2023 IEEE International Conference on Data Mining (ICDM) | 979-8-3503-0788-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICDM58522.2023.00176

Fine-grained Urban Flow Inference with Unobservable Data via Space-Time Attraction Learning

Ruifeng Wang[1], Yuansheng Liu[2], Yongshun Gong[1,*], Wei Liu[3], Meng Chen[1], Yilong Yin[1,*], Yu Zheng[4]
[1]*Shandong University, China*
[2]*Hunan University, China*
[3]*University of Technology Sydney, Australia*
[4]*JD Intelligent Cities Research, China*
wangzheaos@163.com, yuanshengliu@hnu.edu.cn, {ysgong,mchen,ylyin}@sdu.edu.cn,
Wei.Liu@uts.edu.au, msyuzheng@outlook.com

***Abstract*—Fine-grained urban flow inference focuses on inferring fine-grained urban flows based solely on coarse-grained observations, which is essential for the city management and transportation services. However, most of the existing methods assume that partial urban flows in coarse-grained regions cannot be observable. In this study, we propose a multi-task framework known as UrbanSTA with space-time attraction learning to estimate missing values in coarse-grained urban flow map and forecast fine-grained urban flows simultaneously. Specifically, UrbanSTA comprises two parts: the flow completion network STA and the fine-grained flow inference network FIN. STA captures space-time features with a separable space-time attention encoder and recovers the missing flow features with a decoder. FIN directly uses complete coarse-grained flow features for further decoding, and reconstructs fine-grained flow features based on the complex associations between coarse- and fine-grained urban flows, relying on upsampling constraints. Extensive experiments conducted on two real-world datasets demonstrate that our proposed model yields the best results compared to other state-of-the-art methods. The source code has been provided at https://github.com/Wangzheaos/UrbanSTA.**

***Index Terms*—fine-grained urban flow inference, multi-task learning, urban flow prediction, space-time prediction, time and space attention**

I. Introduction

With the rapid process of smart city construction [1], urban flow prediction has become an essential component of urbanization and smart city development. Accurate urban flow prediction plays a huge role for city managers, which is helpful for future smart city construction [2]. However, considering the influence of implementation factors, obtaining urban flow only by deploying sensors is often insufficient [3]–[5], and there are partial missing situations because it is not possible to install sensors in all areas and intersections, which will result in high human and material costs [6].

To date, there have been limited studies on fine-grained urban flow inference considering missing values [7], [8]. The majority of researchers focus on predicting missing values and fine-grained flow inference separately, neglecting the fact that in practical scenarios, not only is it necessary to complete the flow, but further fine-grained flow inference is required [6]. There are several intrinsic challenges lying ahead.

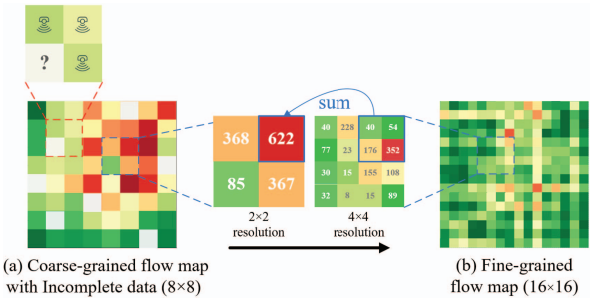


Fig. 1: Urban flow of two levels of granularities.

• *Incomplete urban flow maps.* As shown in Fig. 1(a), owing to uneven sensor distribution coverage and regional constraints, the obtained flow have partial missing values.

• *The complex correlation between coarse- and fine-grained urban flows.* A region in the coarse-grained flow is correlated with multiple adjacent regions in the fine-grained flow, posing a challenge in fine-grained urban flow inference.

• *The spatial attributes within urban flow maps are interdependent.* Typically, the flow variation in a region is primarily influenced by the flow in its neighboring and similar regions.

To address these challenges, we propose a method for fine-grained urban flow inference with missing values, named as UrbanSTA. UrbanSTA comprises two network models. One is the STA module with an asymmetric encoder-decoder architecture, which predicts missing values by extracting spatial and temporal features of urban flow. The other is a fine-grained decoder with spatial attention, which infers fine-grained features from the predicted coarse-grained features and restores the original structural constraint relationship using distributional upsampling. The main contributions and innovations of this

979-8-3503-0788-7/23/$31.00 ©2023 IEEE
DOI 10.1109/ICDM58522.2023.00176

paper are summarized as follows:

- We propose a two-stage framework with spatiotemporal attention learning to address the problem of fine-grained flow inference with incomplete data.
- Besides, a distributional upsampling strategy was designed to address the complex correlation between coarse- and fine-grained urban flows.
- Considering the spatial rotational invariance and periodicity in time, we propose a space-time attraction constraint loss to further improve the space-time representations.

II. Related Work

A. Urban Flow Data Prediction

Urban flow data prediction occupies an important position in the field of urban computing, and its research methods are numerous [9]–[11]. In the method based on deep learning, a deep neural network prediction model DeepST [12] is proposed. This model can not only capture the characteristics of time and space, but also extract the characteristics of external factors. Another typical deep neural network model is ST-ResNet [13], which includes residual units and fusion components for learning more complex spatiotemporal features.

Different from the urban flow data prediction task, the problem studied in this paper aims to infer fine-grained urban flows based on coarse-grained observations with missing values.

B. Urban Flow Data Super-Resolution

Deep learning methods are the main approach for fine-grained urban flow inference [14]. The UrbanFM model [15] is designed to predict city traffic flow by introducing an external factor fusion network to extract external feature information related. To further improve the fine-grained inference performance of UrbanFM, Ouyang et al. [16] proposed an UrbanPy model, which comprises two parts. The first part is an inference network that generates fine-grained flow distribution. The second part is a general fusion subnetwork.

However, the above study is in an ideal situation, and there are missing data in real-world scenarios. Li et al. [6] for the first time studied the problem and proposed a Multi-Task network model known as MT-CSR. They designed a data completion network CMPNet to complete the coarse-grained urban flows by considering both the local spatial dependencies and the global POI similarities. Moreover they proposed a data super-resolution network SRNet to capture the complex associations between fine- and coarse-grained data.

III. Notations and Problem Definition

Definition 1. **Coarse- and fine-grained urban flow map.** In general, the $I \times J$ data obtained based on the original sensor are used as a coarse-grained urban flow map $\mathcal{X}_c^t \in \mathbb{R}_+^{I \times J}$, and an unknown fine-grained urban flow map $\mathcal{X}_f^t \in \mathbb{R}_+^{MI \times MJ}$ is obtained by inference with an upscaling factor M in time slot t.

Definition 2. **Incomplete coarse-grained urban flow map.** In reality, coarse-grained urban flow maps are incomplete with some values missing, and the missing values positions remain fixed over time. We denote the incomplete coarse-grained urban flow maps as $\mathcal{X}_{uc}^t \in \mathbb{R}_+^{I \times J}$.

Definition 3. **Structural constraint.** Each coarse-grained region flow $x_{c,i,j}^t$ is strictly equal to the sum of the corresponding super-resolution fine-grained region flow $x_{f,i',j'}^t$. This constraint is expressed as:

$$x_{c,i,j}^t = \sum_{i',j'} x_{f,i',j'}^t \ s.t.\ i = \left\lfloor \frac{i'}{M} \right\rfloor, j = \left\lfloor \frac{j'}{M} \right\rfloor, \quad (1)$$

where $i' = 1, 2, \ldots, MI$ and $j' = 1, 2, \ldots, MJ$.

Objective. Given an upscaling factor $M \in \mathbb{Z}$, a set of historical and current incomplete coarse-grained urban flow maps $\{\mathcal{X}_{uc}^{t-k}, \mathcal{X}_{uc}^{t-k+1}, \ldots, \mathcal{X}_{uc}^t\} \in \mathbb{R}_+^{I \times J}$, the goal in this study is to infer the current complete fine-grained urban flow maps $\mathcal{X}_f^t \in \mathbb{R}_+^{MI \times MJ}$.

IV. Proposed Method

We propose a fine-grained urban flow inference model UrbanSTA with unobservable data shown in Fig. 2.

A. Urban Flow Data Completion Network

1) Initialization: **Mask.** To conduct urban flow data completion over the regions where the data are missing, we define the mask operation as 1 for the missing area and 0 for the remaining areas.

Split into patches. We divide map $\mathcal{X}_{uc}^t$ into regular non-overlapping $2D$ patches $\mathcal{X}_{uc,p}^t$ according to each grid. Then we take the randomly missing patch $\mathcal{X}_{uc,mp}^t$ as a mask and the remaining patches $\mathcal{X}_{uc,rp}^t$ are in order.

2) STA-Encoder: **Linear Embedding.** Our encoder embeds unmasked patches to map to D dimensions by a trainable linear projection $E \in \mathbb{R}^{D \times C}$:

$$Z_{uc,rp}^{\tau,0} = E\left[\mathcal{X}_{uc,rp}^{t-k}, \mathcal{X}_{uc,rp}^{t-k+1}, \ldots, \mathcal{X}_{uc,rp}^t\right] + E^{pos} + E^{time}, \quad (2)$$

where E^{pos} denotes a position embedding, and E^{time} denotes a time dimension embedding.

Divided Space-Time Attention. First, STA-Encoder processes the embedding vectors $Z_{uc,rp}^{\tau,0}$ via a series of divided space-time Transformer blocks, where temporal and spatial attention are separated one after the other. At each block l, a query/key/value vector is computed for $Z_{uc,rp}^{\tau,l-1}$:

$$Q_{uc,rp}^{\tau,l,a} = W_Q^{l,a} LN(Z_{uc,rp}^{\tau,l-1}),\ Q_{uc,rp}^{\tau,l,a} \in \mathbb{R}^{D_h}, \quad (3)$$

$$K_{uc,rp}^{\tau,l,a} = W_K^{l,a} LN(Z_{uc,rp}^{\tau,l-1}),\ K_{uc,rp}^{\tau,l,a} \in \mathbb{R}^{D_h}, \quad (4)$$

$$V_{uc,rp}^{\tau,l,a} = W_V^{l,a} LN(Z_{uc,rp}^{\tau,l-1}),\ V_{uc,rp}^{\tau,l,a} \in \mathbb{R}^{D_h}, \quad (5)$$

where $LN(\cdot)$ denotes LayerNorm, $a = 1, \ldots, A$ is an index over multiple attention heads, A denotes the number of

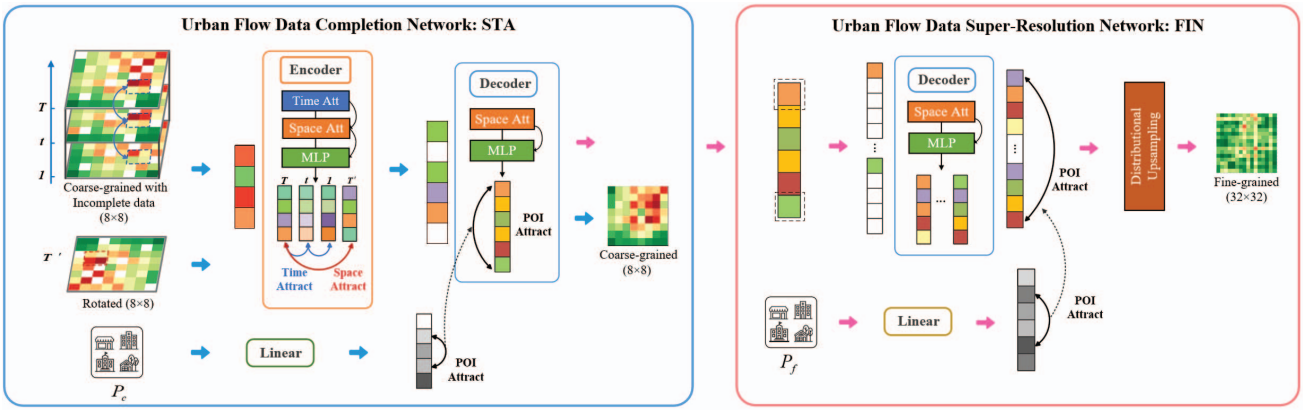


Fig. 2: Framework of UrbanSTA. The proposed method comprises two main parts, the space-time attention completion network (STA) and the fine-grained inference network (FIN). The STA comprises an asymmetric encoder and decoder, while the FIN only comprises a decoder. The UrbanSTA adopts a three-stage architecture, decomposing the multi-task problem into space-time feature encoding and decoding steps, to accomplish traffic completion and super-resolution inference.

attention heads, and its latent dimensions are divided into $D_h = D./A$.

Divided space-time attention weights are computed via the dot-product for each vector. For each block l, we first compute the temporal attention of each patch in t and all patches at the same spatial location at other times as follows:

$$T^{t,l,a}_{uc,rp} = SM(\frac{(Q^{t,l,a}_{uc,rp})^T}{\sqrt{D_h}} \cdot \left[\{K^{\tau,l,a}_{uc,rp}\}_{\tau=t-k,\dots,t}\right]) \cdot \left[\{V^{\tau,l,a}_{uc,rp}\}_{\tau=t-k,\dots,t}\right], \tag{6}$$

where $SM(\cdot)$ denotes the softmax activation function, and $T^{t,l,a}_{uc,rp}$ denotes that each patch is compared $(k+1)$ times in the time dimension. Then, these vectors from all heads are concatenated together as follows:

$$\tilde{Z}^{t,l}_{uc,rp} = W_a \left[T^{t,l,1}_{uc,rp}, \dots, T^{t,l,A}_{uc,rp}\right]^T + Z^{t,l-1}_{uc,rp}, \tag{7}$$

Second, the result of Eq. 7 is used as a new encoding, fed into the spatial attention module. A new $Q/K/V$ is obtained from $\tilde{Z}^{t,l}_{uc,rp}$ according to Eq. 3 - 5. Then, we compute the spatial attention of each patch in time slot t and all patches at different locations simultaneously as follows:

$$S^{t,l,a}_{uc,rp} = SM(\frac{(Q^{t,l,a}_{uc,rp})^T}{\sqrt{D_h}} \cdot \left[\{K^{t,l,a}_{uc,rp}\}_{rp=1,\dots,L_r}\right]) \cdot \left[\{V^{t,l,a}_{uc,rp}\}_{rp=1,\dots,L_r}\right], \tag{8}$$

where $S^{t,l,a}_{uc,rp}$ denotes that each patch is compared L_r times in the spatial dimension. Note that divided attention performs only $(L_r + k + 1)$ comparisons per patch, which is more efficient and effective compared to other self-attention calculation methods [17].

Finally, these vectors $S^{t,l,a}_{uc,rp}$ are concatenated in a way similar to Eq. 7 and passed through an MLP, using residual connections after each operation as $Z^{t,l}_{uc,rp}$.

Time and Space Attract.

Here we devise a time period loss $\mathcal{L}_T$ to extract time period information in the temporal dimension. For current timestamp t, we can obtain three temporal properties: closeness, period, and trend [13] from historical incomplete coarse-grained urban flow maps. The time period loss is calculated as follows:

$$\mathcal{L}_T = \left\| Z^t_{uc,rp} - \frac{1}{nt+np} \sum_{ht=1}^{nt} \sum_{ht=1}^{np} Z^{ht}_{uc,rp} \right\|^2_F, \tag{9}$$

where $Z^{ht}_{uc,rp}$ denotes periodic and trending flow map features, and nt and np denote the number of periodic and trending time slices, respectively.

To enhance the local feature capture capability of the Transformer, we propose a space rotation loss $\mathcal{L}_S$ based on a certain degree of rotation invariance.

The coarse-grained urban flow maps $\mathcal{X}^t_{uc}$ are randomly rotated by an angle of $[0°, 90°, 180°, 270°]$ to obtain $\bar{\mathcal{X}}^t_{uc}$. Then, it is split into patches, and expressed as high-dimensional features $\bar{Z}^t_{uc,rp}$ through spatial attention calculation:

$$\mathcal{L}_S = \left\| Z^t_{uc,rp} - R(\bar{Z}^t_{uc,rp}) \right\|^2_F, \tag{10}$$

where $R(\cdot)$ denotes a reverse rotation function that restores the rotated flow maps to their original position.

3) STA-Decoder: **Reconstruction target.**

The input to the STA-Decoder is the full set of patches $Z^t_{uc,p}$ for coarse-grained urban flow maps, which include encoded patches $Z^t_{uc,rp}$ and masked patches $Z^t_{uc,mp}$:

$$Z^t_{uc,p} = E_d \cdot O_r([Z^t_{uc,rp}, Z^t_{uc,mp}]) + E^{pos}_d, \tag{11}$$

where E_d denotes a smaller linear projection, $O_r(\cdot)$ is a sorting function that arranges all patches in the initial $\mathcal{X}^t_{uc,p}$ order, and E^{pos}_d denotes a position embedding of the decoder.

Then, vectors $Z^{t,0}_{uc,p}$ are input to the space attention module to capture spatial information $Z^t_{uc,p}$. The decoder features are

linearly projected and reshaped to form a complete coarse-grained urban flow map $\hat{\mathcal{X}}_c^t$:

$$\hat{\mathcal{X}}_c^t = M_{uc}(\mathcal{X}_{uc}^t) \cdot R_s(E_x \cdot LN(Z_{uc,p}^t)) + \mathcal{X}_{uc}^t, \tag{12}$$

where $R_s(\cdot)$ denotes a reshape function that restructures the patch into a $2D$ map, E_x denotes a linear projection.

POI Attract. We collect Point of Interest category distribution (e.g.,shopping, medical, education, and residents) in each region, and denote coarse and fine-grained POI features as $P_c \in \mathbb{R}^{I \times J \times K}$ and $P_f \in \mathbb{R}^{MI \times MJ \times K}$ respectively, where K denotes the POI categories [18].

We use linear projection to fuse K POI categories, and then, as shown at the bottom of the blue box in Fig. 2, for the anchor region p^a in coarse-grained POI maps, N^1 positive regional samples $\{p_{n^1}^+\}_{n^1=1}^{N^1}$ and N^2 negative samples $\{p_{n^2}^-\}_{n^2=1}^{N^2}$ are selected from candidate regions $p_{i,j}$ according to the samples distance and threshold:

We propose the POI contrast loss $\mathcal{L}_{cpoi}$ to pull in flow features $z_{n_1}^+$ corresponding to positive sample regions, and make features $z_{n_2}^-$ of negative samples far away as follows:

$$\mathcal{L}_{cpoi} = -\log \frac{\sum_{n_1=1}^{N_1} \exp sim(z^a, z_{n_1}^+)}{\sum_{n_1=1}^{N_1} \exp sim(z^a, z_{n_1}^+) + \sum_{n_2=1}^{N_2} \exp sim(z^a, z_{n_2}^-)}, \tag{13}$$

where $sim(\cdot,\cdot)$ is the similarity function between two embedding features (e.g., inner product).

B. Urban Flow Data Super-Resolution Network

Fine-grained Refactoring. The coarse-grained feature $Z_{uc,p}^t$ relationship needs to be reconstructed into a fine-grained feature relationship as follows:

$$Z_{f,p}^t = \left\{ \left[z_{uc,p}^t, z_{f,1}^t, \ldots, z_{f,M^2}^t \right] \right\}_{p=1}^{L}, \tag{14}$$

where $z_{f,*}^t$ denotes learned vectors, which are represented as missing patches. We supplement each coarse-grained patch vector with M^2 fine-grained patch vectors to form a set of fine-grained relational maps. A total of L groups (M^2+1) patch vectors are used as the inputs.

Super-resolution Inference. Then, the features are further extracted by the spatial attention module. And the coarse-grained patches are discarded, and the remaining L groups of patches are reorganized into a $2D$ map $\hat{\mathcal{X}}_f^t$:

$$\hat{\mathcal{X}}_f^t = R_f(E_f \cdot LN(Z_{f,p}^t - \left\{ \left[z_{uc,p}^t \right] \right\}_{p=1}^{L})), \tag{15}$$

where $R_f(\cdot)$ denotes a reorganization function that combines $1D$ patches into a $2D$ map according to the location of the region. Meanwhile, we choose a M^2-Normalization that makes the sum of fine-grained flow equal to their corresponding coarse-grained flow, which is described as:

$$W_f^t = \frac{\hat{x}_{f,i,j}^t}{\sum_{\substack{i' \in \left[\left\lfloor \frac{i}{M} \right\rfloor M, \left(\left\lfloor \frac{i}{M} \right\rfloor + 1 \right) M \right) \\ j' \in \left[\left\lfloor \frac{j}{M} \right\rfloor M, \left(\left\lfloor \frac{j}{M} \right\rfloor + 1 \right) M \right)}} \hat{x}_{f,i',j'}^t}, \tag{16}$$

TABLE I: Statistics of datasets.

Dataset	TaxiBJ	BikeNYC
Time span	7/1/2013 - 10/31/2013	1/1/2019 - 3/31/2019
Time interval	30 mintures	1 hour
Coarse-grained size	8×8	8×8
Fine-grained size	16×16 / 32×32	16×16 / 32×32
Upscaling factor(M)	2 / 4	2 / 4
POI information	$\surd$	$\surd$

$$\hat{\mathcal{X}}_f^t = W_f^t \odot \hat{\mathcal{X}}_c^t, \tag{17}$$

where $W_f^t \in [0,1]$ denotes the distribution probability of the fine-grained urban flow.

C. Final Objective Function

To jointly perform urban flow data completion and super-resolution, we propose a multi-task learning network with a two-stage training strategy.

The first step pre-training adopts pixel-level MSE, time period, space rotation, POI contrast losses as follows to pretrain the data completion network:

$$\mathcal{L}_C = \lambda_1 \left\| \hat{\mathcal{X}}_c^t - \mathcal{X}_c^t \right\|_F^2 + \mathcal{L}_T + \mathcal{L}_S + \mathcal{L}_{cpoi}, \tag{18}$$

where λ_1 denotes hyper-parameters and $\mathcal{X}_c^t$ denotes ground truth of coarse-grained urban flow map at time t.

In the second stage, we constrain the FIN module with pixel-level MSE loss and POI contrast loss, while training the pre-trained STA module with $\mathcal{L}_C$, and finally achieve the optimal overall UrbanSTA as follows:

$$\mathcal{L}_F = \lambda_2 \left\| \hat{\mathcal{X}}_f^t - \mathcal{X}_f^t \right\|_F^2 + \mathcal{L}_{fpoi} + \mathcal{L}_C, \tag{19}$$

where λ_2 denotes hyper-parameters and $\mathcal{X}_f^t$ denotes the ground truth of fine-grained urban flow map in t.

V. Experiments

A. Experimental Settings

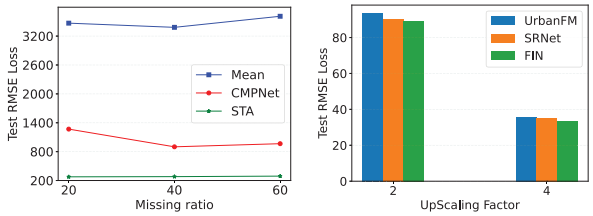
1) Datasets: The statistics for datasets are summarized in Table I. In our experiments, we partitioned the data into non-overlapping training, validation, and test data by a ratio of 2:1:1, respectively.

2) Baselines: We compare UrbanSTA with six baseline methods. Among them, UrbanFM, UrbanPy, FODE are fine-grained flow inference baselines, MT-CSR, STA-UP, STA-FD are fine-grained flow inference baselines with missing values. It is worth noting that STA-UP and STA-FD are the baselines of STA combined with UrbanPy and FODE respectively.

3) Training Details & Hyperparameters: The patch size of $\mathcal{X}_{uc,p}^t$ must be 1. Pretrained STA model has a learning rate $lr = 1e-2$ and batch size $bz = 256$. Trained UrbanSTA model with learning rate $lr = 1e-4$ and batch size $bz = 128$.

TABLE II: Comparison results over TaxiBJ dataset. 2 and 4 are upscaling factor. 20%, 40%, and 60% denote the rates of the incomplete data. The best results are bold and the second best are underlined.

Model				UrbanFM	UrbanPy	FODE	MT-CSR	STA-UF	STA-FD	UrbanSTA
TaxiBJ	2	20%	MAE	73.25	118.87	74.39	75.81	67.62	**62.99**	63.03
			RMSE	116.20	190.13	116.75	123.35	98.84	92.49	**92.36**
		40%	MAE	113.65	174.04	116.19	117.06	72.86	66.95	**65.06**
			RMSE	174.22	255.15	176.04	186.86	105.26	98.32	**96.01**
		60%	MAE	132.48	190.76	133.62	134.32	73.64	**68.89**	69.91
			RMSE	202.88	274.95	201.69	214.16	108.43	**102.99**	103.95
	4	20%	MAE	25.89	31.02	30.77	26.74	25.23	25.71	**23.19**
			RMSE	40.67	48.26	46.54	44.71	38.69	40.03	**37.11**
		40%	MAE	35.53	43.50	39.59	36.34	25.84	24.84	**24.36**
			RMSE	52.73	63.14	57.24	58.75	40.00	39.36	**38.98**
		60%	MAE	39.71	45.48	42.94	39.80	26.09	25.20	**24.59**
			RMSE	60.65	66.24	62.45	65.20	40.87	40.49	**39.95**



(a) STA module analysis. (b) FIN module analysis.

Fig. 3: Effectiveness on two tasks of UrbanSTA model.

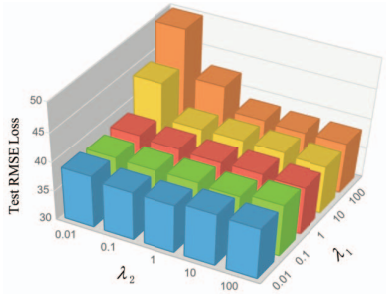


Fig. 4: The model performance with different parameter settings over TaxiBJ dataset with 20% missing data and a upscaling factor of 4.

B. Comparison with Baselines

Performance comparison. The calculation results of different methods are summarized in Table II and III. The results demonstrate that UrbanSTA achieves the best results in almost on the TaxiBJ dataset. UrbanFM, UrbanPy, and FODE both ignore data loss issues. UrbanSTA achieves the second best performance in most cases on the BikeNYC dataset. Because of the Transformer network of the UrbanSTA model, compared with the CNN model, it performs worse on small-scale datasets.

Model effectiveness analysis on the two tasks. We compare the STA module with average completion (Mean) and CMPNet on the TaxiBJ dataset. The results in Fig. 3(a) show that STA significantly outperforms both baselines. The FIN is an super-resolution module that is performed without data. We compare FIN with UrbanPy and SRNet using the TaxiBJ dataset, and the results are shown in Fig. 3(b). The results show that FIN consistently slightly outperforms UrbanPy and SRNet under different upscaling factor conditions.

C. Ablation Analysis

We compare UrbanSTA and its variants on the TaxiBJ datasets. The calculation results are summarized in Table IV. We observe that STA outperforms these seven variants. It is stated that the space rotation, time period, and POI contrast losses of coarse-grained are all useful.

Simultaneously, we also conducted experiments to compare the impact of fine-grained POI constraints on the FIN model. As summarized in Table IV, adding fine-grained POI constraints improves the performance of super-resolution tasks.

D. Parameter Analysis

We study the sensitivity of the model to the two main RMSE functions with different parameters. As summarized in Fig. 4, we study two hyperparameters, λ_1 and λ_2, in the final objective function formula (19). We set λ_1 and λ_2 ranging from $1e-2$ to $1e2$, respectively. The results show that the FIN module is more important than the STA module. Therefore, in our experiments, to increase the constraints on the FIN module, slightly fine-tuning the STA module, we set $\lambda_1 = 1e-2$ and $\lambda_2 = 1e2$, which achieves the best performance.

VI. Conclusion

In this study, we proposed a network for fine-grained urban traffic flow inference with missing values. UrbanSTA addresses three significant challenges in the fine-grained urban traffic flow inference process, including incomplete coarse-grained urban flow maps, complex correlations between coarse- and fine-grained urban flow data, and spatial properties

TABLE III: Comparison results over BikeNYC dataset. 2 and 4 are upscaling factor. 20%, 40%, and 60% denote the rates of the incomplete data. The best results are bold and the second best are underlined.

Model				UrbanFM	UrbanPy	FODE	MT-CSR	STA-UF	STA-FD	UrbanSTA
BikeNYC	2	20%	MAE	0.88	0.91	0.83	0.80	0.81	0.82	**0.79**
			RMSE	2.17	2.24	**1.82**	1.89	2.05	1.84	1.87
		40%	MAE	1.07	1.20	1.09	1.02	1.08	**0.87**	1.44
			RMSE	2.31	2.84	2.33	2.49	2.75	**1.92**	3.15
		60%	MAE	1.15	1.31	1.17	1.12	1.04	0.91	**0.89**
			RMSE	2.48	3.23	2.49	2.58	2.24	**2.00**	2.09
	4	20%	MAE	0.27	0.28	0.27	**0.26**	0.33	**0.26**	0.36
			RMSE	**0.94**	1.07	**0.94**	1.02	0.97	0.95	1.03
		40%	MAE	0.33	0.35	0.33	0.30	0.34	**0.28**	0.30
			RMSE	1.04	1.34	1.04	1.16	1.00	0.99	**0.98**
		60%	MAE	0.35	0.35	0.35	0.33	0.35	**0.32**	**0.32**
			RMSE	1.10	1.34	1.10	1.25	**1.02**	1.06	**1.02**

TABLE IV: Ablation Studies. We report the results with different constraint in same models on TaxiBJ dataset with 20% missing data and a upscaling factor of 4. The best results are bold and the second best are underlined.

Space			√				√	√		√				√	√
Time				√			√		√	√				√	√
POI_c					√			√	√	√				√	√
POI_f															√
Data Completion	MAE	226.72	210.55	214.86	209.31	212.22	210.21	212.27	**204.02**	Super Resolution	MAE	24.20	**24.10**
	RMSE	307.90	289.40	290.67	287.84	288.76	287.21	288.07	**274.80**		RMSE	38.47	**38.29**

influence on each other in urban flow maps. Extensive evaluations on two large-scale urban traffic flow datasets demonstrate that the proposed model significantly improves performance and outperforms state-of-the-art models.

ACKNOWLEDGMENT

This work was supported partly by NSFC (62202270); Natural Science Foundation of Shandong Province (ZR2021QF034); Shandong Excellent Young Scientists Fund (Oversea) (2022HWYQ-044); Taishan Scholar Project of Shandong Province (tsqn202306066); the Open Fund of Beijing Key Laboratory of Traffic Data Analysis and Mining.

REFERENCES

[1] K. Su, J. Li, and H. Fu, "Smart city and the applications," in *2011 international conference on electronics, communications and control (ICECC)*. IEEE, 2011, pp. 1028–1031.

[2] H. R. Kwon, H. Cho, J. Kim, D. Lee, and S. K. Lee, "International case studies of smart cities: Namyangju, republic of korea," 2016.

[3] F. Zhou, L. Li, T. Zhong, G. Trajcevski, K. Zhang, and J. Wang, "Enhancing urban flow maps via neural odes," in *Proceedings of the Twenty-Ninth IJCAI*, 2020.

[4] Y. Gong, Z. Li, J. Zhang, W. Liu, B. Chen, and X. Dong, "A spatial missing value imputation method for multi-view urban statistical data," in *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 2021, pp. 1310–1316.

[5] Y. Gong, X. Dong, J. Zhang, and M. Chen, "Latent evolution model for change point detection in time-varying networks," *Information Sciences*, vol. 646, p. 119376, 2023.

[6] J. Li, S. Wang, J. Zhang, H. Miao, J. Zhang, and P. Yu, "Fine-grained urban flow inference with incomplete data," *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[7] Y. Gong, Z. Li, J. Zhang, W. Liu, Y. Yin, and Y. Zheng, "Missing value imputation for multi-view urban statistical data via spatial correlation learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 1, pp. 686–698, 2021.

[8] Y. Gong, Z. Li, W. Liu, X. Lu, X. Liu, I. W. Tsang, and Y. Yin, "Missingness-pattern-adaptive learning with incomplete data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[9] P. Xie, T. Li, J. Liu, S. Du, X. Yang, and J. Zhang, "Urban flow prediction from spatiotemporal data using machine learning: A survey," *Information Fusion*, vol. 59, pp. 1–12, 2020.

[10] Y. Gong, Z. Li, J. Zhang, W. Liu, and Y. Zheng, "Online spatio-temporal crowd flow distribution prediction for complex metro system," *IEEE Transactions on knowledge and data engineering*, vol. 34, no. 2, pp. 865–880, 2020.

[11] L. Zhao, M. Gao, and Z. Wang, "St-gsp: Spatial-temporal global semantic representation learning for urban flow prediction," in *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, 2022, pp. 1443–1451.

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

[13] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Thirty-first AAAI conference on artificial intelligence*, 2017.

[14] H. Qu, Y. Gong, M. Chen, J. Zhang, Y. Zheng, and Y. Yin, "Forecasting fine-grained urban flows via spatio-temporal contrastive self-supervision," *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[15] Y. Liang, K. Ouyang, L. Jing, S. Ruan, Y. Liu, J. Zhang, D. S. Rosenblum, and Y. Zheng, "Urbanfm: Inferring fine-grained urban flows," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 3132–3142.

[16] K. Ouyang, Y. Liang, Y. Liu, Z. Tong, S. Ruan, D. Rosenblum, and Y. Zheng, "Fine-grained urban flow inference," *IEEE transactions on knowledge and data engineering*, 2020.

[17] G. Bertasius, H. Wang, and L. Torresani, "Is space-time attention all you need for video understanding?" in *Proceedings of the International Conference on Machine Learning (ICML)*, July 2021.

[18] L. Wang, X. Geng, X. Ma, F. Liu, and Q. Yang, "Cross-city transfer learning for deep spatio-temporal prediction," in *International Joint Conference on Artificial Intelligence*, 2019, p. 1893.