Article:
Xing, J, Liu, R orcid.org/0000-0003-0627-3184, Anish, K et al. (1 more author) (2023) A Data Fusion CANDECOMP-PARAFAC Method for Interval-wise Missing Network Volume Imputation. IEEE Transactions on Intelligent Transportation Systems. ISSN 1524-9050

https://doi.org/10.1109/TITS.2023.3289193

A Data Fusion CANDECOMP-PARAFAC Method for Interval-wise Missing Network Volume Imputation

Jiping Xing, Ronghui Liu, Khadka Anish, and Zhiyuan Liu

Abstract—Traffic missing data imputation is a fundamental demand and crucial application for real-world intelligent transportation systems. The wide imputation methods in different missing patterns have demonstrated the superiority of tensor learning by effectively characterizing complex spatiotemporal correlations. However, interval-wise missing volume scenarios remain a challenging topic, in particular for long-term continuous missing and high-dimensional data with complex missing mechanisms and patterns. In this paper, we propose a customized tensor decomposition framework, named the data fusion CANDECOMP/PARAFAC (DFCP) tensor decomposition, to combine vehicle license plate recognition (LPR) data and cellphone location (CL) data for the interval-wise missing volume imputation on urban networks. Benefiting from the unique advantages of CL data in the wide spatiotemporal coverage and correlates highly with real-world traffic states, it is fused into vehicle license plate recognition (LPR) data imputation. They are regarded as data types dimension, combined with other dimensions (different segments, time, days), we innovatively design a 4-way low-n-rank tensor decomposition for data reconstruction. Furthermore, to deal with the diverse disturbances in different data dimensions, we derive a regularization penalty coefficient in data imputation. Different from existing regularization schemes, we further introduce Bayesian optimization (BO) to enhance the performance in the non-convexity of the objective function in our regularized hyperparametric solutions during tensor decomposition. Numerical experiments highlight that our proposed method, combining CL and LPR data, significantly outperforms the imputation method using LPR data only. And a sensitivity analysis with varying missing length and rate scenarios demonstrates the robustness of model performance.

***Index Terms*— Interval-wise missing volume imputation; DFCP tensor method; Data fusion; Bayesian optimization; LPR and cellphone data.**

I. Introduction

Successful deployment of Intelligent Transportation Systems and the Internet of Vehicles Systems is largely dependent on the availability of accurate, reliable, and timely traffic information. For example, real-time traffic information (e.g., volume, speed, density) could be used as navigation for devising urban traffic management measures and formulating urban policies. Among different traffic metrics, traffic volume is one of the most intuitive and widely recognized performance measures in transportation services. Traditional approaches for traffic volume acquisition are mainly based on data from various stationary road-based sensors [1]. However, the problem associated with missing data poses a major challenge in their application. Such issues persist mainly due to hardware or software failure, scheduled maintenance, disruption in power supply, and communication network problems. Traffic data with missing values are usually either discarded or averaged inappropriately with other data, making less effective use of the traffic sensor data on transport networks, which causes the evaluation of the performance measures such as traffic flow and travel time reliability, and vehicle emission and noise would be substantially underestimated [2]. Thus, missing data imputation is crucial for enhancing data quality and supporting downstream applications.

In the existing literature, there are three categories of missing patterns: missing completely at random (MCR), missing at random (MR), and not missing at random (NMR). And then, point-wise, interval-wise, and slice-wise missing comprises three different types of missing mechanisms [8, 9]. Despite recent efforts and advances in traffic data imputation, existing studies still focus on point-wise or slice-wise missing scenarios from varying missing patterns and mechanisms [10, 11], and it remains a challenge to perform efficient imputation on interval-wise missing. The latter will result in two challenging issues, *diverse disturbances,* and *long-term missing data*. Furthermore, as we note in the following, interval-wise missing is the highly common missing mechanism in vehicle license plate recognition (LPR) detectors for the addressed problem.

In recent years, data driven-based approach has been gaining popularity in missing data imputation. Unfortunately, due to the limitation of parameterization and the potential occurrence of overfitting in the machine and deep learning method [13], they can not fully utilize both spatial and temporal dimensions simultaneously in a unified framework when the missing rate is high. In this context, a tensor decomposition-based method has demonstrated superiority in imputation tasks by exploiting the inner correlations of traffic flow at some dimensions of the day, week, and road [5]. Though tensor-based imputation approaches are found to be useful in capturing global information in certain special cases, in the complex urban network, road segments under different usage functions and grades still have different traffic volume patterns. For example, traffic volume on an arterial road varies from that on a trunk road or branch road. However, these inner correlations from

This research is supported by Innovate UK (No. RSSB/494204565/aVSTP) and the National Science Foundation of China (No. 71922007, 52131203).

Jiping Xing and Zhiyuan Liu are with the School of Transportation, Southeast University, Nanjing 211189, China (Email: xingjiping@seu.edu.cn; zhiyuanl@seu.edu.cn)

Ronghui Liu is with Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK. (Corresponding author. Email: R.Liu@its.leeds.ac.uk)

Khadka Anish is with Department of Civil and Environmental Engineering, University of Auckland, New Zealand. (Email: aahk996@aucklanduni.ac.nz).

varying segment levels cannot be fully utilized [14, 15]. The diverse disturbances from available data dimensions can significantly affect the imputation performance. Even if the addition of regularization items has been used to alleviate this impact [16-18], it is still limited by the non-convex problems in regularized parameter solving.

Furthermore, as the studies on adding data dimensions in spatiotemporal traffic data have proven that it can effectively enhance the imputation performance [19], the fusion of multiple feasible data types in tensor can be also regarded as an emerging means. Recently, the heterogeneous residual [14] and traffic speed [20] have been fused into traffic volume imputation successively. Meanwhile, other homogeneous data fusion methods with their unique advantages deserve to be studied. This is particularly the case in interval-wise missing scenarios, where the long-term missing traffic data need to describe the large fluctuating trends.

This study attempts to tackle two technical issues: 1) how to achieve high imputation accuracy when dealing with interval-wise missing data scenarios? 2) how to ensure stable imputation performance even with diverse disturbances in urban networks? More specifically, we propose a data fusion-based CANDECOMP/PARAFAC (DFCP) tensor decomposition framework, to realize the interval-wise missing network volume data imputation. The spatiotemporal CL data is used as a secondary data source for LPR data imputation. These integrate correlated data from different day-of-week, segment, and time slots to comprise a 4-way tensor. To overcome the diverse disturbances from the different road segment types, the regularization penalty terms are set for improving the performance of imputation during tensor decomposition. The overall contribution of this work is threefold:

1) We integrate homogeneous data fusion into the CP tensor decomposition framework for dealing with the long-term missing in interval-wise data imputation. By benefiting from the similar real-world traffic dynamic and wide spatiotemporal coverage, we can better characterize the missing traffic data, which is a unique property of CL data.
2) We develop Bayesian optimization (BO) to update regularized penalty coefficients in reducing the disturbances from the inter variability of urban networks. It can capture more reliable regularize hyperparameters in the solution of non-convex problems.
3) We conduct extensive numerical experiments in the urban network of Nanjing, China. Imputation performances under different interval-wise missing lengths, rates, and data patterns illustrate the superiority of DFCP over recent state-of-art models.

The remainder of this paper is organized as follows: in section II, we review the relevant literature on missing network volume imputation. The problem statement is listed in section III. The mathematical model formulations are introduced in section IV, while the case study is introduced in section V. Finally, section VI summarizes the conclusions and discusses future research directions.

II. Literature Review

Over the last decade, many researchers have investigated missing traffic flow data. We summarize these methods from the aspect of modeling into two categories: model-based and data-driven methods [1, 21]. For model-based data imputation methods, when there are only a few missing data, some straightforward filling methods such as replacing by a constant value could be used. When dealing with a high proportion (>40%) of missing data, more sophisticated methods such as the compression sensing approach (CSA) [12], probabilistic principal component analysis (PPCA) [4], and other regression approaches [22] are generally used. However, these approaches are principally based on some assumptions that are practically irrelevant to real traffic data. Moreover, the form of the function should be pre-determined including some parameters during the modeling. For data-driven related methods, they usually employ the performance of cutting-edge artificial intelligence (AI) methods (e.g. Artificial Neural Networks (ANN) [6], Stacked Denoising Autoencoder (SDA) [23]) for missing value imputation or estimation. Even though these machine learning and deep learning methods have the advantages of deep structure and hyper-parametric processing, it is usually not clear to interpret intrinsic mechanisms of influencing factors, and cannot utilize multi-dimensional data correlation.

Thus, the other kind of data-driven approach, the tensor-based method, has risen with the multi-correlation description for data imputation [5]. The two main categories of tensor methods include rank minimization tensor completion (e.g. HaLRTC), and low-rank tensor factorization (e.g. CP decomposition, and Tucker decomposition) [24]. The details about the differences between these two methods can be found in the following Section IV.A. Both types of tensor models have shown superior performance in the application of data imputation. For instance, Tan, et al. [5] applied tensor decomposition to utilize the strong spatiotemporal correlation patterns for missing imputation by the multidimensional data

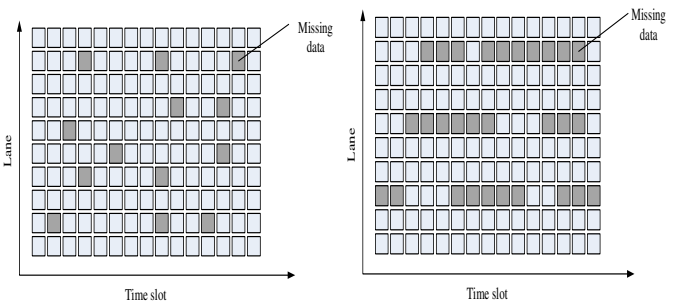
TABLE I Studies in Missing Traffic Data Imputation

Scenario	Missing mechanism and patterns	Basic approach	Data fusion	Key references
Highway	MCR/Point-wise	PPCA& ANN&LR	No	Qu, et al. [4],Zhong, et al. [6],Van Lint, et al. [3]
Highway	MCR&MR/ Point-wise, Interval-wise	Tensor &cokring & SDA	No	Tan, et al. [5], Bae, et al. [7]
Urban	MCR/Point-wise	Pattern clusters &CSA	No	Laña, et al. [8],Zhu, et al. [12]
Urban	NMR/Slice-wise	TransL&GTM	Yes	Liu, et al. [11], Zhang, et al. [2]
Urban	MR&NMR/ Interval-wise	Tensor with BO	Yes	**This paper**

structure. As to further enhance the robust performance of the tensor model, both Chen, et al. [25] and Tang, et al. [10] proposed bayesian-based tensor decomposition approaches. The former introduced conjugate priors to derive the posterior distributions, and then evaluated the performance of different data dimensions as model input. The latter applied the Bayesian probabilistic-based tensor method for considering the effect of the multi-modal distribution of data and found it performs well even under extremely missing conditions. Furthermore, adding the regularisation terms has been regarded as a feasible approach that considers diverse disturbances of data itself in tensor methods in some cases. For example, Said and Erradi [26] and Deng, et al. [16] applied the regularized tensor method for the validation of factorization by urban context data, which could overcome the effect of non-unique solutions caused by non-convexity in tensor decomposition. However, to our knowledge, despite the decent performance in the aforementioned Bayesian-based tensor approach, there are still fewer studies that considered the non-convexity effect in the regularization cases of tensor, and the merits of Bayesian optimization in regularized hyperparameter solution should be fully exploited. Especially when there are significant differences in each factorization matrix caused by variations in the urban network.

In studies of different missing data scenarios, data imputation on the interval-wise missing scenario is more challenging due to the long-term missing of traffic profile information. Especially when the interval size and rate of missing data exceed a certain threshold. Existing studies in missing interval-wise scenarios mainly focused on the effect of missing length on interpolation performance. For instance, Van Lint, et al. [3] developed linear regression (LR) as an interpolation method for processing missing intervals of length up to 30 samples. However, they have less considered homogeneous data fusion to upgrade the performance of long-term missing imputation. Especially with the innovation of digital information technology, the characteristics of the full spatiotemporal coverage and low-cost acquisition of some probe data emerging multi-source floating vehicle data (i.e. mobile phone data, taxi Global Positioning System (GPS) data) have great potential to be fused into traffic estimation and prediction with their full spatiotemporal coverage and low-cost acquisition [27]. With the fusion of homogeneous [28] or heterogeneous [29] floating vehicle data, the performance of network-wide slice-wise missing flow estimation can be effectively improved by the transfer learning (TransL) method or graph theory model (GTM), respectively. Hence, the study on the data fusion-based tensor approach for interval-wise missing deserves to be attempted.

In Table I, we summarize the development in missing traffic data imputation, characterized in terms of research scenario, missing mechanism and patterns, basic approach, and/or imputation using a single or multiple data sources. In each row, we only list typical methods and references, and more details of relevant methods can be found in the literature review of Sun, et al. [30] and Xing, et al. [1]. Note that the last row in Table I highlights the novel contribution of this research compared to existing literature.



(a) Missing completely at random. (b) Missing at random.

Fig. 1. Classification of missing value by missing mechanism.

III. Problem Statements

A. Problem Formulation

The interval-wise missing traffic volume data is constructed into a 4-way tensor, notate that $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$ comprising 'time mode', 'day mode', 'segment mode', and 'data type mode'. Here, I_1, I_2, I_3 and I_4 depicts the number of time intervals, days, segments, and data types, respectively. Meanwhile, we define a tensor $\mathcal{W}$ of identical size $\mathcal{A}$, representing the indices of observed entries. For a missing traffic volume data entry, each element of w, i.e. $w_{i_1 \times i_2 \times i_3 \times i_4}$, $i_n \in [1, I_n]$ equals 0, while a value equals 1 for an observed volume data entry.

Let $A^{(n)}$ denote factor matrices which are factorized by tensor $\mathcal{A}$, and T_n represent the calibration variables along with n th factor mode. The objective function of missing data imputation can be simplified as follows

$$f_w(A^{(1)}, A^{(2)}, \cdots, A^{(N)}) \equiv \frac{1}{2}\left\|\mathcal{Y} - \mathcal{Z}\right\|^2 + \sum_{n=1}^{N} \frac{\lambda_n}{2}\left\|A^{(n)}\right\|^2 \quad (1)$$

where $\mathcal{Y} = \mathcal{W} * \mathcal{X}$ and $\mathcal{Z} = \mathcal{W} * \left[\!\left[A^{(1)}, A^{(2)}, \cdots, A^{(N)} \right]\!\right]$. The tensor $\mathcal{Y}$ can be pre-computed for a given $\mathcal{X}$ as neither $\mathcal{W}$ and $\mathcal{X}$ changes during the iterations, while the tensor $\mathcal{Z}$ can also be computed efficiently using tensor completion. In this tensor decomposition approaches, the latter term of equation (1) adds the regularization term. On this basis, inspired by Acar, et al. [31] and Jiang, et al. [17], we extra add customized mode averaging in the original regularization term. This adjustment is used to reduce the impact of diverse disturbances caused by different segment grades and different missing lengths and make full use of the unique characteristics of CL data as an auxiliary role in process of our interval-wise missing scenario. As such, an new objective function for interval-wise missing data imputation in urban network is listed as follows

$$f_w(A^{(1)}, A^{(2)}, \cdots, A^{(N)}) \equiv \frac{1}{2}\left\|\mathcal{Y} - \mathcal{Z}\right\|^2 + \sum_{n=1}^{N} \frac{\lambda_n}{2}\left\|A^{(n)} - M_n\right\|^2 \quad (2)$$

where M_n is the average mode matrix that is used to weigh and

reduce the diverse disturbances in the urban network. And these details can be found in the following Section IV.

Furthermore, regularization parameters λ can be viewed as a tradeoff between the approximation errors and the fitting error. It can be solved by Bayesian optimization, which is composed as $\lambda^* = \arg\max_{\lambda \in \mathcal{L}} G(\lambda)$, where $G(\lambda)$ denote the error function of missing traffic volume imputation.

B. *Classification of Missing Pattern*

Based on the difference in missing mechanisms, missing traffic data can be classified into three types [32]: the missing completely at random (MCR), missing at random (MR), and not missing at random (NMR).

In MCR, the reason for this missing pattern is mainly due to instantaneous lower voltage power or communication fault. Herein, these missing points are randomly scattered over all periods, and each missing location is completely independent of the other. As shown in Fig. 1(a), in MCR, missing entries are often caused by physical damage or maintenance backlog, while the missing values are related to the readings of their temporal or spatial neighbors. And thus, missing patterns tend to appear at a particular sequential missing point at the same time or on the same sensor. As shown in Fig. 1(b), in MR**,** the missing pattern may occur due to long-term detector malfunction. The missing data occur in random patterns, and the missing values are like blocks. The first two (MCR and MR) imply that there is no underlying mechanism for generative models in the description of missing data, while NMR assumes the dependence of missing data distribution on the complete dataset.

Furthermore, the missing data can also be further classified by the type of missing length. As the spatial dimension and temporal dimension of traffic flow data can be viewed as a two-dimensional matrix. To determine whether the missing values appear sequentially in one dimension or two dimensions, we can divide them into two classes: point-wise missing, interval-wise missing, and slice-wise missing [8]. In point-wise missing, the traffic flow data in the road segment is missing only at a certain instance, while the data adjacent to the missing point is complete. In interval-wise missing, sequential missing values appear at consecutive periods in a specific road segment or some different space point at the same instance. Slice-wise missing is caused by the lack of detectors installed in entire road segments. The last type of missing pattern estimation generally belongs to coarse-grained imputation or estimation, which is a different research field from the former. Compared existing traffic studies, particularly that on interval-wise missing data, usually take less consideration on the distinction of these aforementioned missing patterns [9].

This study is focused on the interval-wise missing volume data extracted from LPR detector with different missing length, and compare their performance in different missing ratios and two types of missing scenarios: missing at random (MR) and not missing at random (NMR).

C. *Study Frameworks*

Fig. 3 presents the proposed framework. The objective is to apply CL data and LPR data for characterization in our proposed tensor-based model and thus allow us to impute

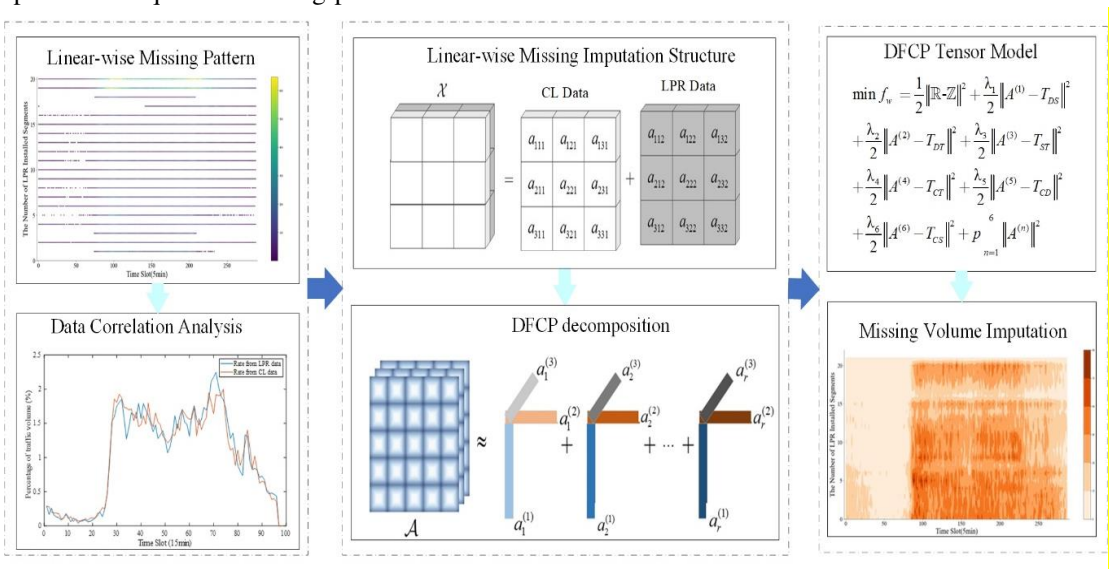


Fig. 2. Framework of interval-wise missing volume imputation.

interval-wise missing volume. The steps of missing traffic data imputation are described as follows:

Step 1: The set of similar segments from the study area is selected to estimate missing traffic volume values. Correlation analysis is used for assessing segment similarity.

Step 2: The structure of the DFCP tensor-based model is constructed, with the 4-way tensor model representing the dimensions of day by day, segment by segment, time slot by time slot, and data by data. Additionally, the size of the input data from each dimension is determined.

Step 3: The regularization penalty coefficients from each dimension are calculated. The hyper-parameter method by Bayesian optimization is applied to solve the optimal value.

Step 4: The imputation results both including and excluding multi-source data into the tensor model are compared. Also, at

different missing types and different missing rates, and evaluated with other baseline methods.

IV. Mathematical Model Formulations

In this work, our goal is to extract the underlying factors via our proposed DFCP tensor model from the presence of the multi-way nature of the traffic data and recover interval-wise missing traffic volume. We assume that four exist certain correlations from time series, days, segments, and data type modes among each segment on the urban transport network. To combine and utilize these four modes well, a tailored four-way tensor completion method is developed for interval-wise missing data imputation in Section 4.2. And Section 4.3 introduces a Bayesian optimization method for the hyperparameter tuning of regularization penalty coefficients in the proposed tensor model. Before that, we first introduce the notations and the basic principles of tensor in Section 4.1.

A. DFCP Tensor Theory and Notations

1) DFCP Tensor

Tensor, as the higher-order generalization of an array or matrix, can provide a natural way to represent multi-dimensional data whose entries are indexed by several coordinate variables. More formally, a first-order tensor is a vector, a second-order tensor is a matrix, and tensors of order three or higher are called higher-order tensors. It has proven to be a powerful tool for extracting the underlying factors and obtaining latent properties between each data mode of global structure [33]. Due to the advantages of strong interpretability, fast convergence, high accuracy, and small storage space, tensor-based approaches have been widely applied in the field of psychometrics, chemometrics, signal processing, computer vision, data mining, and elsewhere [34].

Tensor-based methods (e.g., tensor completion) have been imported to missing traffic data imputation. Tensor completion is defined as a problem of completing an N -th order tensor from its known elements based on tensor decomposition and low-rank approximation. For example, the global information of traffic data can be simultaneously taken into account by the underlying multi-mode correlation (day, week, time, and space mode). In recent literature, the successful recovery of tensor completion mainly relies on its low-rank assumption. The methods for tensor completion include two kinds of approaches. One is based on a given rank and updates factors in tensor completion. The other one directly minimizes the tensor rank and updates the low-rank tensor. To improve the computational efficiency in the large missing rate of traffic data which has low-rank performance, we select the given rank-based tensor completion method that is based on the first-order optimization as our initial basic method. This given rank-based tensor completion method is to approximate the tensor decomposition with an estimated low rank or low-n-rank when only parts entries of a tensor are observed. And then, the generated factorizations from a given rank-based tensor decomposition can be used to impute the missing entries, and this procedure can be repeated to iteratively determine suitable values for the missing entries.

According to the different research needs, the typical tensor decomposition methods can be divided into CANDECOMP/PARADAC (CP) decomposition [35] and Tucker decomposition [36]. The CP model can be regarded as a special case of the Tucker model, which is a high-order generalization of singular value decomposition and principal component analysis. The rank number determined can affect imputation performance in the process of tensor decomposition. Too few numbers rank would reduce the accuracy of tensor completion. In practice, the rank is generally not known and is not easily determined, and the tensor rank is difficult to minimize in general since it is a non-convex function. However, results from Acar, et al. [37] indicate that CP-based decomposition methods have an advantage over other decomposition approaches when the rank is over-estimated. Both of their optimization methods are NP-hard problems. Meaning that the factors in each mode are necessarily linear dependent. In practice, the rank is generally not known and is not easily determined [34], and tensor rank is difficult to minimize in general since it is a non-convex function. But the rank of CP decomposition is allowed to be greater than the largest data dimension. Furthermore, the tensor composed of spatiotemporal traffic flow data is proved to be a low-rank property data [38]. Herein, to improve the performance, the CP decomposition is regarded as a promising method in our research. Computing CP decompositions by applying the alternating least squares method, which computes the factor matrices one at a time, and iterative imputation is quite effective and has the advantages of often being simple and fast. Furthermore, the CP tensor factorizations that can exploit the global information, have more obvious underlying physical interpretability modes than the Tucker tensor. Furthermore, to ensure that the value of the un-missing data in the original tensor remains unchanged, CP decomposition with weighted optimization (CP-WOPT) for dealing with a large amount of missing value has been testified to provide a good imputation performance [37], which could significantly reduce the storage and computation costs.

With the summary of the aforementioned existing tensor-based methods, to improve the imputation efficiency, we select the given rank-based and CP-WOPT-based tensor decomposition methods as our basic methods. Furthermore, our goal is to interpolate interval-wise missing data with long-term missing, considering the effect of diverse disturbances in complex urban networks. In our research, our proposed DFCP model can well solve this problem by adding full spatiotemporal coverage of CL data and a regularization penalty term. With the application of the Bayesian optimization for solving, our data imputation performance can be enhanced by accurate regularized hyperparametric. In this process, we simplify the difficulty of the non-uniqueness/non-convexity problem of regularized hyperparametric solutions caused by the rank determination of tensor decomposition. The details of this method are listed in the following section.

2) Notations

In this paper, the notations denoted in Acar et al. [37] and

Kolda and Bader [34] were partially adopted. Note that the third-order tensors are denoted by calligraphic letters $(\mathcal{A},\mathcal{B},...)$, matrices by capitals $(A,B,...)$, vectors by lower-case letters $(a,b,...)$, and special scalars by capital letters $(I_1,I_2,I_3,N,R,...)$. For instance, the element of a third-order tensor $\mathcal{A}$ is denoted as $a_{ijk}=\mathcal{A}_{ijk}$. The Cartesian product of two sets is represented by the symbol "$\times$".

The elements of a third-order tensor are referred to with a set of three indices. For third-order tensors, mode-1 and mode-2 vectors are their columns and rows respectively. In general, we can obtain model-n. vectors $(n=1,2,3)$ by varying the n th index while keeping other indices constant. The number of the linearly independent model-n vectors are called mode-n rank. It is a generalization of the row rank and column ranks of a matrix. Unlike that in the case of a matrix, different model-n ranks aren't necessarily equal to each other.

The model-n product $\mathcal{A}_{\times n}M^{(n)}$, $n=1,2,3$ of a tensor $A\in\mathbb{R}^{I_1\times I_2\times I_3}$ with matrices $M^{(n)}\in\mathbb{R}^{J_n\times I_n}$ is defined as follows

$$\begin{aligned}(\mathcal{A}_{\times 1}M^{(1)})_{J_1 i_2 i_3}&=\sum_{i_1}a_{i_1 i_2 i_3}m^{(1)}_{j_1 i_1},\\ (\mathcal{A}_{\times 2}M^{(2)})_{i_1 J_2 i_3}&=\sum_{i_2}a_{i_1 i_2 i_3}m^{(2)}_{j_2 i_2},\\ (\mathcal{A}_{\times 3}M^{(3)})_{i_1 i_2 J_3}&=\sum_{i_3}a_{i_1 i_2 i_3}m^{(3)}_{j_3 i_3},\end{aligned}\tag{3}$$

where $1\le i_n\le I_n$, $1\le j_n\le J_n$.

An N-way tensor can be rearranged as a matrix. This process is called matricization and is also known as unfolding or flattening. The model-n matricization of a tensor, $\mathcal{A}\in\mathbb{R}^{I_1\times I_2\cdots\times I_N}$ can be denoted as

$$(\mathcal{A}_{(1)})_{i_1(i_2-1)I_3+i_3}=(\mathcal{A}_{(2)})_{i_2(i_3-1)I_1+i_1}=(\mathcal{A}_{(3)})_{i_3(i_1-1)I_2+i_2}=a_{i_1 i_2 i_3},\tag{4}$$

where $1\le i_1\le I_1$, $1\le i_2\le I_2$, $1\le i_3\le I_3$. The matricization process arranges the mode-n one-dimensional "fibers" as columns of the resulting matrix, see [37] for details.

B. DFCP Tensor Factorization

In this section, we present a customized tensor decomposition-based method called data fusion-based CP (DFCP) factorization with weighted regularization. This method enables the imputation of interval-wise missing data $\mathcal{A}_{\text{est}}$ with high performance, even when the missing traffic volume data appear over large periods.

Based on Kolda and Bader [34], an N-way tensor $\mathcal{A}\in\mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$ with missing values can be factorized into N-factor matrices $A^{(n)}\in\mathbb{R}^{I_n\times R}$. It can be denoted as

$$\min f_w\left(A^{(1)},A^{(2)},\cdots,A^{(N)}\right)=\frac{1}{2}\left\|\mathcal{W}*(\mathcal{A}-\left[\!\left[A^{(1)},A^{(2)},\cdots,A^{(N)}\right]\!\right])\right\|^2\tag{5}$$

$$\left[\!\left[A^{(1)},A^{(2)},\cdots,A^{(N)}\right]\!\right]\equiv\sum_{r=1}^{R}a_r^{(1)}\circ a_r^{(2)}\circ\cdots\circ a_r^{(N)},\tag{6}$$

where R denotes the estimated rank, '$*$' denotes the Hadamard product of tensors, and '$\circ$.' denotes the outer product of vectors. $\mathcal{W}$, which is of the same size as $\mathcal{A}$, is a non-negative indicator tensor and can be defined as

$$\mathcal{W}_{i_1 i_2\cdots i_N}=\begin{cases}1 & \text{if } x_{i_1 i_2\cdots i_N}\text{ is known}\\ 0 & \text{if } x_{i_1 i_2\cdots i_N}\text{ is missing}\end{cases}.\tag{7}$$

For all $i_n\in\{1,2,\cdots,I_n\}$ and $n\in\{1,2,\cdots,N\}$, $a_r^{(n)}\in\mathbb{R}^{I_n}$ is the r th column vector of $A^{(n)}$. Acquiring the calculated factor matrices $A^{(1)},A^{(2)},\cdots,A^{(N)}$, the missing entries $\mathcal{A}$ can be estimated to generate a complete tensor, given as follows

$$\mathcal{A}_{est}=\mathcal{W}*\mathcal{A}+(1-\mathcal{W})*\left[\!\left[A^{(1)},A^{(2)},\cdots,A^{(N)}\right]\!\right]\tag{8}$$

where 1 is a tensor of all ones with the same size $\mathcal{W}$. The representation of CP tensor decomposition is described in Fig. 3.

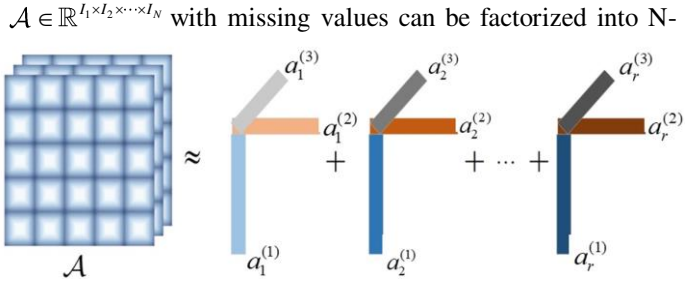


Fig. 3. Representation of CP-based tensor decomposition.

Our goal is to find matrices $A^{(n)}\in\mathbb{R}^{I_n\times R}$, $n\in\{1,2,\cdots,N\}$ that minimize the objective function in equation (5). In the solution of this function, we fix the factor matrix in two directions to optimize the factor matrix in the other direction, and convergence is achieved by alternating. The derivation of the gradient in the weighted case is given in Acar, et al. [37]; here we just report the formula. In matrix notation, using $\mathcal{A}$ and $\mathcal{A}_{est}$ from equation (8), we can rewrite the gradient equation as

$$\frac{\partial f_{\mathcal{W}}}{\partial A^{(n)}}=\left(\mathcal{A}_{est(n)}-\mathcal{A}_{(n)}\right)A^{(-n)},\tag{9}$$

where $A^{(-n)}=A^{(N)}\odot\cdots\odot A^{(n+1)}\odot A^{(n-1)}\odot\cdots\odot A^{(1)}$, for $n=1,\ldots,N$. The symbol $\odot$ denotes the Khatri-Rao product.

However, the aforementioned approach could suffer from scaling and permutation indeterminacy when the input data from each dimension has a large variant. For example, in large-scale study areas, due to differences in road levels and commuting times, the degree of correlation from day mode and segment mode varies continuously. In this study area, according to the correlation description presented in section 5.3, we found that the spatiotemporal correlations between day mode and segment mode are fluctuant. In addition, some parts with high correlation have almost similar trends, and vice versa. Hence, to address this issue, we need to utilize these correlations well by weighing the similarities between day mode and segment mode.

Furthermore, in the interval-wise missing data problems, an increase in the length of interval-wise missing data will have a considerable effect on the imputation performance. In this context, based on the similar volume trends and high correlation between CL data and LPR data, the part of interval-wise missing volume from LPR data could be effectively interpolated with the complete CL data. Hence, we established

a four-way tensor model that includes day mode, segment mode, time slot mode, and data type mode. The representation of the structure of the four-way tensor model is shown in Fig. 4.

As the amount of missing data increases, the performance of the algorithm may suffer since the initialization and the intermediate models used to impute the missing values will increase the risk of converging to a less optimal solution.

To model and perform the afore-mentioned analysis, a Tikhonov regularization loss function is added to the objective function as follows

$$\min f_w(A^{(1)},A^{(2)},\cdots,A^{(N)})=\frac{1}{2}\left\|\mathcal{W}*(\mathcal{A}-[\![A^{(1)},A^{(2)},\cdots,A^{(N)}]\!])\right\|^2 +\sum_{n=1}^{N}\frac{\lambda_n}{2}\left\|A^{(n)}-M_n\right\|^2 \quad (10)$$

Assume that the four-way tensor $\mathbb{R}$ is the actual traffic volume value, the four-way tensor $\mathbb{Z}$ is the recovered traffic volume value with the same size as $\mathbb{R}$, T is the time slot mode, S is the road segment mode, D is the day mode, and C is the data category mode. In equation (10), the Tikhonov regularization term is added, and the original objective function is transformed as follows

$$\min f_w =\frac{1}{2}\left\|\mathbb{R}\text{-}\mathbb{Z}\right\|^2+\frac{\lambda_1}{2}\left\|A^{(1)}-M_{DS}\right\|^2+\frac{\lambda_2}{2}\left\|A^{(2)}-M_{DT}\right\|^2 +\frac{\lambda_3}{2}\left\|A^{(3)}-M_{ST}\right\|^2+\frac{\lambda_4}{2}\left\|A^{(4)}-M_{CT}\right\|^2+p\sum_{n=1}^{N}\left\|A^{(n)}\right\|^2+\xi \quad ,(11)$$

where $A^{(n)}\in\mathbb{R}^{I_n\times R}$ is a factor matrix of mode- unfolding, and represents matricization of constructed tensor in the corresponding mode after tensor decomposition, R is the number of modes. Furthermore, M_n is the average mode matrix, which is averaged from the combination of any two traffic volume modes in these four data dimensions. And the subscripts denote the corresponding four data dimensions: time slot mode (T), segment mode (S), day mode (D), and data category model (C). For instance, M_{DS} represents the two-dimensional matrix composed of days and segments, in that the data category and time slot mode are eliminated by averaging these two data dimensions, and the remaining days and segments form this new matrix. As the size of the matrix is composed of data category and days, or data category and segments is too small and has little impact. Thus, we do not consider the regularization of these two matrices. When dealing with the diverse disturbances in the urban network, such averaging within each dimension and calculating the difference between the factorization matrix and it can be used to determine the severity of the disturbance in this factorization matrix. Then, these disturbances are weighted and reduced by setting the value of the coefficients λ_i. Among them, coefficients λ_i for $i=1,\cdots,4$ representing the weights of the regularized items in each dimension, and are used to control the appropriate importance of different modes in the tensor decomposition method. When its value is larger, it means that the greater the calibration of the specified mode is given, and vice versa. For example, when in an urban transport network with a large difference in segment grades, the segment mode brings a weaker calibration impact, while when long-term missing information occurs, the data mode brings a stronger calibration impact.

Based on the aforementioned matricization, the 4-way tensor could be rearranged into six matrices by flattening it in six different directions. We can use regularization terms to calibrate interval-wise missing parts by the time-varying characteristics of traffic volume from spatiotemporal modes (day mode, and segment mode) and similar fluctuation trends between CL data and LPR data. The four regularization terms of equation (11), represent the calibration for the different combinations of modes, e.g., $\left\|A^{(1)}-M_{DS}\right\|$ denotes the calibration for time slot mode in all data categories. Furthermore, to prevent overfitting of the recovered data, the regularization penalty term is set to $p\sum_{n=1}^{N}\left\|A^{(n)}\right\|^2$. To obtain the solution of this algorithm, the established DFCP function is solved using the gradient descent method, which is similar to the CP-WOPT approach. A comprehensive discussion on the computation of tensor decomposition can be found in the reference [37].

C. *Scheme of Regularization Penalty Terms*

To avoid the effect of the ill-posed problem caused by non-uniqueness in the process of tensor completion [39], the rule of Tikhonov regularization has been applied in our study [40]. The effect of regularization is to keep the permutation and scaling of the factor matrix in tensor decomposition unchanged, and then the lack of a unique solution can be corrected by modifying

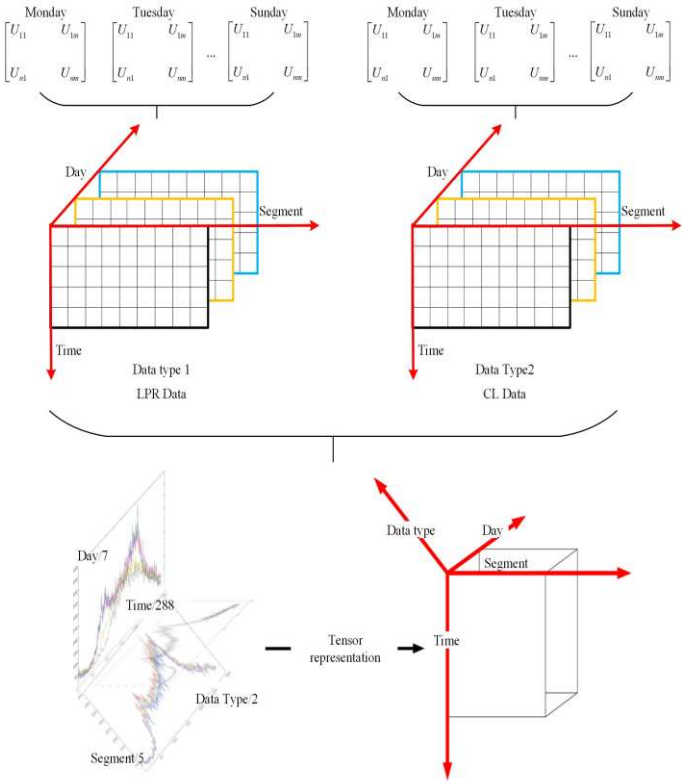


Fig. 4. Representation of the structure of the four-way tensor model.

the objective function to include a Tikhonov regularization term [31]. In our interval-wise missing traffic imputation scenario, the regularization penalty terms are set in equation (11). Its function is to make the tensor completion result unique. In particular, after introducing the factor matrices T as a priori information, it can be uniquely used for rectifying the diverse disturbances caused by the difference of segment grades in the urban transport networks and providing auxiliary in the long-term missing information with CL data.

Clearly, the solution of regularization penalty coefficients is also crucial in the performance of the regularization term. And the solution of regularization value $\boldsymbol{\lambda}$ is the process of hyper-parameters tunning. It can be divided into priority selection rules and posterior selection rules, and the latter is primary. For example, Morozov's discrepancy principle, generalized cross-validation, and the L-curve method [41]. However, these methods are empirical. And when multi-dimension hyperparameters exist, the function structures are difficult to quantify the uncertainty and obtain the optimal hyperparametric solution. Note that we set 4 regularization penalty coefficients in our studies. As Bayesian optimization can rely on previous empirical values and accelerate to obtain hyper-parameter values in the fewest iterations. It is an approach to optimizing objective functions which are "expensive to evaluate", even lacking known special structures like concavity or linearity, which can be seen as optimizing a black-box function [42]. Since then, we set Bayesian optimization (BO) as our hyperparameter tunning approach which has rarely been applied in previous studies on the regularization of tensor decomposition.

Regardless of the complexity of the structure of the tensor decomposition function $\min f_w\left(A^{(1)}, A^{(2)}, \cdots, A^{(N)}\right)$ and the uncertainty of factor matrices $A^{(1)}, A^{(2)}, \cdots, A^{(N)}$, our goal is just to build a surrogate $\boldsymbol{\lambda}^* = \arg\min_{\boldsymbol{\lambda}\in\mathcal{G}} G\left(f_w(\boldsymbol{\lambda})\right)$ for replacement. This is a process that delicately converts the tensor decomposition error under each hyperparameter into the calculating of the mean and covariance functions. And assuming that the hyperparameters obey a Gaussian distribution, a Gaussian process is then used as a surrogate function and the new hyperparameter values in the posterior can be found based on the prior distribution. It could save computational effort by reducing the complex tensor decomposition process.

This process is before the tensor decomposition that the hyperparameters are set before each determination of the factor matrix, and the imputation error function G is used to evaluate each hyperparameter setting. Among them, the hyperparameter here is a vector $\boldsymbol{\lambda}^i(\lambda_1^i, \lambda_2^i, \lambda_3^i, \lambda_4^i)$. BO is applied to find a solution by making a series of evaluations $\boldsymbol{\lambda}^1, \boldsymbol{\lambda}^2, \cdots, \boldsymbol{\lambda}^N$ of function G, and it is to solve the global optimization problem. We can divide this process into two steps: 1) fit a Gaussian process (GP) from the regularization coefficient set $\mathcal{D}_n (n = 1, \cdots N)$ and obtain $\boldsymbol{\lambda}^* = \arg\min G\left(f_w(\boldsymbol{\lambda}_n)\right)$; 2) apply an acquisition function to evaluate $G\left(f_w(\boldsymbol{\lambda}_n)\right)$ and $\mathcal{D}_n = \mathcal{D}_{n-1} \cup (\boldsymbol{\lambda}_n, G_n)$, and decide where to sample. The details are listed as follows.

1) Gaussian Process

In our research, the surrogate function we selected to perform Bayesian optimization $G\left(f_w(\boldsymbol{\lambda}_n)\right)$ is realized by building a Gaussian process. The flexible distribution allows us to associate a normally distributed random variable at every point in the continuous input space. As the distribution for a new observation of the regularization penalty coefficient vector $\boldsymbol{\lambda}$ is obtained that follows a Gaussian distribution [43, 44]. Since then, the joint distribution of G and G_* are set as

$$\begin{pmatrix} G \\ G_* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m(\boldsymbol{\lambda}) \\ m(\boldsymbol{\lambda}_*) \end{pmatrix}, \begin{pmatrix} \boldsymbol{K} & \boldsymbol{K}_* \\ \boldsymbol{K}_*^T & \boldsymbol{K}_{**} \end{pmatrix}\right), \tag{12}$$

where $K = k(\boldsymbol{\lambda}, \boldsymbol{\lambda})$, $K_* = k(\boldsymbol{\lambda}, \boldsymbol{\lambda}_*)$ and $K_{**} = k(\boldsymbol{\lambda}_*, \boldsymbol{\lambda}_*)$. A joint distribution over G and G_* is used for modeling $p(G, G_* | \boldsymbol{\lambda}, \boldsymbol{\lambda}_*)$, and then we have the following

$$p(G_* | \boldsymbol{\lambda}_*, \boldsymbol{\lambda}, G) = \mathcal{N}(G_* | \mu_*, \sigma_*), \tag{13}$$

$$\mu_* = m(\boldsymbol{\lambda}_*) + K_*^T K^{-1}(G - m(\boldsymbol{\lambda})), \tag{14}$$

$$\sigma_* = K_{**} - K_*^T K^{-1} K_*. \tag{15}$$

2) Acquisition Functions

This process is building an acquisition function, which is to replace the surrogate model to determine the next point $\lambda_{i+1} = \arg\min_{\lambda\in\mathcal{H}} G_i\left(f_w(\lambda)\right)$ to evaluate. The acquisition function should be carefully designed to trade-off between exploration of the search space and exploitation of current promising regions. Herein, the expected improvement acquisition function is applied in our experiment. We defined the expected improvement as

$$\boldsymbol{\lambda}_{i+1} = \arg\min_{\boldsymbol{\lambda}\in\mathcal{L}} E[\max(G_i(f_w(\boldsymbol{\lambda})) - G^+, 0)], \tag{16}$$

$$\boldsymbol{\lambda}_{i+1} = \arg\min_{\boldsymbol{\lambda}\in\mathcal{L}} \begin{pmatrix} \left(G^+ - \mu_i(\boldsymbol{\lambda})\right)\Phi\left(\dfrac{G^+ - \mu_i(\boldsymbol{\lambda})}{\sigma_i(\boldsymbol{\lambda})}\right) \\ +\sigma_i(\boldsymbol{\lambda})\phi\left(\dfrac{f_w^+ - \mu_i(\boldsymbol{\lambda})}{\sigma_i(\boldsymbol{\lambda})}\right) \end{pmatrix}, \tag{17}$$

where Φ and ϕ denote cumulative distribution function and probability density function, respectively.

TABLE II DETAILS OF THE DATA TYPES USED

Type of data	Data label	Amount of data	Period
CL data	Mobile Phone ID, Longitude, Latitude, Timestamp	6.8 million data/each day	01/10/2016~07/10/2016
LPR data	The encrypted vehicle plate ID, Timestamp, the vehicle type (Large vehicle Medium vehicle; Small vehicle), The segment ID with LPR detector installed	9 million data	01/10/2016~07/10/2016

V. Case Study

In this section, we conduct comprehensive experiments to evaluate the performance of the proposed DFCP tensor model on real-world datasets under different missing mechanisms and rates. First, we give a brief description of the study area and datasets. Then we introduce our experiment setup and three state-of-the-art imputation models as baselines. Finally, we give some discussions about the corresponding results in detail.

A. Data Preparation

1) Study Area

We obtain our applied traffic data from the city of Nanjing, China. This study area covers the road network around Nanjing south train station, located in the center of the city. It comprises 36 bidirectional road segments, and there are 20 road segments for that an LPR detector is to be installed. The location of which is shown by the blue thumbtack in Fig. 5 (a). In addition, Fig. 5 (b) shows a snapshot of the number of CL users on the study network. Our research aims to realize interval-wise missing data imputation from the LPR detector installed road segment with the assistance of CL data. The CL data has the same spatiotemporal coverage as the LPR data. It is scattered homogeneously and continuously in this area, and there are no missing data. In our selected network, there are 7-arterial segments, 5-second trunk segments, and 8-branch segments. The topology of the research road network is shown in Fig. 6.

Fig. 5. (a) Study area with locations of LPR detectors marked as thumbtack.

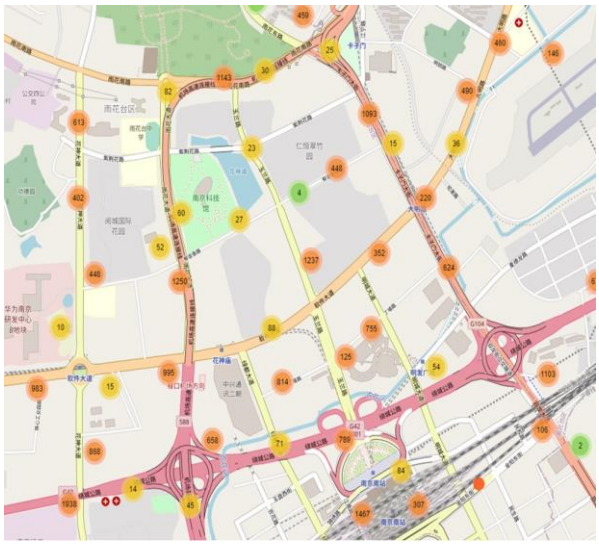


Fig. 5. (b) Snapshot of the number of CL users on the study network.

2) Spatiotemporal Traffic Datasets

The LPR detector system uses cameras installed along the road segment to capture images of vehicle license plate numbers and identifies them using image processing techniques. In China, the LPR system reportedly achieves over 95% recognition accuracy during the daytime and no less than 90% during the nighttime. The volume aggregated from LPR data can be viewed as the actual traffic volume as it also detects vehicle type. Furthermore, note that LPR detectors are mainly used for traffic safety monitoring, and it is generally deployed on arterial or second-trunk roads.

The CL data is extracted by identifying the location information of cellphone signal transmissions using the time difference of arrival (TDOA) positioning technique [27]. Its positioning accuracy is around 50 to 100 meters. The unique advantage of CL data is that it has a relatively high coverage rate as compared to other flexible detector data such as GPS data, floating car data, etc. CL data transmission rate is also relatively consistent as it is generated whenever the mobile information exchange occurs. Hence, it can be regarded as high-quality data with whole spatiotemporal coverage. Existing studies also have demonstrated that there is a high correlation between CL data with actual traffic volume extracted by LPR data [45]. As such, we utilize this correlation of both types of

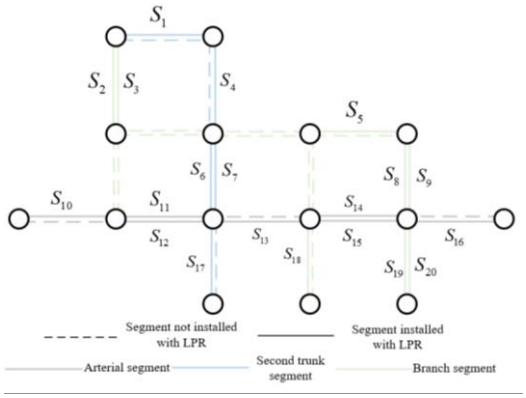


Fig. 6. The topology of research road network.

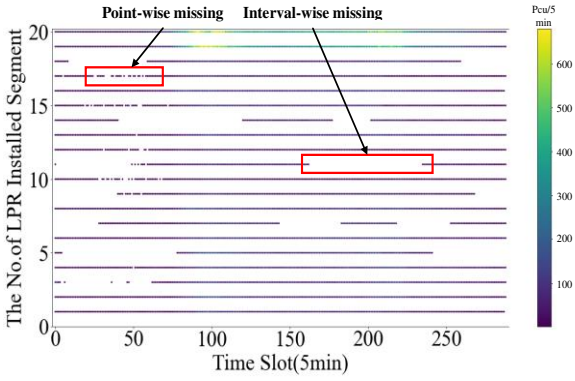


Fig. 7. Visualization of missing LPR data in our study area.

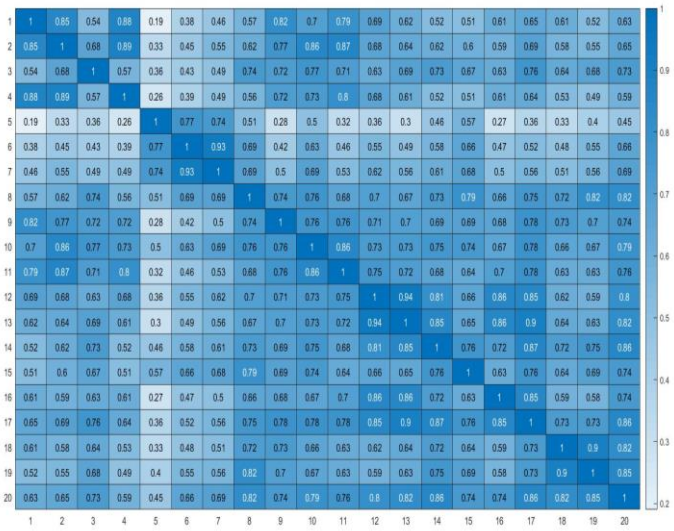


Fig. 8. Value of PCC between each research segment.

data, and the homogeneity of the CL data to realize the interval-wise missing LPR volume imputation.

Table II illustrates the details of the two data types used in the study. No discussions are presented here regarding the data preprocessing technique used. Readers interested in understanding the relevant methods in detail could refer to Xing et al. [27]. The aggregated traffic volume from cellphone location data and LPR data for this paper can be found in the Dryad platform.[1]

To justify that this interval-wise missing in our used LPR data is genuine, we select one day on segments with LPR installed to visualize this missing status in Fig. 7. We can find that there are both interval-wise and point-wise missing data on these detected segments. And most of the interval-wise missing data can be found to occur mainly at nighttime. This poor data quality is due to the challenges in processing low-light images, which is the focus of our study.

B. Experiment Setup

1) Research Segment Pre-processing

This multi-dimension correlation of traffic data is critical in enhancing the performance of missing data imputation in the process of tensor decomposition. In our study of urban network, it includes arterial roads, secondary trunk roads, and branch roads, which is shown in Fig.6. Each type of road segment has different traffic operation functions, and this variance between road grades would lead to the interference of imputation and unnecessary computation workload. Hence, following existing studies in Ran, et al. [38] and Li, et al. [46], we select pre-processing and pick a set of similar segments in advance. To measure the segment similarity with traffic volume, we adopted the Pearson correlation coefficient (PCC) as the standard. The formula for PCC can be written as follows

$$PCC_{i,i+1} = \frac{\sum_{j=1}^{n}(S_{i,j}-\overline{S}_i)(S_{i+1,j}-\overline{S}_{i+1})}{\sqrt{\sum_{j=1}^{n}(S_{i,j}-\overline{S}_i)^2}\sqrt{\sum_{j=1}^{n}(S_{i+1,j}-\overline{S}_{i+1})^2}}, \quad (18)$$

where n denotes the number of samples, $S_{i,j}$ and $S_{i+1,j}$ denotes the traffic volume from any two segments in i th time interval respectively, and $\overline{s}_i$ is the mean value of $S_{i,j}$, $j=1,2,...,n$ and $\overline{S}_{i+1}$ is the mean value of $S_{i+1,j}$, $j=1,2,...,n$. The PCC value ranges from +1 to -1, where a positive PCC value depicts a positive correlation and vice versa. In practical application, the value of PPC can indicate the correlation level.

As introduced earlier, there are 20 road segments with LPR detectors installed in the study area. The Pearson correlation between each pair of segments is calculated and displayed in Fig. 8. On the one hand, the number of segments selected can have an effect on the rank number in the process of tensor decomposition that has been mentioned in setction IV. On the other hand, due to the divergence from different grades of segments, the selection of too many segments would also affect the interpolation, particularly in the complex urban network. Herein, limited by the scale of our study area, we set the number of research segments of the input model to five in this study. To establish the DFCP tensor structure, we select five segments with the top 5 highest PCC values as the research segments. For example, we select road S_{11} as target interval-wise missing research segment, based on the value of PCC in Fig. 8, the corresponding four correlated segments are selected sequentially for S_{12}, S_{10}, S_{17}, S_2 in Fig. 6. Note that a larger number of road segments in the large-scale network used for our proposed model are also feasible.

In Fig. 9, the time-variation curves in average traffic volume on the five selected segments show their fluctuation trends over

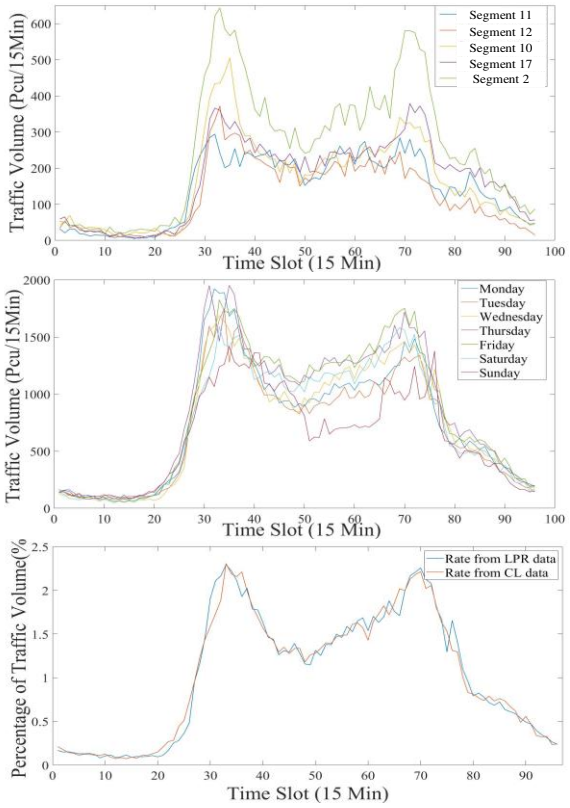


Fig. 9. Correlation analysis from traffic volume on road segments, days, percentage between CL data and LPR data, respectively.

[1] From https://doi.org/10.5061/dryad.qz612jmkz.

the entire research period. The traffic volume in selected five research segments by calculating the average reveals a certain correlation during 5 weekdays, while a relative difference is observed in average traffic volume between weekdays and weekends. Moreover, a similar fluctuation trend between the CL data and LPR data in the percentage of average traffic volume is shown in Fig. 9.

2) Missing Patterns Generation

To evaluate the performance of the DFCP model for missing data imputation sufficiently, we consider two missing data mechanisms as aforementioned in Fig. 1, which include missing at random (MR) and not missing at random (NMR) under different interval-wise missing rates. Following these two missing mechanisms, we manually mask certain entries of observation values as missing data used for our missing cases and then compare the imputed values with our masked ones. The remaining partial values are regarded as input data for learning a well-behaved model. Herein, with equal intervals (set at 10% for one interval), from the minimum missing rate of 10% to the maximum missing rate of 50%.

The evaluation metrics we use are mean absolute percentage error (MAPE) and the root means squared error (RMSE). MAPE measures the average magnitude of the errors by percentage absolute error, which is also normally regarded as an intuitive evaluation metric. RMSE assigns relatively high weights to large errors, and further amplifies and severely punishes large errors. Both the evaluation metrics are widely used error criteria and are defined as

$$MAPE = \frac{1}{|\Omega_m|}\sum_{i\in\Omega_m}\frac{q_i - q_i^{est}}{q_i}, \tag{19}$$

$$RMSE = \sqrt{\frac{1}{|\Omega_m|}\sum_{i\in\Omega_m}\left(q_i - q_i^{est}\right)^2}, \tag{20}$$

where $\Omega_m = \{i_1, i_2, i_3\}$ when q_{i_1,i_2,i_3} is masked and not missing in the original tensor is the total number of missing values from the segment, days, and time slot dimensions. Note that we only calculate imputation error on these masked entries whose ground-truth value is available, which is used to measure the real performance. Besides, we set the missing unit as 5min. The different missing units are viewed as different lengths of interval-wise missing data. Four different missing lengths, at 3, 6, 12, and 18 missing units (i.e. 15min, 30min, 60min, and 90min) are selected and are defined as type 1, type 2, type 3, and type 4 respectively.

3) Baseline Models

Existing studies have testified that tensor decomposition models have greater superiority over other matrix-based models [5] and hybrid approaches [8] in traffic data imputation. Meanwhile, our study's purpose is to analyze the performance of interval-wise missing data imputation based on a multi-source dataset, and whether there is improved efficiency in optimizing the solution of our regularization scheme. Therefore, to assess the accuracy of imputation performance, this section performs a comparison with three state-of-the-art tensor learning models as follows. For a fair comparison, all of them are based on tensor data recovery with precise-designed rank approximation and apply the same data types and tensor dimensions as input.

- *HaLRTC* (Liu, et al. [47]): High-accuracy low-rank tensor completion. This is smartly transformed into a non-convex problem by low-rank approximation, which minimizes the tensor nuclear norm via ADMM. In comparison with HaLRTC, we can verify the low rankness of the two datasets we selected.
- *CP-ALS-NR* (Comon, et al. [48]): CP decomposition-based model in alternating least squares method with normal regularization case. This is a nonconvex minimization model based on the defined tensor nuclear norm, and regularizes only the factorization matrix without any other correction terms. Compared with CP-ALS-NR, the advantage of uniqueness and authenticity of tensor decomposition in our proposed modeling is demonstrated.
- *CP-WOPT-CR* (Acar, et al. [37]): CP tensor model in weighted optimization method with added correction term (e.g. the average value in each dimension) regularization case. This model aims to compare a new tensor completion with our same regularization term, but solving by grid search algorithm, which shows competitive performance Bayesian optimization in solving for regularized parameters.

All aforementioned models are learning-based data-driven approaches and learn model parameters in an iterative way. Both CP-ALS-NR and CP-WOPT-CR apply stochastic gradient descent with an adaptive learning rate is applied for taining. And the main differences are whether to preserve the unmissing

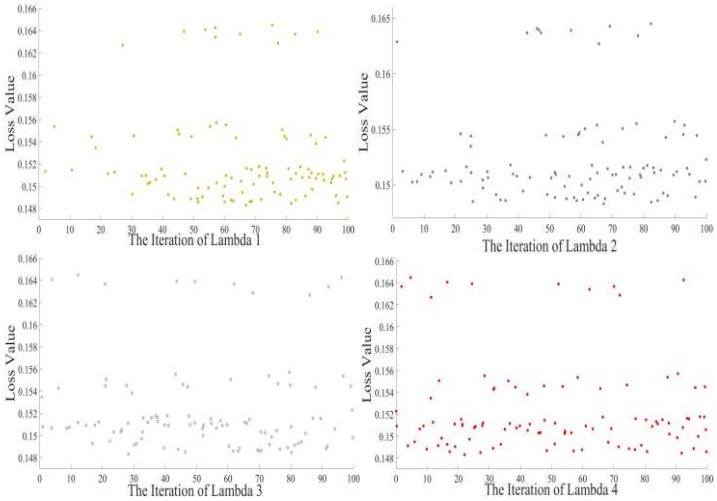


Fig. 10. The iterative process of four hyperparameters.

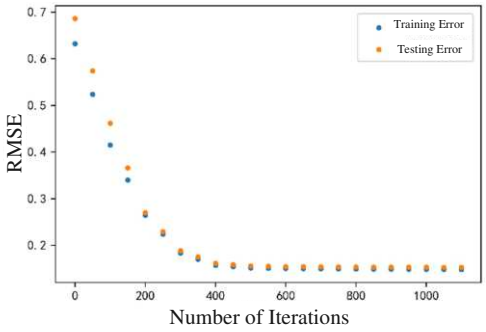


Fig. 11. Error variation in different iterations.

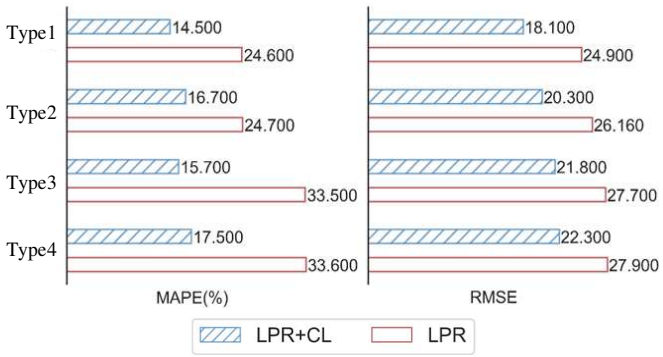


Fig. 12. Comparison of the performance when including/excluding CL data.

values and to use the average value in each dimension to calibrate the factorization matrix. Since the structure of traffic volume data in time series is relatively simple, referring to the imputation result of rank number setting in our proposed tensor model, we set the same rank number in each models. The number of rank is set to 6. Furthermore, we also carefully tuned the hyperparameters in these baseline models. The missing rate from 10% to 50% is set for all missing types. The convergence criterion for all models is set to 0.5×10^{-3}. The maximum number of iteration is set to 1200. And we set the regularization

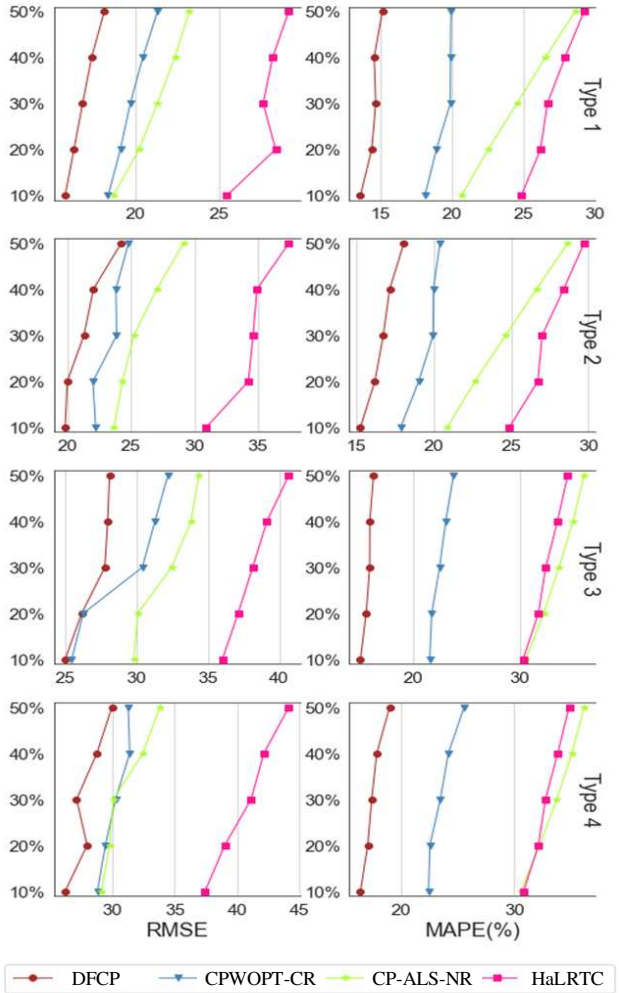


Fig.13. MAPE and RMSE of baseline in different missing rates.

parameter $\lambda_1=\lambda_2=\lambda_3=\lambda_4$ with the initial value being 10^{-3} and updating in each iteration with $\lambda=\min\{1.05\times\lambda,10^5\}$

C. Tensor Structure And Parameter Settings

Based on the similarity and periodicity from different day-of-week, segment, time slot, and data type, this study proposes a 4-way tensor structure as shown in Fig. 4, and then formulate a tensor completion problem.

To verify the effectiveness of missing data completion, a real-world experiment is conducted based on week-long CL data and LPR data from the selected 5 similar segments in the study area. Furthermore, each time slot is set as a 5-min interval, and a day is divided into 288-time slots. A 4-way tensor which includes time slots × days × road segments × data category is formed, with the final size of $144\times7\times20\times2$.

We applied the hyperparameter tuning method for the implementation of the proposed data fusion tensor-based decomposition approach. The tensorly toolkit in Python is applied for this process, and the experiments are carried out on a computer with an Intel(R) Core i7-6500 CPU@2.50GHz and 8GB of RAM. To ensure the reliability of the experimental results, each experiment was run 10 times and the average was taken as the final result. All code used during this study is available in a repository online on the GitHub platform. [2]

Furthermore, we apply Python software to optimize the parameters including the regularization penalty coefficient, penalty, and the number of ranks. In this work, we consider this problem through the framework of Bayesian optimization, in which a learning algorithm's generalization performance is modeled as a sample from a Gaussian process (GP) [49]. And the power of the Gaussian process to express a rich distribution of functions rests solely on the shoulders of the covariance function [50]. Usually, we use some simple kernel functions to represent the covariance matrix, such as the Radial Basis Function (RBF) kernel function and the Matern kernel function:

$$k_{RBF}(\boldsymbol{\lambda}_i,\boldsymbol{\lambda}_j)=\exp\left(-\frac{\left\|\boldsymbol{\lambda}_i-\boldsymbol{\lambda}_j\right\|_2^2}{2l^2}\right) \tag{21}$$

$$k_{Matern}(\boldsymbol{\lambda}_i,\boldsymbol{\lambda}_j)=\frac{2^{1-v}}{\Gamma(v)}\left(\sqrt{2v}\frac{\left\|\boldsymbol{\lambda}_i-\boldsymbol{\lambda}_j\right\|_2}{l}\right)^v H_v\left(\sqrt{2v}\frac{\left\|\boldsymbol{\lambda}_i-\boldsymbol{\lambda}_j\right\|_2}{l}\right) \tag{22}$$

where $\|\cdot\|_2^2$ is the Euclidean distance, l is a fixed length parameter, and v is a smoothing factor. As RBF kernel function is the general form of the Matern kernel function. Hence, we select the RBF function as our covariance function in our Gaussian process. We use the recovered data as the original input data into the iterative process of parameter optimization and set the maximum number of iterations to 1200. The optimal result of the parameter setting is listed as follows, λ_1:16.802, λ_2:9.514, λ_3:6.052, λ_4:8.636, ξ:0.117, rank:6. Fig. 10 shows the iterative process of the four hyperparameters. Since the four parameters are iterated simultaneously, the optimal combination is selected from them. Herein, the iterative plot of the four parameters can be viewed as a tangent line from a four-dimensional curve. Furthermore, while determining the optimal

[2] From https://github.com/xingjiping/DFCPtensor

value of each hyper-parameter, it is necessary to balance the effect of the setting of rank. The improvement effect is most significant when the rank is set to 6 and then tends to be flat as it increases. The training error and testing error are shown in Fig. 11. It can be observed that convergence is reached when the number of iterations approaches 400, and the final verification error (RMSE) is 0.148.

D. Imputation Performance

1) The Performance with/without Fusing CL Data

To demonstrate the superiority of fusing CL data to improve the performance of interval-wise interpolation, we attempt to compare the performance of a 3-way tensor model in time slots × days × segments with/without adding the data type dimensions. Herein, this section presents a comparison by establishing our proposed model that uses only LPR data. It is a 3-way tensor, with dimensions time slots × days ×segments. The number of each dimension of this 3-way tensor is 288×7×5. And the same missing length and ratio are set with our proposed 4-way tensor.

Comparing with the data model between "LPR data + CL data" and "only LPR data", the imputation performance under different missing types are calculated, which are shown in Fig. 12. For all these missing modes, we can clearly see that both the value of MAPE and RMSE are found to increase with the increase in the missing length. The performance of MAPE and RMSE in the "LPR data + CL data" is significantly superior compared to that using LPR data only. Among different missing types, there is a significant increase in imputation error in the missing type 3. The MAPE from applying two data categories changed from 16.7 to 15.7 and showed a downward trend. On the contrary, the MAPE from using only LPR data changed from 24.7 to 37.5, showing a significant increase. Therefore, the imputation performance using data fusion is significantly better than those obtained using only a single data, which is because the continuity of CL data can further better reflect the real traffic state trend [11, 51].

2) Performance Comparison with Baseline Models

The MAPE and RMSE results for baseline experiments are displayed in Fig. 13. It is clear to see that the proposed DFCP model outperforms the other baseline models on four different interval-wise missing lengths and rates. The accuracies of CP-WOPT-CR, CP-ALS-NR, and HaLRTC are found to gradually declining. As HaLRTC is a typical benchmark used in many data imputation works, in comparison with it, we can directly show the strength of our applied CP tensor-based framework. This demonstrates the feasibility of tensor factorization methods in the application of low-rankness traffic data [10]. To directly presents the effect of a unique regularization scheme in our proposed model, we compare sequentially a normal regularization scheme in the CP-ALS-NR, a modified regularization scheme in the CP-WOPT-CR, and a Bayesian optimization-based regularization scheme in the DFCP. Moreover, the basic structure of the three models is similar in that they all iterate by stochastic gradient descent, and the latter two models only improve computational efficiency by calculating fewer non-interpolated values. From the RMSE and MAPE in Fig.12, the performance in the CP-WOPT-CR is overall better than that in the CP-ALS-NR, which shows the advantages of using the average of each dimension to correct the factorization matrix in the regularization term. This is because the variance disturbances in the urban network can be partially offset by weighting each factorization matrix. Furthermore, the performance of DFCP outperforms that of the CP-WOPT-CR, this indicates that the BO-based regularized hyperparametric solution can achieve more reliable performance than grid search.

On the one hand, as the interval-wise missing length increases from type 1 to type 4, the errors keep increasing in all four models. As such, the values of RMSE in the three baselines is ranging from around 25 to 35, and the corresponding values in the DFCP varies slightly, which is ranging from around 25 to 35, this displays that when the interval-wise missing length gets longer, the error does not shift as much as the baseline, which has a stability performance. On the other hand, with the increase in missing rate, the change in slope of the linearity in Fig. 13 shows the extent of the increase in imputation error. Among them, the slope of DFCP is greater than that of other baselines, this highlights that DFCP is stable even when the missing rate is high, and performs satisfactorily and is advantageous. This is achieved by taking advantage of the full spatiotemporal

TABLE III IMPUTATION RESULTS UNDER DIFFERENT MISSING PATTERNS

Missing length	Missing rates	MR		NMR	
		MAPE (%)	RMSE	MAPE (%)	RMSE
Type 1	10%	13.63	15.76	17.86	20.31
	20%	14.42	16.29	19.03	21.24
	30%	14.74	16.81	19.96	21.74
	40%	14.59	17.35	19.98	22.63
	50%	15.16	18.11	20.42	23.67
Type 2	10%	15.26	19.84	20.69	22.36
	20%	16.19	20.01	22.54	23.37
	30%	16.74	21.32	24.55	24.17
	40%	17.15	22.06	26.53	25.23
	50%	18.07	24.24	28.61	26.99
Type 3	10%	15.02	25.17	24.85	27.64
	20%	15.56	26.16	26.19	28.94
	30%	15.89	27.78	26.67	30.09
	40%	15.93	27.98	27.91	30.74
	50%	16.25	29.11	29.23	31.79
Type 4	10%	16.35	24.24	30.58	27.54
	20%	17.12	25.03	32.25	28.57
	30%	17.35	26.16	33.78	29.47
	40%	17.82	27.78	35.11	30.82
	50%	18.94	27.98	36.21	31.42

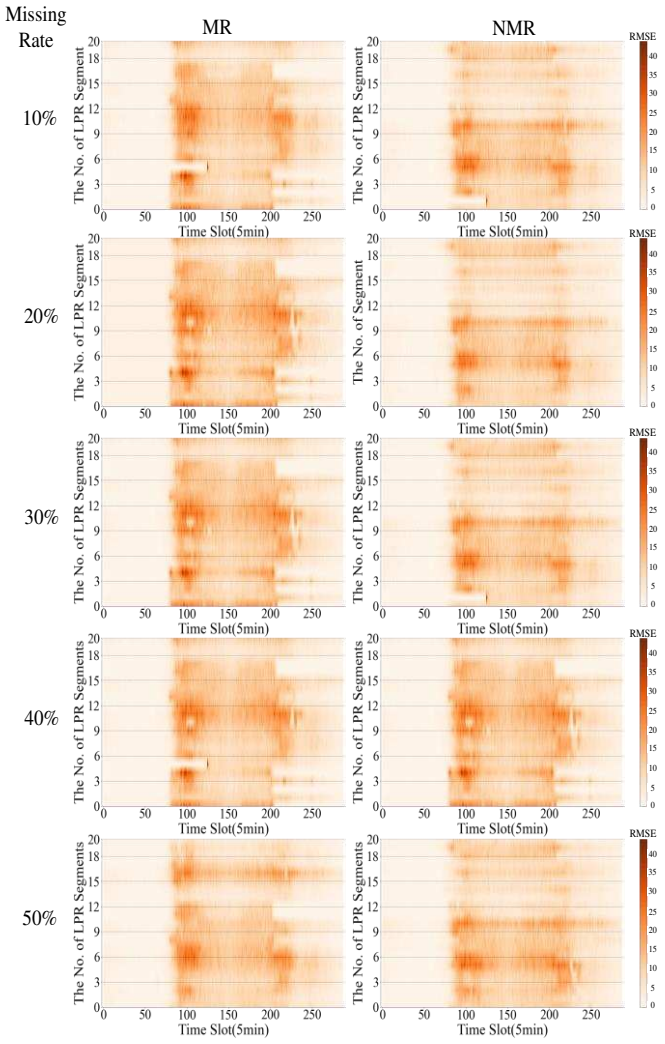


Fig. 14. The performance of RMSE in different missing patterns and rates.

coverage and continuity of CL data, particularly when long interval-wise missing lengths cause continuous missing traffic information, and affected by variance disturbance in the road network, the multi-internal correlation does not work well. Meanwhile, the advantages of BO based regularization scheme in our proposed method can also be found by comparison among other baselines.

3) Sensitivity Analysis of DFCP Model

The MAPE and RMSE of results of random missing patterns and non-random missing patterns are compared and shown in Table III. Furthermore, the performance of RMSE in different missing patterns and rates are also visualized in Fig. 14. For all interval-wise missing lengths and rates, we can clearly see that the accuracy of missing at random (MR) is better than that of the not missing at random (NMR). This is because, for example, when scenarios with traffic missing data occur simultaneously at the same time in all research segments or all days, this blank missing correlation information can only be obtained by CL data in the dimension of the data types

VI. Conclusions

This paper realized the interval-wise missing data imputation from the LPR detector by combining it with the CL data. More specifically, this paper explored how to combine multi-source data into a tensor-based method to solve long-term missing data in the urban area, and also performed a comprehensive analysis of the proposed regularization scheme to handle diverse disturbances by comparing various kinds of missing types and rates with baselines. Model imputation performance was evaluated and compared under different scenarios. For different classical tensor methods and the proposed DFCP model, we applied the result of MAPE and RMSE to compare how different kinds of missing types and missing rates influence their performance. We also analyzed the accuracy of random missing patterns and non-random missing patterns in interval-wise missing scenarios. Based on the experiments conducted, we found that the performance of our proposed method has advantages over other methods, especially when the missing length is long and there are diverse disturbances in the urban network. Overall, both theoretical and empirical evidence indicate that our proposed method has robust and efficient performance in the internal-wise missing data imputation.

Despite the promising results the proposed DFCP model has shown, there are still some directions worth investigating in our future research. Limited by the application of data sources in our selected research urban network, the requirements in the same spatiotemporal coverage should be satisfied. Future efforts are needed to explore the efficiency of this imputation model for different traffic scenarios in the large-scale urban network. Furthermore, different surrogate function schemes in Bayesian optimization and combinations of graph theory and tensor could be further attempted to handle the sensitivity analysis in our added regularization and automatic selection of correlation road, respectively.

References

[1]J. Xing, W. Wu, Q. Cheng, and R. Liu, "Traffic state estimation of urban road networks by multi-source data fusion: Review and new insights," *Physica A: Statistical Mechanics and its Applications,* vol. 595, p. 127079, 2022, doi: https://doi.org/10.1016/j.physa.2022.127079.

[2] Y. Zhang, Q. Cheng, Y. Liu, and Z. Liu, "Full-scale spatio-temporal traffic flow estimation for city-wide networks: a transfer learning based approach," *Transportmetrica B-Transport Dynamics,* Article; 2022, doi: 10.1080/21680566.2022.2143453.

[3] J. W. C. Van Lint, S. P. Hoogendoorn, and H. J. van Zuylen, "Accurate freeway travel time prediction with state-space neural networks under missing data," *Transportation Research Part C-Emerging Technologies,* vol. 13, no. 5-6, pp. 347-369, Oct-Dec 2005, doi: 10.1016/j.trc.2005.03.001.

[4] L. Qu, L. Li, Y. Zhang, and J. Hu, "PPCA-Based Missing Data Imputation for Traffic Flow Volume: A Systematical Approach," *IEEE Transactions on Intelligent Transportation Systems,* Article vol. 10, no. 3, pp. 512-522, Sep 2009, doi: 10.1109/tits.2009.2026312.

[5] H. Tan, G. Feng, J. Feng, W. Wang, Y.-J. Zhang, and F. Li, "A tensor-based method for missing traffic data completion," *Transportation Research Part C-Emerging Technologies,* vol. 28, pp. 15-27, 2013.

[6] M. Zhong, P. Lingras, and S. Sharma, "Estimation of missing traffic counts using factor, genetic, neural, and regression techniques," *Transportation Research Part C-Emerging Technologies,* vol. 12, no. 2, pp. 139-166, Apr 2004, doi: 10.1016/j.trc.2004.07.006.

[7] B. Bae, H. Kim, H. Lim, Y. Liu, L. D. Han, and P. B. Freeze, "Missing data imputation for traffic flow speed using spatio-temporal cokriging," *Transportation Research Part C-Emerging Technologies,* vol. 88, pp. 124-139, 2018.

[8] I. Laña, I. I. Olabarrieta, M. Vélez, and J. Del Ser, "On the imputation of missing data for road traffic forecasting: New insights and novel techniques," *Transportation research part C-Emerging Technologies,* vol. 90, pp. 18-33, 2018.

[9] T. Nie, G. Qin, and J. Sun, "Truncated tensor Schatten p-norm based approach for spatiotemporal traffic data imputation with complicated missing patterns," *Transportation Research Part C-Emerging Technologies,* Article vol. 141, 2022, Art no. 103737, doi: 10.1016/j.trc.2022.103737.

[10] K. Tang, S. Chen, Z. Liu, and A. J. Khattak, "A tensor-based Bayesian probabilistic model for citywide personalized travel time estimation," *Transportation Research Part C-Emerging Technologies,* vol. 90, pp. 260-280, May 2018, doi: 10.1016/j.trc.2018.03.004.

[11] Z. Liu, Y. Liu, Q. Meng, and Q. Cheng, "A tailored machine learning approach for urban transport network flow estimation," *Transportation Research Part C-Emerging Technologies,* vol. 108, pp. 130-150, 2019, doi: 10.1016/j.trc.2019.09.006.

[12] Y. Zhu, Z. Li, H. Zhu, M. Li, and Q. Zhang, "A compressive sensing approach to urban traffic estimation with probe vehicles," *IEEE Transactions on Mobile Computing,* vol. 12, no. 11, pp. 2289-2302, 2013.

[13] X. Wu, M. Xu, J. Fang, and X. Wu, "A multi-attention tensor completion network for spatiotemporal traffic data imputation," *IEEE Internet of Things Journal,* Article vol. 9, no. 20, pp. 20203-20213, Oct 15 2022, doi: 10.1109/jiot.2022.3171780.

[14] J. Li, L. Xu, R. Li, P. Wu, and Z. Huang, "Deep spatial-temporal bi-directional residual optimisation based on tensor decomposition for traffic data imputation on urban road network," *Applied Intelligence,* Article vol. 52, no. 10, pp. 11363-11381, Aug 2022, doi: 10.1007/s10489-021-03060-4.

[15] X. Chen, M. Lei, N. Saunier, and L. Sun, "Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation," *IEEE Transactions on Intelligent Transportation Systems,* Article vol. 23, no. 8, pp. 12301-12310, Aug 2022, doi: 10.1109/tits.2021.3113608.

[16] L. Deng, X.-Y. Liu, H. Zheng, X. Feng, and Y. Chen, "Graph spectral regularized tensor completion for traffic data imputation," *IEEE Transactions on Intelligent Transportation Systems,* Article vol. 23, no. 8, pp. 10996-11010, Aug 2022, doi: 10.1109/tits.2021.3098637.

[17] J. Jiang, F. Sanogo, and C. Navasca, "Low-CP-rank tensor completion via practical regularization," *Journal of Scientific Computing,* vol. 91, no. 1, Apr 2022, Art no. 18, doi: 10.1007/s10915-022-01789-9.

[18] A. Said and A. Erradi, "Spatiotemporal tensor completion for improved urban traffic imputation," *IEEE Transactions on Intelligent Transportation Systems,* 2021, doi: 10.1109/tits.2021.3062999.

[19] X. Chen, Y. Chen, N. Saunier, and L. Sun, "Scalable low-rank tensor learning for spatiotemporal traffic data imputation," *Transportation Research Part C-Emerging Technologies,* vol. 129, Aug 2021, Art no. 103226, doi: 10.1016/j.trc.2021.103226.

[20] Y. Zhu, J. Wang, J. Wang, and Z. He, "Multitask neural tensor factorization for road traffic speed-volume correlation pattern learning and joint imputation," *IEEE Transactions on Intelligent Transportation Systems,* pp. 1-11, 2022, doi: 10.1109/TITS.2022.3205917.

[21] T. Seo, A. M. Bayen, T. Kusakabe, and Y. Asakura, "Traffic state estimation on highway: A comprehensive survey," *Annual Reviews in Control,* vol. 43, pp. 128-151, 2017 2017, doi: 10.1016/j.arcontrol.2017.03.005.

[22] J. Haworth and T. Cheng, "Non-parametric regression for space-time forecasting under missing data," *Computers Environment and Urban Systems,* vol. 36, no. 6, pp. 538-550, Nov 2012, doi: 10.1016/j.compenvurbsys.2012.08.005.

[23] C. K. Wei, G. R. Jagadeesh, A. Prakash, and T. Srikanthan, "A clustering-based approach for data-driven imputation of missing traffic data," in *Integrated & Sustainable Transportation Systems*, 2016, vol. 7, pp. 112-132.

[24] Z. Long, Y. Liu, L. Chen, and C. Zhu, "Low rank tensor completion for multiway visual data," *Signal Processing,* vol. 155, pp. 301-316, Feb 2019, doi: 10.1016/j.sigpro.2018.09.039.

[25] X. Chen, Z. He, and L. Sun, "A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation," *Transportation Research Part C-Emerging Technologies,* vol. 98, pp. 73-84, Jan 2019, doi: 10.1016/j.trc.2018.11.003.

[26] A. Said and A. Erradi, "Spatiotemporal tensor completion for improved urban traffic imputation," *IEEE Transactions on Intelligent Transportation Systems,* 2021, doi: 10.1109/tits.2021.3062999.

[27] J. Xing, Z. Liu, C. Wu, and S. Chen, "Traffic volume estimation in multimodal urban networks using cellphone location data," *IEEE Intelligent Transportation Systems Magazine,* vol. 11, no. 3, pp. 93-104, Fal 2019, doi: 10.1109/mits.2019.2919593.

[28] P. Wang, J. Lai, Z. Huang, Q. Tan, and T. Lin, "Estimating traffic flow in large road networks based on multi-source traffic data," *IEEE Transactions on Intelligent Transportation Systems,* vol. 22, no. 9, pp. 5672-5683, Sep 2021, doi: 10.1109/tits.2020.2988801.

[29] Z. Zhang, X. Lin, M. Li, and Y. Wang, "A customized deep learning approach to integrate network-scale online traffic data imputation and prediction," *Transportation Research Part C-Emerging Technologies,* vol. 132, pp. 67-80, Nov 2021, Art no. 103372, doi: 10.1016/j.trc.2021.103372.

[30]T. Sun, S. Zhu, R. Hao, B. Sun, and J. Xie, "Traffic missing data imputation: A selective overview of temporal theories and algorithms," *Mathematics,* Review vol. 10, no. 14, Jul 2022, Art no. 2544, doi: 10.3390/math10142544.

[31]E. Acar, D. M. Dunlavy, and T. G. Kolda, "A scalable optimization approach for fitting canonical tensor decompositions," *Journal of Chemometrics,* vol. 25, no. 2, pp. 67-86, Feb 2011, doi: 10.1002/cem.1335.

[32]D. F. H. Basu and Srabashi, "Distinguishing “missing at random” and “missing completely at random”," *The American Statistician,* vol. 10, pp. 34-56, 1996.

[33]D. Wang, Z. Cai, Y. Cui, and X. Chen, "Nonnegative tensor decomposition for urban mobility analysis and applications with mobile phone data," *Transportmetrica A-Transport Science,* 2019, 3 123-149. doi: 10.1080/23249935.2019.1692961.

[34]T. G. Kolda and B. W. Bader, "Tensor Decompositions and Applications," *Siam Review,* vol. 51, no. 3, pp. 455-500, 2009.

[35]J. Carroll and J. J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition," *Psychometrika,* vol. 35, pp. 121-134, 1970.

[36]L. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika,* vol. 31, no. 3, pp. 279-311, 1966.

[37]E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Morup, "Scalable tensor factorizations for incomplete data," *Chemometrics and Intelligent Laboratory Systems,* vol. 106, no. 1, pp. 41-56, Mar 15 2011, doi: 10.1016/j.chemolab.2010.08.004.

[38]B. Ran, H. Tan, Y. Wu, and P. J. Jin, "Tensor based missing traffic data completion with spatial-temporal correlation," *Physica A,* vol. 446, pp. 54-63, Mar 15 2016, doi: 10.1016/j.physa.2015.09.105.

[39]H. Tan, J. Feng, Z. Chen, F. Yang, and W. Wang, "Low multilinear rank approximation of tensors and application in missing traffic data," *Advances in Mechanical Engineering,* 2014 2014, Art no. 157597, doi: 10.1155/2014/157597.

[40]R. C. Aster, B. Borchers, and C. H. Thurber, "Tikhonov Regularization," *Parameter Estimation and Inverse Problems (Second Edition),* pp. 93-127, 2013.

[41]A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problem*, New York: John Wiley ed. New York: John Wiley, 1977.

[42]P. I. Frazier, *A tutorial on bayesian optimization*, Stanford University Press ed. Stanford University Press, 2018.

[43]C. E. Rasmussen and C. Williams, "Gaussian Process for Machine Learning,"*International Journal of Neural Systems.* 2006.

[44]F. Rodrigues and F. C. Pereira, "Heteroscedastic Gaussian processes for uncertainty modeling in large-scale crowdsourced traffic data," *Transportation Research Part C-Emerging Technologies,* vol. 95, pp. 636-651, 2018, doi: https://doi.org/10.1016/j.trc.2018.08.007.

[45]Z. Wang, S. Y. He, and Y. Leung, "Applying mobile phone data to travel behaviour research: A literature review," *Travel Behaviour and Society,* vol. 11, pp. 141-155, 2018.

[46]L. Li, J. Zhang, Y. Wang, and B. Ran, "Missing Value Imputation for Traffic-Related Time Series Data Based on a Multi-View Learning Method," *IEEE Transactions on Intelligent Transportation Systems,* vol. 20, no. 8, pp. 2933-2943, 2019, doi: 10.1109/tits.2018.2869768.

[47]J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 35, pp. 208-220, 2013, doi: 10.1109/TPAMI.2012.39.

[48]P. Comon, X. Luciani, and A. de Almeida, "Tensor decompositions, alternating least squares and other Tales," *Journal of Chemometrics,* vol. 23, 2009, doi: 10.1002/cem.1236.

[49]J. Bergstra, R. Bardenet, B. Kégl, and Y. Bengio, "Algorithms for Hyper-Parameter Optimization," presented at the Advances in Neural Information Processing Systems, 2011.

[50]J. Snoek, H. Larochelle, and R. Adams, "Practical bayesian optimization of machine learning algorithms," *Advances in Neural Information Processing Systems,* vol. 4, 06/13 2012.

[51]F. Zhao *et al.*, "Exploratory Analysis of a Smartphone-Based Travel Survey in Singapore," *Transportation Research Record,* Article no. 2494, pp. 45-56, 2015, doi: 10.3141/2494-06.