



Dynamic adaptive generative adversarial networks with multi-view temporal factorizations for hybrid recovery of missing traffic data

Jinlong Li¹ · Ruonan Li² · Zilin Huang³ · Pan Wu¹ · Lunhui Xu¹

Received: 16 March 2022 / Accepted: 16 November 2022 / Published online: 20 December 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract

Making reliable recovery of missing traffic data facilitates diverse applications of data-driven intelligent transportation system. But faced with correlation and heterogeneity along spatial-temporal dimensions, most existing works lack sufficient capability to capture these complex properties, resulting in suboptimal imputation performance. By addressing this challenge, we propose a hybrid framework TFs-DGAN consisting of dynamic adaptive generative adversarial networks (DA-GAN) with multi-view temporal factorizations (TFs), which can efficiently repair missing data by modeling those spatial-temporal correlations. Of these, DA-GAN model can generate traffic data from noise distribution and keep iterating dynamically to extract the global consistency. To further exploit the local consistency, TFs model drives the continual reduction in local elements in residuals by a novel truncation mechanism. Unlike the single model computation, TFs-DGAN integrates all stage-optimized residuals by local feedback and finally outputs the best repair results. In fact, our intention for this strategy is that DA-GAN module produces data but inaccurately, while TFs module refines its imperfections by modeling multi-view temporal properties. From the numerical analysis, the empirical evaluation relying on two publicly available traffic datasets suggests that our TFs-DGAN significantly outperforms the state-of-the-art baseline models in terms of accuracy, stability and computational efficiency.

Keywords Missing traffic data · Generative adversarial networks · Temporal factorizations · Truncation mechanism

1 Introduction

With the advent of modern intelligent transportation system (ITS), substantial studies on traffic data quality (e.g., integrity) have sparked considerable interests in academic/industrial communities [1]. However, sparse, low-quality or lost traffic data (e.g., flow, speed, density) remains an inevitable and ubiquitous problem caused by the potential software/hardware failures [2]. Once the issues occur at a critical moment, the interrupted traffic data analysis may induce ITS collapse. According to the reports, nearly 50%

of missing data exists from Alberta, Canada's database [3]; Texas Transportation Institute acknowledges the typical range of 16%–93% missing rates for its varied traffic systems [4]. Despite the increasingly sophistication of technologies for collecting, transmitting and storing traffic data, the missing rates have unavoidably risen due to the complexity of ITS [1, 3]. A number of factors can account for data loss, but the typical missing scenario mainly includes the random missing (RM), cluster missing (CM) and hybrid missing (HM) [2]. When faced with variety of serious predicaments, the artificial deleting or filling in data may provide faulty results for target system. Thus, performing an accurate recovery of traffic data has been a critical step before starting traffic activities [5].

Literature and scientific studies provide various recovery algorithms, which execute the regular solutions for repair tasks by exploiting the spatiotemporal correlations and environmental factors [4]. As the availability of computational resources improved, these algorithms have enabled research on large-scale traffic data repair. At

✉ Lunhui Xu
lhxu@scut.edu.cn

¹ School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510641, China

² School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, Shenzhen 518055, China

³ Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA

present, all relevant repair models roughly fall into three categories: traditional interpolation, statistical learning and machine learning (ML) [6] models. Specifically, the interpolation-based models fail to achieve satisfactory results because of their insufficient capability to capture the identical spatiotemporal properties from the historical, adjacent or regression data [1, 2]. By contrast, the statistical methods realize reliable fixes owing to the rational mathematical structure with clear interpretability [6]. Particularly, the matrix-based recovery models [7, 8] work well in various repair tasks owing to their exhibited robust nonlinear fitting ability. Further, the tensor preserves the inherent spatial-temporal features of data at a maximum due to its structural advantage, which makes it feasible to capture global and local information [9]. For example, Chen et al. [10] elevated the matrix factorization (MF) to higher Bayesian Gaussian CANDECOMP/PARAFAC (BGCP) decomposition model to learn the essential traffic patterns. Yet, tensor-based statistical models face some challenges, like a specific decomposition process, low scalability and hundreds of iterations to converge [11, 12].

Aside from statistical models, the ML-related models present excellent repair effects with their well-fitting ability. But compared to classical ML models (e.g., support vector regression (SVR) [6]), deep learning (DL) can provide more effective and accurate repair results with its stackable structure. An example is Li et al. [13] built a multi-modal DL framework for heterogeneous traffic data recovery. For this, a lot of scholars have creatively introduced generative adversarial networks (GAN) for data repair since it was proposed by Goodfellow et al. [14]. Zhang et al. [15] introduced a GAN-based travel times imputation (TTI-GAN) model to recover link travel time with GPS trajectories. Usually, GAN can achieve better performance by estimating the generative model under an adversarial process [16]. Compared to network-based algorithms, GAN requires less parameter selection and computational regulation. Even it can produce specified data from noise distribution, which provides the possibility for traffic data repair in extreme cases. However, the standard GAN lacks stability and tends to fall into collapse during the training process [17]. Besides, the GANs clearly lack a dynamic mechanism for self-adapting their loss iterative process. So far, each repair technique poses its own drawbacks, but the hybrid modeling can break through such constraints and reinforce its performance by taking full advantage of multiple individual methods. Inspired by the recent studies, we integrate the GAN and MF to jointly deliver the desirable repair performance for missing traffic data.

Before fully building the target model, we develop a novel dynamic adaptive GAN (DA-GAN) model to determine the ideal number of iterations during generation.

Considering the downsides of DA-GAN, we further propose a multi-view temporal factorizations (TFs) model. The joint paradigm TFs-DGAN combines the advantages of two modules: strong global generation capability and adept at tackling local temporal properties. Overall, the main contributions of this paper are as follows:

- (1) We propose a hybrid repair framework TFs-DGAN for modeling multivariate time series, which integrates the residuals from DA-GAN and TFs model outputs by means of the local feedback for adequately capturing the spatial-temporal dependencies of traffic data;
- (2) We design a DA mechanism embedded in the GAN algorithm with Savitzky–Golay (S–G) smoothing and 2nd-order derivation (2D) to dynamically identify the number of generator iterations;
- (3) We consider multiple temporal correlations (e.g., closeness, periods, trends) and construct a multi-view TFs model with truncation mechanism. By fully exploiting the underlying periodicities, our TFs perform the rolling recovery for driving the reduction in local elements in residuals;
- (4) We verify the effectiveness of TFs-DGAN on historical traffic datasets of Guangzhou and Hangzhou, China, and conduct the ablation study, in-depth analysis of extreme cases and complexity calculation. The empirical results prove the superiority (e.g., accuracy, stability) of TFs-DGAN over other state-of-the-art baseline models under various missing situations.

The rest of this article is organized as follows. In Sect. 2, we briefly review the relevant work on matrix/GAN-based repair methods; Sects. 3 and 4 describe the preliminaries and structure of TFs-DGAN model, respectively; in Sect. 5, we conduct the empirical study on two traffic datasets and accomplish the comparison with baseline models; Finally, Sect. 6 draws the conclusion.

2 Related literature

As mentioned earlier, most of the popular approaches for traffic data recovery broadly classify as discriminative methods (e.g., SVR [6]) and generative ones (e.g., GAN) [17]. Using different strategies, a range of literature give specific restoration results relying on adjacent data (spatially or temporally). In this section, we start with a review of traffic data recovery under typically encountered constraints (e.g., spatial/temporal dependencies), as well as compare the repair performance of the matrix-based probabilistic models and GAN-based generative models. From the technical perspective, the below subsections

detail the related literature of various topics studied by this paper.

Generally, urban areas are divided into irregular road networks because of natural environment and human activities, and the resulting spatial–temporal dependencies dominate the traffic operation. Many classical methods consider single spatial or temporal relationships but neglect the role of their joint modeling for repair tasks. For example, Tang et al. [6] built a recovery model to estimate traffic data, but it discarded the spatial information from different detectors. In terms of spatial correlation, the observations derived from adjacent sensors (e.g., loop detector and GPS) interact with each other. While this local coherence evolves accordingly with road environment, they contribute to data repair. Similarly, multiple types of temporal dependencies (e.g., closeness, periods, trends) among different time intervals can directly influence the observations. Yu et al. [18] established a temporal regularized MF (TRMF) model for a variety of time series problems, but it lacked an account for spatial features. Chen et al. [19] built a Bayesian temporal MF (BTMF) model to recover traffic data. In fact, modeling such spatial–temporal dependencies and then fusing their results with multi-views can achieve better repair accuracy and stability because of fully considering all inherent factors of traffic data [6].

Recently, the low-rank matrix completion (LRMC) has yielded promising results for data repair [20, 21]. Sure et al. [22] provided two LRMC models for traffic matrix reconstruction by adopting spatial and temporal constraints. Chen et al. [23] created a correlation-based LRMC model and its advanced version with ensemble learning (CLRMC-EN), which performed superior over LRMC in most cases. To repair incomplete time series with time-varying nonlinear latent structures, Fan et al. [24] proposed the dynamic nonlinear MC model with high accuracy and efficiency. However, the described LRMC cannot repair high rank or full-rank incomplete matrices [25], and Fan et al. [26] developed a high rank MC based on kernelized factorization that successfully solved this task. Unlike MCs that only consider partial features, Jia et al. [27] further applied the joint MF for modeling the periodicity, road similarity and temporal coherence in a 3D traffic data. But faced with the gradual optimization of the ensemble or joint algorithms, MFs always fail to achieve an effective fit since a 2D matrix cannot carry spatial–temporal properties of 3D data [8]. In general, the default MF can avoid overfitting caused by global low-rank structures [23] after designing appropriate regularization terms. For example, Wen et al. [28] presented a graph-regularized MF to fully exploit the local geometric similarities of incomplete data. Both Yang et al. [11] and Li et al. [29] designed different MF models with spatial–temporal regularizers for traffic

data repair. As an extension of MF, Fan et al. [30] further proposed multi-mode deep matrix and tensor factorizations to exploit the full nonlinearity of incomplete data to solve matrix and tensor completion problems. However, MFs still fail to provide an optimal solution for repair tasks because they cannot deeply exploit the heterogeneity of data from different temporal standpoints at low costs.

Currently, increasingly scholars have introduced the GAN into multi-dimensional data analysis and described its advantages in traffic data repair. For example, Yoon et al. [16] utilized the generative adversarial imputation nets (GAIN) with hint mechanism to recover several real-world datasets. Han et al. [12] established the content-aware traffic data completion method base on the GAN (CA-GAN) to efficiently repair the elements with consecutively missing cases. Inspired by the success of GANs, Xu et al. [31] applied GAN to process the information of adjacent links to estimate road traffic states. After modeling several features at different levels, Yang et al. [32] built the spatial–temporal learnable bidirectional attention GAN (ST-LBAGAN) for traffic data imputation. However, the disadvantages of GAN training difficulty and instability require to be tackled especially for high missing rate cases. Moreover, the available GANs still lack an early-stop mechanism when the loss of GANs decreases with the increase in the number of iterations. To the extent of our knowledge, the hybrid learning of multi-view spatial–temporal properties stored in scalable matrices can greatly facilitate data recovery. Sun et al. [33] employed the graph convolutional network (GCN) to model multiple types of temporal correlations and all outputs were fed into a fusion module. Liu et al. [34] applied a multi-view spatial–temporal model (MVSTM) for capturing the mutual dependencies of nonlinear trajectory features. Motivated by the global–local variations of spatial–temporal properties, Li et al. [4] proposed a multi-view learning method (MVLM) to recover missing traffic values. To make up for above deficiencies, we develop a hybrid repair model TFs-DGAN, which can generate traffic data by an optimized GAN and ensure the optimal performance by fusing the imputation results of multiple TF models.

3 Preliminaries

3.1 Topology of the traffic network

In the traffic tasks, the complex time series can be organized as a matrix [11, 19]. Taking Fig. 1 as instance, we define a real traffic network as 2D topology $\mathcal{G}(I, J)$, where the I is the number of sensors and the J represents the product of the days t and the number of traffic data collected by each sensor during a day. For the matrix

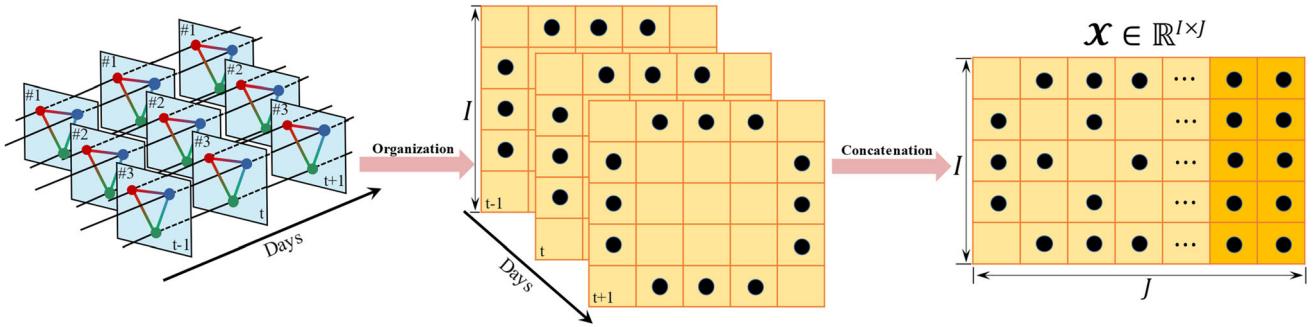


Fig. 1 Topology of the traffic network. Note that the t denotes the days of traffic data collection

$\mathcal{X} \in \mathbb{R}^{I \times J}$, modeling the spatial-temporal correlations precisely provides the natural and effective solutions for traffic data repair tasks.

3.2 Missing spatial-temporal traffic data

As stated above, the traffic data matrix of $\mathcal{X} \in \mathbb{R}^{I \times J}$ describes the spatial-temporal relationships of the entities in traffic network, where $\mathcal{X}_j \in \mathbb{R}^I$ shows the traffic states of all I sensors at the j th step. At the element level, the x_{ij} represents the observation collected by the i th sensor at the j th timestep, $i \in \{1, 2, 3, \dots, I\}, j \in \{1, 2, 3, \dots, J\}$. In Eq. (1), we define the mask matrix \mathcal{M} with the same shape as \mathcal{X} , which takes values in $\{0, 1\}$ according to the actual elements of \mathcal{X} .

$$m_{ij} = \begin{cases} 1, & \text{if } x_{ij} \in \mathcal{X}_{\Omega} \\ 0, & \text{if } x_{ij} \text{ is missing} \end{cases} \quad (1)$$

where \mathcal{X}_{Ω} is the partially observed dataset, which satisfies $\mathcal{X}_{\Omega} = \mathcal{X} \odot \mathcal{M}$; Ω is the index set and the “ \odot ” stands for the Hadamard product between matrices. Considering the different forms of missing data, we define the x_{ij} as missing value when it is abnormal. For the repair tasks, we aim to fill these outliers by learning the algebraic structure of \mathcal{X}_{Ω} .

3.3 Problem formulation

With the formulations of data above, we realize that the traffic data recovery relies on modeling the spatial-temporal correlations of these explanatory variables. Many available models suffer from insufficient granularity to address traffic features alone. In terms of methodology, we describe those variables that show a distinction across space/time as spatial/temporal attributes [35], respectively, and the repair task can be formulated as follows:

$$\tilde{\mathcal{X}} = f_x((\mathcal{X}_s, \mathcal{X}_t, \mathcal{X}_e; \mathcal{G}, \mathcal{M} | x_{ij}, (i, j) \notin \Omega)) \quad (2)$$

where $\mathcal{X}_s \in \mathbb{R}^I$ stands for the spatial variables indicating the dependencies between different sensors; the temporal

variable $\mathcal{X}_t \in \mathbb{R}^J$ contains the adjacent time, time-of-day, days-of-week and so on; \mathcal{X}_e is the external variable; $\tilde{\mathcal{X}} \in \mathbb{R}^{I \times J}$ represents the imputed traffic matrix; \tilde{x}_{ij} is the missing traffic data. Although all the variable information can contribute to the traffic data recovery, the spatial-temporal correlations dominate in it, and later we will discuss the details of their modeling in repair tasks.

4 Methodology

4.1 Spatial-temporal generation model optimization

4.1.1 Generative adversarial networks

Architecturally, both the GAN and its variants satisfy a basic framework formed by a generator (G) and discriminator (D), where G produces samples from noise while D judges their authenticity. After training, GAN meets an equilibrium at which the generative ability of G and the discriminative ability of D both reach the maximum. Motivated by the success of GAN in the generative tasks [12, 36], we intend to train GAN to learn the intrinsic distribution p_{real} of traffic data to impute the lost entries. Specifically, the repair task can be viewed as an optimization procedure for a min–max game by:

$$\min_G \max_D V(D, G) = \underbrace{\mathbb{E}_{x \sim p_{\text{real}}(x)} [\log D(\mathcal{X}; \mathcal{M})]}_{L_G} + \underbrace{\mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(\mathcal{Z}; \mathcal{M}); \mathcal{M}))]}_{L_D} \quad (3)$$

where L_G and L_D denote the generative and adversarial processes of GAN separately; p_z is a random noise distribution and \mathcal{Z} is a noise (i.e., $\mathcal{Z} \sim p_z$); $D(\mathcal{X}; \mathcal{M})$ and $G(\mathcal{Z}; \mathcal{M})$ are the neural networks with mask matrix \mathcal{M} in this study. When \mathcal{Z} is fed into G for calculation, we implicitly define a probability distribution p_g for G and our

aim is to approximate p_{real} by p_g . In theory, any differentiable function is allowed as D and G . For a G fixed, we can maximize the quantity $V(D, G)$ and find an optimal D_G^* .

$$D_G^*(\mathcal{X}; \mathcal{M}) = \frac{p_{\text{real}}(\mathcal{X}; \mathcal{M})}{p_{\text{real}}(\mathcal{X}; \mathcal{M}) + p_g(\mathcal{X}; \mathcal{M})} \quad (4)$$

where the training goal is to maximize the log-likelihood for estimating the conditional probability $P(Y = y|\mathcal{X})$, where Y is whether \mathcal{X} comes from p_{real} or p_g . The above issue can be reformulated as:

$$\begin{aligned} \mathcal{C}(G) &= \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{real}}} [\log D_G^*(\mathcal{X}; \mathcal{M})] \\ &\quad + \mathbb{E}_{x \sim p_g} [\log (1 - D_G^*(\mathcal{Z}; \mathcal{M}))] \\ &= \mathbb{E}_{x \sim p_{\text{real}}} [\log D_G^*(\mathcal{X}; \mathcal{M})] + \mathbb{E}_{x \sim p_g} [\log (1 - D_G^*(\mathcal{X}; \mathcal{M}))] \\ &= \mathbb{E}_{x \sim p_{\text{real}}} \left[\log \frac{p_{\text{real}}(\mathcal{X}; \mathcal{M})}{p_{\text{real}}(\mathcal{X}; \mathcal{M}) + p_g(\mathcal{X}; \mathcal{M})} \right] \\ &\quad + \mathbb{E}_{x \sim p_g} \left[\log \frac{p_g(\mathcal{X}; \mathcal{M})}{p_{\text{real}}(\mathcal{X}; \mathcal{M}) + p_g(\mathcal{X}; \mathcal{M})} \right] \end{aligned} \quad (5)$$

Next, Eq. (5) reaches its global optimum (at $p_g = p_{\text{real}}$) when considering the Jensen–Shannon divergence among p_g and p_{real} is nonnegative and G perfectly replicates the data generating process.

4.1.2 Dynamic adaptive mechanism of GAN

Despite this great success, the standard GAN faces its challenges [14, 17, 37], and thus reaps various variants by taking different strategies [31, 38]. Inherited from the classical repair models [16], we further consider that the training of GAIN lacks an adaptive process that dynamically adjusts its number of iterations. Actually, the fixed iteration does restrict the ability of GAIN rather than always reducing the errors as desired. Therefore, we design an improved GAN algorithm with DA mechanism, which can auto-adapt the number of iterations of the GAN by monitoring the loss of G . In Eq. (6), G outputs all components of $\tilde{\mathcal{X}}$ whether they are absent or not.

$$\tilde{\mathcal{X}} = G(\mathcal{X}_\Omega, \mathcal{M}, (1 - \mathcal{M}) \odot \mathcal{Z}) \quad (6)$$

Similar to the GAIN setups, we then splice the observed data within \mathcal{X}_Ω and the generated data from $\tilde{\mathcal{X}}$ to form the complete matrix $\hat{\mathcal{X}}$ in Eq. (7).

$$\hat{\mathcal{X}} = \mathcal{M} \odot \mathcal{X}_\Omega + (1 - \mathcal{M}) \odot \tilde{\mathcal{X}} \quad (7)$$

Next, we introduce the D with a hint mechanism [16] in GAIN to verify “real” or “fake” samples. We input the well-defined random variable \mathcal{H} into D to adjust the amount of information contained in \mathcal{H} about \mathcal{M} . As the objective function defined in Eq. (8) for this min–max

problem, we train G to minimize the probability of D distinguishing \mathcal{M} while ensuring the best discrimination of D .

$$\begin{aligned} \min_{G} \max_{D} \mathbb{E} \left[\sum_{i=1}^I \sum_{j=1}^J \left[\mathbf{m}_{ij} \log \left(\left(D_G^*(\hat{\mathcal{X}}; \mathcal{H}) \right)_{ij} \right) \right. \right. \\ \left. \left. + (1 - \mathbf{m}_{ij}) \log \left(1 - \left(D_G^*(\hat{\mathcal{X}}; \mathcal{H}) \right)_{ij} \right) \right] \right] \end{aligned} \quad (8)$$

Although the GAIN generates data without competitive precision, it provides a promising basis for us to introduce a DA mechanism. In Fig. 2, we monitor the train_loss/test_loss value of G instead of the loss of $G(G_loss)$ and $D(D_loss)$ because the former relates directly to the repair results.

During the iterations of GAN, the train_loss/test_loss update over time and we collect them by Eqs. (9) and (10), respectively, and the obtained $\mathcal{S}_{\text{train}}/\mathcal{S}_{\text{test}}$ are retained.

$$\mathcal{L}_{\text{train}} = \frac{1}{|\Omega|} \sum_{(i,j) \in \Omega} \left((\mathcal{M} \odot \mathcal{X})_{ij} - (\mathcal{M} \odot \hat{\mathcal{X}})_{ij} \right)^2 \xrightarrow{\text{yields}} \mathcal{S}_{\text{train}} \quad (9)$$

$$\begin{aligned} \mathcal{L}_{\text{test}} = \frac{1}{|1 - \Omega|} \sum_{(i,j) \in (1 - \Omega)} \left(((1 - \mathcal{M}) \odot \mathcal{X})_{ij} \right. \\ \left. - ((1 - \mathcal{M}) \odot \hat{\mathcal{X}})_{ij} \right)^2 \xrightarrow{\text{yields}} \mathcal{S}_{\text{test}} \end{aligned} \quad (10)$$

In fact, the noise is inevitable for time series. In Eq. (11), we use $\mathcal{S}_{\text{train}}$ and $\mathcal{S}_{\text{test}}$ for identifying the number of GAN iterations. Note that the $\hat{\mathcal{S}}()$ denotes the smoothed loss vectors.

$$\mathcal{S}(y) \xrightarrow{\text{Savitzky–Golay}} \hat{\mathcal{S}}(y) \xrightarrow{\nabla} \hat{\mathcal{S}}''(y) = \frac{d^2 \hat{\mathcal{S}}(y)}{dy^2} \quad (11)$$

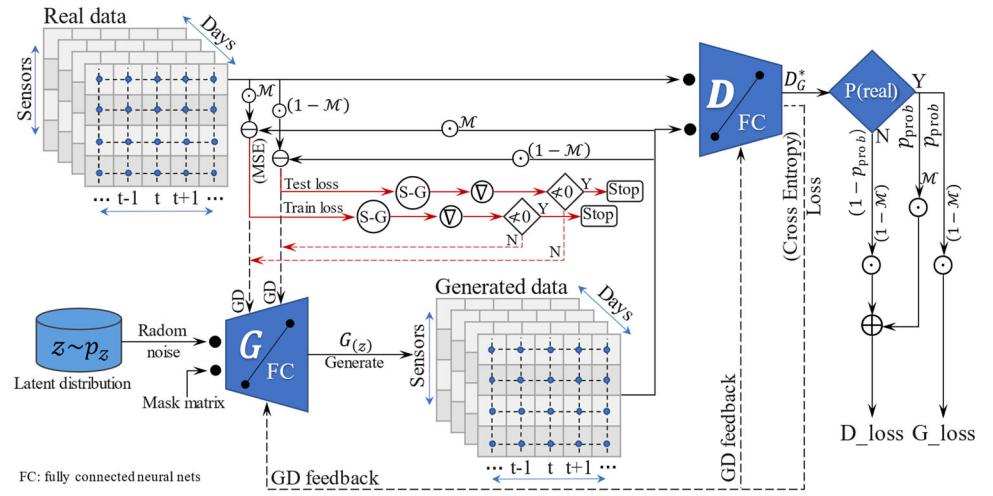
In particular, the DA mechanism incorporates S–G smoothing and 2D. As a low-pass filter, the S–G is heavily applied for data denoising [39] and we adopt it to smooth the $\mathcal{S}_{\text{train}}/\mathcal{S}_{\text{test}}$ of G to observe the trend of GAN iteration. As in Eq. (12), we construct a polynomial of order n for $\mathcal{S}_{\text{train}}/\mathcal{S}_{\text{test}}$ with $(2w+1)$ loss values, where w is half of the series’ width and $[c_{n0}, c_{n1}, c_{n2} \dots c_{nn}]^T$ is its coefficient.

$$[\mathcal{S}(-w), \dots, \mathcal{S}(0), \dots, \mathcal{S}(w)] \in \mathcal{S}_{\text{train}}/\mathcal{S}_{\text{test}}$$

$$\begin{aligned} f_y &= c_{n0} + c_{n1}y + c_{n2}y^2 + \dots + c_{nn}y^n \\ &= \sum_{k=0}^n c_{nk}y^k, n \leq 2w+1 \end{aligned} \quad (12)$$

Next, we calculate the MSE function $\mathfrak{T}(w; n)$ between the original values and corrected values in Eq. (13) and then fit them by the least square method.

Fig. 2 Illustration of the structure of the GAN model with DA mechanism



$$\begin{aligned} \mathfrak{T}(w; n) &= \sum_{y=-w}^w (f_y - \mathcal{S}(y))^2 \\ &= \sum_{y=-w}^w \left(\sum_{k=0}^n c_{nk} y^k - \mathcal{S}(y) \right)^2 \end{aligned} \quad (13)$$

To minimize $\mathfrak{T}(w; n)$, we calculate the partial derivative for its coefficients in Eq. (14):

$$\begin{aligned} \frac{\partial \mathfrak{T}(w; n)}{\partial c_{nr}} &= 2 \sum_{y=-w}^w \left(\sum_{k=0}^n c_{nk} y^k - \mathcal{S}(y) \right) y^r, \\ &= (0, 1, 2 \dots n) \end{aligned} \quad (14)$$

As a result, we find the solution as follows:

$$\sum_{y=-w}^w y^r + \sum_{k=0}^n c_{nk} = \sum_{y=-w}^w \mathcal{S}(y) y^r \quad (15)$$

Since the training of GAN is a dynamic process, $\mathcal{S}(y)$ and $\widehat{\mathcal{S}}(y)$ keep being updated. But note that the smoothing result on $\mathcal{S}(y)$ depends on w/n values, which can usually be determined empirically. Further, we complete the 2D calculation of the smoothed $\mathcal{S}(y)$ in Eq. (11), which aims to locate the turning point where the loss of $\mathcal{S}(y)$ starts to counter-increase. Similar to an early-stop strategy, our DA mechanism is capable of identifying an excellent iteration before the training of GAN undergoes turbulence, which provides a robust basis for subsequent in-depth traffic data imputation.

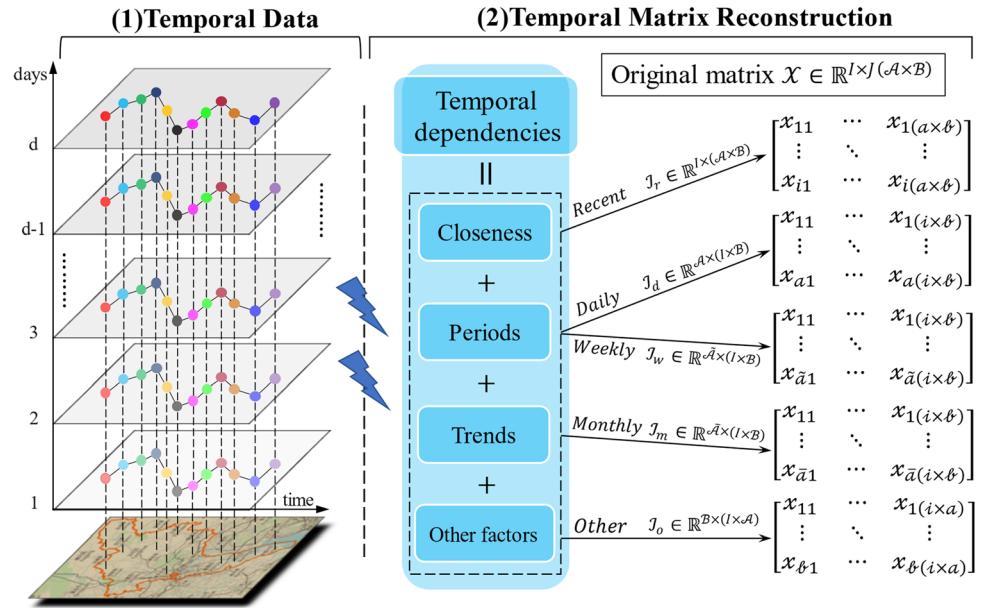
4.2 Multi-view temporal dependencies learning

While the large improvement of DA-GAN over GAIN, our analysis reveals that DA-GAN still lacks sufficient ability for repair tasks as evidenced by its inferior results to non-GAN methods [2, 10]. In fact, the temporal factors are vital but not fully considered. Next, we will exploit finer information by temporal modeling within a hybrid framework to further reduce repair errors. As shown in Fig. 3, we display the representation and matrix reconstruction of the multi-view temporal properties.

Of these, the closeness depicts the relationship between adjacent times; the periods indicate the relationships of the consecutive workdays (daily period)/weekends (weekly period); the trends (e.g., monthly) rely on factors like environment, which in turn affects traffic systems. As for each attribute, we reshape the matrix of $\mathcal{X} \in \mathbb{R}^{I \times J}$ within equal intervals and all periodicities will naturally diminish with increasing time. In particular, we assume that all collecting days is re-defined as \mathcal{A} , the number of traffic data collected in one day at a fixed resolution is \mathcal{B} and the matrix $\mathcal{X} \in \mathbb{R}^{I \times J}$ can be rewritten as $\mathcal{X} \in \mathbb{R}^{I \times (\mathcal{A} \times \mathcal{B})}$. Thus, we reconstruct five temporal matrices for recent, daily, weekly, monthly and other properties as $\mathcal{I}_r \in \mathbb{R}^{I \times (\mathcal{A} \times \mathcal{B})}$, $\mathcal{I}_d \in \mathbb{R}^{\mathcal{A} \times (I \times \mathcal{B})}$, $\mathcal{I}_w \in \mathbb{R}^{\widetilde{\mathcal{A}} \times (I \times \mathcal{B})}$, $\mathcal{I}_m \in \mathbb{R}^{\overline{\mathcal{A}} \times (I \times \mathcal{B})}$ and $\mathcal{I}_o \in \mathbb{R}^{\mathcal{B} \times (I \times \mathcal{A})}$, separately. Note the $\widetilde{\mathcal{A}}$ and $\overline{\mathcal{A}}$ are the sets of days in the neighboring week and month, respectively. Further, we establish a scalable TFs model by the vector autoregressive (VAR) process to model the temporal matrixes. In Eq. (16), the matrix \mathcal{X} (for \mathcal{I}_r , \mathcal{I}_d , \mathcal{I}_w , \mathcal{I}_m and \mathcal{I}_o) is approximately decomposed into $\mathcal{Q} \in \mathbb{R}^{\mathcal{R} \times I}$ and $\mathcal{T} \in \mathbb{R}^{\mathcal{R} \times J}$, where \mathcal{R} is far smaller than $\min\{I, J\}$.

$$\mathcal{X} \approx \mathcal{Q}^T \mathcal{T} \quad (16)$$

Fig. 3 Representation and matrix reconstruction of multi-view temporal properties



Unlike in TRMF [18], VAR can avoid excessive regularized parameters when modeling temporal matrixes, and we employ it to describe these temporal dependencies in \mathcal{T} as follows in Eq. (17).

$$\boldsymbol{\ell}_{j+1} = \sum_{\varphi=1}^{\delta} \mathcal{D}_{\varphi} \boldsymbol{\ell}_{j+1-\varphi} + \boldsymbol{\epsilon}_j \quad (17)$$

where the size of each coefficient matrix \mathcal{D}_{φ} is of $\mathcal{R} \times \mathcal{R}$, which stands for the partially latent space; δ is the order of VAR; $\{\varphi_1, \varphi_2, \dots, \varphi_{\delta}, \dots, \varphi_{\delta}\}$ denotes the time lag set; $\boldsymbol{\epsilon}_j$ is a Gaussian noise vector.

We assume that all x_{ij} are independent and satisfy a Gaussian distribution (GD) of precision τ_i .

$$x_{ij} \sim \mathcal{N}(\boldsymbol{\varphi}_i^T \boldsymbol{\ell}_j, \tau_i^{-1}), (i, j) \in \Omega \quad (18)$$

where τ_i complies with the Gamma prior of $\tau_i \sim \text{Gamma}(\alpha, \beta)$ that α/β are the shape/rate parameters,

respectively; the $\boldsymbol{\varphi}_i$ satisfies the multivariate GD of $\boldsymbol{\varphi}_i \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\varphi}}, \boldsymbol{\Lambda}_{\boldsymbol{\varphi}}^{-1})$. Additionally, in Eq. (19), the conjugate Gaussian–Wishart prior is applied to solve the hyper-parameters $(\boldsymbol{\mu}_{\boldsymbol{\varphi}}, \boldsymbol{\Lambda}_{\boldsymbol{\varphi}})$.

$$\boldsymbol{\mu}_q | \boldsymbol{\Lambda}_q \sim \mathcal{N}(\boldsymbol{\mu}_0, (\omega_0 \boldsymbol{\Lambda}_q)^{-1}), \boldsymbol{\Lambda}_q \sim W(W_0, v_0) \quad (19)$$

where $\boldsymbol{\mu}_0 \in \mathbb{R}^{\mathcal{R}}$ is a mean vector; $W(W_0, v_0)$ is a Wishart distribution with a $\mathcal{R} \times \mathcal{R}$ scale matrix W_0 and v_0 degrees of freedom. For the matrix \mathcal{T} , we place the conjugate matrix normal inverse Wishart (MNIW) prior [19] on matrix \mathcal{D} which satisfies the $\mathcal{D} \sim MN_{(\mathcal{R}d) \times \mathcal{R}}(\mathbf{M}_0, \Phi_0, \Psi)$.

Of these, Ψ meets the $\Psi \sim IW(\mathcal{P}_0, v_0)$, we present the probability density function of the matrix \mathcal{D} as:

$$p(\mathcal{D} | \mathbf{M}_0, \Phi_0, \Psi) = (2\pi)^{-\mathcal{R}^2 d/2} |\Phi_0|^{-\mathcal{R} \odot 2} |\Psi|^{-\mathcal{R} d/2} \times \exp\left(-\frac{1}{2} \text{tr}[\Psi^{-1} \mathcal{O}^T \Phi_0^{-1} \mathcal{O}]\right) \quad (20)$$

where the matrices $\Psi \in \mathbb{R}^{\mathcal{R} \times \mathcal{R}}$ and $\Phi_0 \in \mathbb{R}^{(\mathcal{R}d) \times (\mathcal{R}d)}$; the \mathcal{O} equal to the difference between \mathcal{D} and \mathbf{M}_0 .

In doing so, we apply the non-informative priors [18, 40] to determine all the constant parameters. We employ Gibbs sampling [10] for all of the variables in each iteration, such as $\boldsymbol{\varphi}_i$, $\boldsymbol{\ell}_j$, $\boldsymbol{\mu}_{\boldsymbol{\varphi}}$, $\boldsymbol{\Lambda}_{\boldsymbol{\varphi}}$, \mathcal{D} , Ψ . Thus, each TF module provides an efficient solution to model the temporal views in the latent space, and the outputs of all modules are fed into the multi-view fusion module (see Sect. 4.4) followed by the local residual feedback to obtain the final repair result.

4.3 Overview of TFs-DGAN framework

Figure 4 illustrates the spatial–temporal hybrid framework TFs-DGAN. As clearly available that the recovery model mainly consists of both co-trained modules: an optimized generation model DA-GAN which can perform the repair tasks imperfectly and a multi-view TFs model which can further model the temporal properties. In fact, the first part equals the data preparation phase, while the later temporal modeling of the resulting residuals is the core of TFs-DGAN model. Thus, we assume that the output of DA-GAN is a residual matrix $\mathcal{X}_{\mathcal{E}}$, which satisfies $\mathcal{X}_{\mathcal{E}} = \tilde{\mathcal{X}} - \hat{\mathcal{X}}$ and we reshape it as \mathcal{I}_r . For the temporal matrices (like \mathcal{I}_r , \mathcal{I}_d , \mathcal{I}_w), TFs model with truncation mechanism further

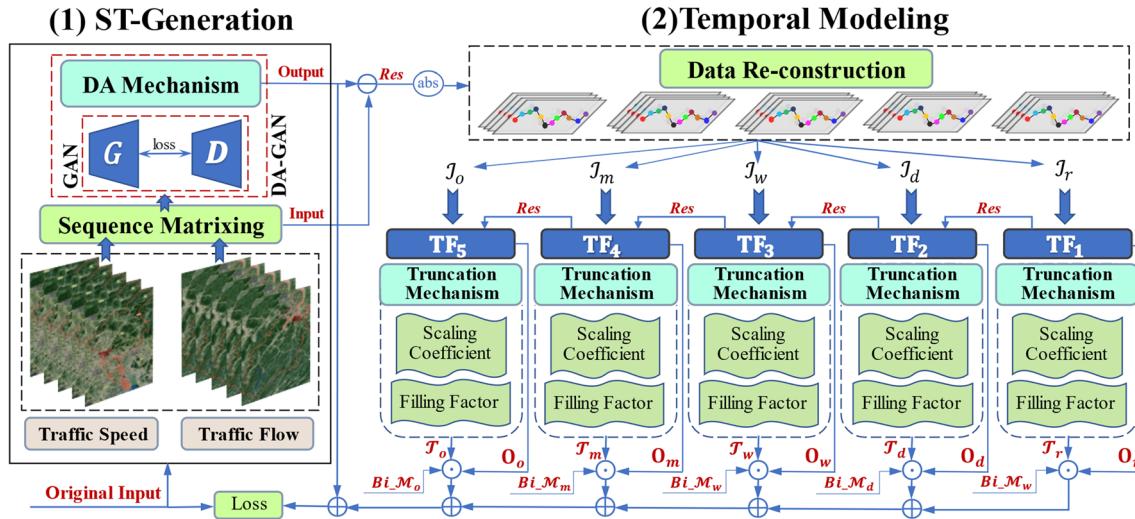


Fig. 4 Spatial-temporal (ST) hybrid architecture of the TFs-DGAN model

reduces its residuals as per Sect. 4.2. Specifically, the final result of the residual learning is as follows:

$$\mathcal{X} = \hat{\mathcal{X}} + \sum \mathcal{F}_{\text{TF}}(\mathcal{I}_{\ell}; \theta_{\ell}) \quad (21)$$

where \mathcal{F}_{TF} is the TF module; ℓ denotes the temporal properties; θ includes all learnable parameters (e.g., T_{ℓ} , O_{ℓ} , $\mathcal{B}_i\mathcal{M}_{\ell}$) in each temporal modeling and they come with the residual matrix \mathcal{X}_{ℓ} .

In particular for TF model, although the stochastic gradient descent method motivates it to satisfy Eq. (16), it fails to control the reduction in the elements in \mathcal{X}_{ℓ} . Thus, in Fig. 4, the novel truncation mechanism involving scaling coefficient φ and filling factors ($\Phi = \{\xi_r, \xi_d, \xi_w, \xi_o\}$) is introduced to reduce their local residuals \mathcal{X}_{ℓ} . Specifically, for φ , we calculate Eq. (22):

$$\mathcal{X}_{\ell} = \mathcal{I}_{\ell} - \varphi \times O_{\ell} \quad (22)$$

where we resize the elements in \mathcal{X}_{ℓ} by altering the coefficient φ , especially for the values in O_{ℓ} that vary greatly relative to the values in \mathcal{I}_{ℓ} . For all generated data, if they satisfy $\mathcal{X}_{\ell} \geq O_{\ell}$, we uniformly replace them with the filling factors Φ . In reality, the different φ and Φ impact the filling amount in \mathcal{X}_{ℓ} , and we apply a parametric-matrix-based fusion method [33] to merge all TF outputs as below:

$$\sum \mathcal{F}_{\text{TF}}(\mathcal{I}_{\ell}; \theta_{\ell}) = \sum O_{\ell} \odot T_{\ell} \odot \mathcal{B}_i\mathcal{M}_{\ell} \quad (23)$$

where O_{ℓ} refers to the results of TFs modeling for ℓ features; T_{ℓ} is the matrix of the same size as O_{ℓ} ; and we extract the direction matrix $\mathcal{B}_i\mathcal{M}_{\ell}$ for each \mathcal{X}_{ℓ} . By

stacking multiple such TF residual units, we can perform more refined network-scale traffic data repair. Of note is that we can train all models using only the back propagation and dropout algorithms [41], and we employ the adaptive momentum (Adam) algorithm [12, 32] to update all parameters of the TFs-DGAN model.

5 Numerical experiments

5.1 Datasets preparation

5.1.1 Spatial-temporal traffic datasets

To practically evaluate the performance of TFs-DGAN model for traffic data repair, we conduct substantial experiments on two real-world traffic datasets [10, 42] in Guangzhou and Hangzhou, China:

- Dataset(G): Guangzhou urban traffic speed. The large-scale dataset collected traffic speed from the navigation apps distributed on 214 road segments at a 10-min resolution over 61 days (August 1 to September 30, 2016). There are about 1.8 million data, but nearly 1.29% of them are unavailable.
- Dataset(H): Hangzhou metro passenger flow. The dataset registered incoming passenger flow from 80 metro stations at a 10-min resolution over 25 days (January 1 to January 25, 2019). Since the interval 0:00 a.m.–6:00 a.m. without services is dropped, there are 216,000 elements.

5.1.2 Outliers, missing rates and missing scenarios

As a supervised learning task, the outliers from diverse sources may impair model performance [43]. In Fig. 5, we can find that the traffic data from the sensors $\{47, 54, 122, 145, 146\}$ and $\{9, 15\}$ are not normal, where the former are all zero and the latter are clearly above the natural range. Although the outliers have been neglected in related studies, the proper treatment facilitates the model training. In this study, we replace outliers in the dataset(G) by using the average of observations from multiple timestamps or adjacent sensors, while excluding outliers in the dataset(H). As in Table 4, we provide the comparison of the impact of whether to eliminate outliers on the repair results.

After removing outliers, we consider three types of missing scenarios without losing generality. In particular, the HM scenario which we should emphasize derives precisely from the joint calculation of RM and CM modes. As in Table 1, the combinations of three patterns and the missing rates (10%–99.99%) cover essentially all missing cases, which also satisfy the experimental requirements.

5.2 Experimental settings

5.2.1 Benchmark models

We compare the repair performance of TFs-DGAN with the following benchmark models:

- HaLRTC [44]: *High accuracy Low-Rank Tensor Completion* (HaLRTC) is a well-known model which has been widely applied for missing traffic data recovery.

- GAIN [16]: *Generative Adversarial Imputation Nets* (GAIN) proves its superiority at completing the imputation tasks by the generative process with an adversarial mechanism.
- BGCP [10]: *Bayesian Gaussian CANDECOMP/PARAFAC* (BGCP) model repairs traffic data by extending the Bayesian probabilistic MF model, and we adjust its CP ranks to adapt to our datasets.
- FNNTEL [1]: *Tensor heterogeneous Ensemble Learning* based on *Fuzzy Neural Network* firstly introduces FNN into the heterogeneous ensemble learning of tensor decomposition for accuracy and robust imputation of missing traffic data.
- LRTC-TNN [42]: *Low-Rank Tensor Completion* with *Truncated Nuclear Norm* (LRTC-TNN) model delivers accurate repair results after using the alternating direction method of multipliers.
- LATC [9]: *Low-rank Autoregressive Tensor Completion* (LATC) model confirms its superiority via numerical experiments after well capturing the global-local consistency of traffic data.
- ST-BiRT [2]: *Spatial-Temporal Bi-directional Residual optimization* (ST-BiRT) based on *tensor decomposition* maximally captures the spatial-temporal dependencies of traffic data by dynamically stacking massive residual units.

5.2.2 Implementation details

Before the model training, we update the experimental settings for all baselines properly based on the datasets(G) and (H). For HaLRTC, we specify the learning parameters of three scenarios (i.e., RM, CM, HM) as 10^{-4} , 10^{-4} , 10^{-5} ,

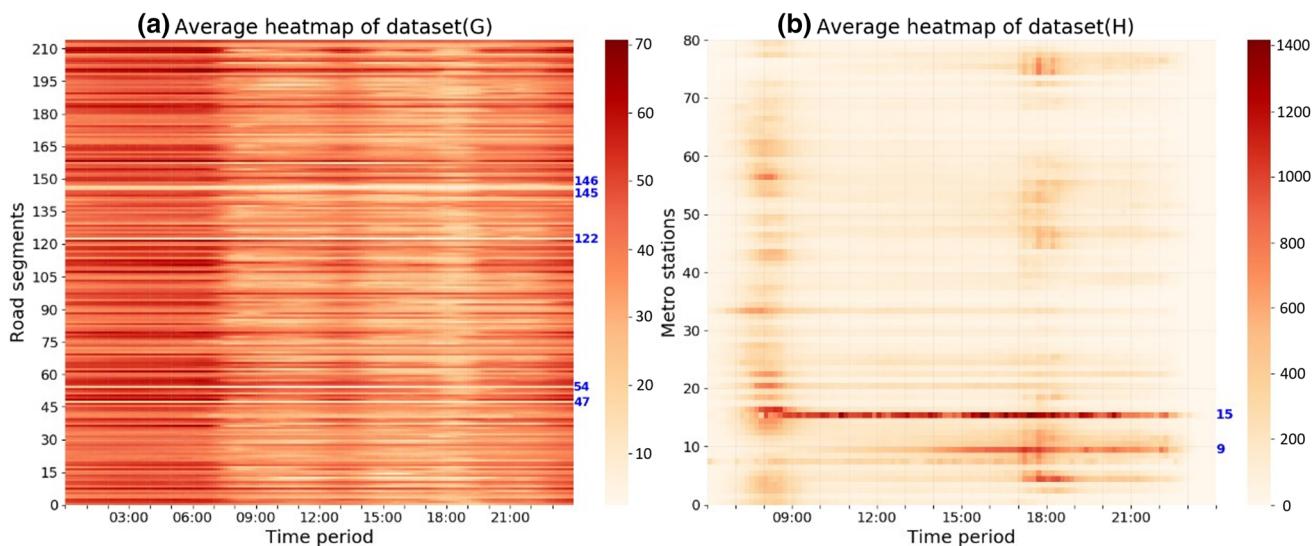


Fig. 5 Average heatmaps of the datasets (G) and (H)

Table 1 The calculated missing rates for the datasets (G) and (H) in the HM scenario

Datasets/Modes	MisRates (%)								
	HM(Determinate)								
	10	20	30	40	50	60	70	80	
(G)	RM/CM	5.1112	10.3780	16.2550	22.5719	29.2351	36.6875	45.2173	55.3090
	HM (Accurate)	9.9999	19.9999	29.9945	39.9999	50.0000	59.9959	69.9997	79.9999
(H)	RM/CM	3.8945	9.7245	15.1350	21.8740	28.6427	35.7250	43.6135	53.5746
	HM (Accurate)	10.0000	20.0000	30.0000	40.0000	50.0093	60.0037	70.0000	80.0000
Datasets/Modes	MisRates (%)								
	HM(Determinate)								
	90	91	93	95	97	99	99.9	99.99	
(G)	RM/CM	68.1478	69.8200	73.4893	77.5625	82.6285	89.9178	96.6949	98.8970
	HM (Accurate)	89.9981	91.0006	92.9991	94.9998	96.9999	98.9999	99.8998	99.9899
(H)	RM/CM	67.6157	69.3735	72.9200	77.0750	81.7106	89.8559	96.8700	98.9976
	HM (Accurate)	89.9981	91.0000	93.0000	95.0000	96.9977	99.0019	99.9000	99.9903

respectively, and the number of iterations are all 200; for GAIN, the iteration count of adversarial process is 1000 and the hyperparameter of loss function of G is 10; for BGCP, its CP ranks for three patterns are 110, 80, 40, separately; for LRTC-TNN, the truncation parameters of three modes are 0.3, 0.05, 0.05, accordingly, and the learning parameter and maximum iterations are 10^{-5} , 200, respectively; for LATC, the learning rate, truncation parameter and trade-off coefficient are 10^{-5} , $\{5, 10, 15, 20, 25, 30\}$ and $\{1/10, 1/5, 1, 5, 10\}$, separately; for ST-BiRT, the optimal tensor ranks for three scenarios are 110, 80, 50, respectively, and the residual units are 200; for FNNTEL, we can provide test results on the dataset (G) (RM mode only, others are not available).

5.2.3 Performance metrics

We employ three performance metrics (i.e., MAPE, RMSE and MAE) to evaluate and compare the repair results of all models. The detailed formulations of the three metrics are written as:

$$\text{MAPE} = \frac{1}{|\Omega|} \sum_{(i,j) \in \Omega} \left| \frac{x_{ij} - \hat{x}_{ij}}{x_{ij}} \right| \times 100 \quad (24)$$

$$\text{RMSE} = \sqrt{\frac{1}{|\Omega|} \sum_{(i,j) \in \Omega} (x_{ij} - \hat{x}_{ij})^2} \quad (25)$$

$$\text{MAE} = \frac{1}{|\Omega|} \sum_{(i,j) \in \Omega} |x_{ij} - \hat{x}_{ij}| \quad (26)$$

where $|\Omega|$ is the size of the index set Ω of the \mathcal{X}_Ω ; x_{ij} and \hat{x}_{ij} are the actual value and its estimation, respectively. In general, the smaller the three metrics, the better the model repair performance [2, 29].

5.3 Performance analysis of TFs-DGAN model

5.3.1 Effects of dynamic iteration on DA-GAN model

At the training stage, G and D minimize their loss values by iterations to enable the true spatial-temporal distribution learned by GAN. In fact, over-fitting or under-fitting occurs irregularly due to the fixed iterative calculation. As drawn in Fig. 6, although both train_loss/test_loss have decreased substantially overall, the latter begins to grow inversely with increasing number of iterations, which indicates that the repair performance of GAIN starts to diminish. Simultaneously, this phenomenon exhibits an accelerated incremental trend with the changes in missing rates/scenarios. For one aspect, the high missing rate impairs their repair ability as it only captures less information from traffic data. For another aspect, the sensitivity of RM, HM, CM to data loss rises sequentially since they corrupt the data structure increasingly. Besides, GAN also lacks sufficient ability to repair data exactly under the combined effects of missing rates/scenarios, which manifests itself in periodic and right-shifting jagged variation in the test_loss curve. Hence, it is crucial for GAN to abort its training early in light of its own loss of information, such as the test_loss(red) of G .

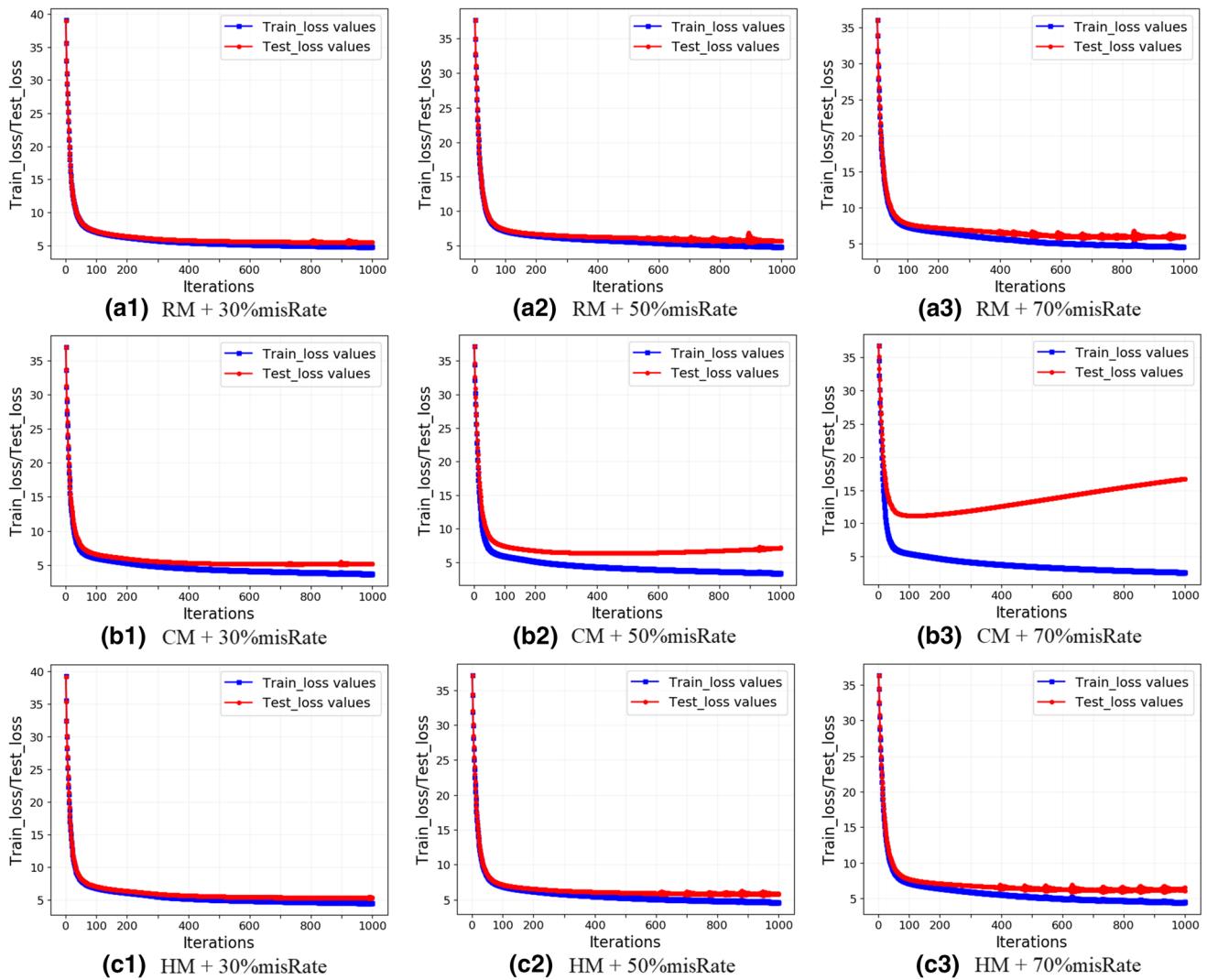


Fig. 6 Variation of the loss values of the GAN model with the number of iterations on the dataset (G)

An effective solution is that we introduce a DA mechanism to monitor the variation of test_loss of G , which enables the GAN model to automatically determine an excellent number of iterations. Specifically, DA mechanism adopts S-G smoothing with 2D to iteratively deal with the accumulated test_loss of S_{test} , where the former filters out the effect of interference factors to reflect the dominant trend of S_{test} reduction, while 2D can precisely detect the inflection point of S_{test} curve. As a control, we select the optimal number of iterations ((G):100; (H):200) for GAN based on the transitive region in Fig. 6, and we set the fixed iteration of 1000 for DA-GAN for its dynamic selection during training. As in Table 2, we report the evaluation metrics and final loss values for GAN and DA-GAN models on both datasets. Of these, we realize that the DA-GAN delivers a large improvement over GAN in terms of

repair accuracy for most missing cases. And this DA mechanism significantly enhances the modeling efficiency over the fixed iteration setting. Further, our DA-GAN also exceeds GAN model from the ratio of the test_loss to train_loss, which is more over-fitted. As an advance, our DA-GAN model will further strengthen its repair performance by multi-view temporal modeling.

5.3.2 Influences of truncated mechanism on TFs-DGAN model

For a complex repair task, it is crucial to upgrade its model by exploiting and fusing the intrinsic temporal attributes. However, DA-GAN lacks more flexible repair ability due to the lack of multiple temporal information. We propose a TFs model with truncation mechanism to alleviate this

Table 2 Results comparison of the GAN and the optimized DA-GAN models

Results			DA-GAN												
Datasets/Modes			GAN		Metrics		Loss		Metrics/Index		DA-GAN		Loss		
(G)	MAPE	RMSE	MAE	Train	Test	MAPE	Index	RMSE	Index	MAE	Index	Train	Test		
RM	30%	17.5252	6.9685	5.3497	6.8918	6.9983	13.1209	743	5.4874	743	4.1172	743	4.9094	5.5069	
	70%	19.0852	7.7338	6.0738	7.3514	7.7597	16.1844	385	6.6079	385	5.0836	385	5.8428	6.6268	
CM	30%	16.9379	6.7200	5.1332	6.1152	6.7697	12.4260	675	5.1483	675	3.7763	651	4.0303	5.2058	
	70%	25.6175	11.0459	11.0459	5.4789	11.0622	25.2118	113	10.8948	110	8.6686	110	5.3049	10.9134	
HM	30%	17.5456	6.9546	5.3507	6.7770	7.0411	12.7099	845	5.3069	845	3.9543	845	4.6556	5.4071	
	70%	18.7806	7.6027	5.9686	7.1362	7.6378	16.0173	388	6.5590	388	5.0683	388	5.6126	6.5882	
(H)	RM	30%	42.4451	105.6567	43.8977	43.8181	104.1009	54.6721	85	60.8091	85	41.0739	85	60.3133	104.3647
	70%	49.2713	124.2781	56.2038	42.0959	122.4183	54.5522	86	59.9900	86	41.4647	86	59.4981	134.5107	
CM	30%	49.3004	99.4003	45.3864	34.9280	97.9699	43.8744	118	47.0213	118	32.3098	118	46.5341	98.0204	
	70%	77.1806	141.9468	76.3376	26.8851	139.8496	48.3422	67	50.5214	67	34.2135	67	50.3554	141.9799	
HM	30%	43.4570	104.7998	43.1709	40.4787	99.7394	51.8306	93	55.4240	93	38.4209	93	55.6910	96.4283	
	70%	50.4350	122.2361	56.3287	40.6246	119.7305	50.6555	99	56.6905	99	39.2785	99	56.9669	126.2861	

Best results are highlighted in bold fonts

problem by mining multi-view temporal correlations (i.e., closeness, periods, trends) in phases. Of these, this truncation mechanism comprises of two parts: the local filling and scaling coefficient. For the former, we list the evaluation metrics in Table 3 after filling in the given constants. As can be found that the under-sized values have little contribution to TFs model, but the over-sized values render the repair task non-convergent. In fact, the suitable filling values can indeed cut the metrics of each TF module, especially for the set [2, 1, 1/2, 1/4], the RMSE of GAN is 3.8424 times higher than the TFs-DGAN. As compared with Table 2, the final filling effect of TFs-DGAN also outperforms the best results of DA-GAN model. Note that GAN gets introduced as the comparison of the temporal modeling results. Moreover, we enrich the sources of filling values, where the values in $[\hat{\xi}_1, \hat{\xi}_2, \hat{\xi}_3, \hat{\xi}_4]$ are chosen from {-1,1}; the values in $[\tilde{\xi}_1, \tilde{\xi}_2, \tilde{\xi}_3, \tilde{\xi}_4]$ come from [-1,1]; the values in $[\bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3, \bar{\xi}_4]$ refer to the average of the matrices for four TF modeling in the traffic time series.

As for the scaling treatment, it makes a detailed attempt for the TF to seek the best filling range. After such elastic filling, the residuals at each stage get regularized and reduced, which ensures that the observations close to the true values, while the missing data end up at the optimal values. Usually, a small coefficient performs very badly as it affects the majority of normal data, while the large one causes the loss of the optimized function. As in Fig. 7, we exhibit the metrics variation of TFs-DGAN model with different scaling coefficients at 70% missing rate, CM scenario and dataset (G). Overall, the scaling treatment is

favorable for TFs-DGAN. In particular, the performance of TFs-DGAN model varies drastically in the local region [1, 2], where the best coefficient of 2 is marked by the red ellipse. By this setting, TFs-DGAN model gives satisfactory results and its MAPE, RMSE and MAE values decrease to 4.5652, 2.9448 and 1.5679, respectively. Note that the coefficient of 1 is a blank set.

5.4 Ablation study

For further analysis, we conduct the ablation study on both datasets to verify the effect of inset components on the final outputs. Here, we name TFs-DGAN without diverse components as follows: 1) GAN, which acts as a reference for other methods; 2) DA-GAN (with outliers), which checks the validity of DA mechanism when it contains outliers; 3) DA-GAN (without outliers), which reflects the necessity and advantage of removing outliers; 4) TFs-DGAN (G_f), which examines the effect of the global feedback on TFs-DGAN model; 5) TFs-DGAN ($L_f -$), which validates the impact of the staged multi-view temporal correlations learning with local feedback. Specifically, the TFs-DGAN ($L_f - I_r / I_d / I_w$) reflects the effects of recent time/consecutive day/adjacent week on the final results. For TFs-DGAN ($L_f - I_o$), it covers the effect of other temporal factors on the final output. In Table 4, we show the RMSE values for all variants of TFs-DGAN at the missing rate of 50% and three modes. As compared to GAN, we can find that DA mechanism remarkably enhances the repair performance (e.g., accuracy, efficiency) of DA-GAN model

Table 3 Combination tests of filling values of TFs-DGAN model based on truncation mechanism under 70% missing rate, CM scenario and dataset(G)

Groups	Results												
	GAN			Temporal correlations									
	MAPE	RMSE	MAE	Recent		Daily		Weekly		Total (Other)			
				MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAE	
[0, 0, 0, 0]	25.3400	10.9765	8.7334	3.7525	8.2773	15.5094	5.3570	13.0956	3.8874	6.0520	3.9904	2.0863	
[1, 1, 1, 1]	25.2205	10.9628	8.7104	4.0702	8.3042	14.3028	5.1832	8.6846	3.4583	6.0969	3.5807	2.1367	
[1, 1/2, 1/4, 1/8]	25.3841	10.9740	8.7097	6.1192	8.2980	13.9370	5.2075	13.9222	3.6715	5.2221	3.6287	1.7681	
[2, 2, 2, 2]	25.6644	11.0175	8.7883	14.9995	8.3475	22.4811	5.3491	10.3259	4.8734	12.8106	6.0602	4.5124	
[2, 1, 1/2, 1/4]	25.4499	10.9703	8.7628	3.6327	8.2486	12.3518	5.3106	6.2253	3.3834	4.1967	2.8354	1.4405	
[4, 4, 4, 4]	25.0624	10.9422	8.6559	3.1141	8.2819	23.3830	7.2020	19.9842	12.7534	149.1620	71.4008	51.3553	
[4, 2, 1, 1/2]	25.1948	10.9656	8.7209	3.9473	8.2664	17.9907	7.2109	10.1573	5.9784	6.0918	2.8814	2.0744	
$[\hat{\xi}_1, \hat{\xi}_2, \hat{\xi}_3, \hat{\xi}_4]$	25.2306	10.9242	8.6828	5.0656	8.2423	12.2350	5.1580	9.5107	3.4919	8.6113	4.4061	3.0229	
$[\tilde{\xi}_1, \tilde{\xi}_2, \tilde{\xi}_3, \tilde{\xi}_4]$	24.9634	10.8021	8.5803	5.5264	8.1276	14.3315	5.1705	6.2898	3.5269	6.6361	3.5361	2.1085	
$[\bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3, \bar{\xi}_4]$	25.2385	10.9646	8.6937	4.1654	8.3017	105.8530	5.3230	22.8108	3.8945	5.8368	3.9905	2.0111	

Best results are highlighted in bold fonts

Fig. 7 Determination of scaling coefficient of the TFs-DGAN under 70% missing rate, CM scenario and dataset (G)

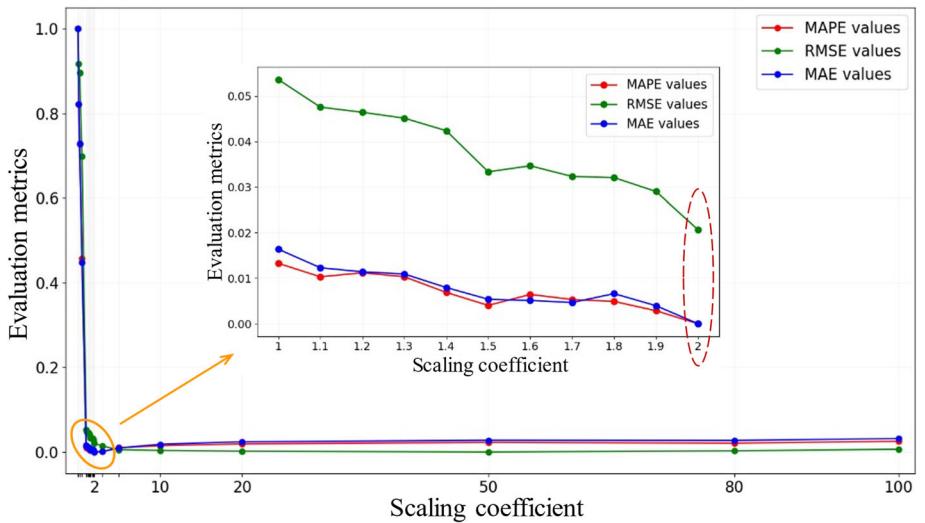


Table 4 Ablation study for RMSE values among different components of the TFs-DGAN

Models	Results					
	Dataset(G)			Dataset(H)		
	RM	CM	HM	RM	CM	HM
GAN	7.1058(+ 0)	7.2738(+ 0)	7.0642(+ 0)	111.4510(+ 0)	115.3500(+ 0)	113.6310(+ 0)
DA-GAN (With outliers)	6.5005(+ 0.6053)	6.6203(+ 0.6535)	6.0294(+ 1.0348)	52.3603(+ 59.0907)	41.5980(+ 73.7520)	55.0255(+ 58.6055)
DA-GAN (Without outliers)	6.1064(+ 0.9994)	6.3338(+ 0.9400)	5.7685(+ 1.2957)	38.0696(+ 73.3814)	28.3243(+ 87.0257)	41.5970(+ 72.0340)
TFs-DGAN (G_f)	5.3200 (+ 1.7858)	5.5294(+ 1.7444)	5.1051(+ 1.9591)	104.7053(+ 6.7457)	110.5989(+ 4.7511)	106.3118(+ 7.3192)
TFs-DGAN ($L_f - \mathcal{I}_r$)	4.8375(+ 2.2683)	4.8465(+ 2.4273)	4.5457(+ 2.5185)	34.7270(+ 76.7240)	26.8562(+ 88.4938)	35.7116(+ 77.9194)
TFs-DGAN ($L_f - \mathcal{I}_d$)	3.4113(+ 3.6945)	3.4156(+ 3.8582)	3.2104(+ 3.8538)	20.9622(+ 90.4888)	16.3765(+ 98.9735)	22.0848(+ 91.5462)
TFs-DGAN ($L_f - \mathcal{I}_w$)	2.2311(+ 4.8747)	2.4101(+ 4.8637)	2.1060(+ 4.9582)	20.3142(+ 91.1368)	15.5935(+ 99.7565)	21.5661(+ 92.0649)
TFs-DGAN ($L_f - \mathcal{I}_o$)	1.5764(+ 5.5294)	1.8579(+ 5.4159)	1.5239(+ 5.5403)	17.0038(+ 94.4472)	12.7070(+ 102.6430)	18.4686(+ 95.1624)

Best results are highlighted in bold fonts and the values in bracket “(+)” represent the gain relative to GAN

because of its dynamic iterative process. Further, the data pre-processing is necessary for data recovery as the outliers have a negative effect on DA-GAN. Compared with the local feedback, the global feedback of TFs-DGAN (G_f) elicits the fluctuation of repair results so that we pick the target model from TFs-DGAN ($L_f -$). In addition, the comparative results reveal that all the temporal correlations (e.g., closeness, periods) are valuable for repair tasks, but the more similar temporal attributes (e.g., closeness) contribute more to the TFs-DGAN. Overall, the proposed TFs-DGAN model achieves the best results since its optimization mechanism captures and integrates the spatial-

temporal dependencies of the traffic data to the greatest extent possible.

5.5 Experimental results and discussion

5.5.1 Multiple model comparison

To verify the superiority of TFs-DGAN in traffic data recovery, we compare it against the state-of-the-art models (e.g., LATC, GAIN, BGCP and ST-BiRT) under the combinations of missing rates of 10%–90%, RM, CM and HM scenarios. In Fig. 8, we present all evaluation metrics

of TFs-DGAN and all baseline models on the dataset(G). As for the missing rate, it satisfies the trend that all models can successfully fulfill the repair tasks at low rates, but their performance decreases gradually as the rate increases. Similarly, the three missing scenarios have different effects on recovery performance. Overall, all metric curves have similar trends under RM, CM and HM modes, but the metrics in Fig. 8(b) are the largest because CM mode corrupts the data structure at the most. Under the same criteria, the recovery models also exhibit dramatic differences. Among them, the classical GAIN model lacks competition altogether, but it deserves efforts to make further studies. Moreover, we cancel the tests of the HaLRTC model in CM and HM modes because of its poorer repair results in RM mode, which also reveals that the traditional statistical methods are incapable of dealing with complex repair tasks. For those excellent tensor/DL-based repair models (e.g., BGCP, FNNTEL, LATC and LRTC-TNN), we can explicitly find that TFs-DGAN consistently outperforms the state-of-the-art baseline models. In particular for ST-BiRT, our TFs-DGAN still prevails on the dataset (G).

Subsequently, we proceed with our experiments on the dataset(H) to provide additional insights into the effectiveness of TFs-DGAN model. Due to the different traffic forms, the models yield varying results for repair tasks. At that, the evaluation metrics in Fig. 9 show more disorder than in Fig. 8 as the missing rate increases. Compared with RM mode, the CM and HM modes have more destructive effect

on metro traffic data. But overall, the TFs-DGAN achieves the optimal recovery results so far over variety of state-of-the-art baseline models under complex missing situations. In particular for the combination of 90% missing rate and CM mode, the TFs-DGAN surpasses the latest ST-BiRT model. In Sect. 5.5.2, we will further evaluate the performance of TFs-DGAN model in extreme cases.

5.5.2 Analysis of extreme cases

While the extreme lack of data rarely happens in practice, that is difficult to repair when it does. Most of the recovery models have neglected to account for these extreme cases, and thus we conduct additional experiments on both datasets with over 90% missing rate for validating the scalability of TFs-DGAN model. Table 5 shows the evaluation metrics of TFs-DGAN model for the missing rates of 91%, 93%, 95%, 97%, 99%, 99.9%, 99.99%. The compared results reveal that all metrics of TFs-DGAN model exhibit an accelerated increase with the increase in missing rates, which indicates the sharp decrease in the repair performance of TFs-DGAN model. In particular, when the rate exceeds a certain level (e.g., 99%), the TFs-DGAN almost defeats to accomplish the repair tasks. For the three scenarios, TFs-DGAN yields different imputation results. Among them, our TFs-DGAN gives more satisfactory results for HM mode, while it fails for CM mode under the same conditions. In addition, the modeling effect of TFs-DGAN for dataset (G) is superior to dataset (H) when

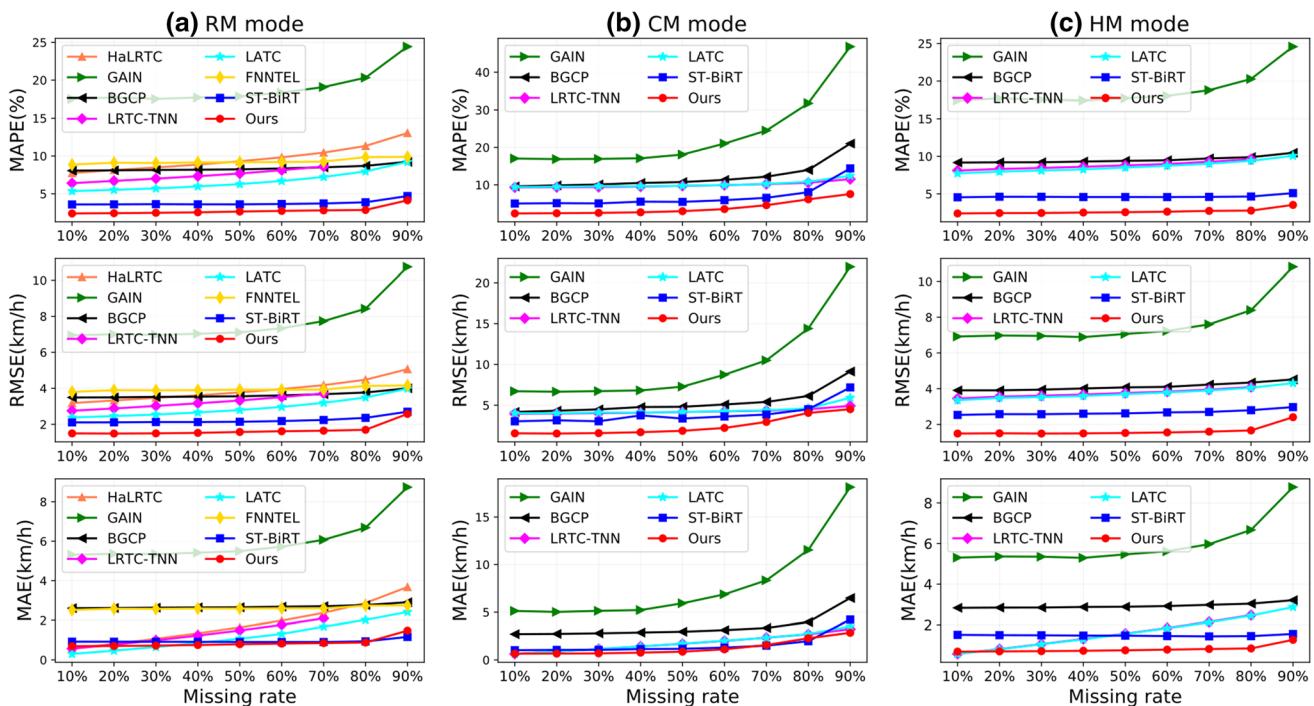


Fig. 8 Performance comparison of the TFs-DGAN and baseline models on the dataset (G)

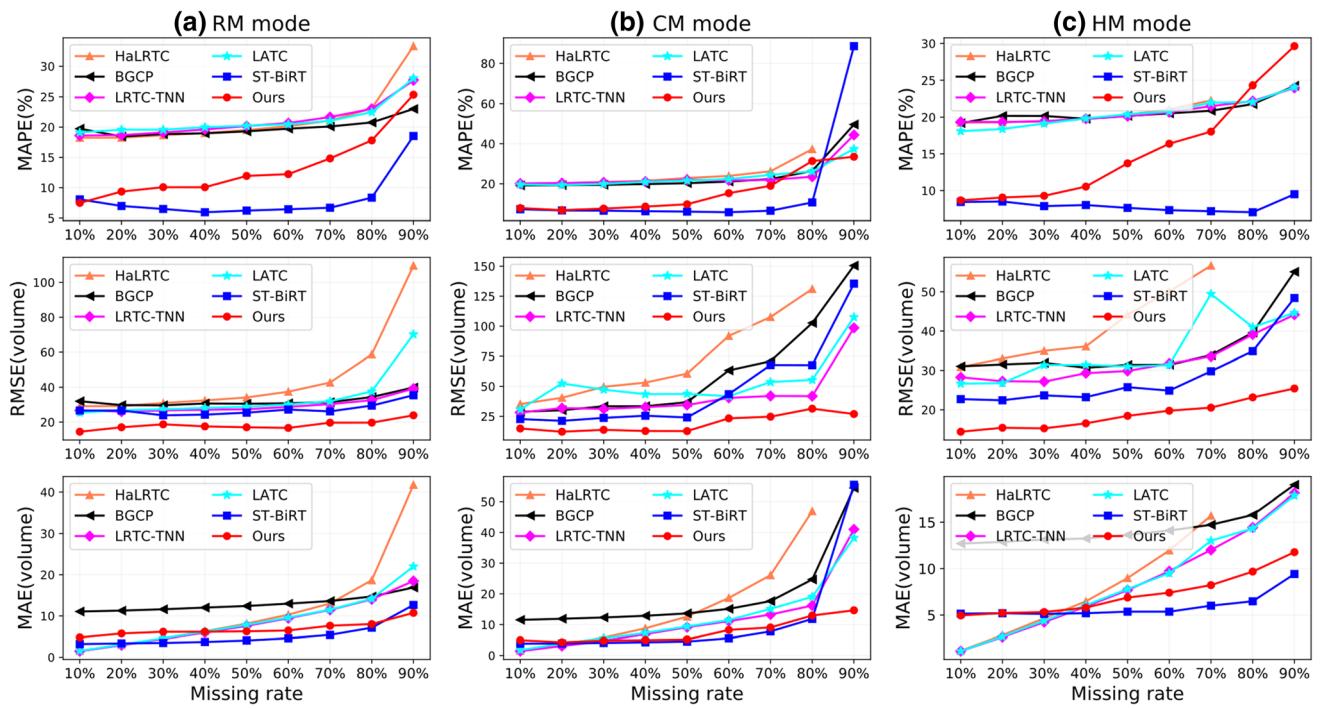


Fig. 9 Performance comparison of the TFs-DGAN and baseline models on the dataset (H)

Table 5 Imputation performance of the TFs-DGAN under extreme cases

MisRates	Results	Dataset(G) & Scenarios											
		RM				CM				HM			
		MAPE	RMSE	MAE	Mean	MAPE	RMSE	MAE	Mean	MAPE	RMSE	MAE	Mean
91%	5.5846	3.3875	1.9856	38.9974	7.7739	4.5746	2.8715	38.4516	4.1400	2.6088	1.5067	38.8906	
93%	5.8770	3.5755	2.0881	38.8856	8.2920	4.6760	3.0873	38.4255	5.0771	3.0859	1.8123	38.8803	
95%	6.0676	3.7506	2.2191	38.6267	9.0523	4.9672	3.3223	38.5402	4.9105	3.0465	1.7967	38.7999	
97%	7.0877	4.4168	2.6759	38.2016	9.7059	5.1955	3.4644	38.7960	5.2518	3.2836	1.9733	38.6154	
99%	29.1116	14.2084	10.9934	34.0893	22.8232	10.3034	7.7240	38.8343	8.1710	4.5203	3.0072	38.5183	
99.9%	93.6381	38.6884	36.8739	2.3956	99.5615	40.5250	39.0547	0.1694	9.7367	4.9240	3.1934	39.1100	
99.99%	99.2817	40.4149	38.9474	0.2630	\	\	\	\	19.7009	8.8348	6.4888	38.8837	
MisRates	Results	Dataset(H) & Scenarios											
		RM				CM				HM			
		MAPE	RMSE	MAE	Mean	MAPE	RMSE	MAE	Mean	MAPE	RMSE	MAE	Mean
91%	27.1715	24.7192	11.0556	117.8168	34.6936	27.4455	14.8994	111.6561	31.6056	26.1260	12.4486	114.6738	
93%	30.9202	23.7122	11.4428	115.3654	38.4510	31.9952	17.1109	108.7884	39.3824	29.0430	13.9989	115.0630	
95%	31.9870	23.8834	12.2781	111.3794	32.5463	31.5206	15.5621	94.4243	43.6291	31.0101	15.9297	112.6548	
97%	38.2563	28.8725	14.9063	100.7466	47.0096	34.4641	17.2542	84.4399	50.6791	33.2291	18.3210	103.6067	
99%	88.6986	40.9458	28.6312	61.0925	232.8831	181.3552	102.4039	76.8015	72.3728	40.7597	28.6753	67.9341	

"\\" stands for not available

accounting for the data attributes, volume and the potential spatial-temporal regularity. Moreover, relative to ST-BiRT, TFs-DGAN model provides better recovery results for the challenging CM and HM scenarios. Thus, we can conclude that TFs-DGAN has stronger imputation performance for extreme cases compared with the well-established recovery models.

5.6 Algorithm complexity analysis

In this section, we investigate the computational complexity of the TFs-DGAN on both datasets.

In theory, the time consumption of our algorithm mainly comprises two parts: the time for DA-GAN iterations and TFs modeling. For a matrix $\mathcal{X} \in \mathbb{R}^{I \times J}$, the complexity of DA-GAN is $\mathcal{O}\left(2K_1^2 \sum_{l=1}^3 I_l J_l\right)$, where G and D are 3-layer FC neural nets and K_1 is the number of dynamic iterations of DA-GAN. The result of multi-view TFs modeling includes a linear combination of several TF models with the same structure so that the complexity is written as $\mathcal{O}(K_2 I^2 J \log \mathcal{R})$, and K_2 is the number of each TF sampling and the \mathcal{R} is very small. Hence, the total complexity of TFs-DGAN is $T(\mathcal{O}(K_1^2 IJ + K_2 I^2 J))$. Indeed, the cost time covers the training and inference time, both of which relate to the configuration of the experiment platform. In the study, the computing time explicitly refers to the former since the inference time of TFs-DGAN is relatively too small to be ignored. Table 6 provides the comparison of time costs for all repair models on both datasets. Of these, the conventional models (e.g., HaLRTC) consume less time than the latest tensor/DL-based models (i.e., LATC, BGCP, ST-BiRT) due to their relatively fewer parameters and poorer repair precision. At fact, it is these excellent recovery models to achieve higher performance. For the ST-BiRT referenced in Sect. 5.5.1, our TFs-DGAN model compares favorably with it in terms of time consumption.

Furthermore, a further analysis finds that the tensor-based completion methods have inevitably increased model complexity since they require a large rank to deliver a guarantee for the repair tasks. However, our hybrid model TFs-DGAN improves the repair efficiency over the tensor-based models by setting a small rank ($\mathcal{R} = 3$), while ensuring the optimal accuracy against the GAN-based models. On that basis, we present the time consumed by TFs-DGAN in Fig. 10 for the missing rates of 10%–90% and three scenarios. Because of its similarity, TFs-DGAN model maintains the acceptable time consumption across all tested ranges for two traffic datasets, which also complies with the basic law of data repair. Besides, algorithm entails a trade-off between time complexity and recovery

Table 6 Comparison of computational costs under 50% missing rate and three scenarios (Time unit: seconds)

Models	Time(s)					
	Dataset(G) & Scenarios			Dataset(H) & Scenarios		
	RM	CM	HM	RM	CM	HM
HaLRTC	10	25	26	4	3	3
GAIN	47	42	48	9	8	8
LRTC-TNN	65	59	61	7	5	4
LATC	796	635	648	71	73	72
BGCP	1496	890	262	617	376	99
ST-BiRT	3730	3086	354	656	380	107
DA-GAN	285	237	321	6	6	5
TFs-DGAN	734	696	764	127	118	117

precision, and we will further strengthen the TFs-DGAN to ensure less running time with better accuracy.

5.7 Visualization verification

Instead of intuitively presenting the whole recovery effect of the TFs-DGAN by Figs. 8 and 9, we select several specific road segments and time periods for visual validation as in Figs. 11 and 12. Under the missing rate of 50% and three scenarios, Figs. 11 and 12 displays the compared results by TFs-DGAN model on Road segment #1 on dataset(G) (Metro station #1 on dataset(H)) from August 1 to August 10, 2016 (January 1 to January 10, 2019). It can be seen that the TFs-DGAN provides a well-repaired for missing data. Even if there are fluctuations in the observed data, TFs-DGAN model can still accurately acquire the time-evolving traffic patterns from incomplete data. But it is difficult for CM mode to capture information from other dimensions due to the loss of temporal dependencies, which accounts for the TFs-DGAN being less effective than RM and HM scenarios in repair tasks. In addition, the recovery of mutated time series remains very challenging, as evidenced by the raised data points in Figs. 11 and 12. Overall, the visual verification again confirms that TFs-DGAN model has the consistent and robust capability for missing traffic data imputation.

6 Conclusion

Traditional probability-based/GAN-based recovery algorithms lack successful cases of jointly exploiting spatial-temporal features of traffic data. Here, we develop a novel hybrid framework TFs-DGAN, which consists of DA-GAN module and TFs module, aimed at its adaptation to the

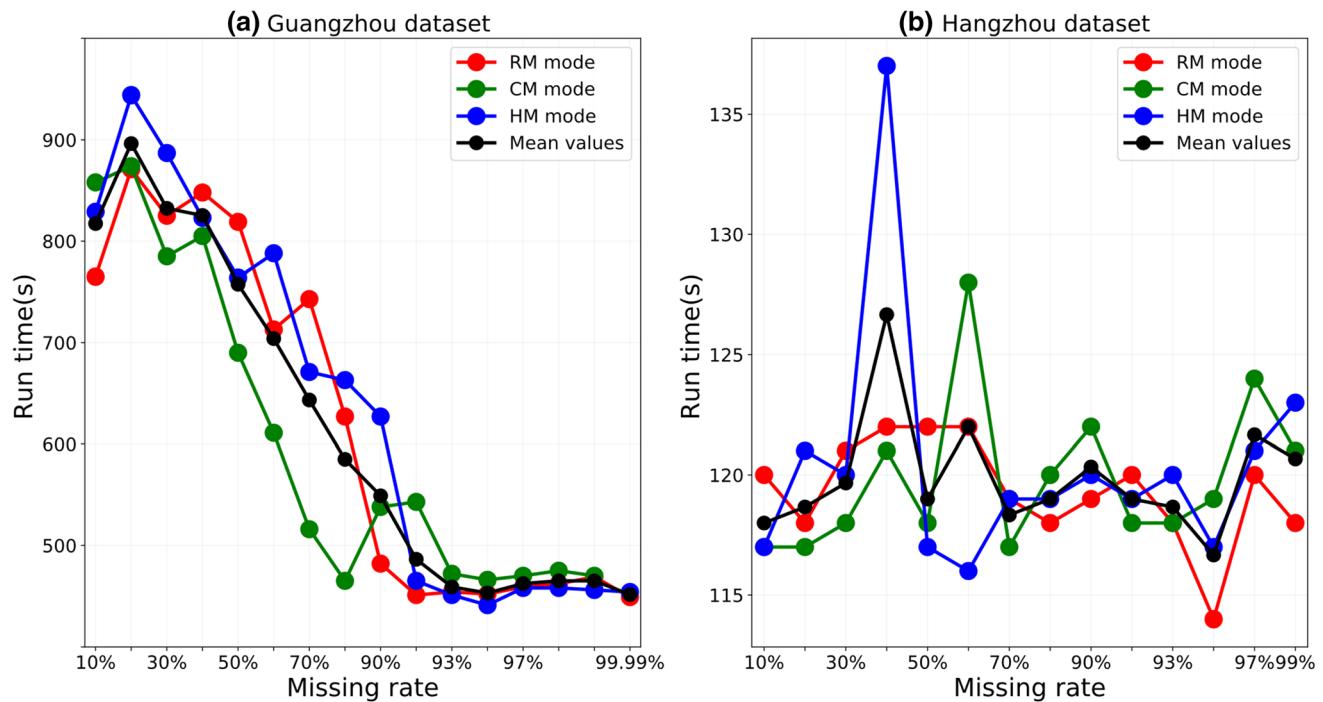


Fig. 10 Computation time of the TFs-DGAN under the missing rates of 10%–90% and three scenarios

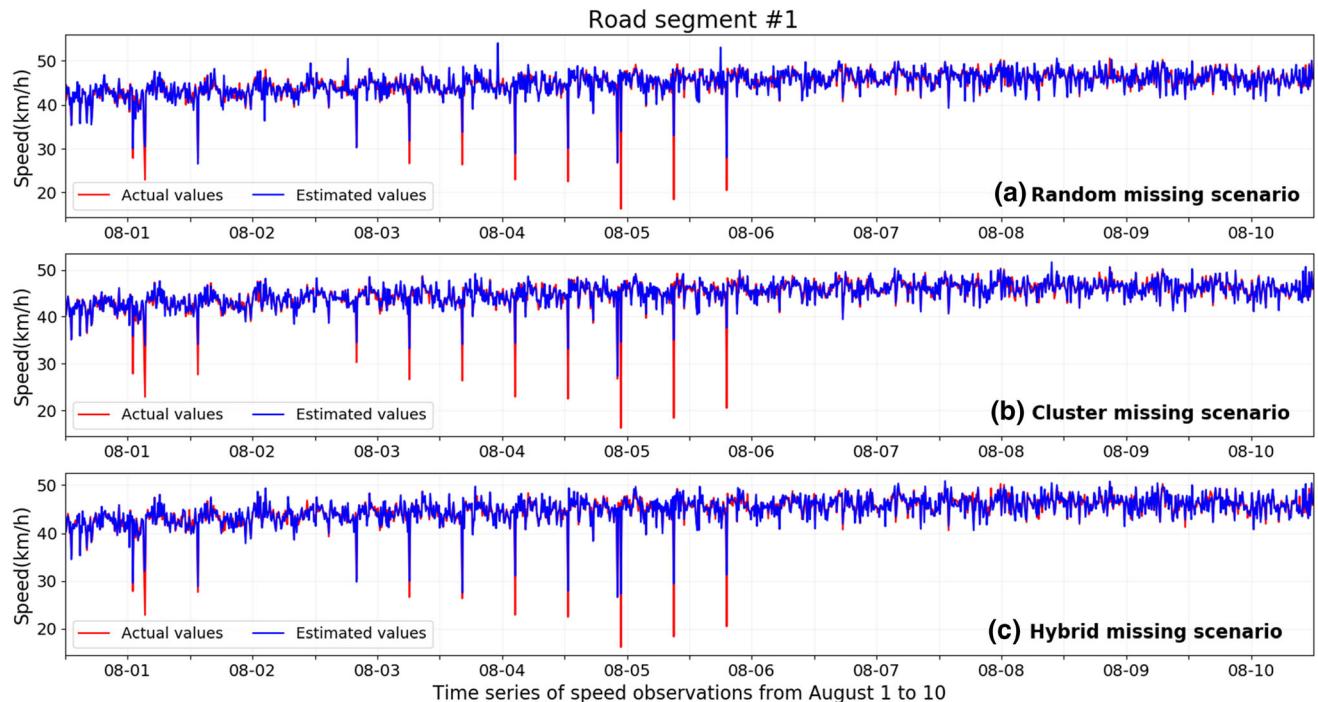


Fig. 11 Comparison of the ground truth and repaired values by TFs-DGAN model tested on the dataset (G) under the missing rate 50% with RM, CM and HM scenarios

tasks of missing traffic data repair. Of these, our improved DA-GAN module is capable of delivering hopeful but not competitive results since the introduced DA mechanism enables the generator to dynamically identify the optimal

number of iterations. Compared with traditional GAN models, DA-GAN yields better performance gains with the combinations of S-G smoothing and 2D processing. Nevertheless, the complex temporal properties make GAN

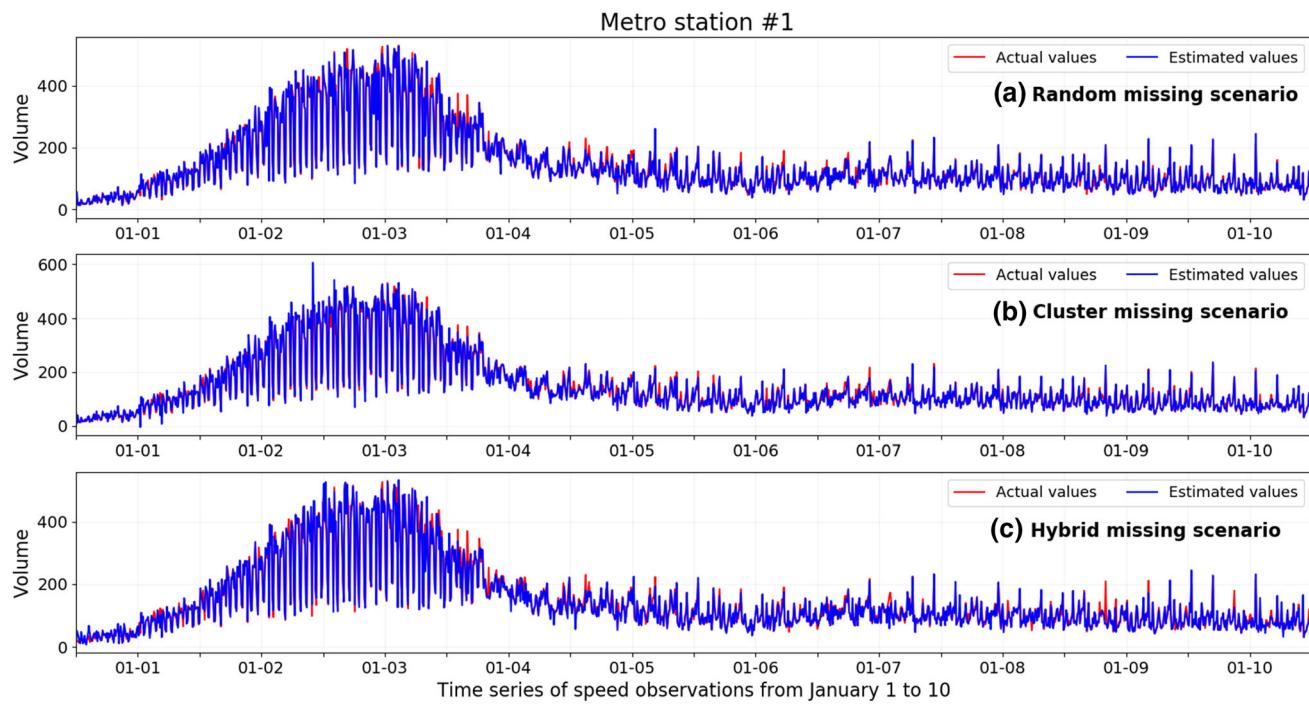


Fig. 12 Comparison of the ground truth and repaired values by TFs-DGAN model tested on the dataset (H) under the missing rate 50% with RM, CM and HM scenarios

difficult to learn the realistic distribution of traffic data. Considering the obvious periodicities, four TF modules with truncation mechanism explicitly model multiple forms of temporal attributes. Each TF module employs Gibbs sampling for model inference, and applies filling and scaling controls to automatically drive the reduction in the resulting residuals. All optimized residuals are finally fused by means of the staged local feedback and the output serves as the best repair result. Under the hybrid framework, both the constructed models work together to capture the global-local information. Using the spatial-temporal traffic data collected in Guangzhou and Hangzhou, China, we evaluate the TFs-DGAN model for the missing rates of 10%–99.99% and for the RM, CM and HM patterns. The empirical analyses reveal that our TFs-DGAN performs well and consistently outperforms other state-of-the-art baseline models.

In future studies, we will consider more additional factors (e.g., weather) for TFs-DGAN model and further optimize its structure and focus on its scalability and efficiency.

Acknowledgements The authors would like to thank anonymous referees for their valuable comments. This study is supported by the project of the National Natural Science Foundation of China, Youth Fund Project (No. 11702099) and the National Natural Science Foundation of China (No. 52072130).

Declarations

Conflict of interest No conflict of interest exists in the submission of this paper.

Data availability statement The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

- Zhang T, Zhang DG, Yan HR, Qiu JN, Gao JX (2021) A new method of data missing estimation with FNN-based tensor heterogeneous ensemble learning for internet of vehicle. Neurocomputing 420:98–110
- Li J, Xu L, Li R, Wu P, Huang Z (2022) Deep spatial-temporal bi-directional residual optimisation based on tensor decomposition for traffic data imputation on urban road network. Appl Intell 52:1–19
- Kaur M, Singh S, Aggarwal N (2022) Missing traffic data imputation using a dual-stage error-corrected boosting regressor with uncertainty estimation. Inf Sci 586:344–373
- Li L, Zhang J, Wang Y, Ran B (2018) Missing value imputation for traffic-related time series data based on a multi-view learning method. IEEE Trans Intell Transp Syst 20(8):2933–2943
- Acar E, Dunlavy DM, Kolda TG, Mørup M (2011) Scalable tensor factorizations for incomplete data. Chemom Intell Lab Syst 106(1):41–56
- Tang J, Zhang X, Yu T, Liu F (2021) Missing traffic data imputation considering approximate intervals: a hybrid structure

- integrating adaptive network-based inference and fuzzy rough set. *Physica A* 573:125776
7. Fan J, Cheng J (2018) Matrix completion by deep matrix factorization. *Neural Netw* 98:34–41
 8. Deng L et al (2021) Graph spectral regularized tensor completion for traffic data imputation. *IEEE Trans Intell Transport Sys.* 23:10996
 9. Chen X, Lei M, Saunier N, & Sun L (2021) Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation. *IEEE Trans Intell Transport Syst*
 10. Chen X, He Z, Sun L (2019) A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation. *Transport Res Part C Emerging Technol* 98:73–84
 11. Yang JM, Peng ZR, Lin L (2021) Real-time spatiotemporal prediction and imputation of traffic status based on LSTM and Graph Laplacian regularized matrix factorization. *Transport Res Part C Emerging Technol* 129:103228
 12. Han L, Zheng K, Zhao L, Wang X, Wen H (2020) Content-aware traffic data completion in ITS based on generative adversarial nets. *IEEE Trans Veh Technol* 69(10):11950–11962
 13. Li L, Du B, Wang Y, Qin L, Tan H (2020) Estimation of missing values in heterogeneous traffic data: application of multimodal deep learning model. *Knowl-Based Syst* 194:105592
 14. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, & Bengio Y (2014). Generative adversarial nets. *Adv Neural Inf Process Syst* 27
 15. Zhang K, He Z, Zheng L, Zhao L, Wu L (2021) A generative adversarial network for travel times imputation using trajectory data. *Comput-Aid Civ Infrastruct Eng* 36(2):197–212
 16. Yoon J, Jordon J, & Schaar M (2018). Gain: missing data imputation using generative adversarial nets. In: international conference on machine learning (pp. 5689–5698). PMLR
 17. Wang Y, Li D, Li X, Yang M (2021) PC-GAIN: Pseudo-label conditional generative adversarial imputation networks for incomplete data. *Neural Netw* 141:395–403
 18. Yu HF, Rao N, & Dhillon IS (2016) Temporal regularized matrix factorization for high-dimensional time series prediction. *Adv Neural Inf Process Syst* 29:
 19. Chen X, & Sun L (2021) Bayesian temporal factorization for multidimensional time series prediction. *IEEE Trans Pattern Anal Mach Intell*
 20. Fan J, Ding L, Chen Y, & Udell M (2019) Factor group-sparse regularization for efficient low-rank matrix recovery. *Adv Neural Inf Process Syst* 32
 21. Candès EJ, Recht B (2009) Exact matrix completion via convex optimization. *Found Comput Math* 9(6):717–772
 22. Sure P, Srinivasan CP, & Babu CN (2021). Spatio-temporal constraint-based low rank matrix completion approaches for road traffic networks. *IEEE Trans Intell Transport Syst*
 23. Chen X, Wei Z, Li Z, Liang J, Cai Y, Zhang B (2017) Ensemble correlation-based low-rank matrix completion with applications to traffic data imputation. *Knowl-Based Syst* 132:249–262
 24. Fan J (2022) Dynamic nonlinear matrix completion for time-varying data imputation. *J AAAI Conf Artif Intell* 36:6587–6896
 25. Fan J, Zhang Y, Udell M (2020) Polynomial matrix completion for missing data imputation and transductive learning. *Proc AAAI Conf Artif Intell* 34:3842–3849
 26. Fan J, & Udell M (2019) Online high rank matrix completion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8690–8698)
 27. Jia X, Dong X, Chen M, Yu X (2021) Missing data imputation for traffic congestion data based on joint matrix factorization. *Knowl-Based Syst* 225:107114
 28. Wen J, Zhang Z, Zhang Z, Fei L, Wang M (2020) Generalized incomplete multiview clustering with flexible locality structure diffusion. *IEEE Trans Cybern* 51(1):101–114
 29. Li J, Wu P, Li R, Pian Y, Huang Z, Xu L, Li X (2022) ST-CRMF: compensated residual matrix factorization with spatial-temporal regularization for graph-based time series forecasting. *Sensors* 22(15):5877
 30. Fan J (2021). Multi-mode deep matrix and tensor factorization. In: international conference on learning representations
 31. Xu D, Wei C, Peng P, Xuan Q, Guo H (2020) GE-GAN: a novel deep learning framework for road traffic state estimation. *Transport Res Part C Emerging Technol* 117:102635
 32. Yang B, Kang Y, Yuan Y, Huang X, Li H (2021) ST-LBAGAN: spatio-temporal learnable bidirectional attention generative adversarial networks for missing traffic data imputation. *Knowl-Based Syst* 215:106705
 33. Sun J, Zhang J, Li Q, Yi X, Liang Y, & Zheng Y (2020) Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks. *IEEE Trans Knowl Data Eng*
 34. Liu Z, Wu Z, Wang M, & Zhang R (2021) Multi-view spatial-temporal model for travel time estimation. In: Proceedings of the 29th international conference on advances in geographic information systems (pp. 646–649)
 35. Wu P, Xu L, Li J, Guo H, Huang Z (2022) Recognizing real-time transfer patterns between metro and bus systems based on spatial-temporal constraints. *J Transport Eng Part A Syst* 148(9):04022065
 36. Liu MY, Huang X, Yu J, Wang TC, Mallya A (2021) Generative adversarial networks for image and video synthesis: algorithms and applications. *Proc IEEE* 109(5):839–862
 37. Liu Y, Fan H, Yuan X, Xiang J (2022) GL-GAN: adaptive global and local bilevel optimization for generative adversarial network. *Pattern Recogn* 123:108375
 38. Zhang W, Zhang P, Yu Y, Li X, Biancardo SA, & Zhang J (2021). Missing data repairs for traffic flow with self-attention generative adversarial imputation net. *IEEE Trans Intell Transport Syst*
 39. Li J, Sun L, Li R (2020) Nondestructive detection of frying times for soybean oil by NIR-spectroscopy technology with Adaboost-SVM (RBF). *Optik* 206:164248
 40. Chen X, Chen Y, Saunier N, Sun L (2021) Scalable low-rank tensor learning for spatiotemporal traffic data imputation. *Transport Res Part C Emerging Technol* 129:103226
 41. Wu P, Huang Z, Pian Y, Xu L, Li J, & Chen K (2020) A combined deep learning method with attention-based LSTM model for short-term traffic speed forecasting. *J Adv Transport*
 42. Chen X, Yang J, Sun L (2020) A nonconvex low-rank tensor completion model for spatiotemporal traffic data imputation. *Transport Res Part C Emerging Technol* 117:102673
 43. Li J, Sun L, Li Y, Lu Y, Pan X, Zhang X, Song Z (2019) Rapid prediction of acid detergent fiber content in corn stover based on NIR-spectroscopy technology. *Optik* 180:34–45
 44. Liu J, Musialski P, Wonka P, Ye J (2012) Tensor completion for estimating missing values in visual data. *IEEE Trans Pattern Anal Mach Intell* 35(1):208–220

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.