. Feng, S. Pan, T. Qin, R. Xu, C. Wang, Adarnn: adaptive learning and forecasting of time series, in: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, 2021, pp. 402–411.

[35] X. You, M. Zhang, D. Ding, F. Feng, Y. Huang, Learning to learn the future: modeling concept drifts in time series prediction, in: Proceedings of the 30th ACM International Conference on Information &, Knowledge Management, 2021, pp. 2434–2443.

[36] H. Yu, X. Xu, T. Zhong, F. Zhou, Overcoming forgetting in fine-grained urban flow inference via adaptive knowledge replay, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, 2023, pp. 5393–5401.

[37] Z. Cai, R. Jiang, X. Yang, Z. Wang, D. Guo, H. Kobayashi, X. Song, R. Shibasaki, Memda: forecasting urban time series with memory-based drift adaptation, arXiv preprint, arXiv:2309.14216, 2023.

[38] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, arXiv preprint, arXiv:1409.0473, 2014.

[39] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, 2017, pp. 5998–6008.

[40] X. Jia, B. De Brabandere, T. Tuytelaars, L.V. Gool, Dynamic filter networks, Adv. Neural Inf. Process. Syst. (2016) 667–675.

[41] B. Yang, G. Bender, Q.V. Le, J. Ngiam, Condconv: conditionally parameterized convolutions for efficient inference, Adv. Neural Inf. Process. Syst. (2019) 1307–1318.

[42] J. Zhou, V. Jampani, Z. Pi, Q. Liu, M.-H. Yang, Decoupled dynamic filter networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 6647–6656.

[43] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.-X. Wang, X. Yan, Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting, Adv. Neural Inf. Process. Syst. 32 (2019) 5243–5253.

[44] D. Cao, Y. Wang, J. Duan, C. Zhang, X. Zhu, C. Huang, Y. Tong, B. Xu, J. Bai, J. Tong, Q. Zhang, Spectral temporal graph neural network for multivariate time-series forecasting, in: Advances in Neural Information Processing Systems, 2020, pp. 17766–17778.

[45] L. Byron, M. Wattenberg, Stacked graphs–geometry & aesthetics, IEEE Trans. Vis. Comput. Graph. 14 (6) (2008) 1245–1252.

[46] D. Ha, A. Dai, Q.V. Le, Hypernetworks, arXiv preprint, arXiv:1609.09106, 2016.