Transportation Research Part C 193 (2026) 105967

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

A unified spatio-temporal and coupled multi-view tensor framework for simultaneous traffic data reconstruction and event detection

Quan Yu[a], Yu-Hong Dai[b,c], Lu-Bin Cui[d], Shahin Gelareh[e], Minru Bai[a,*]

[a] School of Mathematics, Hunan University, Changsha, 410082, China
[b] State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China
[c] School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China
[d] School of Mathematics and Statistics, Henan Normal University, Xinxiang, 453007, Henan, China
[e] Département R&T, IUT de Béthune, Université d'Artois, Béthune, F-62000, France

ARTICLE INFO

Keywords:
Traffic data reconstruction
Traffic event detection
Multi-view traffic data
Tensor self-representation
Shared subspace learning
Joint sparsity

ABSTRACT

Reliable traffic state sensing is a fundamental prerequisite for Intelligent Transportation Systems (ITS). In real-world scenarios, however, multi-view traffic observations are often simultaneously affected by missing data and anomalous events. These issues can substantially degrade the reliability of downstream traffic analytics. Existing methods typically focus on either traffic data reconstruction or event detection in isolation. Even when low-rank-plus-sparse decompositions are adopted, they often fail to jointly characterize two key properties of multi-view traffic data: the shared latent structure of normal traffic states and the synchronized manifestation of incident-driven deviations across views. To address this limitation, we propose a Unified Spatio-temporal and Coupled Multi-view Tensor (USCMT) framework for simultaneous traffic data reconstruction and event detection. The central idea of USCMT is a traffic-oriented two-level cross-view coupling mechanism. At the background level, cross-view consistency is captured by imposing a joint low-rank constraint on the concatenation of subspace tensors learned via t-product based self-representation. At the event level, common traffic incidents are characterized through a cross-view coupled sparsity-inducing norm applied to anomaly tensors across different views. A unified optimization formulation with shared variables jointly models cross-view background structures and synchronized anomaly patterns. To achieve tighter approximations of the underlying low-rank and sparse structures, generalized nonconvex relaxation functions are incorporated. In addition, explicit spatio-temporal regularization terms are introduced to preserve the physical continuity and temporal evolution patterns of traffic flows. The resulting nonconvex optimization problem is efficiently solved using an Alternating Direction Method of Multipliers (ADMM) algorithm, and the convergence of the generated iterates to a Karush–Kuhn–Tucker (KKT) point is theoretically guaranteed. Extensive experiments on real-world traffic datasets show that USCMT consistently outperforms representative baselines in both traffic data reconstruction and event detection. The source code is publicly available at https://github.com/quanyumath/USCMT.

* Corresponding author.
E-mail addresses: quanyu@hnu.edu.cn (Q. Yu), dyh@lsec.cc.ac.cn (Y.-H. Dai), cuilubin@htu.edu.cn (L.-B. Cui), shahin.gelareh@univ-artois.fr (S. Gelareh), minru-bai@hnu.edu.cn (M. Bai).

https://doi.org/10.1016/j.trc.2026.105967
Received 12 March 2026; Received in revised form 9 July 2026; Accepted 17 August 2026
Available online 22 August 2026

1. Introduction

The effective management and operation of modern Intelligent Transportation Systems (ITS) increasingly rely on large volumes of spatio-temporal traffic data collected from urban road networks (Li et al., 2020b; Zhang et al., 2022). These data commonly include multiple views, such as traffic volume, lane occupancy, and speed, which together provide complementary descriptions of the underlying traffic process. They are fundamental for monitoring urban mobility patterns, optimizing traffic signal control, enhancing road safety, and supporting transportation planning. In practice, however, real-world traffic measurements are frequently affected by data quality issues, most notably missing values caused by sensor malfunctions, communication disruptions, and transmission errors (Chen et al., 2018; Deng et al., 2022). At the same time, ITS applications also require timely detection of critical traffic events, such as crashes, abnormal congestion, and major disturbances, which appear as anomalous patterns in the data (Wang and Sun, 2021). These two challenges are intrinsically coupled in real-world traffic sensing: severe anomalies can distort the estimation of normal traffic states, while inaccurate recovery of missing entries can in turn obscure or weaken event signatures. This interaction becomes even more pronounced in multi-view settings, where normal traffic states exhibit shared latent structure across views, while traffic events give rise to synchronized deviations. Therefore, an effective traffic sensing framework should address reconstruction and detection jointly rather than treating them as two isolated tasks.

1.1. Literature review

Data reconstruction: A primary challenge in traffic sensing is data incompleteness, which makes traffic data reconstruction a foundational task. If not properly addressed, missing data can significantly degrade the performance of subsequent analyses and applications. Existing approaches generally fall into three categories: interpolation-based, probabilistic, and low-rank representation methods.

Interpolation-based approaches estimate missing values by leveraging spatially proximate neighbors, such as the closest one (Chen and Shao, 2000) or the top-K nearest neighbors (Tak et al., 2016). Although computationally simple, these methods struggle to capture long-range spatio-temporal dependencies and tend to perform poorly when the proportion of missing data is high (Hu et al., 2024). Probabilistic models recover data by modeling its generative process with predefined probability distributions and then using statistical inference to estimate unknown parameters and entries. Representative models include Probabilistic Principal Component Analysis (PPCA), Bayesian Gaussian CANDECOMP/PARAFAC (BGCP) decomposition (Chen et al., 2019b), and Bayesian Augmented Tensor Factorization (BATF) (Chen et al., 2019a). However, because traffic data are inherently heterogeneous, such predefined assumptions often fail to generalize well across diverse scenarios. Low-rank models represent traffic data as matrices or tensors and exploit a global low-rank assumption to uncover latent patterns. These models are commonly categorized into Matrix Completion (MC) and Tensor Completion (TC) approaches. MC methods, such as matrix factorization (Wang et al., 2019) and nuclear norm minimization (Chen et al., 2017), focus on two-dimensional data recovery. In contrast, TC methods, including tensor factorization (Tan et al., 2013; Xing et al., 2023) and tensor nuclear norm minimization (Chen et al., 2020), naturally extend to higher-order data. As a result, they can capture complex multi-way correlations and often deliver more accurate reconstructions. To complement this global structure, both MC and TC approaches can incorporate local constraints to preserve spatio-temporal properties. Representative temporal constraints include Toeplitz matrix regularization (Chen et al., 2021) and autoregressive processes (Chen et al., 2023), while Laplacian regularization (Wang et al., 2019) is commonly used to enforce spatial proximity.

A key limitation of many reconstruction methods is that they are single-view and treat each traffic modality in isolation. This prevents them from using the complementary information shared across traffic volume, occupancy, and speed. Early multi-view frameworks were proposed to address this issue (Du and Chen, 2019; He et al., 2024). However, they typically rely on matrix-based representations that disrupt the natural high-order structure of traffic data, and their cross-view fusion is often relatively coarse.

Event detection: Beyond data integrity, another core ITS task is traffic event detection (Jin et al., 2014), which aims to identify anomalous patterns corresponding to incidents such as crashes, road closures, or major public gatherings. Timely and accurate detection is critical for enabling rapid emergency response, rerouting traffic flow, and enhancing public safety (Huang et al., 2020). Existing methodologies mainly fall into two categories: supervised and unsupervised approaches.

Supervised classification frameworks train models such as support vector machines (Parsa et al., 2019; Xiao, 2019) or neural networks (Li et al., 2020a) to distinguish between typical and anomalous patterns using labeled examples. While powerful, their performance is fundamentally constrained by the scarcity of labeled event data and by limited scalability to large networks, where false alarm rates can become unmanageable (Sun et al., 2022). This limitation has motivated strong interest in unsupervised approaches, which detect anomalies without relying on labels. Early methods in this category were often statistical, such as those based on Bayesian inference (Levin and Krause, 1978) or Standard Normal Deviate (SND) algorithms (Culli and Hall, 1997), which identify incidents by analyzing deviations from expected traffic flow. While simple, they are often limited by rigid assumptions. More recent unsupervised paradigms include isolation-based strategies, exemplified by the isolation forest (Liu et al., 2012). Although computationally efficient, its performance can be sensitive to the complex and multimodal patterns of normal traffic. Another major paradigm is reconstruction-based detection, represented by autoencoders such as the multi-layer perceptron autoencoder (Coursey et al., 2024). A key drawback of this line of work is the need for a separate training phase on anomaly-free data. A third paradigm, which is more aligned with our work, is decomposition-based detection. These methods decompose traffic data into regular patterns and sparse anomalies (Hu and Work, 2020), and can identify incidents without prior training.

However, most decomposition-based detection models are still single-view. As a result, they cannot effectively capture subtle events whose signatures become clear only when multiple traffic streams are analyzed jointly. Although multi-view models such as

BRPCA (Yang et al., 2014) and coupled SBRTF (Li et al., 2021) address this issue by enforcing a common sparse structure, they usually emphasize anomaly coupling while giving limited attention to the shared latent structure of normal traffic states.

1.2. Research gaps and contributions

The above review reveals three main gaps. First, many existing traffic sensing methods remain confined to single-view analysis, which leads to information silos and prevents the effective exploitation of latent cross-view correlations. Second, even among multi-view approaches, reconstruction-oriented models mainly focus on shared background patterns, whereas detection-oriented models mainly focus on sparse event coupling. As a result, they often capture only one side of the multi-view traffic structure. Third, although robust tensor decomposition methods based on low-rank-plus-sparse modeling can jointly separate regular and abnormal components, they usually do not explicitly encode two traffic-specific cross-view properties: the shared latent structure of normal traffic states and the synchronized emergence of incident-driven deviations across multiple traffic measurements, such as volume, occupancy, and speed.

To address the above limitations, we propose a Unified Spatio-temporal and Coupled Multi-view Tensor (USCMT) framework for simultaneous traffic data reconstruction and event detection. The main novelty of USCMT lies not merely in integrating these two tasks, but in developing a traffic-oriented two-level cross-view coupling strategy within a tensor self-representation framework. At the first level, USCMT captures the shared cross-view structure of normal traffic backgrounds. At the second level, it characterizes the synchronized cross-view manifestation of event-driven anomalies. By unifying these two levels of coupling, USCMT enables the joint recovery of shared traffic backgrounds and anomaly patterns. The main contributions of this paper are summarized as follows:

(1) **A traffic-oriented two-level cross-view coupling mechanism:** We introduce a coupled tensor modeling strategy tailored to multi-view traffic sensing, where shared structures are characterized at both the background and event levels. For normal traffic states, cross-view consistency is captured by imposing a joint low-rank constraint on the concatenated subspace tensors learned from each view. For traffic events, cross-view synchronization is characterized through a coupled sparsity-inducing regularization on the anomaly tensor, which suppresses unstructured deviations while preserving shared event-related patterns across views.
(2) **A unified self-representation framework for simultaneous reconstruction and detection:** Based on the above coupling mechanism, we formulate traffic data reconstruction and event detection within a unified tensor self-representation model with shared variables. Rather than treating the two tasks as a generic combination, the proposed formulation links them through jointly learned cross-view background representations and coupled anomaly patterns.
(3) **Traffic-aware structural regularization and flexible nonconvex modeling:** USCMT incorporates explicit spatio-temporal regularization for each view to preserve temporal smoothness and spatial correlation in traffic flows. It also adopts generalized nonconvex surrogate functions for both low-rank and coupled sparse components, enabling a more accurate characterization of latent background structures and event-driven deviations.
(4) **A provably convergent optimization algorithm:** We develop an efficient ADMM-based algorithm to solve the resulting nonconvex optimization problem. Despite the inherent nonconvexity, we rigorously establish that the sequence generated by the proposed algorithm converges to a Karush–Kuhn–Tucker (KKT) point.

2. Preliminaries

To prepare for the discussions that follow, we introduce the notation, tensor operations, and basic properties used throughout this paper. For a more comprehensive review, readers are referred to Kolda and Bader (2009).

For a positive integer n, we define the set $[n] := \{1, 2, \dots, n\}$. Scalars, vectors, matrices and tensors are denoted by lowercase letters, boldface lowercase letters, uppercase letters and calligraphic letters, respectively; for example, x, $\boldsymbol{x}$, X, $\mathcal{X}$. The inner product between two tensors $\mathcal{X}$ and $\mathcal{Y}$ is defined as $\langle \mathcal{X}, \mathcal{Y} \rangle = \mathrm{vec}(\mathcal{X})^T \mathrm{vec}(\mathcal{Y})$, where the $\mathrm{vec}(\cdot)$ operator vectorizes the tensor by stacking all its elements into a single column vector. The Frobenius norm is then defined as $\|\mathcal{X}\|_F = \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}$. The kth frontal slice of a tensor $\mathcal{X}$ is denoted by $X^{(k)}$. The mode-n unfolding of a tensor $\mathcal{X}$ is denoted by $X_{(n)}$. Its inverse operation, the mode-n folding operator $\mathrm{fold}_n(\cdot)$, reconstructs the tensor via $\mathrm{fold}_n(X_{(n)}) = \mathcal{X}$. The mode-$n$ product of $\mathcal{X}$ with a matrix M is written $\mathcal{X} \times_n M$, and satisfies $(\mathcal{X} \times_n M)_{(n)} = M\, X_{(n)}$. The Fast Fourier Transform (FFT) applied along the third dimension of $\mathcal{X}$ is represented as $\bar{\mathcal{X}} = \mathrm{fft}(\mathcal{X}, [], 3)$. Conversely, $\mathcal{X}$ can be recovered from $\bar{\mathcal{X}}$ using the inverse FFT, expressed as $\mathcal{X} = \mathrm{ifft}(\bar{\mathcal{X}}, [], 3)$.

The following definitions, based on the t-product, establish an algebraic framework for tensors that is analogous to matrix algebra.

Definition 2.1. (t-product Kilmer and Martin, 2011) Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{Y} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$. The t-product $\mathcal{Z} = \mathcal{X} * \mathcal{Y} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$ is defined by the circular convolution of tube fibers:

$$\mathcal{Z}(i, j, :) = \sum_{k=1}^{n_2} \mathcal{X}(i, k, :) \bullet \mathcal{Y}(k, j, :),$$

where $\bullet$ denotes the circular convolution operator.

Definition 2.2. (conjugate transpose Kilmer and Martin, 2011) The conjugate transpose of a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is denoted by $\mathcal{X}^T \in \mathbb{R}^{n_2 \times n_1 \times n_3}$. It is obtained by transposing each frontal slice and then reversing the order of the transposed slices from 2 to n_3.

Definition 2.3. (identity tensor Kilmer and Martin, 2011) The identity tensor $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ is a tensor whose first frontal slice is the $n \times n$ identity matrix and whose subsequent frontal slices are zero matrices.

Definition 2.4. (tensor inverse Kilmer and Martin, 2011) A tensor $\mathcal{X} \in \mathbb{R}^{n \times n \times n_3}$ is invertible if there exists an inverse tensor $\mathcal{X}^{-1}$ such that $\mathcal{X} * \mathcal{X}^{-1} = \mathcal{X}^{-1} * \mathcal{X} = \mathcal{I}$.

Definition 2.5. (orthogonal tensor Kilmer and Martin, 2011) A tensor $\mathcal{X} \in \mathbb{R}^{n \times n \times n_3}$ is orthogonal if it satisfies $\mathcal{X}^T * \mathcal{X} = \mathcal{X} * \mathcal{X}^T = \mathcal{I}$.

These definitions provide the foundation for the Tensor Singular Value Decomposition (t-SVD).

Theorem 2.1. *(t-SVD Kilmer and Martin, 2011) Any third-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ can be decomposed as:*

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T,$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors, and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an f-diagonal tensor, meaning each of its frontal slices is a diagonal matrix.

Definition 2.6. (tensor tubal rank Kilmer et al., 2013) Let $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$ be the t-SVD of $\mathcal{X}$. The tensor tubal rank of $\mathcal{X}$, denoted as $\mathrm{rank}_t(\mathcal{X})$, is defined as the number of non-zero singular tubes in $\mathcal{S}$, i.e., $\mathrm{rank}_t(\mathcal{X}) = \#\{i \mid \mathcal{S}(i,i,:) \neq \mathbf{0}\}$.

3. A USCMT model for simultaneous traffic data reconstruction and event detection

In this section, we present the Unified Spatio-temporal and Coupled Multi-view Tensor (USCMT) model for joint traffic data reconstruction and event detection in a multi-view setting. Specifically, Section 3.1 introduces the unified coupled multi-view multi-task tensor framework, Section 3.2 describes the spatio-temporal regularity constraints, and Section 3.3 integrates these components into a unified optimization formulation.

3.1. A unified coupled multi-view multi-task tensor framework

In multi-view traffic sensing, variables such as traffic volume, occupancy, and speed reflect the same underlying traffic process from complementary perspectives. These complementary variables therefore exhibit shared latent structures across different views. They also exhibit synchronized deviations when significant traffic events occur. Therefore, our goal is not merely to combine reconstruction and detection in a single formulation. Instead, we aim to jointly characterize two traffic-specific cross-view properties: the consistency of normal traffic backgrounds and the synchronization of event-induced anomalies. To motivate our formulation, we start from the following generalized multi-view self-representation model (Du and Chen, 2019; He et al., 2024):

$$\begin{aligned} \underset{X^v, Z^v, E^v}{\arg\min} \quad & R_1(Z^1, \dots, Z^V) + \gamma \textstyle\sum_{v=1}^{V} R_2(E^v) \\ \text{s.t.} \quad & X^v = X^v Z^v + E^v,\ \mathcal{P}_{\Omega_v}(X^v) = \mathcal{P}_{\Omega_v}(M^v),\ v = 1, \dots, V. \end{aligned} \tag{1}$$

Here, $X^v \in \mathbb{R}^{DT \times N}$ denotes the vth traffic data matrix to be recovered, where the rows correspond to time steps (with T intervals per day over D days) and the columns correspond to the N road segments. $M^v \in \mathbb{R}^{DT \times N}$ is the vth partially observed traffic data matrix with observed entries indexed by Ω_v. The operator $\mathcal{P}_{\Omega_v}$ keeps entries in Ω_v unchanged and sets all others to zero. The variables Z^v and E^v represent the subspace representation matrix and the reconstruction error matrix, respectively.

In our model, we introduce three key enhancements.

Tensor self-representation: Unlike model (1), which represents each traffic data instance X^v as a matrix, we model the data as a third-order tensor $\mathcal{X}^v \in \mathbb{R}^{N \times D \times T}$. This tensor representation naturally captures the high-dimensional spatio-temporal structure of traffic flows. Accordingly, we reformulate model (1) as

$$\begin{aligned} \underset{\mathcal{X}^v, \mathcal{Z}^v, \mathcal{E}^v}{\arg\min} \quad & \mathcal{R}_1(\mathcal{Z}^1, \dots, \mathcal{Z}^V) + \gamma \textstyle\sum_{v=1}^{V} \mathcal{R}_2(\mathcal{E}^v) \\ \text{s.t.} \quad & \mathcal{X}^v = \mathcal{X}^v * \mathcal{Z}^v + \mathcal{E}^v,\ \mathcal{P}_{\Omega_v}(\mathcal{X}^v) = \mathcal{P}_{\Omega_v}(\mathcal{M}^v),\ v = 1, \dots, V. \end{aligned} \tag{2}$$

By generalizing from matrices to tensors, the matrix self-representation $X^v = X^v Z^v + E^v$ in model (1) becomes the tensor self-representation $\mathcal{X}^v = \mathcal{X}^v * \mathcal{Z}^v + \mathcal{E}^v$ in model (2). This shift is fundamental, as illustrated by the following two theorems.

Theorem 3.1. *(Zhou et al., 2021, Theorem 1) Suppose $X \in \mathbb{R}^{n_1 n_3 \times n_2}$ and $Z \in \mathbb{R}^{n_2 \times n_2}$ satisfy $X = XZ$. Then there exist two third-order tensors $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{Z} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ such that $\mathcal{X} = \mathcal{X} * \mathcal{Z}$, where $\mathcal{X}$ can be constructed from X by setting $\mathcal{X} = \mathrm{fold}_2(X^T)$ and $\mathcal{Z}$ can be computed as $\mathcal{Z}(:,t,:) = \mathrm{ifft}\left(\bar{\mathcal{Z}}(:,t,:),[],3\right)$ in which $\bar{\mathcal{Z}}(:,t,j) = Z(:,t)$ for $t \in [n_2]$ and $j \in [n_3]$. The converse, however, does not hold in general.*

Theorem 3.2. *Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{Z} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ satisfy $\mathcal{X} = \mathcal{X} * \mathcal{Z}$, with the frontal slices of $\mathcal{Z}$ given by $Z^{(1)} = Z,\ Z^{(2)} = \dots = Z^{(n_3)} = 0$. Then there exists a matrix $X \in \mathbb{R}^{n_1 n_3 \times n_2}$ such that $X = XZ$, where X can be constructed from $\mathcal{X}$ by setting $X = X_{(2)}^T$.*

Remark 3.1. The two theorems above indicate that any data admitting self-representation after matricization will also satisfy tensor self-representation, but not vice versa unless restrictive conditions are imposed on the subspace-representation tensor. In contrast, tensor self-representation naturally preserves multi-way relationships without forcing the data into a flattened matrix form. This flexibility makes tensor-based self-representation more expressive than its matrix-based counterpart for modeling complex, high-dimensional traffic data.

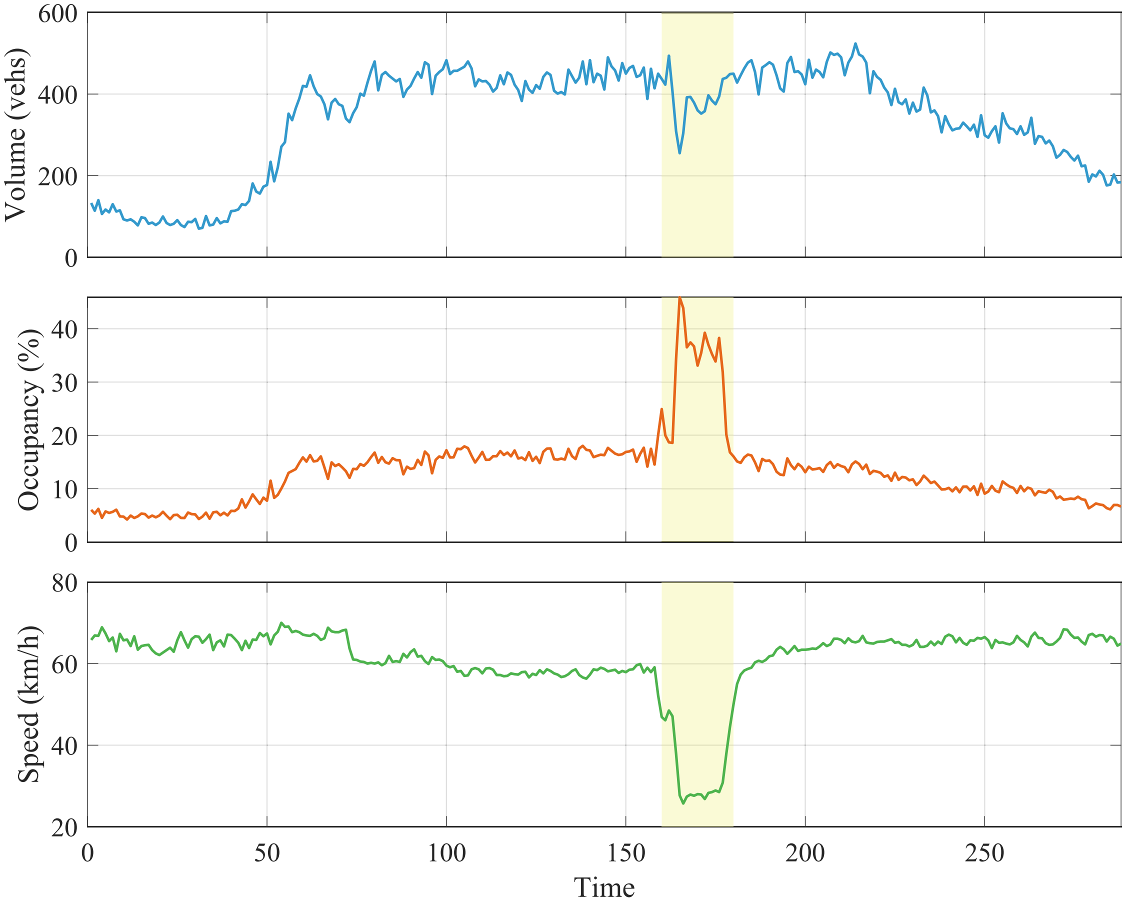

Fig. 1. Relationships among Volume, Occupancy, and Speed for sensor #1 from the PeMS-D8 dataset. The highlighted region (yellow) shows a congestion event, characterized by a sharp decrease in Speed and Volume, and a corresponding spike in Occupancy.

Remark 3.2. The specific tensor orientation of $\mathcal{X}^v \in \mathbb{R}^{N\times D\times T}$ is critical to the t-product formulation $\mathcal{X}^v = \mathcal{X}^v * \mathcal{Z}^v + \mathcal{E}^v$. Because the t-product implicitly applies the FFT along the third dimension, we designate the intra-day time dimension (T) as this mode to align the transformation with the strong daily periodicity of traffic data. This strategy maximizes energy concentration in the frequency domain and facilitates a more effective low-rank representation through $\mathcal{Z}^v$. The resulting performance benefit is demonstrated in Section 5.9.

Two-level cross-view coupling: To achieve a deeper fusion of multi-view traffic data, we propose a two-level coupling strategy that simultaneously captures structural interdependencies in both the traffic background and anomalous events.

To characterize the consistency of underlying traffic patterns, we adopt a low-rank tensor approximation to uncover high-order correlations among the subspace representation tensors $\mathcal{Z}^v$. By constructing the concatenated tensor $\mathcal{Z}$ such that $\mathcal{Z}(:,:,v) = Z^v_{(3)}$, we introduce the joint low-rank penalty $\mathcal{R}_1(\mathcal{Z}) = \mathrm{rank}_t(\mathcal{Z})$. This penalty encourages consistent structural patterns across all $\mathcal{Z}^v$ while avoiding overly rigid magnitude constraints, thereby preserving the distinctive features of each view.

Conventional reconstruction models typically treat the residual term $\mathcal{E}^v$ as independent noise. This assumption neglects the intrinsic spatio-temporal correlations among anomalies. In contrast, traffic flow theory (Zeng et al., 2018) suggests that random sensor noise is idiosyncratic and independent, whereas significant incidents tend to manifest concurrently across multiple traffic views. This physical co-occurrence is visually corroborated in Fig. 1, where a single event induces temporally synchronized anomalies across all sensing streams. Consequently, rather than treating residuals as isolated errors, the second component of our framework focuses on disentangling and synchronizing these coupled anomalies.

To translate this physical intuition into a rigorous mathematical constraint, we consolidate the residuals from all views into a fourth-order tensor $\mathcal{E} \in \mathbb{R}^{N\times D\times T\times V}$, where $\mathcal{E}(:,:,:,v) = \mathcal{E}^v$. To capture the above synchronization pattern, we enforce a coupled sparsity structure by regularizing $\mathcal{E}$ with the following mixed-norm penalty:

$$\mathcal{R}_2(\mathcal{E}) = \|\mathcal{E}\|_{F,0} = \sum_{i,j,k} \|\mathcal{E}(i,j,k,:)\|_F^0.$$

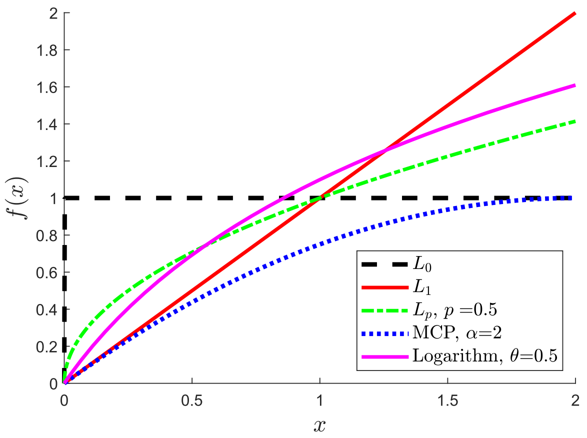

Fig. 2. Comparison of the convex L_1-norm with several nonconvex penalties as surrogates for the L_0-norm.

Minimizing this penalty acts as a group-wise hard-thresholding operator. By evaluating the combined energy of each cross-view vector $\mathcal{E}(i, j, k, :)$, the operator suppresses low-energy components that are characteristic of random representation errors. In contrast, it preserves high-energy vectors corresponding to significant and synchronized anomalies. This structural constraint effectively filters out idiosyncratic noise and keeps the model focused on robust detection of cross-view consistent events. This modeling choice is further supported by the ablation study in Section 5.9.

Generalized nonconvex surrogates: Most existing traffic data analysis frameworks rely solely on nonconvex approximations for low-rank structures. In contrast, our approach applies such nonconvex approximations to both low-rank and sparse components. Furthermore, we employ a general surrogate to approximate both the low-rank and sparsity constraints, formulated as follows:

$$\|\mathcal{Z}\|_{\phi} := \mathcal{R}_1(\mathcal{Z}) = \frac{1}{V}\sum_{k=1}^{V}\sum_{i=1}^{\min\{T,D^2\}} \phi\big(\sigma_i\big(\bar{Z}^{(k)}\big)\big),$$

$$\|\mathcal{E}\|_{F,\psi} := \mathcal{R}_2(\mathcal{E}) = \sum_{i=1}^{N}\sum_{j=1}^{D}\sum_{k=1}^{T} \psi(\|\mathcal{E}(i,j,k,:)\|_F),$$

where $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ and $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ are nonconvex penalty functions. These surrogates provide tighter approximations that enhance the recovery of complex traffic patterns by more closely modeling the true tensor rank and joint sparsity, as illustrated in Fig. 2. Our formulation is versatile and can accommodate various choices for ϕ and ψ, with several notable examples including the following:

(1) L_1 (Tibshirani, 1996): $f^{L_1}(x) = x$;

(2) L_p (Chen et al., 2010; Marjanovic and Solo, 2012): $f^{L_p}(x) = x^p$ with $p \in (0, 1)$;

(3) MCP (Zhang, 2010): $f^{\mathrm{MCP}}(x) = \begin{cases} x - \frac{x^2}{2\alpha}, & 0 \le x \le \alpha; \\ \frac{\alpha}{2}, & x > \alpha \end{cases}$ with $\alpha > 0$;

(4) Logarithm (Friedman, 2012; Gong et al., 2013): $f^{\log}(x) = \log\left(\frac{x}{\theta} + 1\right)$ with $\theta > 0$.

For the specific forms of ϕ and ψ discussed above, their proximal mappings often admit closed-form solutions, which are summarized in Appendix A.

3.2. Spatio-temporal regularization

In transportation networks, traffic data typically exhibit gradual changes across spatial segments, over weeks, and across time intervals. Specifically, there is minimal variation in measured values between adjacent segments, consecutive weeks, and successive time intervals. To quantify this stability, we compute the differences between data points at corresponding positions in adjacent spatial segments, consecutive weeks, and successive time intervals. Let $\mathcal{X}(n, d, t)$ represent the traffic data at segment n, day d, and time interval t. We define the normalized differences as follows:

$$\Delta_N(n,d,t) = \frac{|\mathcal{X}(n,d,t) - \mathcal{X}(n+1,d,t)|}{\max_{n,d,t} |\mathcal{X}(n,d,t) - \mathcal{X}(n+1,d,t)|},$$

$$\Delta_D(n,d,t) = \frac{|\mathcal{X}(n,d,t) - \mathcal{X}(n,d+7,t)|}{\max_{n,d,t} |\mathcal{X}(n,d,t) - \mathcal{X}(n,d+7,t)|},$$

$$\Delta_T(n,d,t) = \frac{|\mathcal{X}(n,d,t) - \mathcal{X}(n,d,t+1)|}{\max_{n,d,t} |\mathcal{X}(n,d,t) - \mathcal{X}(n,d,t+1)|}.$$

We plot the Cumulative Distribution Functions (CDFs) of Δ_N, Δ_D, and Δ_T in Fig. 3. From the figure, we observe that over 80% of the Δ_N values are less than 0.4, approximately 20% of the Δ_D values are exactly 0, and over 90% of the Δ_T values are less than 0.4.

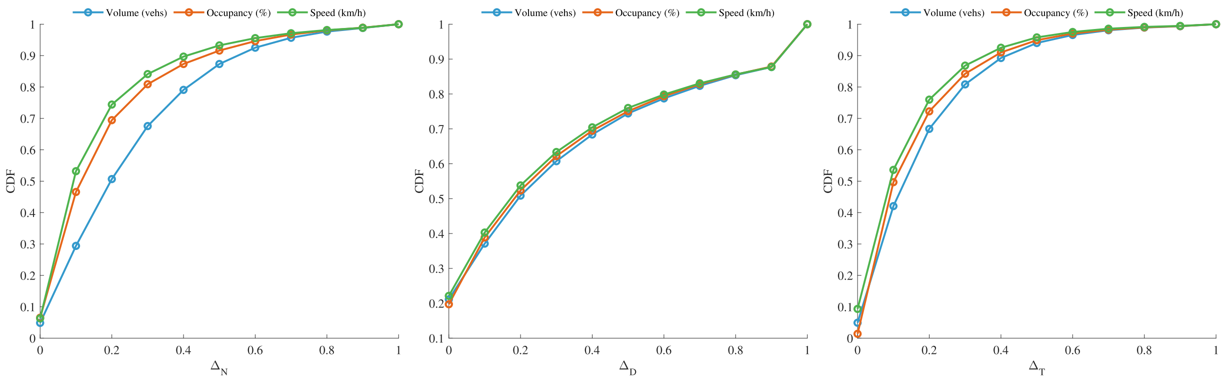

Fig. 3. CDFs of the normalized differences Δ_N, Δ_D, and Δ_T on the PeMS-D8 dataset.

This stability can be captured by the following regularization term:

$$\sum_{v=1}^{V}\sum_{u=1}^{3}\|\mathcal{X}^v \times_u D_u\|_F^2,$$

where D_u for $u \in [3]$ are Toeplitz matrices designed to compute differences along the spatial, weekly, and temporal dimensions, respectively. Specifically, D_1 has 1 on the main diagonal and -1 on the first upper diagonal, capturing differences between adjacent spatial segments; D_2 has 1 on the main diagonal and -1 on the seventh upper diagonal, capturing differences between the same day and time interval in consecutive weeks; and D_3 has 1 on the main diagonal and -1 on the first upper diagonal, capturing differences between consecutive time intervals.

Remark 3.3. In contrast to LRTC-3DST (Shu et al., 2024), which applies the nuclear norm to transformed modes for global low-rankness, we minimize the Frobenius norm of the spatiotemporal differences. This formulation aligns with the statistical stability revealed in Fig. 3, where minimizing local gradients effectively captures data smoothness. Crucially, this approach bypasses the computationally expensive SVD operations required by the nuclear norm. We further differentiate our model by constructing the regularization matrix D_2 to explicitly encode the weekly periodicity characteristic of traffic data.

3.3. Proposed USCMT model

By integrating the discussions from the two preceding subsections, we formulate the overall objective function of the proposed USCMT model as follows:

$$\begin{aligned} \underset{\mathcal{X}^v,\mathcal{Z}^v,\mathcal{E}^v}{\arg\min} \quad & \|\mathcal{Z}\|_\phi + \gamma\|\mathcal{E}\|_{F,\psi} + \tfrac{\lambda}{2}\textstyle\sum_{v=1}^{V}\sum_{u=1}^{3}\|\mathcal{X}^v \times_u D_u\|_F^2 + \sum_{v=1}^{V}\delta_{\mathbb{S}^v}(\mathcal{X}^v) \\ \text{s.t.} \quad & \mathcal{X}^v = \mathcal{X}^v * \mathcal{Z}^v + \mathcal{E}^v,\ v = 1,\dots,V. \end{aligned} \tag{3}$$

Here $\mathcal{X}^v \in \mathbb{R}^{N\times D\times T}$, $\mathcal{Z}^v \in \mathbb{R}^{D\times D\times T}$, $\mathcal{E}^v \in \mathbb{R}^{N\times D\times T}$. Moreover, $\delta_{\mathbb{S}^v}(\mathcal{X}^v)$ denotes the indicator function of the feasible set $\mathbb{S}^v = \{\mathcal{X}^v \mid \mathscr{P}_{\Omega_v}(\mathcal{X}^v) = \mathscr{P}_{\Omega_v}(\mathcal{M}^v)\}$.

Remark 3.4. We note that MVLR (He et al., 2024) imposes an explicit low-rank constraint on $\mathcal{X}^v$. However, given the tensor self-representation $\mathcal{X}^v = \mathcal{X}^v * \mathcal{Z}^v$, it follows from (Zhou et al., 2018, Lemma 2) that $\operatorname{rank}_t(\mathcal{X}^v) = \operatorname{rank}_t(\mathcal{X}^v * \mathcal{Z}^v) \le \operatorname{rank}_t(\mathcal{Z}^v)$. Since the regularization term $\|\mathcal{Z}\|_\phi$ explicitly promotes the low-rank structure of $\mathcal{Z}^v$, the low-rankness of $\mathcal{X}^v$ is implicitly enforced through this upper bound. Therefore, imposing an additional low-rank constraint on $\mathcal{X}^v$ is redundant in our model and is omitted to avoid unnecessary computational complexity. This design choice is further supported by the ablation results reported in Section 5.9, where adding an explicit low-rank regularization on $\mathcal{X}^v$ brings only marginal reconstruction gains.

4. Proposed algorithm and convergence analysis

In this section, we develop an Alternating Direction Method of Multipliers (ADMM) algorithm to solve the USCMT model and then establish its convergence properties.

4.1. ADMM algorithm

By introducing the auxiliary tensor variables $\mathcal{G}$, $\mathcal{Y}$, and $\mathcal{K}$, problem (3) can be rewritten as:

$$\begin{aligned} \underset{\mathcal{X}^v,\mathcal{Z}^v,\mathcal{E}^v}{\arg\min} \quad & \|\mathcal{G}\|_\phi + \gamma\|\mathcal{E}\|_{F,\psi} + \tfrac{\lambda}{2}\textstyle\sum_{v=1}^{V}\sum_{u=1}^{3}\|\mathcal{K}^v \times_u D_u\|_F^2 + \sum_{v=1}^{V}\delta_{\mathbb{S}^v}(\mathcal{X}^v) \\ \text{s.t.} \quad & \mathcal{X}^v = \mathcal{Y}^v * \mathcal{Z}^v + \mathcal{E}^v,\ \mathcal{G} = \mathcal{Z},\ \mathcal{Y} = \mathcal{X},\ \mathcal{K} = \mathcal{X}. \end{aligned} \tag{4}$$

Then the augmented Lagrangian function of (4) is

$$\mathscr{L}_{\rho_i}(\mathcal{X},\mathcal{Z},\mathcal{E},\mathcal{G},\mathcal{Y},\mathcal{K};\mathcal{P},\mathcal{Q},\mathcal{R},\mathcal{J}) = \|\mathcal{G}\|_\phi + \gamma\|\mathcal{E}\|_{F,\psi} + \frac{\lambda}{2}\sum_{v=1}^{V}\sum_{u=1}^{3}\|\mathcal{K}^v \times_u D_u\|_F^2 + \sum_{v=1}^{V}\delta_{\mathbb{S}^v}(\mathcal{X}^v)$$
$$+\sum_{v=1}^{V}\Big(\langle\mathcal{P}^v,\mathcal{X}^v-\mathcal{Y}^v*\mathcal{Z}^v-\mathcal{E}^v\rangle+\frac{\rho_1}{2}\|\mathcal{X}^v-\mathcal{Y}^v*\mathcal{Z}^v-\mathcal{E}^v\|_F^2\Big)+\langle\mathcal{Q},\mathcal{G}-\mathcal{Z}\rangle+\frac{\rho_2}{2}\|\mathcal{G}-\mathcal{Z}\|_F^2$$
$$+\langle\mathcal{R},\mathcal{Y}-\mathcal{X}\rangle+\frac{\rho_3}{2}\|\mathcal{Y}-\mathcal{X}\|_F^2+\langle\mathcal{J},\mathcal{K}-\mathcal{X}\rangle+\frac{\rho_4}{2}\|\mathcal{K}-\mathcal{X}\|_F^2,$$

where $\rho_i > 0$ for $i \in [4]$ are penalty parameters, and $\mathcal{P}$, $\mathcal{Q}$, $\mathcal{R}$, and $\mathcal{J}$ are Lagrange multipliers. Within the ADMM framework, we update each variable alternately:

$$\begin{cases}
\mathcal{X}^{l+1} \in \operatorname{argmin}_{\mathcal{X}} \mathscr{L}_{\rho_i^l}\big(\mathcal{X},\mathcal{Z}^l,\mathcal{E}^l,\mathcal{G}^l,\mathcal{Y}^l,\mathcal{K}^l;\mathcal{P}^l,\mathcal{Q}^l,\mathcal{R}^l,\mathcal{J}^l\big);\\
\mathcal{Z}^{l+1} \in \operatorname{argmin}_{\mathcal{Z}} \mathscr{L}_{\rho_i^l}\big(\mathcal{X}^{l+1},\mathcal{Z},\mathcal{E}^l,\mathcal{G}^l,\mathcal{Y}^l,\mathcal{K}^l;\mathcal{P}^l,\mathcal{Q}^l,\mathcal{R}^l,\mathcal{J}^l\big);\\
\mathcal{E}^{l+1} \in \operatorname{argmin}_{\mathcal{E}} \mathscr{L}_{\rho_i^l}\big(\mathcal{X}^{l+1},\mathcal{Z}^{l+1},\mathcal{E},\mathcal{G}^l,\mathcal{Y}^l,\mathcal{K}^l;\mathcal{P}^l,\mathcal{Q}^l,\mathcal{R}^l,\mathcal{J}^l\big);\\
\mathcal{G}^{l+1} \in \operatorname{argmin}_{\mathcal{G}} \mathscr{L}_{\rho_i^l}\big(\mathcal{X}^{l+1},\mathcal{Z}^{l+1},\mathcal{E}^{l+1},\mathcal{G},\mathcal{Y}^l,\mathcal{K}^l;\mathcal{P}^l,\mathcal{Q}^l,\mathcal{R}^l,\mathcal{J}^l\big);\\
\mathcal{Y}^{l+1} \in \operatorname{argmin}_{\mathcal{Y}} \mathscr{L}_{\rho_i^l}\big(\mathcal{X}^{l+1},\mathcal{Z}^{l+1},\mathcal{E}^{l+1},\mathcal{G}^{l+1},\mathcal{Y},\mathcal{K}^l;\mathcal{P}^l,\mathcal{Q}^l,\mathcal{R}^l,\mathcal{J}^l\big);\\
\mathcal{K}^{l+1} \in \operatorname{argmin}_{\mathcal{K}} \mathscr{L}_{\rho_i^l}\big(\mathcal{X}^{l+1},\mathcal{Z}^{l+1},\mathcal{E}^{l+1},\mathcal{G}^{l+1},\mathcal{Y}^{l+1},\mathcal{K};\mathcal{P}^l,\mathcal{Q}^l,\mathcal{R}^l,\mathcal{J}^l\big),
\end{cases}$$

and

$$\begin{cases}
\mathcal{P}^{v,l+1} = \mathcal{P}^{v,l} + \rho_1^l\big(\mathcal{X}^{v,l+1} - \mathcal{Y}^{v,l+1} * \mathcal{Z}^{v,l+1} - \mathcal{E}^{v,l+1}\big);\\
\mathcal{Q}^{l+1} = \mathcal{Q}^l + \rho_2^l\big(\mathcal{G}^{l+1} - \mathcal{Z}^{l+1}\big);\\
\mathcal{R}^{l+1} = \mathcal{R}^l + \rho_3^l\big(\mathcal{Y}^{l+1} - \mathcal{X}^{l+1}\big);\\
\mathcal{J}^{l+1} = \mathcal{J}^l + \rho_4^l\big(\mathcal{K}^{l+1} - \mathcal{X}^{l+1}\big);\\
\rho_i^{l+1} = \beta_i\rho_i^l.
\end{cases} \tag{5}$$

Here, l denotes the iteration index, and each $\beta_i > 1$ for $i \in [4]$ is a constant. In the following, we detail the solution strategy for each subproblem.

In the detailed iteration that follows, we drop the superscript l from the iteration updates for notational simplicity and to enhance readability.

Computing $\mathcal{X}$: The sub-problem to update $\mathcal{X}^v$ is

$$\begin{aligned}\underset{\mathcal{X}^v}{\arg\min}\quad &\langle\mathcal{P}^v,\mathcal{X}^v-\mathcal{Y}^v*\mathcal{Z}^v-\mathcal{E}^v\rangle+\tfrac{\rho_1}{2}\|\mathcal{X}^v-\mathcal{Y}^v*\mathcal{Z}^v-\mathcal{E}^v\|_F^2+\langle\mathcal{R}^v,\mathcal{Y}^v-\mathcal{X}^v\rangle\\ &+\tfrac{\rho_3}{2}\|\mathcal{Y}^v-\mathcal{X}^v\|_F^2+\langle\mathcal{J}^v,\mathcal{K}^v-\mathcal{X}^v\rangle+\tfrac{\rho_4}{2}\|\mathcal{K}^v-\mathcal{X}^v\|_F^2+\delta_{\mathbb{S}^v}(\mathcal{X}^v).\end{aligned} \tag{6}$$

Clearly, the optimal solution to the $\mathcal{X}^v$ sub-problem is given by

$$\mathcal{X}^v = \mathscr{P}_{\Omega_v^c}\left(\frac{\rho_1\mathcal{Y}^v*\mathcal{Z}^v+\rho_1\mathcal{E}^v-\mathcal{P}^v+\rho_3\mathcal{Y}^v+\mathcal{R}^v+\rho_4\mathcal{K}^v+\mathcal{J}^v}{\rho_1+\rho_3+\rho_4}\right)+\mathscr{P}_{\Omega_v}(\mathcal{M}^v). \tag{7}$$

Computing $\mathcal{Z}$: The sub-problem to update $\mathcal{Z}^v$ is

$$\underset{\mathcal{Z}^v}{\arg\min}\ \langle\mathcal{P}^v,\mathcal{X}^v-\mathcal{Y}^v*\mathcal{Z}^v-\mathcal{E}^v\rangle+\frac{\rho_1}{2}\|\mathcal{X}^v-\mathcal{Y}^v*\mathcal{Z}^v-\mathcal{E}^v\|_F^2+\langle\mathcal{Q}^v,\mathcal{G}^v-\mathcal{Z}^v\rangle+\frac{\rho_2}{2}\|\mathcal{G}^v-\mathcal{Z}^v\|_F^2. \tag{8}$$

By taking the gradient of the above objective with respect to $\mathcal{Z}^v$ and setting it to zero, we obtain the update for $\mathcal{Z}^v$ by solving the following system of linear equations:

$$\Big(\rho_1\mathcal{Y}^{vT}*\mathcal{Y}^v+\rho_2\mathcal{I}\Big)*\mathcal{Z}^v=\mathcal{Y}^{vT}*\big(\rho_1\mathcal{X}^v-\rho_1\mathcal{E}^v+\mathcal{P}^v\big)+\rho_2\mathcal{G}^v+\mathcal{Q}^v. \tag{9}$$

Computing $\mathcal{E}$: The sub-problem to update $\mathcal{E}$ is

$$\begin{aligned}&\underset{\mathcal{E}}{\arg\min}\ \ \gamma\|\mathcal{E}\|_{F,\psi}+\langle\mathcal{P},\mathcal{X}-\mathcal{Y}*\mathcal{Z}-\mathcal{E}\rangle+\frac{\rho_1}{2}\|\mathcal{X}-\mathcal{Y}*\mathcal{Z}-\mathcal{E}\|_F^2\\ =&\underset{\mathcal{E}}{\arg\min}\ \ \gamma\|\mathcal{E}\|_{F,\psi}+\frac{\rho_1}{2}\|\mathcal{X}-\mathcal{Y}*\mathcal{Z}-\mathcal{E}+\frac{\mathcal{P}}{\rho_1}\|_F^2.\end{aligned} \tag{10}$$

Here $[\mathcal{Y}*\mathcal{Z}](:,:,:,v) := \mathcal{Y}^v*\mathcal{Z}^v$. By (Yu and Bai, 2024, Theorem 4.1), the optimal solution of (10) is given by

$$\mathcal{E}(i,j,k,:)=\begin{cases}\operatorname{Prox}_{\gamma/\rho_1}\psi\big(\hat{e}_{ijk}\big)\hat{\mathcal{E}}(i,j,k,:)/\hat{e}_{ijk}, & \hat{e}_{ijk}\neq 0;\\ 0, & \hat{e}_{ijk}=0,\end{cases} \tag{11}$$

where $\hat{\mathcal{E}}=\mathcal{X}-\mathcal{Y}*\mathcal{Z}+\mathcal{P}/\rho_1$ and $\hat{e}_{ijk}=\|\hat{\mathcal{E}}(i,j,k,:)\|_F$.

Computing $\mathcal{G}$: The sub-problem to update $\mathcal{G}$ is

$$\begin{aligned}&\arg\min_{\mathcal{G}} \|\mathcal{G}\|_{\phi} + \langle \mathcal{Q}, \mathcal{G} - \mathcal{Z} \rangle + \frac{\rho_2}{2}\|\mathcal{G} - \mathcal{Z}\|_F^2 \\ =&\arg\min_{\mathcal{G}} \|\mathcal{G}\|_{\phi} + \frac{\rho_2}{2}\|\mathcal{G} - \mathcal{Z} + \frac{\mathcal{Q}}{\rho_2}\|_F^2.\end{aligned} \tag{12}$$

Denote the t-SVD of $\mathcal{Z} - \mathcal{Q}/\rho_2$ as $\mathcal{U} * \mathcal{C} * \mathcal{V}^T$. Then, by (Qiu et al., 2021, Theorem 2), the update (12) admits the closed-form solution

$$\mathcal{G} = \mathcal{U} * \mathcal{F} * \mathcal{V}^T, \tag{13}$$

where $\mathcal{F}$ is a tensor such that $\bar{F}_{i,i}^{(k)} = \mathrm{Prox}_{1/\rho_2}\,\phi(\bar{C}_{i,i}^{(k)})$ and $\bar{F}_{i,j}^{(k)} = 0,\ i \neq j$.

Computing $\mathcal{Y}$: The sub-problem to update $\mathcal{Y}^v$ is

$$\arg\min_{\mathcal{Y}^v} \langle \mathcal{P}^v, \mathcal{X}^v - \mathcal{Y}^v * \mathcal{Z}^v - \mathcal{E}^v \rangle + \frac{\rho_1}{2}\|\mathcal{X}^v - \mathcal{Y}^v * \mathcal{Z}^v - \mathcal{E}^v\|_F^2 + \langle \mathcal{R}^v, \mathcal{Y}^v - \mathcal{X}^v \rangle + \frac{\rho_3}{2}\|\mathcal{Y}^v - \mathcal{X}^v\|_F^2. \tag{14}$$

By differentiating the objective above with respect to $\mathcal{Y}^v$ and enforcing a zero gradient, we obtain the update for $\mathcal{Y}^v$ by solving the following system of linear equations:

$$\mathcal{Y}^v * \left(\rho_1 \mathcal{Z}^v * \mathcal{Z}^{vT} + \rho_3 \mathcal{I}\right) = (\rho_1 \mathcal{X}^v - \rho_1 \mathcal{E}^v + \mathcal{P}^v) * \mathcal{Z}^{vT} + \rho_3 \mathcal{X}^v - \mathcal{R}^v. \tag{15}$$

Computing $\mathcal{K}$: The sub-problem to update $\mathcal{K}^v$ is

$$\arg\min_{\mathcal{K}^v} \ \tfrac{\lambda}{2} \textstyle\sum_{u=1}^{3} \|\mathcal{K}^v \times_u D_u\|_F^2 + \langle \mathcal{J}^v, \mathcal{K}^v - \mathcal{X}^v \rangle + \tfrac{\rho_4}{2}\|\mathcal{K}^v - \mathcal{X}^v\|_F^2. \tag{16}$$

Taking the derivative for (16) with respect to $\mathcal{K}^v$ and equating the result to zero, we obtain the linear system

$$\rho_4 \mathcal{K}^v + \lambda \sum_{u=1}^{3} \mathcal{K}^v \times_u (D_u^T D_u) = \rho_4 \mathcal{X}^v - \mathcal{J}^v.$$

Applying Lemma 1 in Shu et al. (2024) yields the closed-form update

$$\mathcal{K}^v = \frac{(\rho_4 \mathcal{X}^v - \mathcal{J}^v) \times_1 V_1^T \times_2 V_2^T \times_3 V_3^T}{\rho_4 \mathbf{1} + \lambda \sum_{u=1}^{3} \mathbf{1} \times_u S_u} \times_1 V_1 \times_2 V_2 \times_3 V_3, \tag{17}$$

where $\mathbf{1}$ is a tensor with all entries as 1, and each $D_u^T D_u = V_u S_u V_u^T$ is the spectral decomposition defining S_u and V_u.

We summarize the algorithm for USCMT in Algorithm 4.1.

Algorithm 4.1 The ADMM-based solver for the proposed USCMT model.

Input: Incomplete multi-view traffic dataset $\mathscr{P}_{\Omega_v}(\mathcal{M}^v)$, regularization parameters γ and λ.

Output: Multi-view traffic tensor $\mathcal{X} = [\mathcal{X}^1, \ldots, \mathcal{X}^V]$ and event detection tensor $\mathcal{E} = [\mathcal{E}^1, \ldots, \mathcal{E}^V]$.

```
l ← 0;
while not converged do
    Update X^{l+1} by computing (7);
    Update Z^{l+1} by solving the linear system in (9);
    Update E^{l+1} by computing (11);
    Update G^{l+1} by computing (13);
    Update Y^{l+1} by solving the linear system in (15);
    Update K^{l+1} by computing (17);
    Update Lagrangian multipliers and the penalty parameter via (5);
    Check the convergence conditions
```

$$max\left\{ \frac{\|\mathcal{X}^{l+1} - \mathcal{Y}^{l+1} * \mathcal{Z}^{l+1} - \mathcal{E}^{l+1}\|_F}{\|\mathcal{X}^{l+1}\|_F + \|\mathcal{Y}^{l+1} * \mathcal{Z}^{l+1}\|_F + \|\mathcal{E}^{l+1}\|_F}, \frac{\|\mathcal{G}^{l+1} - \mathcal{Z}^{l+1}\|_F}{\|\mathcal{G}^{l+1}\|_F + \|\mathcal{Z}^{l+1}\|_F}, \frac{\|\mathcal{Y}^{l+1} - \mathcal{X}^{l+1}\|_F}{\|\mathcal{Y}^{l+1}\|_F + \|\mathcal{X}^{l+1}\|_F} \right\} < \epsilon;$$

```
    l ← l + 1;
end
```

4.2. Convergence analysis

This subsection provides a theoretical guarantee for the convergence of our proposed Algorithm 4.1. To simplify the notation, we let $\mathcal{W} = (\mathcal{W}_1, \mathcal{W}_2)$, where $\mathcal{W}_1 = (\mathcal{X}, \mathcal{Z}, \mathcal{E}, \mathcal{G}, \mathcal{Y}, \mathcal{K})$ and $\mathcal{W}_2 = (\mathcal{P}, \mathcal{Q}, \mathcal{R}, \mathcal{J})$. Furthermore, set $\Delta\mathcal{W}^l = \mathcal{W}^{l+1} - \mathcal{W}^l$.

Lemma 4.1. *Let $\{\mathcal{W}^l\}_{l=1}^{\infty}$ be the sequence generated by Algorithm 4.1, then*

$$\mathscr{L}_{\rho_i^{l+1}}(\mathcal{W}^{l+1}) - \mathscr{L}_{\rho_i^l}(\mathcal{W}^l) \le -\frac{\tau}{2}\|\mathcal{W}_0^{l+1} - \mathcal{W}_0^l\|_F^2 + \frac{1+\beta_1}{2\rho_1^l}\|\Delta\mathcal{P}^l\|_F^2 + \frac{1+\beta_2}{2\rho_2^l}\|\Delta\mathcal{Q}^l\|_F^2 + \frac{1+\beta_3}{2\rho_3^l}\|\Delta\mathcal{R}^l\|_F^2 + \frac{1+\beta_4}{2\rho_4^l}\|\Delta\mathcal{J}^l\|_F^2, \tag{18}$$

where $\mathcal{W}_0 = (\mathcal{X}, \mathcal{Z}, \mathcal{Y}, \mathcal{K})$ and $\tau = \min_{i\in[4]} \rho_i^0$.

Lemma 4.2. *Suppose the sequences $\{\mathcal{W}_2^l\}_{l=1}^{\infty}$ and $\{\mathcal{X}^l\}_{l=1}^{\infty}$ produced by Algorithm 4.1 are bounded. Then the following statements hold:*

(1) The full sequence $\{\mathcal{W}^l\}_{l=1}^{\infty}$ is bounded.
(2) $\lim_{l\to\infty} \|\mathcal{W}_0^{l+1} - \mathcal{W}_0^l\|_F = 0$, where $\mathcal{W}_0 = (\mathcal{X}, \mathcal{Z}, \mathcal{Y}, \mathcal{K})$.

The following principal conclusion is established concerning the convergence of Algorithm 4.1 applied to problem (4).

Theorem 4.1. *Let $\{\mathcal{W}^l\}_{l=1}^{\infty}$ be the sequence generated by Algorithm 4.1. If the sequences $\{\mathcal{W}_2^l\}_{l=1}^{\infty}$ and $\{\mathcal{X}^l\}_{l=1}^{\infty}$ are bounded, then every accumulation point of $\{\mathcal{W}^l\}_{l=1}^{\infty}$ satisfies the Karush–Kuhn–Tucker (KKT) conditions of problem (4).*

4.3. Computation complexity

We now analyze the computational complexity of Algorithm 4.1. During each iteration, the update of $\mathcal{X}^v$ requires computing the t-product, which has a computational complexity of $\mathcal{O}(D(N+D)T\log T + D^2NT)$. For updating $\mathcal{Z}^v$ and $\mathcal{Y}^v$, the primary computational costs are associated with the t-product and an inverse operation, each costing $\mathcal{O}(D(N+D)T\log T + D^2NT + D^3T)$. The update of $\mathcal{E}$ mainly involves t-product computations, resulting in a complexity of $\mathcal{O}(D(N+D)T\log T + D^2NT)$. The computation of $\mathcal{G}$ primarily involves the t-SVD, which has a time complexity of $\mathcal{O}(V(D^2T\log T + D^3T))$. The computation of $\mathcal{K}^v$ involves the mode product, componentwise division, and spectral decomposition, with a total time complexity of $\mathcal{O}(NDT(N+D+T) + N^3 + D^3 + T^3)$. Therefore, the overall computational complexity per iteration of Algorithm 4.1 is $\mathcal{O}(D^2T\log T + NDT(N+D+T))$, assuming $N \le DT$, $D \le NT$, and $T \le ND$.

5. Experiments

In this section, we comprehensively evaluate the proposed USCMT model on several real-world datasets through both quantitative and qualitative experiments. The evaluation focuses on three aspects: traffic data reconstruction, event detection, and the joint performance of both tasks. All experiments were conducted in MATLAB 2022a on a system equipped with an Intel Core i5-12500H CPU (2.50 GHz) and 16 GB of RAM.

5.1. Traffic datasets

We evaluate the performance of USCMT on three real-world multi-view traffic datasets. To ensure a consistent scale across views, we normalize the occupancy data from its original [0, 1] range to [0, 100]. The datasets are described below.

- **PeMS-D8:** Provided by the Caltrans Performance Measurement System (PeMS) [1], this dataset contains traffic volume, occupancy, and speed data from 170 sensors in District 8 (San Bernardino, USA). The data span July 1 to August 31, 2016, with a 5-min temporal resolution. We use a 15-day subset in our experiments, forming a tensor of size $(170 \times 15 \times 288) \times 3$.
- **PeMS-D4:** Also from PeMS, this dataset records traffic metrics from 307 sensors in District 4 (San Francisco Bay Area, USA) between January 1 and February 28, 2018, at a 5-min resolution. We select a 15-day period for evaluation, resulting in a tensor of size $(307 \times 15 \times 288) \times 3$.
- **FT-AED:** The Freeway Traffic Anomalous Event Detection (FT-AED) dataset (Coursey et al., 2024) contains traffic volume, occupancy, and speed data from 49 sensor locations along the four-lane Interstate 24 (I-24) corridor in Tennessee. The data are recorded on a per-lane basis at a high temporal resolution of 30 sec. They cover the morning peak period (4:00 AM–12:00 PM) across 20 workdays in October 2023. The resulting tensor has dimensions $(49 \times 20 \times 960) \times 3$. A key advantage of this dataset is its detailed two-tier ground-truth labeling system. It provides an official "crash record," a binary flag indicating whether a crash was formally reported at a given timestamp. In addition, it includes a "human label," derived from reports compiled by officers responsible for monitoring road conditions and identifying a broader range of unusual events.

The distinct characteristics of these datasets support a multi-faceted evaluation of our model. Because PeMS-D4 and PeMS-D8 do not contain official anomaly labels, we use them to benchmark traffic data reconstruction performance. In contrast, FT-AED provides verified event logs and therefore serves as the primary benchmark for evaluating anomaly detection accuracy and, more importantly, the synergistic performance of our unified reconstruction and detection framework.

[1] https://pems.dot.ca.gov/

5.2. Baseline models

To provide a comprehensive evaluation, we benchmark USCMT against nine representative baseline models.

- BGCP (Bayesian Gaussian CP decomposition Chen et al., 2019b). A fully Bayesian tensor factorization model that employs Markov Chain Monte Carlo (MCMC) to infer latent factor matrices representing the underlying low-rank structure of the data.
- BATF (Bayesian Augmented Tensor Factorization Chen et al., 2019a). An augmented tensor factorization model that leverages generic forms of domain knowledge from transportation systems.
- LRTC-TNN (Low-Rank Tensor Completion with Truncation Nuclear Norm Chen et al., 2020). A low-rank tensor completion method that utilizes Truncated Nuclear Norm (TNN) minimization to better preserve the principal low-rank patterns within traffic data.
- LATC (Low-rank Autoregressive Tensor Completion Chen et al., 2022). A low-rank tensor completion model that integrates an autoregressive process to explicitly model and enforce local temporal consistency.
- iForest (Isolation Forest Liu et al., 2012). An anomaly detection algorithm that identifies outliers based on their susceptibility to isolation, without requiring distance or density calculations.
- MVLR (Multi-View learning and Low-rank Representation He et al., 2024). A multi-view recovery framework that models intra-view patterns using a low-rank representation while capturing inter-view consistency through a shared subspace tensor built from single-view matrix self-representations.
- RTC (Robust Tensor Recovery Hu and Work, 2020). A model that formulates the recovery task as the decomposition of the observed data tensor into a low-rank component and a fiber-sparse corruption component. This decomposition enables the simultaneous imputation of missing data and the detection of extreme events.
- MLP (Multi-Layer Perceptron autoencoder Coursey et al., 2024). A standard autoencoder that processes the features of each node as an independent sample.
- IGNNK (Inductive Graph Neural Network Kriging Wu et al., 2021). An inductive graph neural network framework designed for spatiotemporal kriging that recovers unobserved signals by aggregating information from neighboring nodes based on the underlying graph topology.

5.3. Performance metrics

To quantitatively evaluate all models, we adopt a set of widely used metrics for the two primary tasks: data reconstruction and event detection.

Metrics for data reconstruction. Reconstruction accuracy is measured by the Mean Absolute Percentage Error (MAPE) and the Root Mean Square Error (RMSE), defined as follows:

$$\mathrm{MAPE} = \frac{1}{|\Omega^c|} \sum_{(ijk)\in\Omega^c} \left| \frac{\mathcal{X}_{ijk} - \hat{\mathcal{X}}_{ijk}}{\mathcal{X}_{ijk}} \right| \times 100\%, \qquad \mathrm{RMSE} = \sqrt{\frac{1}{|\Omega^c|} \sum_{(ijk)\in\Omega^c} \left(\mathcal{X}_{ijk} - \hat{\mathcal{X}}_{ijk}\right)^2},$$

where Ω^c denotes the set of missing-entry indices, and $\mathcal{X}$ and $\hat{\mathcal{X}}$ denote the ground-truth tensor and the reconstructed tensor, respectively.

Metrics for event detection. Event detection performance is evaluated from three aspects: the rate of missed events, the overall classification ability, and the timeliness of detection. For USCMT, the event score is computed by aggregating the multi-view anomaly tensor along the view dimension, i.e., $\sum_{v=1}^{V} |\mathcal{E}^v|$. Therefore, the reported detection metrics are identical across the three view blocks.

First, to assess the model's reliability in identifying critical events, we compute the miss percentage. It is defined as the percentage of actual crashes that the model fails to detect as anomalies within a 15-min window after the official report time:

$$\text{Miss Percentage} = \frac{\text{number of missed crashes}}{\text{total number of crashes}} \times 100\%.$$

Second, to measure overall classification performance across all possible thresholds, we use the Receiver Operating Characteristic (ROC) curve and report the Area Under the Curve (AUC). A higher AUC indicates a better trade-off between detecting true positives and avoiding false positives.

Finally, to measure timeliness, we use the detection delay, defined as the time difference between the model's detection time and the official incident report time. A negative value indicates desirable early detection.

$$\text{Detection Delay} = t_{\text{detected}} - t_{\text{incident}},$$

where t_{detected} is the timestamp of the model's detection and t_{incident} is the official report time of the incident.

5.4. Performance on data reconstruction

In this subsection, we systematically evaluate the reconstruction performance of USCMT. To assess its capability, we compare it with several baselines under complementary data loss scenarios, including one random pattern and two structured patterns:

Table 1
Performance comparison for data reconstruction on RM scenarios. Each result cell for the baseline models displays performance in the format of MAPE / RMSE. The running time is also reported for each method.

Dataset	Scenario	View / Metric	BGCP	BATF	LRTC-TNN	LATC	IGNNK	MVLR	USCMT
PeMS-D8	RM-40%	Volume	11.05 / 25.97	12.14 / 26.36	8.86 / 21.90	8.62 / **20.87**	14.26/28.83	8.74 / 21.57	**8.58** / 20.94
		Occupancy	15.43 / 1.71	16.51 / 1.55	10.76 / 1.21	10.63 / 1.14	20.69/1.54	10.83 / 1.18	**9.64 / 1.07**
		Speed	2.93 / 2.94	3.01 / 2.97	1.79 / 1.86	1.61 / 1.67	2.56/2.43	1.71 / 1.80	**1.28 / 1.41**
		Time (s)	257.45	215.63	96.19	247.96	74.12	69.83	**16.27**
	RM-60%	Volume	10.85 / 26.35	12.59 / 27.49	9.19 / 23.38	8.85 / 21.74	17.74/33.71	9.08 / 22.78	**8.65 / 21.44**
		Occupancy	15.49 / 1.74	16.67 / 1.96	11.54 / 1.34	11.43 / 1.25	25.64/1.64	12.17 / 1.41	**9.84 / 1.12**
		Speed	3.01 / 3.07	3.04 / 3.04	2.15 / 2.30	1.94 / 2.05	3.55/3.13	2.07 / 2.22	**1.49 / 1.69**
		Time (s)	191.14	212.20	92.40	198.14	63.21	86.09	**18.24**
	RM-80%	Volume	11.31 / 27.60	13.53 / 28.57	10.37 / 26.89	9.91 / 24.30	25.65/40.41	10.65 / 26.51	**9.38 / 23.60**
		Occupancy	17.82 / 2.64	18.34 / 2.46	12.87 / 1.53	13.45 / 1.51	32.94/2.05	16.71 / 2.40	**10.86 / 1.37**
		Speed	3.34 / 3.63	3.33 / 3.43	2.72 / 2.87	2.55 / 2.67	4.85/4.13	2.82 / 2.91	**2.02 / 2.27**
		Time (s)	164.77	165.07	75.70	192.28	62.46	57.08	**24.71**
PeMS-D4	RM-40%	Volume	16.33 / 32.18	17.32 / 32.96	12.90 / 26.54	12.68 / 25.88	19.06/36.14	12.89 / 26.43	**12.10 / 25.79**
		Occupancy	27.20 / 1.71	27.31 / 1.72	16.61 / 1.35	16.33 / 1.29	29.74/1.58	17.00 / 1.32	**14.79 / 1.19**
		Speed	3.47 / 3.18	3.56 / 3.23	2.16 / 2.01	2.01 / 1.87	3.22/2.92	2.04 / 1.92	**1.55 / 1.50**
		Time (s)	302.49	328.24	92.01	268.83	123.79	125.52	**27.86**
	RM-60%	Volume	16.54 / 32.20	18.15 / 33.32	13.44 / 27.62	13.16 / 26.70	27.85/42.00	13.48 / 27.43	**12.34 / 26.36**
		Occupancy	27.51 / 1.75	26.47 / 1.80	17.90 / 1.45	17.57 / 1.39	37.14/1.92	18.56 / 1.48	**15.23 / 1.26**
		Speed	3.50 / 3.50	3.61 / 3.28	2.54 / 2.38	2.37 / 2.23	4.72/3.90	2.41 / 2.29	**1.75 / 1.75**
		Time (s)	292.05	279.69	100.94	304.73	78.04	107.99	**31.47**
	RM-80%	Volume	17.03 / 32.95	18.74 / 34.65	14.67 / 30.63	14.31 / 28.69	27.36/43.99	15.18 / 30.09	**13.46 / 28.10**
		Occupancy	30.26 / 2.21	30.33 / 2.13	20.12 / 1.71	20.34 / 1.67	47.37/2.19	24.15 / 2.26	**17.89 / 1.54**
		Speed	3.68 / 3.37	3.73 / 3.40	3.19 / 2.98	3.00 / 2.83	5.67/4.55	3.17 / 2.96	**2.42 / 2.41**
		Time (s)	297.09	309.72	97.57	320.67	119.91	105.58	**37.35**

Best results are highlighted in bold fonts.

- Random Missing (RM): This scenario simulates sporadic data loss, such as that caused by individual sensor malfunctions or transmission errors.
- Cross Detector Missing (CDM): This non-random scenario simulates structured and systematic data loss, where the same time-feature points are missing across all detectors. It represents network-wide data corruption or failure of a common data aggregation unit.
- Continuous Temporal Missing (CTM): This non-random scenario simulates structured and systematic data loss, where selected detector-day pairs contain a consecutive missing time block with a random starting time, to mimic temporary sensor outages or communication interruptions.

The detailed results for each setting are presented below.

5.4.1. *Reconstruction results on RM data*

Table 1 reports the reconstruction performance of USCMT and the baseline methods under RM scenarios with different missing rates. USCMT outperforms all baselines in both MAPE and RMSE in nearly every test case. The only exception is the PeMS-D8 RM-40% scenario, where LATC achieves a slightly lower RMSE for Volume (20.87) than USCMT (20.94). Overall, these results validate the robustness and effectiveness of the proposed model.

As expected, reconstruction errors for all methods generally increase as the missing rate rises, except in a few isolated cases. This trend reflects the increasing difficulty of the imputation task. Among the baselines, LATC and MVLR are often the most competitive, especially at lower missing rates. However, the performance gap between these methods and USCMT widens as data loss becomes more severe, highlighting the stronger robustness of USCMT under challenging conditions.

In addition to its strong reconstruction performance, USCMT also demonstrates notable computational efficiency. It consistently records the shortest running times and is often an order of magnitude faster than baselines such as BGCP, BATF, and LATC. This makes it a practical choice for large-scale traffic data reconstruction under random missing conditions.

5.4.2. *Reconstruction results on CDM data*

We next evaluate model performance under the more challenging CDM scenarios, with detailed results reported in Table 2. USCMT achieves the lowest MAPE and RMSE values across all three CDM test cases on both datasets. The advantage of USCMT is even more pronounced under these structured missing data settings than under the RM scenarios. For example, in the PeMS-D4 CDM-80% scenario, USCMT attains a Volume MAPE of 13.54, substantially outperforming the next-best model, LATC, which records a MAPE of 20.15. This contrasts with the corresponding RM-80% scenario, where the performance gap between the two models is relatively small.

Table 2
Performance comparison for data reconstruction on CDM scenarios. Each result cell for the baseline models displays performance in the format of MAPE / RMSE. The running time is also reported for each method.

Dataset	Scenario	View / Metric	BGCP	BATF	LRTC-TNN	LATC	IGNNK	MVLR	USCMT
PeMS-D8	CDM-40%	Volume	11.60 / 27.86	13.31 / 27.48	10.19 / 24.06	9.39 / 22.26	22.48/47.95	14.27 / 34.34	**8.50 / 20.84**
		Occupancy	15.79 / 1.59	16.89 / 1.49	12.40 / 1.24	11.64 / 1.14	31.72/1.91	16.10 / 1.47	**9.74 / 1.04**
		Speed	3.24 / 3.15	3.08 / 3.03	2.08 / 2.10	1.76 / 1.79	4.64/3.90	2.37 / 2.38	**1.33 / 1.46**
		Time (s)	213.18	120.98	47.59	156.18	73.76	76.77	**12.83**
	CDM-60%	Volume	13.72 / 32.84	14.32 / 29.40	12.51 / 30.56	9.80 / 23.20	26.56/51.65	17.80 / 44.11	**8.79 / 21.65**
		Occupancy	16.81 / 1.65	17.66 / 1.56	14.16 / 1.44	12.69 / 1.27	36.29/1.96	21.72 / 2.05	**9.94 / 1.11**
		Speed	4.41 / 3.88	3.18 / 3.14	3.11 / 5.52	2.08 / 2.11	6.66/5.43	3.09 / 3.09	**1.43 / 1.59**
		Time (s)	164.15	135.34	50.15	161.88	63.25	78.96	**15.75**
	CDM-80%	Volume	31.96 / 85.41	20.31 / 42.94	21.17 / 71.11	11.20 / 26.01	43.65/70.66	27.23 / 80.21	**9.42 / 23.45**
		Occupancy	21.89 / 2.00	24.29 / 2.05	22.53 / 2.40	15.79 / 1.55	48.72/2.57	33.58 / 3.15	**11.14 / 1.40**
		Speed	14.41 / 11.50	3.78 / 3.83	5.75 / 10.00	2.85 / 2.85	7.81/6.08	8.46 / 14.05	**2.01 / 2.27**
		Time (s)	166.61	229.37	59.41	187.42	62.46	113.35	**23.14**
PeMS-D4	CDM-40%	Volume	17.41 / 33.48	18.00 / 33.65	14.35 / 28.89	13.46 / 27.19	27.21/45.56	22.16 / 42.45	**12.03 / 25.91**
		Occupancy	29.45 / 1.75	25.57 / 1.69	18.05 / 1.38	17.35 / 1.28	40.97/1.65	24.07 / 1.75	**14.88 / 1.15**
		Speed	3.65 / 3.27	3.61 / 3.28	2.47 / 2.24	2.13 / 1.97	5.03/4.19	3.04 / 2.70	**1.51 / 1.48**
		Time (s)	400.15	274.82	96.38	282.79	124.06	174.67	**26.69**
	CDM-60%	Volume	19.48 / 35.56	19.35 / 34.48	16.28 / 32.59	13.90 / 27.50	38.48/61.74	25.69 / 50.71	**12.29 / 26.43**
		Occupancy	30.56 / 1.79	28.08 / 1.77	21.16 / 1.72	19.55 / 1.47	48.50/2.16	33.64 / 2.33	**15.46 / 1.26**
		Speed	4.40 / 3.80	3.73 / 3.36	3.29 / 2.96	2.59 / 2.40	8.15/6.66	4.41 / 3.95	**1.72 / 1.70**
		Time (s)	464.47	470.10	172.86	462.05	80.30	146.30	**37.88**
	CDM-80%	Volume	38.93 / 56.12	35.83 / 50.33	26.98 / 66.54	20.15 / 52.16	42.30 / 70.12	39.30 / 97.22	**13.54 / 28.16**
		Occupancy	44.14 / 2.30	42.91 / 2.19	31.12 / 2.43	25.48 / 1.75	60.37 / 2.40	46.15 / 3.32	**18.30 / 1.54**
		Speed	12.70 / 10.24	4.37 / 3.93	6.91 / 10.28	3.60 / 3.26	7.95 / 6.19	9.50 / 12.99	**2.43 / 2.42**
		Time (s)	428.73	726.77	197.53	537.08	108.80	223.39	**64.96**

Best results are highlighted in bold fonts.

Table 3
Performance comparison for data reconstruction on CTM scenarios. Each result cell for the baseline models displays performance in the format of MAPE / RMSE. The running time is also reported for each method.

Dataset	Scenario	View / Metric	BGCP	BATF	LRTC-TNN	LATC	IGNNK	MVLR	USCMT
PeMS-D8	CTM (2h, 40%)	Volume	11.25 / 27.57	12.87 / 28.39	9.87 / 27.08	9.79 / 25.99	133.38 / 258.82	12.83 / 33.76	**9.13 / 25.64**
		Occupancy	16.85 / 3.95	16.95 / 3.08	12.72 / 1.64	13.10 / 1.70	70.13 / 3.69	15.94 / 1.83	**10.84 / 1.60**
		Speed	3.30 / 3.44	3.55 / 3.60	2.74 / 3.04	2.98 / 3.39	24.56 / 18.62	3.88 / 3.76	**2.63 / 3.00**
		Time (s)	127.95	110.70	28.35	98.53	60.40	51.89	**19.36**
PeMS-D4	CTM (2h, 40%)	Volume	16.42 / 30.23	17.43 / 31.23	14.41 / 28.66	14.48 / 28.62	88.96 / 165.37	18.81 / 36.48	**13.91 / 28.61**
		Occupancy	33.66 / 5.96	29.36 / 2.62	20.69 / 1.79	20.99 / 1.90	117.88 / 3.59	27.32 / 1.99	**17.40 / 1.77**
		Speed	4.23 / 3.83	4.42 / 3.99	3.65 / 3.37	3.64 / 3.42	25.59 / 19.34	4.67 / 4.12	**3.39 / 3.22**
		Time (s)	224.14	218.14	52.86	183.76	108.23	104.91	**33.75**

Best results are highlighted in bold fonts.

Methods such as MVLR, which are competitive under RM scenarios, suffer substantial performance degradation in high-rate CDM settings, whereas USCMT remains robust and effective. These findings highlight the strength of the proposed architecture, particularly its ability to handle the complex and structured missing data patterns commonly encountered in real-world transportation systems.

To provide a more intuitive understanding of the reconstruction performance, Fig. 4 visualizes the time series curves from sensor #1 under the most challenging CDM-80% scenario. This qualitative comparison is consistent with the quantitative results reported in Table 2. Most baseline methods struggle to capture the complex daily traffic dynamics. For instance, the reconstructions from BGCP and BATF are noisy and fail to follow the peaks and troughs, while MVLR deviates substantially from the ground truth. By contrast, the curve generated by USCMT aligns closely with the original data, accurately recovering both the periodic patterns and the sharp transient fluctuations. This visual evidence further illustrates USCMT's ability to model intricate spatio-temporal dependencies under extreme data loss conditions.

5.4.3. Reconstruction results on CTM data

Table 3 presents the reconstruction results under the CTM setting, where 40% of detector–day pairs are randomly selected and each selected pair contains a consecutive 2-hour missing segment. As in the previous two missing scenarios, USCMT maintains strong reconstruction performance under the CTM setting on both PeMS-D4 and PeMS-D8. It achieves the best overall results across the three traffic views and remains computationally efficient compared with the baseline methods. These findings further confirm the robustness and effectiveness of USCMT in handling structured temporal missingness.

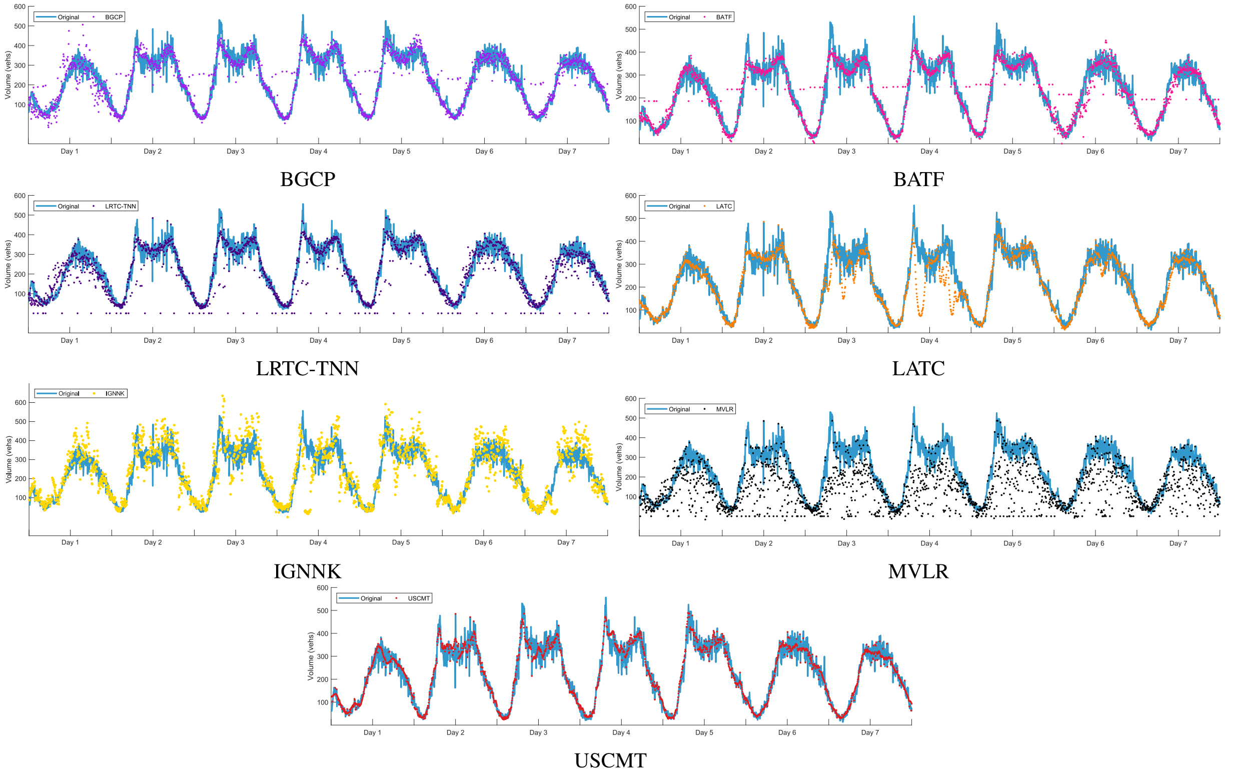

Fig. 4. Visual comparison of one week time series reconstructions of Volume (vehs) by different models for sensor #1 from the PeMS-D4 dataset, under the severe CDM-80% scenario.

Table 4
AUC comparison for event detection.

View	iForest	MLP	RTC	MVLR	USCMT
Volume	0.5469	0.6941	0.5883	0.5124	**0.7270**
Occupancy	0.7077	0.6294	0.7095	0.6590	**0.7270**
Speed	0.5632	0.7084	0.5976	0.6925	**0.7270**

Best results are highlighted in bold fonts.

5.5. *Performance on event detection*

This section presents a comprehensive evaluation of the event detection performance of USCMT against the relevant baselines. The experiments are conducted on the FT-AED dataset (Lane 4), which is well suited for this task because it includes verified ground-truth event labels.

The AUC results are summarized in Table 4. USCMT attains the highest overall AUC of 0.7270, exceeding the best view-specific results of the baseline methods. For USCMT, the same AUC value appears across Volume, Occupancy, and Speed because anomaly detection is based on the aggregated event score $\sum_{v=1}^{V} |\mathcal{E}^v|$, rather than on separate view-specific scores. In contrast, the baseline detectors show noticeable variation across views; for instance, MLP performs relatively well on the Speed view but less strongly on Occupancy, while RTC shows the opposite tendency. This comparison suggests that the aggregated multi-view indicator of USCMT provides stronger overall ranking performance than the single-view detection baselines.

To provide a more detailed visual representation of these results, Fig. 5 plots the corresponding ROC curves for all models on each of the three traffic views. Across all subplots, the ROC curve of USCMT is generally closer to the top-left corner than those of most baseline methods, indicating stronger overall detection performance. This visual trend is consistent with the higher AUC values reported in Table 4 and further supports the quantitative findings. Overall, USCMT achieves a better trade-off between the true positive rate and the False Positive Rate (FPR) when distinguishing between normal and anomalous traffic patterns.

To further quantify performance at practical operating points, we analyze the detection delay and miss percentage at FPR levels of 1%, 2.5%, and 5%, as reported in Table 5. The results show that no single method dominates every metric, but USCMT delivers the most robust and well-balanced performance overall. The baseline models tend to be either inconsistent or specialized. For instance, in the Occupancy view at 5% FPR, RTC achieves an excellent detection delay of -7.39 min but with a relatively high miss percentage of 27.27%; conversely, MVLR attains a lower miss percentage of 15.91% but with a less competitive delay. USCMT, by contrast,

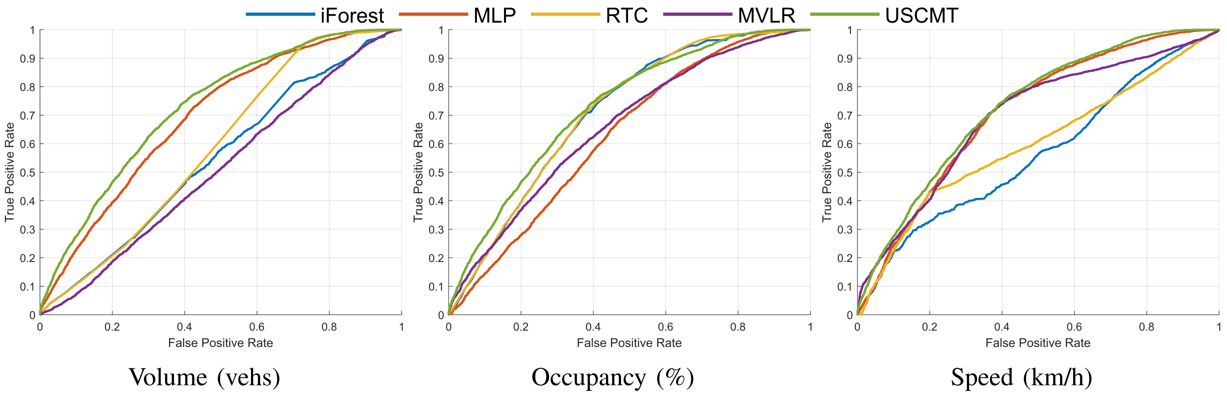

Fig. 5. ROC curves for event detection across the three traffic views.

Table 5
Effect of the FPR on detection delay (mean ± standard deviation) and miss percentage, where missed crashes are defined as crashes not detected within 15 min after the official report time.

FPR (%)	iForest		MLP		RTC		MVLR		USCMT	
	Delay (min)	Miss (%)	Delay (min)	Miss (%)	Delay (min)	Miss (%)	Delay (min)	Miss (%)	Delay (min)	Miss (%)
Volume										
1	**−2.17 ± 5.98**	61.36	−1.66 ± 6.26	**52.27**	−2.09 ± 5.85	61.36	−0.40 ± 5.07	72.73	−1.80 ± 5.44	**52.27**
2.5	**−5.17 ± 6.72**	31.82	−4.14 ± 7.86	**27.27**	−5.08 ± 6.62	31.82	−1.44 ± 6.00	59.09	−4.70 ± 7.09	36.36
5	**−7.05 ± 7.38**	**18.18**	−5.86 ± 8.16	20.45	−7.02 ± 7.36	**18.18**	−3.10 ± 7.66	27.27	−6.97 ± 6.38	25.00
Occupancy										
1	−1.61 ± 6.66	**43.18**	−0.03 ± 4.92	81.82	0.00 ± 0.00	100.00	**−1.82 ± 6.61**	50.00	−1.80 ± 5.44	52.27
2.5	−3.68 ± 6.64	36.36	−0.61 ± 5.61	63.64	0.00 ± 0.00	100.00	−3.95 ± 7.03	**27.27**	**−4.70 ± 7.09**	36.36
5	−6.86 ± 6.90	29.55	−5.01 ± 7.05	**15.91**	**−7.39 ± 7.25**	27.27	−6.13 ± 7.33	**15.91**	−6.97 ± 6.38	25.00
Speed										
1	−0.28 ± 4.17	86.36	−0.14 ± 6.31	63.64	0.42 ± 2.94	81.82	−1.31 ± 4.92	72.73	**−1.80 ± 5.44**	**52.27**
2.5	−0.77 ± 6.65	59.09	−3.31 ± 7.64	43.18	−1.59 ± 6.20	38.64	−1.41 ± 6.20	47.73	**−4.70 ± 7.09**	**36.36**
5	−2.50 ± 7.61	50.00	−5.33 ± 8.00	34.09	−5.14 ± 7.02	**18.18**	−3.45 ± 6.65	34.09	**−6.97 ± 6.38**	25.00

Best results are highlighted in bold fonts.

maintains consistently strong results across the reported operating points. Even in the few instances where it is not the top performer, its results remain close to the best. For example, for the Volume view at 5% FPR, its delay of -6.97 min is only slightly higher than the best score of -7.05 min achieved by iForest. At the 2.5% FPR operating point, USCMT achieves both the best delay and the lowest miss percentage among the compared methods. This balanced performance profile highlights the practical value of USCMT as a robust event detection framework.

Finally, to provide a more traffic-oriented interpretation of the detection results, Fig. 6 presents a visual case study of two reported incidents on Lane 4 of the FT-AED dataset. The first incident was reported at 05:52:28, initially classified as a disabled vehicle and later updated to an incident, with a recorded duration of approximately 1 h 24 min. The second incident was reported at 09:51:11, remained classified as an incident throughout, and lasted approximately 7 h 18 min. A closer visual comparison reveals clear differences in detection behavior across methods. Several baselines, particularly RTC and MVLR, label large portions of the disturbed occupancy region as anomalous, while MLP even produces a long pre-incident alarm segment before the first reported time in the upper detector rows. These patterns suggest that such methods may extend anomaly labels well beyond the incident onset and into broader traffic states affected by the disturbance. In contrast, USCMT does not saturate the entire disturbed region with alarms. Its detections are more concentrated around the reported times and around the onset and propagation boundary of the disturbance, while remaining comparatively sparse over the rest of the affected area. Overall, Fig. 6 suggests that USCMT provides a more selective identification of incident-related disruptions without broadly labeling the full congestion field and its subsequent evolution as anomalous.

5.6. *Performance of joint reconstruction and detection*

This subsection culminates our evaluation by assessing the performance of models capable of joint data reconstruction and anomaly detection across a comprehensive suite of scenarios. The experiments are conducted on the FT-AED dataset (Lane 4), encompassing both RM and CDM missing data patterns at various severities. The detailed results, reporting MAPE, RMSE, AUC, and computation time, are consolidated in Table 6.

The results establish the strong overall performance of USCMT. Across the full range of test conditions, USCMT consistently delivers the best reconstruction accuracy, achieving the lowest MAPE and RMSE in nearly every scenario. More importantly, it remains robust when transitioning from RM to the more challenging structured CDM scenarios. For instance, at the 60% missing rate,

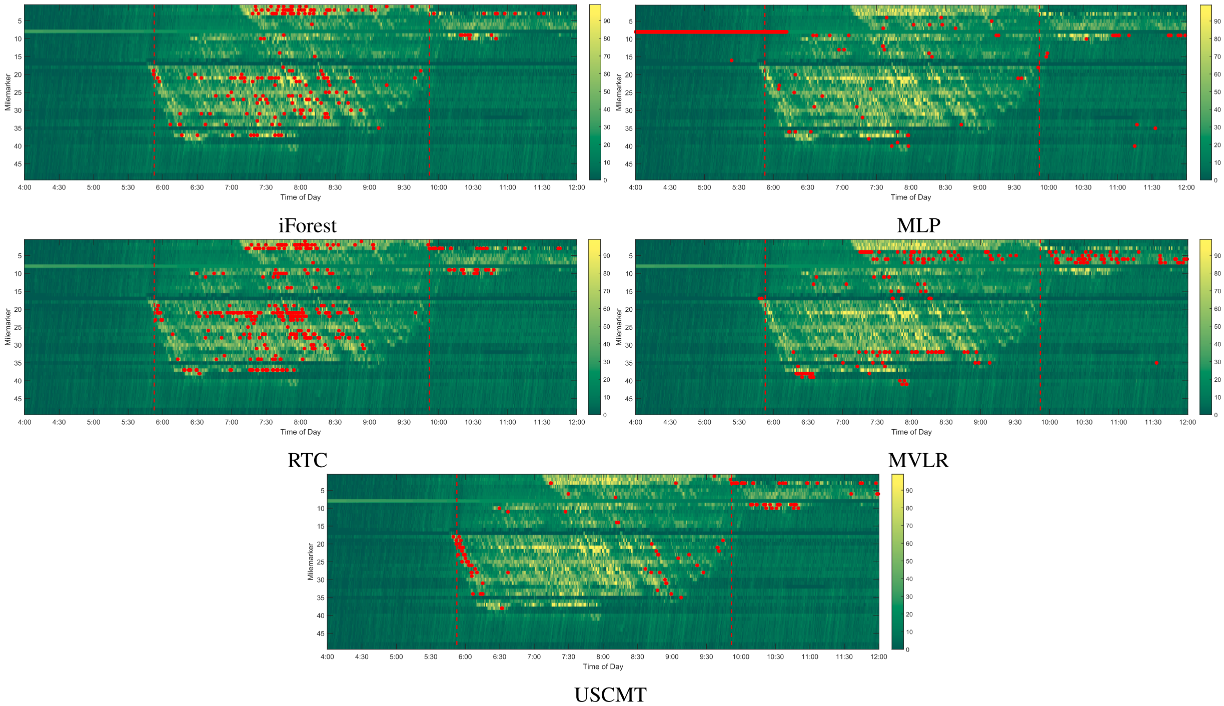

Fig. 6. Visual comparison of incident detection results for a real-world case study with two reported traffic incidents. The spatio-temporal Occupancy (%) map is displayed, with the vertical red line indicating the ground-truth crash report time. The overlaid markers represent the anomalies detected by each model, all operating at a threshold corresponding to a 5% FPR.

Table 6
Comprehensive performance comparison for the joint task of data reconstruction and anomaly detection.

Scenario	View / Metric	RTC			MVLR			USCMT		
		MAPE	RMSE	AUC	MAPE	RMSE	AUC	MAPE	RMSE	AUC
RM-40%	Volume	44.29	3.01	0.5075	43.10	2.63	0.6010	**38.52**	**2.35**	**0.6854**
	Occupancy	61.06	8.31	0.5115	66.96	8.08	0.5947	**51.77**	**7.07**	**0.6854**
	Speed	3.04	2.40	0.5008	3.12	1.88	0.6166	**1.99**	**1.40**	**0.6854**
	Time (s)		99.04			319.79			**54.89**	
RM-60%	Volume	44.94	3.04	0.5041	43.31	2.64	0.5951	**39.61**	**2.42**	**0.6758**
	Occupancy	62.73	8.89	0.5178	67.22	8.40	0.6212	**52.67**	**7.32**	**0.6758**
	Speed	5.68	4.32	0.5040	5.27	3.09	0.6385	**2.56**	**1.78**	**0.6758**
	Time (s)		92.37			258.65			**55.12**	
CDM-40%	Volume	40.71	2.57	0.5348	43.96	2.69	0.5807	**38.60**	**2.41**	**0.6113**
	Occupancy	58.77	8.64	0.5770	68.74	8.37	0.5995	**50.09**	**7.27**	**0.6113**
	Speed	8.61	4.97	0.5959	6.94	4.16	**0.6228**	**2.56**	**1.68**	0.6113
	Time (s)		95.82			394.85			**62.26**	
CDM-60%	Volume	41.77	2.75	0.5234	44.35	2.70	0.5813	**39.90**	**2.44**	**0.5939**
	Occupancy	57.12	9.53	0.5645	68.90	8.85	0.5879	**52.21**	**7.52**	**0.5939**
	Speed	12.48	7.01	0.5260	9.98	5.85	0.5795	**3.92**	**2.48**	**0.5939**
	Time (s)		86.37			295.28			**55.89**	

Best results are highlighted in bold fonts.

while the reconstruction error for most baselines increases substantially when moving from RM to CDM, USCMT degrades much more gracefully. In terms of anomaly detection, USCMT also maintains leading AUC performance in most cases.

Furthermore, this superior accuracy does not come at the cost of efficiency. The computation times reported in the table show that USCMT is significantly faster than its main competitors, RTC and MVLR, making it practical for real-world deployment. This combination of accuracy, robustness across diverse scenarios, and computational efficiency indicates that USCMT is an effective solution for the joint task of traffic data recovery.

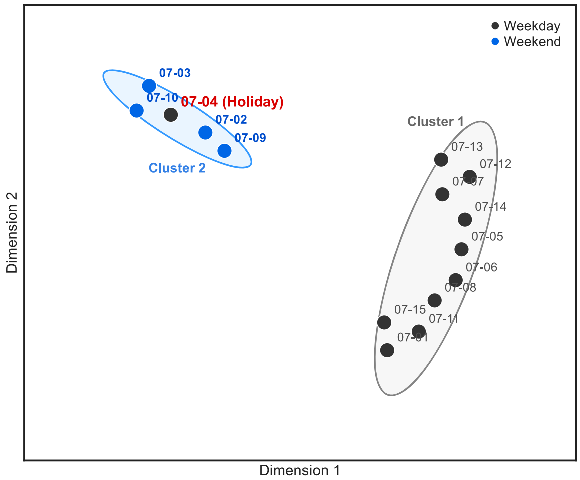

Fig. 7. Visualization of the global affinity structure among daily traffic patterns in the Volume view via spectral t-SNE.

5.7. Pattern analysis

A distinct advantage of USCMT lies in its ability to support interpretable structural analysis through tensor self-representation. In this subsection, we analyze the latent relational patterns and cross-view structural consistency revealed by the learned coefficient tensor.

To characterize the global affinity structure among daily traffic patterns, we construct a symmetric pairwise affinity matrix W defined as:

$$W = \frac{1}{2}\left(\tilde{W} + \tilde{W}^T\right), \quad \text{where} \quad W_{ij} = \frac{1}{T}\sum_{k=1}^{T} |\mathcal{Z}(i,j,k)|.$$

Based on this affinity matrix, spectral clustering is performed to partition daily traffic patterns into two distinct regimes. To visualize the resulting structure, we apply t-SNE (van der Maaten and Hinton, 2008) to the spectral embeddings (spectral t-SNE), projecting the high-dimensional data into a 2D plane as shown in Fig. 7. In the visualization, the unsupervised clustering results are summarized by confidence ellipses (Cluster 1 and Cluster 2), while individual samples are color-coded according to their ground-truth calendar labels (black for weekdays and blue for weekends). The figure suggests that when constrained to two latent regimes, the traffic samples spontaneously organize into groups that align closely with socially defined work/rest categories, driven purely by their intrinsic dynamics. Furthermore, the model successfully captures the underlying “urban rhythm” beyond rigid calendar indices. This is explicitly evidenced by the handling of July 4th (Independence Day). As highlighted in Fig. 7, although this day is a calendar Monday (represented by a black dot), the model correctly assigns it to the blue ellipse of Cluster 2. This indicates that the algorithm recognized the actual traffic demand resembled a rest-day pattern, effectively overriding the misleading calendar label.

Second, moving beyond the global affinity structure, we examine the fine-grained structural consistency across distinct physical views revealed by the tensor $\mathcal{Z}$. Specifically, the magnitude of the tensor entry $\mathcal{Z}(i,j,t,v)$ determines the reconstruction contribution of the historical Day j to the target Day i at time interval t under view v. Fig. 8 visualizes this cross-view synchronization by plotting the joint distribution of Speed and Occupancy coefficient pairs in a 2D Cartesian plane. Notably, the data points tightly adhere to the diagonal line. This structural alignment, quantitatively supported by a high Pearson correlation, confirms that the model captures a unified intrinsic structure regardless of the specific view.

5.8. Robustness analysis

Beyond imputation accuracy under ideal missing data conditions, robustness to real-world data imperfections is a critical indicator of practical utility. This subsection therefore investigates the resilience of USCMT when the observed data are also contaminated by common measurement errors and corruptions. To ensure a controlled and consistent comparison, all robustness tests are conducted on the reconstruction task using the PeMS-D8 dataset under the RM scenario with a 60% missing rate. We consider two main categories of perturbations:

- The first category tests the model’s resilience to common measurement inaccuracies. This category includes two sub-experiments: simulating sensor jitter by adding signal-dependent Gaussian noise at levels of 1%, 2%, and 3% of the true signal value; and simulating calibration drift by applying a systematic multiplicative bias to a random 20% of sensors, with scaling factors selected from the ranges 1 ± 0.1, 1 ± 0.2, and 1 ± 0.3.
- The second category evaluates the model’s ability to handle gross non-Gaussian errors. This is examined through two types of sparse corruptions: spike outliers, where 1% of the data points are perturbed by an additive value equal to 1, 2, or 3 times the

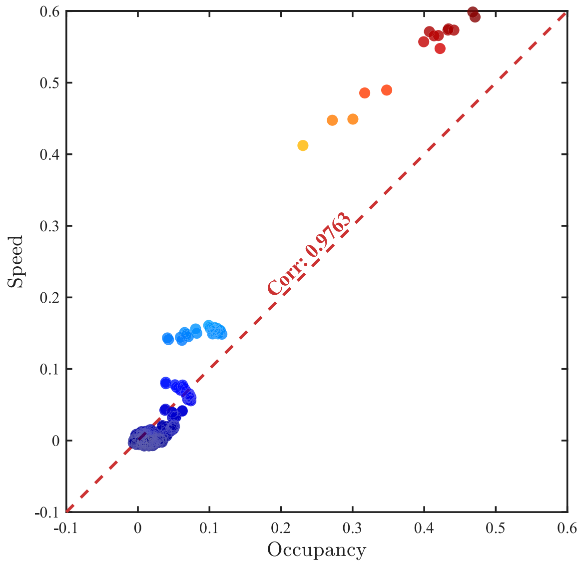

Fig. 8. Visualization of cross-view structural consistency between Speed and Occupancy views.

standard deviation of the corresponding view; and stuck-at corruptions, where 1%, 1.5%, or 2% of the data points are randomly set to either zero or the maximum value of their respective view.

As shown in Fig. 9, we first assess robustness to common measurement inaccuracies. Adding signal-dependent Gaussian noise to simulate sensor jitter causes only a marginal and graceful degradation in USCMT. At the highest noise level (3% of the true signal value), the Volume MAPE increases only slightly, from a baseline of 8.65% to 8.85%. A more challenging test simulates calibration drift by applying a multiplicative bias to 20% of the sensors. Under a severe bias of ±30%, the Volume RMSE of USCMT increases from 21.44 to 27.59. Although this is a noticeable performance drop, USCMT still remains substantially more accurate than all benchmark models, whose error metrics deteriorate more sharply. These results suggest that USCMT can effectively mitigate both high-frequency random noise and low-frequency systematic bias.

We also evaluate the model's ability to handle gross non-Gaussian errors, which are common in traffic sensor data. USCMT demonstrates strong resilience to spike outliers. When 1% of the data points are corrupted by an additive value equal to three times the standard deviation of the corresponding view, the Volume MAPE increases by only 0.55 percentage points, from 8.65% to 9.20%. In contrast, competing models show MAPE increases of up to 2 percentage points under the same condition. The performance gap becomes even more pronounced in the stuck-at corruption test, which mimics complete sensor failure. At a 2% corruption rate, most benchmark models experience severe performance degradation, with Volume RMSE values exceeding 36 in some cases, making them impractical for operational use. By contrast, USCMT maintains its structural robustness and limits the Volume RMSE to 24.03.

In summary, the results across all perturbation scenarios confirm the strong robustness of USCMT. It consistently outperforms benchmark methods and shows particular strength in the presence of gross non-Gaussian errors that typically impair standard imputation algorithms. This high degree of resilience suggests that the model architecture effectively identifies and discounts localized data corruptions while preserving the underlying spatio-temporal traffic patterns. Such capability is critical for developing reliable and deployable intelligent transportation systems that must operate on imperfect real-world data streams.

5.9. Ablation study and component analysis

5.9.1. Ablation on low-rank and anomaly modeling

To further evaluate the design choices in USCMT, we consider three alternative formulations: adding an explicit low-rank regularization on $\mathcal{X}^v$, replacing the coupled group-sparse regularization on $\mathcal{E}$ with an ℓ_1-based sparsity model, and decomposing $\mathcal{E}$ into a sparse outlier error term $\mathcal{E}_1$ and a group-sparse anomaly term $\mathcal{E}_2$. All experiments are conducted on the FT-AED dataset (Lane 4) under the CDM scenario with a missing rate of 20%. Spike outliers are introduced by perturbing 3% of the data points with additive noise equal to one standard deviation of the corresponding view. The results are reported in Table 7.

For the low-rank modeling of $\mathcal{X}^v$, the additional explicit regularization yields only marginal improvements in MAPE and RMSE, while leaving the AUC essentially unchanged. This indicates that explicitly penalizing $\mathcal{X}^v$ provides limited benefit beyond the low-rank structure already induced by the self-representation mechanism. For anomaly modeling, replacing the coupled group-sparse term with an ℓ_1-based formulation leads to substantially worse AUC performance, although the reconstruction results remain broadly comparable. This behavior suggests that element-wise sparsity is not well matched to the anomaly patterns considered here, where traffic events typically manifest jointly across Volume, Occupancy, and Speed rather than as isolated deviations. We further examine the decomposed formulation $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2$. Although this variant yields slight improvements in some reconstruction metrics, it also

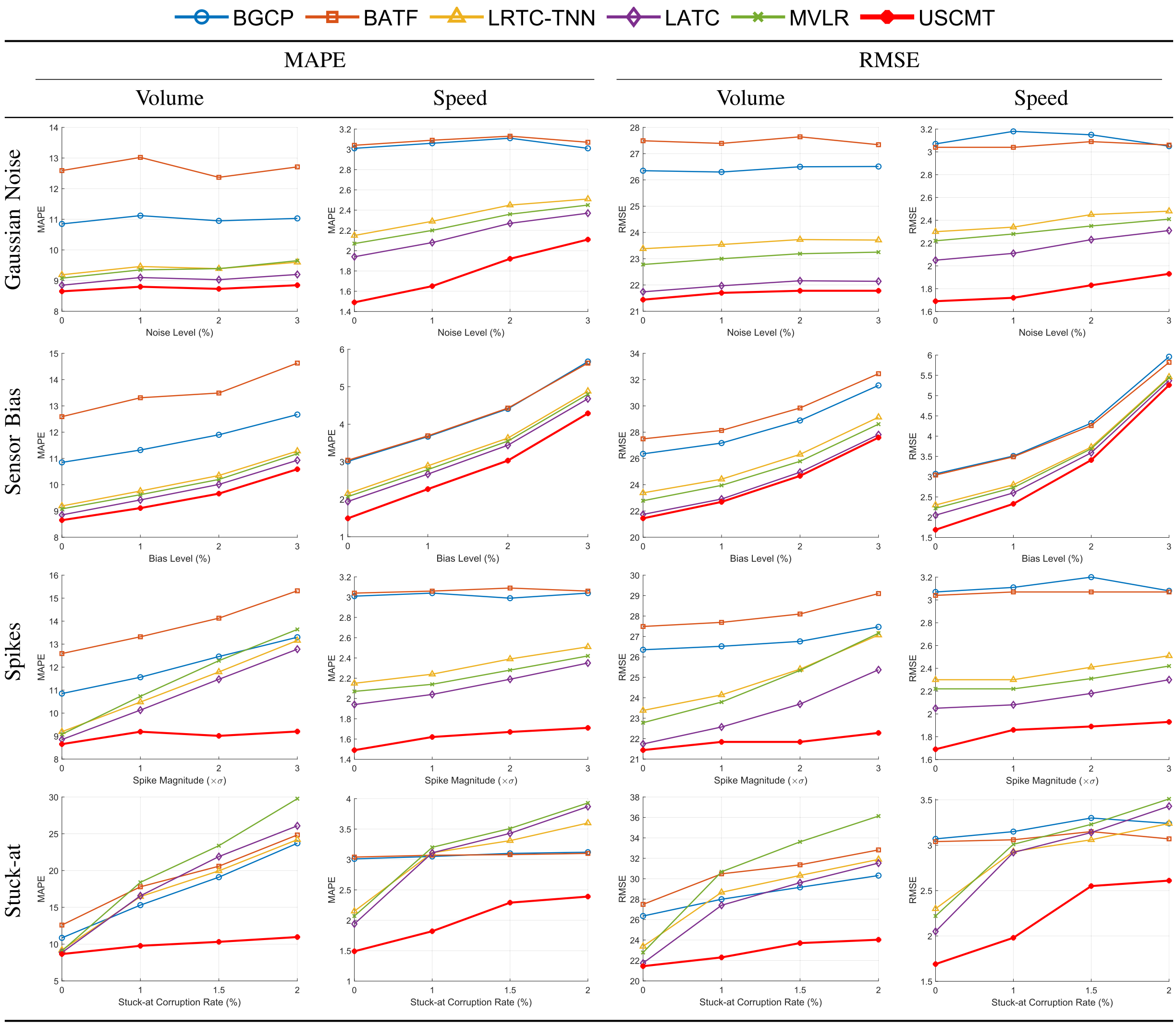

Fig. 9. Comprehensive robustness analysis across four types of data perturbations. Each row evaluates model performance against a specific perturbation: Gaussian noise, sensor bias, spike outliers, and stuck-at corruptions. The columns compare the two error metrics (MAPE and RMSE) for the Volume and Speed traffic views.

Table 7
Ablation study on low-rank and anomaly modeling.

Variant	Volume			Occupancy			Speed		
	MAPE	RMSE	AUC	MAPE	RMSE	AUC	MAPE	RMSE	AUC
USCMT + $\|\mathcal{X}^v\|_\phi$	39.39	2.46	0.5962	58.17	6.91	0.5962	2.23	1.57	0.5962
USCMT w/ $\|\mathcal{E}_1\|_{1,\psi} + \|\mathcal{E}_2\|_{F,\psi}$	39.83	2.50	0.5565	58.60	6.98	0.5565	2.05	1.47	0.5565
USCMT w/ $\|\mathcal{E}\|_{1,\psi}$	39.83	2.50	0.5017	58.88	7.00	0.6331	2.24	1.60	0.5051
USCMT	39.82	2.50	0.5968	58.78	7.00	0.5968	2.27	1.60	0.5968

causes a noticeable reduction in AUC, indicating that separating residual effects into two channels is less effective for anomaly detection. A likely reason is that such a decomposition weakens the structural concentration of event-related deviations and makes the cross-view anomaly pattern less pronounced. Taken together, these results suggest that the proposed model is more suitable for jointly achieving accurate traffic data reconstruction and reliable event detection.

5.9.2. Effectiveness of nonconvex surrogates

A core component of the proposed model is the use of nonconvex surrogate functions to better approximate the rank and joint sparsity terms. To identify suitable surrogates for our framework, we conduct a comparative analysis on the FT-AED dataset (Lane 4,

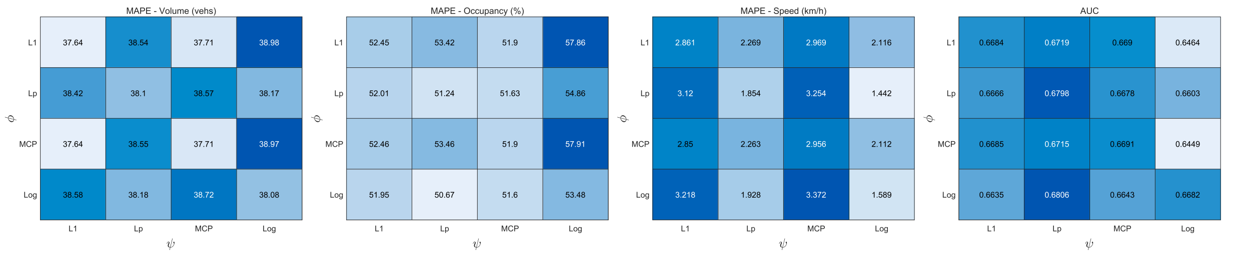

Fig. 10. Sensitivity analysis with respect to functions ϕ and ψ. The first three subplots show the MAPE on three traffic views, while the fourth subplot shows the AUC for event detection.

Table 8
Ablation study on the Spatio-Temporal (ST) module. Performance is compared between our full model (w/ ST) and its ablated variant (w/o ST).

Scenario	Volume				Occupancy				Speed			
	MAPE		RMSE		MAPE		RMSE		MAPE		RMSE	
	w/o ST	w/ ST	w/o ST	w/ ST	w/o ST	w/ ST	w/o ST	w/ ST	w/o ST	w/ ST	w/o ST	w/ ST
RM-40%	12.01	**12.00**	**25.60**	**25.60**	**14.44**	14.48	**1.16**	**1.16**	1.55	**1.54**	**1.51**	**1.51**
RM-60%	12.29	**12.26**	26.40	**26.34**	15.17	**15.15**	**1.25**	**1.25**	1.82	**1.75**	1.77	**1.74**
RM-80%	13.79	**13.48**	28.50	**28.28**	18.87	**18.10**	1.56	**1.52**	2.64	**2.44**	2.54	**2.41**
CDM-40%	12.21	**12.18**	25.38	**25.37**	14.62	**14.60**	**1.16**	**1.16**	1.56	**1.54**	1.50	**1.49**
CDM-60%	12.64	**12.30**	26.65	**26.42**	15.95	**15.52**	**1.23**	**1.23**	2.62	**1.70**	2.74	**1.68**
CDM-80%	14.65	**13.68**	28.73	**28.29**	20.64	**18.21**	1.64	**1.56**	10.13	**2.33**	8.32	**2.32**

Best results are highlighted in bold fonts.

RM-40% scenario), evaluating several representative nonconvex functions against the standard convex ℓ_1-norm baseline (equivalent to using the tensor nuclear norm for ϕ and group lasso for ψ).

The results in Fig. 10 demonstrate the benefit of the nonconvex approach. For event detection, the selected nonconvex surrogates show a clear advantage. In particular, the combination of a logarithmic function for ϕ and an ℓ_p function for ψ achieves the highest AUC score of 0.6806, outperforming the standard convex (ℓ_1, ℓ_1) baseline, which attains 0.6684. For reconstruction, although the results vary across views, the nonconvex functions produce substantially lower errors on the more challenging Occupancy and Speed variables. This study supports the use of nonconvex functions in our model, as they provide tangible performance gains over standard convex formulations. Therefore, we adopt the logarithmic surrogate for ϕ and the ℓ_p surrogate for ψ in the final model.

5.9.3. *Effectiveness of the spatio-temporal module*

To quantify the contribution of the Spatio-Temporal (ST) module, we conduct an ablation study comparing our full model (denoted “w/ ST") against a variant where this module is removed (denoted “w/o ST"). This study compares the models across six challenging data loss scenarios, all conducted on the PeMS-D4 dataset. The results, summarized in Table 8, reveal that the module’s effectiveness is not uniform, but instead grows significantly as the data loss scenario becomes more challenging.

This context-dependent importance is clear when comparing simple and severe scenarios. In the relatively simple RM-40% case, the ST module provides only marginal gains, improving the Speed MAPE by just 0.01 while having almost no impact on the other metrics. However, its contribution becomes indispensable in the most challenging CDM-80% scenario. Here, the inclusion of the ST module leads to substantial improvements across all views: the MAPE for Speed decreases from 10.13 to 2.33, for Occupancy from 20.64 to 18.21, and for Volume from 14.65 to 13.68. This indicates that while the model can rely on other components in relatively mild settings, a strong spatio-temporal prior is crucial for recovering large, structured blocks of missing data. Therefore, the ST module is validated as a key component for ensuring robustness and accuracy in severe real-world conditions.

5.9.4. *Impact of tensor arrangement*

To empirically justify the chosen tensor orientation, we compare it with the permuted arrangement $\mathcal{X}' \in \mathbb{R}^{N\times T\times D}$ on the PeMS-D4 dataset. The study covers six distinct data-loss scenarios. The results in Table 9 confirm that the proposed default arrangement $\mathcal{X} \in \mathbb{R}^{N\times D\times T}$ consistently achieves higher reconstruction accuracy. Moreover, this advantage becomes more pronounced as the missing rate increases and the missing patterns become more complex.

5.9.5. *Effectiveness of the unified multi-view multi-task framework*

To evaluate the role of the unified multi-view multi-task (UM) formulation in USCMT, we compare it with three baseline variants: the sequential single-view multi-task (SS) framework, the unified single-view multi-task (US) framework, and the sequential multi-view multi-task (SM) framework. In the sequential frameworks, reconstruction and anomaly detection are performed as two successive and decoupled stages. The experiments in this subsection follow the same setting as in Section 5.9.1. The quantitative results, measured by MAPE, RMSE, and AUC, are reported in Table 10. The results show that the UM formulation most effectively exploits the proposed

Table 9
Performance comparison between different tensor arrangements. “Perm.” denotes the permuted arrangement $\mathcal{X}' \in \mathbb{R}^{N\times T\times D}$, while “Prop.” denotes the proposed arrangement $\mathcal{X} \in \mathbb{R}^{N\times D\times T}$.

Scenario	Volume				Occupancy				Speed			
	MAPE		RMSE		MAPE		RMSE		MAPE		RMSE	
	Perm.	Prop.	Perm.	Prop.	Perm.	Prop.	Perm.	Prop.	Perm.	Prop.	Perm.	Prop.
RM-40%	14.15	**12.13**	27.71	**25.74**	19.91	**14.51**	1.60	**1.16**	2.51	**1.55**	2.48	**1.52**
RM-60%	15.24	**12.33**	29.02	**26.41**	22.26	**14.97**	1.73	**1.26**	2.85	**1.77**	2.75	**1.77**
RM-80%	18.56	**13.62**	34.05	**28.27**	27.73	**18.07**	2.16	**1.52**	3.95	**2.42**	3.68	**2.41**
CDM-40%	21.32	**11.94**	38.95	**25.99**	27.01	**14.39**	1.90	**1.22**	3.31	**1.52**	3.02	**1.48**
CDM-60%	24.68	**12.46**	45.06	**26.76**	33.90	**15.16**	2.12	**1.24**	5.77	**1.80**	5.42	**1.78**
CDM-80%	31.58	**13.61**	59.93	**28.35**	42.98	**17.98**	2.71	**1.50**	9.68	**2.43**	10.54	**2.41**

Best results are highlighted in bold fonts.

Table 10
Performance comparison of reconstruction and anomaly detection under different framework configurations.

Method	Volume			Occupancy			Speed		
	MAPE	RMSE	AUC	MAPE	RMSE	AUC	MAPE	RMSE	AUC
SS	40.30	2.52	0.5031	61.79	7.20	0.5815	2.88	1.99	0.5018
US	40.30	2.52	0.5017	60.69	7.11	**0.6366**	2.28	**1.60**	0.5020
SM	39.84	**2.50**	0.5569	59.65	7.07	0.5569	2.75	1.93	0.5569
UM	**39.82**	**2.50**	**0.5968**	**58.78**	**7.00**	0.5968	**2.27**	**1.60**	**0.5968**

Best results are highlighted in bold fonts.

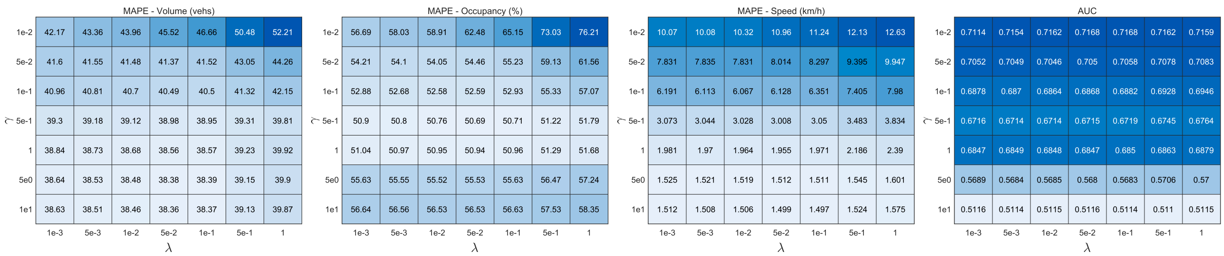

Fig. 11. Sensitivity analysis with respect to hyperparameters λ and γ. The first three subplots show the MAPE on three traffic views, while the fourth subplot shows the AUC for event detection.

cross-view coupling structure. By allowing background recovery and anomaly identification to interact within a shared optimization model, UM consistently achieves the lowest reconstruction errors across all traffic variables while avoiding the error propagation inherent in sequential pipelines. In addition, it delivers competitive and generally superior anomaly detection performance, with the UM variant achieving the highest fused AUC among the multi-view frameworks.

5.10. Parameter sensitivity analysis

This subsection investigates the sensitivity of our model to its two key hyperparameters, γ and λ, in order to identify a robust and effective configuration. As defined in the objective function (Eq. (4)), γ controls the penalty on the sparse anomaly component, while λ weights the spatio-temporal characterization module. The analysis is conducted on the FT-AED dataset (Lane 4, RM-40% scenario), and the results are visualized in Fig. 11.

The results reveal a clear performance trade-off between data reconstruction and event detection, mainly governed by the choice of γ. Reconstruction accuracy for Volume and Speed generally improves as γ increases, whereas Occupancy reaches its best performance at a moderate value of γ. In contrast, the best event detection performance (peak AUC of 0.7168) occurs at the smallest value, $\gamma = 1e\text{-}2$, but this setting yields poor reconstruction quality. To achieve a practical balance, we identify an effective configuration at $\gamma = 1$ and $\lambda = 1e\text{-}2$. This setting provides strong reconstruction accuracy, most notably reducing the MAPE for Speed to below 2.0. Although it entails a moderate reduction in AUC (to approximately 0.685), the substantial gain in overall data fidelity makes it the most balanced configuration for real-world applications where both reliable imputation and event detection are important. Consequently, we adopt $\gamma = 1$ and $\lambda = 1e\text{-}2$ in all subsequent experiments. When detection performance alone is the primary objective, the preferred setting is $\gamma = 1e\text{-}2$ and $\lambda = 1e\text{-}2$.

6. Conclusions and future research

In this paper, we proposed and systematically evaluated the Unified Spatio-temporal and Coupled Multi-view Tensor (USCMT) model for simultaneous traffic data reconstruction and event detection. The main strength of USCMT lies in its traffic-oriented two-level cross-view coupling strategy, which jointly captures shared background structures and synchronized event-driven deviations across multiple traffic views. Specifically, USCMT unifies data reconstruction and event detection in a tensor self-representation framework, enabling the joint learning of shared traffic backgrounds and synchronized anomaly patterns from incomplete and corrupted observations. The model further incorporates explicit spatio-temporal regularization to preserve the physical characteristics of traffic flows, together with generalized nonconvex surrogate functions to more accurately characterize the underlying low-rank and sparse components. Extensive experiments on real-world traffic datasets demonstrate that USCMT consistently outperforms representative baselines in both reconstruction accuracy and event detection performance.

Several promising directions for future research arise from this work. The current USCMT framework operates in an offline batch-processing manner and requires access to complete historical data tensors. An important extension is the development of online or incremental variants that can adapt to streaming traffic data by dynamically updating the low-rank background and sparse anomaly components, thereby supporting real-time traffic monitoring and control. In addition, extending the proposed framework from the current multi-view traffic data setting to more general multi-source sensing scenarios is another promising direction. This includes integrating heterogeneous data sources such as probe trajectories or GPS data, weather observations, and network topology or GIS-based adjacency information, which could further enhance the modeling of complex traffic dynamics.

CRediT authorship contribution statement

Quan Yu: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Yu-Hong Dai:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization; **Lu-Bin Cui:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization; **Shahin Gelareh:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization; **Minru Bai:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

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

Acknowledgments

This research is supported partly by National Key R&D Program of China (2021YFA1000300, 2021YFA1000301), National Natural Science Foundation of China (11971159, 12071399,12021001, 92473208), Hunan Provincial Key Laboratory of Intelligent Information Processing and Applied Mathematics, National Natural Science Foundation of Henan Province (242300420251), Workshop of Outstanding Foreign Scientists in Henan Province (GZS2024007), Key Science and Technology R&D Program of Henan Province (262102520022). The work of Shahin Gelareh is performed using HPC resources from GENCI-IDRIS (Grant 2025-AD010614693R1 Project ComOptIA) .

Data availability

The data and code are publicly available at: https://github.com/quanyumath/USCMT.

Appendix A. Proximal mapping

The proximal operator of a function f at a point z with parameter λ is defined as

$$\mathrm{Prox}_\lambda f(z) := \arg\min_x \lambda f(x) + \frac{1}{2}(x-z)^2.$$

For the particular choices of the nonconvex functions ϕ and ψ adopted in our formulation, the associated proximal mappings often admit closed-form solutions, which are summarized below.

- $L1$: the proximal mapping of f^{L1} is given by Tibshirani (1996)

$$\mathrm{Prox}_\lambda f^{L1}(z) = \mathrm{sign}(z)\max\{|z|-\lambda, 0\}.$$

- Lp: for any $0 < p < 1$, the proximal mapping of f^{Lp} is given by Marjanovic and Solo (2012)

$$\mathrm{Prox}_{\lambda} f^{Lp}(z) = \begin{cases} 0, & \text{if } |z| < \pi_2; \\ \{0, \mathrm{sgn}(z)\pi_1\}, & \text{if } |z| = \pi_2; \\ \mathrm{sgn}(z)\pi_\star, & \text{if } |z| > \pi_2, \end{cases}$$

where $\pi_1 = (2\lambda(1-p))^{\frac{1}{2-p}}$, $\pi_2 = \pi_1 + \lambda p \pi_1^{p-1}$, and $\pi_\star \in (\pi_1, |z|)$ is the solution of $g(\pi) = \pi + \lambda p \pi^{p-1} - |z| = 0$ with $\pi > 0$.
- MCP: for any $\alpha > 0$, the proximal mapping of f^{MCP} is given by Zhang (2010)

$$\mathrm{Prox}_{\lambda} f^{\mathrm{MCP}}(z) = \begin{cases} 0, & \text{if } |z| \leqslant \lambda; \\ \frac{\mathrm{sign}(z)(|z|-\lambda)}{1-\lambda/\alpha}, & \text{if } \lambda < |z| \leqslant \alpha; \\ z, & \text{if } |z| > \alpha. \end{cases}$$

- Logarithm: for any $\theta > 0$, the proximal mapping of $f^{\log}$ is given by Gong et al. (2013)

$$\mathrm{Prox}_{\lambda} f^{\log}(z) = \mathrm{sign}(z) y,$$

where y is an optimal solution of the problem $y = \arg\min_{x \in \mathfrak{C}} \left\{ \lambda f^{\log}(x) + \frac{1}{2}(x-z)^2 \right\}$, and $\mathfrak{C}$ is a set composed of 3 elements or 1 element. If $a := (|z| - \theta)^2 - 4(\lambda - \theta|z|) \geqslant 0$, $\mathfrak{C} = \left\{ 0, \max\left\{ \frac{|z|-\theta+\sqrt{a}}{2}, 0 \right\}, \max\left\{ \frac{|z|-\theta-\sqrt{a}}{2}, 0 \right\} \right\}$, otherwise, $\mathfrak{C} = \{0\}$.

Appendix B. Proof of Theorem 3.2

Proof. By Lemma 2 in Yu et al. (2023), one has $\mathcal{X} = \mathcal{X} * \mathcal{Z} = \mathcal{X} \times_2 Z^T$. Unfolding this equality along mode-2 gives $X_{(2)} = Z^T X_{(2)}$, and transposing both sides yields $X_{(2)}^T = X_{(2)}^T Z$, which completes the proof. $\square$

Appendix C. Proofs for the convergence analysis in Section 4.2

C.1. Proof of Lemma 4.1

Proof. By applying (6), (8) and leveraging the strong convexity of the objective, we immediately obtain

$$\mathscr{L}_{\rho_i^l}(\mathcal{X}^{l+1}, \mathcal{Z}^l, \mathcal{E}^l, \mathcal{G}^l, \mathcal{Y}^l, \mathcal{K}^l; \mathcal{W}_2^l) \leq \mathscr{L}_{\rho_i^l}(\mathcal{W}^l) - \frac{\rho_1^l + \rho_3^l + \rho_4^l}{2} \|\Delta \mathcal{X}^l\|_F^2 \tag{19}$$

and

$$\mathscr{L}_{\rho_i^l}(\mathcal{X}^{l+1}, \mathcal{Z}^{l+1}, \mathcal{E}^l, \mathcal{G}^l, \mathcal{Y}^l, \mathcal{K}^l; \mathcal{W}_2^l) \leq \mathscr{L}_{\rho_i^l}(\mathcal{X}^{l+1}, \mathcal{Z}^l, \mathcal{E}^l, \mathcal{G}^l, \mathcal{Y}^l, \mathcal{K}^l; \mathcal{W}_2^l) - \frac{\rho_2^l}{2} \|\Delta \mathcal{Z}^l\|_F^2. \tag{20}$$

Given that $\mathcal{E}^{l+1}$ and $\mathcal{G}^{l+1}$ are the optimizers of (10) and (12), respectively, we conclude that

$$\mathscr{L}_{\rho_i^l}(\mathcal{X}^{l+1}, \mathcal{Z}^{l+1}, \mathcal{E}^{l+1}, \mathcal{G}^l, \mathcal{Y}^l, \mathcal{K}^l; \mathcal{W}_2^l) \leq \mathscr{L}_{\rho_i^l}(\mathcal{X}^{l+1}, \mathcal{Z}^{l+1}, \mathcal{E}^l, \mathcal{G}^l, \mathcal{Y}^l, \mathcal{K}^l; \mathcal{W}_2^l) \tag{21}$$

and

$$\mathscr{L}_{\rho_i^l}(\mathcal{X}^{l+1}, \mathcal{Z}^{l+1}, \mathcal{E}^{l+1}, \mathcal{G}^{l+1}, \mathcal{Y}^l, \mathcal{K}^l; \mathcal{W}_2^l) \leq \mathscr{L}_{\rho_i^l}(\mathcal{X}^{l+1}, \mathcal{Z}^{l+1}, \mathcal{E}^{l+1}, \mathcal{G}^l, \mathcal{Y}^l, \mathcal{K}^l; \mathcal{W}_2^l). \tag{22}$$

Drawing on (14), (16), and the strong convexity of the objective, we directly infer

$$\mathscr{L}_{\rho_i^l}(\mathcal{X}^{l+1}, \mathcal{Z}^{l+1}, \mathcal{E}^{l+1}, \mathcal{G}^{l+1}, \mathcal{Y}^{l+1}, \mathcal{K}^l; \mathcal{W}_2^l) \leq \mathscr{L}_{\rho_i^l}(\mathcal{X}^{l+1}, \mathcal{Z}^{l+1}, \mathcal{E}^{l+1}, \mathcal{G}^{l+1}, \mathcal{Y}^l, \mathcal{K}^l; \mathcal{W}_2^l) - \frac{\rho_3^l}{2} \|\Delta \mathcal{Y}^l\|_F^2 \tag{23}$$

and

$$\mathscr{L}_{\rho_i^l}(\mathcal{W}_1^{l+1}; \mathcal{W}_2^l) \leq \mathscr{L}_{\rho_i^l}(\mathcal{X}^{l+1}, \mathcal{Z}^{l+1}, \mathcal{E}^{l+1}, \mathcal{G}^{l+1}, \mathcal{Y}^{l+1}, \mathcal{K}^l; \mathcal{W}_2^l) - \frac{\rho_4^l}{2} \|\Delta \mathcal{K}^l\|_F^2. \tag{24}$$

By the multipliers and penalty parameters update in (5), one has

$$\mathscr{L}_{\rho_i^{l+1}}(\mathcal{W}^{l+1}) - \mathscr{L}_{\rho_i^l}(\mathcal{W}_1^{l+1}; \mathcal{W}_2^l) = \frac{1+\beta_1}{2\rho_1^l} \|\Delta \mathcal{P}^l\|_F^2 + \frac{1+\beta_2}{2\rho_2^l} \|\Delta \mathcal{Q}^l\|_F^2 + \frac{1+\beta_3}{2\rho_3^l} \|\Delta \mathcal{R}^l\|_F^2 + \frac{1+\beta_4}{2\rho_4^l} \|\Delta \mathcal{J}^l\|_F^2. \tag{25}$$

By combining Eqs. (19) through (25), we complete the proof. $\square$

C.2. Proof of Lemma 4.2

Proof. By aggregating the results of Lemma 4.1 from $l = 0$ through $L-1$, we obtain

$$\begin{aligned}&\mathscr{L}_{\rho_i^L}(\mathcal{W}^L) - \mathscr{L}_{\rho_i^0}(\mathcal{W}^0) \le -\frac{\tau}{2}\sum_{l=0}^{L-1}\|\mathcal{W}_0^{l+1}-\mathcal{W}_0^l\|_F^2\\&+\sum_{l=0}^{L-1}\frac{1+\beta_1}{2\rho_1^l}\|\Delta\mathcal{P}^l\|_F^2+\sum_{l=0}^{L-1}\frac{1+\beta_2}{2\rho_2^l}\|\Delta\mathcal{Q}^l\|_F^2+\sum_{l=0}^{L-1}\frac{1+\beta_3}{2\rho_3^l}\|\Delta\mathcal{R}^l\|_F^2+\sum_{l=0}^{L-1}\frac{1+\beta_4}{2\rho_4^l}\|\Delta\mathcal{J}^l\|_F^2.\end{aligned}\tag{26}$$

Under the boundedness assumption of $\mathcal{W}_2^l$, a uniform upper bound $M > 0$ can be established such that

$$\max_l \|\Delta\mathcal{W}_2^l\|_F^2 + \|\mathcal{W}_2^l\|_F^2 < M.\tag{27}$$

Observing that $\rho_i^{l+1} = \beta_i\rho_i^l$, we have

$$\sum_{l=0}^{L-1}\frac{1}{\rho_i^l} = \sum_{l=0}^{L-1}\frac{1}{\beta_i^l\rho_i^0} \le \sum_{l=0}^{\infty}\frac{1}{\beta_i^l\rho_i^0} = \frac{\beta_i}{\rho_i^0(\beta_i-1)}.\tag{28}$$

Recalling the expression for $\mathscr{L}_{\rho_i}(\mathcal{W})$, we obtain

$$\begin{aligned}\mathscr{L}_{\rho_i}(\mathcal{W}) =&\|\mathcal{G}\|_\phi + \gamma\|\mathcal{E}\|_{F,\psi} + \frac{\lambda}{2}\sum_{v=1}^{V}\sum_{u=1}^{3}\|\mathcal{K}^v\times_u D_u\|_F^2 + \sum_{v=1}^{V}\delta_{\mathbb{S}^v}(\mathcal{X}^v) + \sum_{v=1}^{V}\Big(\frac{\rho_1}{2}\|\mathcal{X}^v - \mathcal{Y}^v * \mathcal{Z}^v - \mathcal{E}^v + \mathcal{P}^v/\rho_1\|_F^2\Big)\\&+\frac{\rho_2}{2}\|\mathcal{G}-\mathcal{Z}+\mathcal{Q}/\rho_2\|_F^2 + \frac{\rho_3}{2}\|\mathcal{Y}-\mathcal{X}+\mathcal{R}/\rho_3\|_F^2+\frac{\rho_4}{2}\|\mathcal{K}-\mathcal{X}+\mathcal{J}/\rho_4\|_F^2 - \frac{\|\mathcal{P}\|_F^2}{2\rho_1}-\frac{\|\mathcal{Q}\|_F^2}{2\rho_2}-\frac{\|\mathcal{R}\|_F^2}{2\rho_3}-\frac{\|\mathcal{J}\|_F^2}{2\rho_4}.\end{aligned}\tag{29}$$

(1) Combining inequalities (26)–(28) yields the following bound:

$$\mathscr{L}_{\rho_i^L}(\mathcal{W}^L) + \frac{\|\mathcal{P}\|_F^2}{2\rho_1^L}+\frac{\|\mathcal{Q}\|_F^2}{2\rho_2^L}+\frac{\|\mathcal{R}\|_F^2}{2\rho_3^L}+\frac{\|\mathcal{J}\|_F^2}{2\rho_4^L} \le \mathscr{L}_{\rho_i^0}(\mathcal{W}^0) + \sum_{i=1}^{4}\frac{M}{2\rho_i^0} + \frac{(1+\beta_i)\beta_i M}{2\rho_i^0(\beta_i-1)}.\tag{30}$$

In conjunction with (27), (29) and the boundedness of $\{\mathcal{X}^l\}_{l=1}^{\infty}$, this result ensures that, at every iteration, $\|\mathcal{G}^l\|_\phi$, $\|\mathcal{E}^l\|_{F,\psi}$, $\|\mathcal{G}^l - \mathcal{Z}^l + \mathcal{Q}^l/\rho_2^l\|_F$, $\|\mathcal{Y}-\mathcal{X}+\mathcal{R}/\rho_3\|_F$, $\|\mathcal{K}^l - \mathcal{X}^l + \mathcal{J}^l/\rho_4^l\|_F$, and $\|\mathcal{X}^l\|_F$ all remain uniformly bounded. Therefore, the entire sequence of iterates $\{\mathcal{W}^l\}_{l=1}^{\infty}$ is bounded.

(2) According to (29), one has

$$\mathscr{L}_{\rho_i^l}(\mathcal{W}^l) \ge -\frac{\|\mathcal{P}^l\|_F^2}{2\rho_1^l}-\frac{\|\mathcal{Q}^l\|_F^2}{2\rho_2^l}-\frac{\|\mathcal{R}^l\|_F^2}{2\rho_3^l}-\frac{\|\mathcal{J}^l\|_F^2}{2\rho_4^l} \ge -\sum_{i=1}^{4}\frac{M}{2\rho_i^0},$$

combining which with (26)–(28), we derive the bound

$$\frac{\tau}{2}\sum_{l=0}^{L-1}\|\mathcal{W}_0^{l+1}-\mathcal{W}_0^l\|_F^2 \le \mathscr{L}_{\rho_i^0}(\mathcal{W}^0) + \sum_{i=1}^{4}\frac{M}{2\rho_i^0}+\frac{(1+\beta_i)\beta_i M}{2\rho_i^0(\beta_i-1)}.$$

By taking the limit as $L\to\infty$ in the above inequality, we establish the convergence of the series $\sum_{l=0}^{\infty}\|\mathcal{W}_0^{l+1}-\mathcal{W}_0^l\|_F^2$. This, in turn, implies that $\lim_{l\to\infty}\|\mathcal{W}_0^{l+1}-\mathcal{W}_0^l\|_F = 0$, thereby completing the proof of this statement. □

C.3. Proof of Theorem 4.1

Proof. Under the assumption that the sequences $\{\mathcal{W}_2^l\}_{l=1}^{\infty}$ and $\{\mathcal{X}^l\}_{l=1}^{\infty}$ are bounded, Lemma 4.2-(1) ensures the boundedness of the entire sequence $\{\mathcal{W}^l\}_{l=1}^{\infty}$. Consequently, by the Bolzano–Weierstrass theorem, there exists a convergent subsequence of $\{\mathcal{W}^l\}_{l=1}^{\infty}$, which we denote by the same symbol for simplicity, and a limit point $\mathcal{W}^\star$ such that $\lim_{l\to\infty}\mathcal{W}^l = \mathcal{W}^\star$.

Based on the update scheme presented in (5), we obtain the following convergence results:

$$\begin{cases}\lim\limits_{l\to\infty}\|\mathcal{X}^{v,l+1}-\mathcal{Y}^{v,l+1}*\mathcal{Z}^{v,l+1}-\mathcal{E}^{v,l+1}\|_F = \lim\limits_{l\to\infty}\frac{\|\mathcal{P}^{v,l+1}-\mathcal{P}^{v,l}\|_F}{\rho_1^l} = 0;\\ \lim\limits_{l\to\infty}\|\mathcal{G}^{l+1}-\mathcal{Z}^{l+1}\|_F = \lim\limits_{l\to\infty}\frac{\|\mathcal{Q}^{l+1}-\mathcal{Q}^l\|_F}{\rho_2^l} = 0;\\ \lim\limits_{l\to\infty}\|\mathcal{Y}^{l+1}-\mathcal{X}^{l+1}\|_F = \lim\limits_{l\to\infty}\frac{\|\mathcal{R}^{l+1}-\mathcal{R}^l\|_F}{\rho_3^l} = 0;\\ \lim\limits_{l\to\infty}\|\mathcal{K}^{l+1}-\mathcal{X}^{l+1}\|_F = \lim\limits_{l\to\infty}\frac{\|\mathcal{J}^{l+1}-\mathcal{J}^l\|_F}{\rho_4^l} = 0.\end{cases}\tag{31}$$

Consequently, we have

$$\mathcal{X}^{v,\star}-\mathcal{Y}^{v,\star}*\mathcal{Z}^{v,\star}-\mathcal{E}^{v,\star}=0,\ \mathcal{G}^\star-\mathcal{Z}^\star=0,\ \mathcal{Y}^\star-\mathcal{X}^\star=0,\ \mathcal{K}^\star-\mathcal{X}^\star=0.\tag{32}$$

This implies that the accumulation point $\mathcal{W}^{\star}$ satisfies the feasibility conditions of problem (4).

Moreover, by combining (31) with Lemma 4.2-(2), we obtain

$$\lim_{l\to\infty} \mathcal{W}_1^{l+1} - \mathcal{W}_1^{l} = 0. \tag{33}$$

For the $\mathcal{X}$-subproblem, the first-order optimality condition can be expressed as follows

$$0 \in \mathcal{P}^{v,l} + \rho_1^l(\mathcal{X}^{v,l+1} - \mathcal{Y}^{v,l} * \mathcal{Z}^{v,l} - \mathcal{E}^{v,l}) - \mathcal{R}^{v,l} - \rho_3^l(\mathcal{Y}^{v,l} - \mathcal{X}^{v,l+1}) - \mathcal{J}^{v,l} - \rho_4^l(\mathcal{K}^{v,l} - \mathcal{X}^{v,l+1}) + \partial\delta_{\mathbb{S}^v}(\mathcal{X}^{v,l+1}). \tag{34}$$

By (5) and (33), one has

$$\begin{aligned}
&\lim_{l\to\infty} \mathcal{P}^{v,l} + \rho_1^l(\mathcal{X}^{v,l+1} - \mathcal{Y}^{v,l} * \mathcal{Z}^{v,l} - \mathcal{E}^{v,l}) = \lim_{l\to\infty} \mathcal{P}^{v,l} + \rho_1^l(\mathcal{X}^{v,l+1} - \mathcal{Y}^{v,l+1} * \mathcal{Z}^{v,l+1} - \mathcal{E}^{v,l+1}) = \lim_{l\to\infty} \mathcal{P}^{v,l+1},\\
&\lim_{l\to\infty} \mathcal{R}^{v,l} + \rho_3^l(\mathcal{Y}^{v,l} - \mathcal{X}^{v,l+1}) = \lim_{l\to\infty} \mathcal{R}^{v,l} + \rho_3^l(\mathcal{Y}^{v,l+1} - \mathcal{X}^{v,l+1}) = \lim_{l\to\infty} \mathcal{R}^{v,l+1},\\
&\lim_{l\to\infty} \mathcal{J}^{v,l} + \rho_4^l(\mathcal{K}^{v,l} - \mathcal{X}^{v,l+1}) = \lim_{l\to\infty} \mathcal{J}^{v,l} + \rho_4^l(\mathcal{K}^{v,l+1} - \mathcal{X}^{v,l+1}) = \lim_{l\to\infty} \mathcal{J}^{v,l+1}.
\end{aligned}$$

Hence, taking the $l \to \infty$ in (34) yields

$$0 \in \mathcal{P}^{v,\star} - \mathcal{R}^{v,\star} - \mathcal{J}^{v,\star} + \partial\delta_{\mathbb{S}^v}(\mathcal{X}^{v,\star}). \tag{35}$$

For the $\mathcal{Z}$-subproblem, the first-order optimality condition can be expressed as follows

$$0 \in -\mathcal{Y}^{v,l^T} * \mathcal{P}^{v,l} - \rho_1^l \mathcal{Y}^{v,l^T} * (\mathcal{X}^{v,l+1} - \mathcal{Y}^{v,l} * \mathcal{Z}^{v,l+1} - \mathcal{E}^{v,l}) - \mathcal{Q}^{v,l} - \rho_2^l(\mathcal{G}^{v,l} - \mathcal{Z}^{v,l+1}). \tag{36}$$

By (5) and (33), one has

$$\begin{aligned}
&\lim_{l\to\infty} \mathcal{P}^{v,l} + \rho_1^l(\mathcal{X}^{v,l+1} - \mathcal{Y}^{v,l} * \mathcal{Z}^{v,l+1} - \mathcal{E}^{v,l}) = \lim_{l\to\infty} \mathcal{P}^{v,l} + \rho_1^l(\mathcal{X}^{v,l+1} - \mathcal{Y}^{v,l+1} * \mathcal{Z}^{v,l+1} - \mathcal{E}^{v,l+1}) = \lim_{l\to\infty} \mathcal{P}^{v,l+1},\\
&\lim_{l\to\infty} \mathcal{Q}^{v,l} + \rho_2^l(\mathcal{G}^{v,l} - \mathcal{Z}^{v,l+1}) = \lim_{l\to\infty} \mathcal{Q}^{v,l} + \rho_2^l(\mathcal{G}^{v,l+1} - \mathcal{Z}^{v,l+1}) = \lim_{l\to\infty} \mathcal{Q}^{v,l+1}.
\end{aligned}$$

Hence, taking the $l \to \infty$ in (36) yields

$$0 \in -\mathcal{Y}^{v,\star^T} * \mathcal{P}^{v,\star} - \mathcal{Q}^{v,\star}. \tag{37}$$

For the $\mathcal{E}$-subproblem, the first-order optimality condition can be expressed as follows

$$0 \in \gamma\partial\|\mathcal{E}^{l+1}\|_{F,\psi} - \mathcal{P}^l - \rho_1^l(\mathcal{X}^{l+1} - \mathcal{Y}^l * \mathcal{Z}^{l+1} - \mathcal{E}^{l+1}). \tag{38}$$

By (5) and (33), one has

$$\lim_{l\to\infty} \mathcal{P}^l + \rho_1^l(\mathcal{X}^{l+1} - \mathcal{Y}^l * \mathcal{Z}^{l+1} - \mathcal{E}^{l+1}) = \lim_{l\to\infty} \mathcal{P}^l + \rho_1^l(\mathcal{X}^{l+1} - \mathcal{Y}^{l+1} * \mathcal{Z}^{l+1} - \mathcal{E}^{l+1}) = \lim_{l\to\infty} \mathcal{P}^{l+1}.$$

Hence, taking the $l \to \infty$ in (38) yields

$$0 \in \gamma\partial\|\mathcal{E}^{\star}\|_{F,\psi} - \mathcal{P}^{\star}. \tag{39}$$

For the $\mathcal{G}$-subproblem, the first-order optimality condition can be expressed as follows

$$0 \in \partial\|\mathcal{G}^{l+1}\|_{\phi} + \mathcal{Q}^l + \rho_2^l(\mathcal{G}^{l+1} - \mathcal{Z}^{l+1}) = \partial\|\mathcal{G}^{l+1}\|_{\phi} + \mathcal{Q}^{l+1}, \tag{40}$$

where the last equality comes from (5). Taking the $l \to \infty$ in (40) yields

$$0 \in \partial\|\mathcal{G}^{\star}\|_{\phi} + \mathcal{Q}^{\star}. \tag{41}$$

For the $\mathcal{Y}$-subproblem, the first-order optimality condition can be expressed as follows

$$\begin{aligned}
0 \in& - \mathcal{P}^{v,l} * \mathcal{Z}^{v,l+1^T} - \rho_1^l(\mathcal{X}^{v,l+1} - \mathcal{Y}^{v,l+1} * \mathcal{Z}^{v,l+1} - \mathcal{E}^{v,l+1}) * \mathcal{Z}^{v,l+1^T} + \mathcal{R}^{v,l} + \rho_3^l(\mathcal{Y}^{v,l+1} - \mathcal{X}^{v,l+1})\\
&= - \mathcal{P}^{v,l+1} * \mathcal{Z}^{v,l+1^T} + \mathcal{R}^{v,l+1},
\end{aligned} \tag{42}$$

where the last equality comes from (5). Taking the $l \to \infty$ in (42) yields

$$0 \in -\mathcal{P}^{v,\star} * \mathcal{Z}^{v,\star^T} + \mathcal{R}^{v,\star}. \tag{43}$$

For the $\mathcal{K}$-subproblem, the first-order optimality condition can be expressed as follows

$$0 \in \lambda \sum_{u=1}^{3} \mathcal{K}^{v,l+1} \times_u (D_u^T D_u) + \mathcal{J}^v + \rho_4^l(\mathcal{K}^{v,l+1} - \mathcal{X}^{v,l+1}) = \lambda \sum_{u=1}^{3} \mathcal{K}^{v,l+1} \times_u (D_u^T D_u) + \mathcal{J}^{v,l+1}, \tag{44}$$

where the last equality comes from (5). Taking the $l \to \infty$ in (44) yields

$$0 \in \lambda \sum_{u=1}^{3} \mathcal{K}^{v,\star} \times_u (D_u^T D_u) + \mathcal{J}^{v,\star}. \tag{45}$$

Therefore, based on (32), (35), (37), (39), (41), (43), and (45), it follows that $\mathcal{W}^{\star}$ satisfies the KKT conditions of problem (4). □

References

Chen, J., Shao, J., 2000. Nearest neighbor imputation for survey data. J. Off. Stat. 16 (2), 113.

Chen, X., Chen, Y., Saunier, N., Sun, L., 2021. Scalable low-rank tensor learning for spatiotemporal traffic data imputation. Transp. Res. Part C Emerg. Technol. 129, 103226. https://doi.org/10.1016/j.trc.2021.103226

Chen, X., He, Z., Chen, Y., Lu, Y., Wang, J., 2019a. Missing traffic data imputation and pattern discovery with a Bayesian augmented tensor factorization model. Transp. Res. Part C Emerg. Technol. 104, 66–77. https://doi.org/10.1016/j.trc.2019.03.003

Chen, X., He, Z., Sun, L., 2019b. A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation. Transp. Res. Part C Emerg. Technol. 98, 73–84. https://doi.org/10.1016/j.trc.2018.11.003

Chen, X., He, Z., Wang, J., 2018. Spatial-temporal traffic speed patterns discovery and incomplete data recovery via SVD-combined tensor decomposition. Transp. Res. Part C Emerg. Technol. 86, 59–77. https://doi.org/10.1016/j.trc.2017.10.023

Chen, X., Lei, M., Saunier, N., Sun, L., 2022. Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation. IEEE Trans. Intell. Transp. Syst. 23 (8), 12301–12310. https://doi.org/10.1109/tits.2021.3113608

Chen, X., Wang, K., Li, Z., Zhang, Y., Ye, Q., 2023. A novel nonconvex low-rank tensor completion approach for traffic sensor data recovery from incomplete measurements. IEEE Trans. Instrum. Meas. 72, 1–15. https://doi.org/10.1109/tim.2023.3284929

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. Knowledge-Based Syst. 132, 249–262. https://doi.org/10.1016/j.knosys.2017.06.010

Chen, X., Xu, F., Ye, Y., 2010. Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization. SIAM J. Sci. Comput. 32 (5), 2832–2852. https://doi.org/10.1137/090761471

Chen, X., Yang, J., Sun, L., 2020. A nonconvex low-rank tensor completion model for spatiotemporal traffic data imputation. Transp. Res. Part C Emerg. Technol. 117, 102673. https://doi.org/10.1016/j.trc.2020.102673

Coursey, A., Ji, J., Quinones-Grueiro, M., Barbour, W., Zhang, Y., Derr, T., Biswas, G., Work, D.B., 2024. FT-AED: benchmark dataset for early freeway traffic anomalous event detection. Adv. Neural Inf. Process. Syst. 37, 15526–15549.

Culli, M.J., Hall, F.L., 1997. Incident detection on an arterial roadway. Transp. Res. Rec. 1603 (1), 112–118.

Deng, L., Liu, X.-Y., Zheng, H., Feng, X., Chen, Y., 2022. Graph spectral regularized tensor completion for traffic data imputation. IEEE Trans. Intell. Transp. Syst. 23 (8), 10996–11010. https://doi.org/10.1109/tits.2021.3098637

Du, R., Chen, S., 2019. Multi-view low rank representation for multi-source traffic data completion. Int. J. Intell. Transp. Syst. Res. 17 (3), 200–211. https://doi.org/10.1007/s13177-018-0175-5

Friedman, J.H., 2012. Fast sparse regression and classification. Int. J. Forecast. 28 (3), 722–738. https://doi.org/10.1016/j.ijforecast.2012.05.001

Gong, P., Zhang, C., Lu, Z., Huang, J., Ye, J., 2013. A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In: 30th International Conference on Machine Learning, ICML 2013. International Machine Learning Society (IMLS), pp. 696–704. 30th International Conference on Machine Learning, ICML 2013 ; Conference date: 16-06-2013 Through 21-06-2013.

He, Y., Jia, Y., Jia, Y., An, C., Lu, Z., Xia, J., 2024. An integrated intra-view and inter-view framework for multiple traffic variable data simultaneous recovery. IEEE Trans. Intell. Transp. Syst. 25 (11), 17200–17217. https://doi.org/10.1109/tits.2024.3414506

Hu, L., Jia, Y., Chen, W., Wen, L., Ye, Z., 2024. A flexible and robust tensor completion approach for traffic data recovery with low-rankness. IEEE Trans. Intell. Transp. Syst. 25 (3), 2558–2572. https://doi.org/10.1109/tits.2023.3319033

Hu, Y., Work, D.B., 2020. Robust tensor recovery with fiber outliers for traffic events. ACM Trans. Knowl. Discov. Data 15 (1), 1–27. https://doi.org/10.1145/3417337

Huang, T., Wang, S., Sharma, A., 2020. Highway crash detection and risk estimation using deep learning. *Accid. Analy. Prev.* 135, 105392. https://doi.org/10.1016/j.aap.2019.105392

Jin, X., Zhang, Z., Gan, A., 2014. Traffic management centers: Challenges, best practices, and future plans. Technical Report. National Center for Transportation Systems Productivity and Management (U.S.);United States. Dept. of Transportation. Office of the Assistant Secretary for Research and Technology.

Kilmer, M.E., Braman, K., Hao, N., Hoover, R.C., 2013. Third-order tensors as operators on matrices: a theoretical and computational framework with applications in imaging. SIAM J. Matrix Anal. Appl. 34 (1), 148–172. https://doi.org/10.1137/110837711

Kilmer, M.E., Martin, C.D., 2011. Factorization strategies for third-order tensors. Linear Algebra Appl. 435 (3), 641–658. https://doi.org/10.1016/j.laa.2010.09.020

Kolda, T.G., Bader, B.W., 2009. Tensor decompositions and applications. SIAM Rev. 51 (3), 455–500. https://doi.org/10.1137/07070111x

Levin, M., Krause, G.M., 1978. Incident detection: a Bayesian approach. Transp. Res. Rec. 682, 52–58.

Li, L., Lin, Y., Du, B., Yang, F., Ran, B., 2020a. Real-time traffic incident detection based on a hybrid deep learning model. Transp. A Transp. Sci. 18 (1), 78–98. https://doi.org/10.1080/23249935.2020.1813214

Li, Q., Tan, H., Jiang, Z., Wu, Y., Ye, L., 2021. Nonrecurrent traffic congestion detection with a coupled scalable Bayesian robust tensor factorization model. Neurocomputing 430, 138–149. https://doi.org/10.1016/j.neucom.2020.10.091

Li, R., Ruan, S., Bao, J., Li, Y., Wu, Y., Hong, L., Zheng, Y., 2020b. Efficient path query processing over massive trajectories on the cloud. IEEE Trans. Big Data 6 (1), 66–79. https://doi.org/10.1109/tbdata.2018.2868936

Liu, F.T., Ting, K.M., Zhou, Z.-H., 2012. Isolation-based anomaly detection. ACM Trans. Knowl. Discov. Data 6 (1), 1–39. https://doi.org/10.1145/2133360.2133363

Marjanovic, G., Solo, V., 2012. On l_q optimization and matrix completion. IEEE Trans. Signal Process. 60 (11), 5714–5724. https://doi.org/10.1109/tsp.2012.2212015

Parsa, A.B., Taghipour, H., Derrible, S., Mohammadian, A.K., 2019. Real-time accident detection: coping with imbalanced data. *Accid. Anal. Prev.* 129, 202–210. https://doi.org/10.1016/j.aap.2019.05.014

Qiu, D., Bai, M., Ng, M.K., Zhang, X., 2021. Nonlocal robust tensor recovery with nonconvex regularization. Inverse Probl. 37 (3), 035001. https://doi.org/10.1088/1361-6420/abd85b

Shu, H., Wang, H., Peng, J., Meng, D., 2024. Low-rank tensor completion with 3-D spatiotemporal transform for traffic data imputation. IEEE Trans. Intell. Transp. Syst. 25 (11), 18673–18687. https://doi.org/10.1109/tits.2024.3422214

Sun, Y., Mallick, T., Balaprakash, P., Macfarlane, J., 2022. A data-centric weak supervised learning for highway traffic incident detection. *Accid. Anal. Prev.* 176, 106779. https://doi.org/10.1016/j.aap.2022.106779

Tak, S., Woo, S., Yeo, H., 2016. Data-driven imputation method for traffic data in sectional units of road links. IEEE Trans. Intell. Transp. Syst. 17 (6), 1762–1771. https://doi.org/10.1109/tits.2016.2530312

Tan, H., Feng, G., Feng, J., Wang, W., Zhang, Y.-J., Li, F., 2013. A tensor-based method for missing traffic data completion. Transp. Res. Part C Emerg. Technol. 28, 15–27. https://doi.org/10.1016/j.trc.2012.12.007

Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9 (86), 2579–2605.

Wang, X., Sun, L., 2021. Diagnosing spatiotemporal traffic anomalies with low-rank tensor autoregression. IEEE Trans. Intell. Transp. Syst. 22 (12), 7904–7913. https://doi.org/10.1109/tits.2020.3044466

Wang, Y., Zhang, Y., Piao, X., Liu, H., Zhang, K., 2019. Traffic data reconstruction via adaptive spatial-temporal correlations. IEEE Trans. Intell. Transp. Syst. 20 (4), 1531–1543. https://doi.org/10.1109/tits.2018.2854968

Wu, Y., Zhuang, D., Labbe, A., Sun, L., 2021. Inductive graph neural networks for spatiotemporal kriging. Proc. AAAI Conf. Artif. Intelligence 35 (5), 4478–4485. https://doi.org/10.1609/aaai.v35i5.16575

Xiao, J., 2019. SVM and KNN ensemble learning for traffic incident detection. Phys. A Stat. Mech. Appl. 517, 29–35. https://doi.org/10.1016/j.physa.2018.10.060

Xing, J., Liu, R., Anish, K., Liu, Z., 2023. A customized data fusion tensor approach for interval-wise missing network volume imputation. IEEE Trans. Intell. Transp. Syst. 24 (11), 12107–12122. https://doi.org/10.1109/tits.2023.3289193

Yang, S., Kalpakis, K., Biem, A., 2014. Detecting road traffic events by coupling multiple timeseries with a nonparametric Bayesian method. IEEE Trans. Intell. Transp. Syst. 15 (5), 1936–1946. https://doi.org/10.1109/tits.2014.2305334

Yu, Q., Bai, M., 2024. Generalized nonconvex hyperspectral anomaly detection via background representation learning with dictionary constraint. SIAM J. Imaging Sci. 17 (2), 917–950. https://doi.org/10.1137/23m157363x

Yu, Q., Zhang, X., Huang, Z.-H., 2023. Tensor factorization-based method for tensor completion with spatio-temporal characterization. J. Optim. Theory Appl. 199 (1), 337–362. https://doi.org/10.1007/s10957-023-02287-0

Zeng, J., Yu, S., Qian, Y., Feng, X., 2018. Expressway traffic flow model study based on different traffic rules. IEEE/CAA J. Autom. Sin. 5 (6), 1099–1103. https://doi.org/10.1109/jas.2017.7510469

Zhang, C.-H., 2010. Nearly unbiased variable selection under minimax concave penalty. Ann. Stat. 38 (2). https://doi.org/10.1214/09-aos729

Zhang, M., Li, T., Yu, Y., Li, Y., Hui, P., Zheng, Y., 2022. Urban anomaly analytics: description, detection, and prediction. IEEE Trans. Big Data 8 (3), 809–826. https://doi.org/10.1109/tbdata.2020.2991008

Zhou, P., Lu, C., Feng, J., Lin, Z., Yan, S., 2021. Tensor low-rank representation for data recovery and clustering. IEEE Trans. Pattern Anal. Mach. Intell. 43 (5), 1718–1732. https://doi.org/10.1109/tpami.2019.2954874

Zhou, P., Lu, C., Lin, Z., Zhang, C., 2018. Tensor factorization for low-rank tensor completion. IEEE Trans Image Process. 27 (3), 1152–1163. https://doi.org/10.1109/tip.2017.2762595