Spatial-Temporal Attention Model for Traffic State Estimation With Sparse Internet of Vehicles Data

Jianzhe Xue, *Graduate Student Member, IEEE*, Dongcheng Yuan, *Graduate Student Member, IEEE*, Yu Sun, Tianqi Zhang, Wenchao Xu, *Member, IEEE*, Haibo Zhou, *Fellow, IEEE*, and Xuemin Shen, *Fellow, IEEE*

***Abstract*—The rapid growth of connected vehicles creates new opportunities to exploit internet of vehicles (IoV) data for traffic state estimation (TSE), which is a key enabler of intelligent transportation systems (ITS). In this paper, we propose a cost-effective TSE framework that leverages sparse IoV data, which significantly reduces the data collection overhead associated with large-scale IoV datasets. We further analyze the impact of data sparsification and show that the induced estimation errors can be well approximated by Gaussian noise, thereby reformulating sparse IoV-based TSE as a denoising problem. To enhance estimation accuracy, we develop a spatial-temporal attention model, termed the convolutional retentive network (CRNet), which integrates convolutional neural networks (CNNs) for spatial correlation learning with a retentive network (RetNet) for temporal dependency modeling. Extensive experiments conducted on a large-scale real-world IoV dataset validate the feasibility of TSE under sparse IoV sensing conditions. Notably, even when only 5% of the data is available, CRNet achieves a mean absolute error (MAE) below 5 km/h, demonstrating both the high accuracy of the proposed approach and its practical applicability in real-world scenarios.**

***Index Terms*—Traffic state estimation, internet of vehicle, spatial-temporal, deep learning, data mining.**

I. Introduction

TRAFFIC state estimation (TSE) is a cornerstone of intelligent transportation systems (ITS), supporting applications such as traffic management, congestion monitoring, and road planning through the provision of live, precise traffic state information [1], [2], [3]. Traditional TSE relies on roadside sensors, which suffer from limited coverage and high deployment cost [4], [5]. To this end, the mobile crowdsensing based on the internet of vehicles (IoV) has emerged as a data-driven solution, leveraging vehicle-contributed data to provide wide-area, real-time sensing across road networks [6], [7], [8], [9], [10]. This approach, utilizing data such as vehicle speeds and GPS coordinates, offers a new framework for TSE that not only achieves broader coverage but also is more cost-efficient compared to previous methods.

Although IoV-based TSE is promising, acquiring dense and high-quality IoV data at scale is often infeasible due to privacy concerns, resource constraints, and communication overhead [11], [12]. In practice, only sparse IoV data is realistically available [13], [14]. Therefore, there is a need for a TSE framework that can operate effectively with sparse IoV data collected from a limited number of connected vehicles distributed across the city. Existing studies under sparse data conditions typically formulate TSE as a data imputation problem, assuming that traffic states at certain locations are accurately observed and can serve as reliable anchors for inferring missing values [15]. However, this assumption does not hold in realistic sparse-IoV scenarios, where data sparsity originates at the raw sensing stage and no observed state can be regarded as fully accurate. In this work, data sparsification is applied directly to the original vehicle-level driving data, ensuring that the studied scenario faithfully reflects raw sensing sparsity rather than relying on idealized perfect anchors.

A major challenge of using sparse IoV data lies in the reduced stability and accuracy of TSE results compared with those derived from dense datasets. Nevertheless, accurate TSE remains achievable under sparse IoV sensing by exploiting appropriate data processing techniques to compensate for the loss of information. This feasibility stems from the inherent characteristics of transportation systems, in which traffic data exhibit strong spatial and temporal correlations, leading to substantial redundancy. At the micro level, individual vehicle dynamics follow well-established car-following models, whereby the speed of a vehicle is highly correlated with that of its nearby vehicles. As a result, the speed of a single probe vehicle can, to some extent, serve as a representative indicator of the traffic conditions within a local vehicle cluster. At the macro level, urban traffic states exhibit pronounced spatio-temporal correlations [16]. Road segments adjacent to congested areas are likely to experience congestion, and traffic patterns evolve gradually over time [17]. By leveraging these inherent spatio-temporal correlations in urban traffic dynamics, it becomes feasible to achieve accurate TSE even with sparse IoV data.

Received 28 May 2025; revised 19 January 2026; accepted 7 March 2026. Date of publication 11 March 2026; date of current version 6 July 2026. This work was supported by the National Natural Science Foundation of China under Grant 62271244. Recommended for acceptance by W. Wang. *(Corresponding author: Haibo Zhou.)*

Jianzhe Xue, Dongcheng Yuan, Yu Sun, Tianqi Zhang, and Haibo Zhou are with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: jianzhexue@smail.nju.edu.cn; dongchengyuan@smail.nju.edu.cn; yusun@smail.nju.edu.cn; tianqizhang@smail.nju.edu.cn; haibozhou@nju.edu.cn).

Wenchao Xu is with the Division of Integrative Systems and Design, The Hong Kong University of Science and Technology, Hong Kong, China (e-mail: wenchaoxu@ust.hk).

Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: sshen@uwaterloo.ca).

Digital Object Identifier 10.1109/TMC.2026.3673126

Capturing spatio-temporal correlations in traffic states poses significant challenges, especially when accurate estimation must be derived from sparse IoV data [18], [19]. From a temporal perspective, estimations directly obtained from sparse IoV observations often exhibit substantial instability and noise. From a spatial perspective, the inherent sparsity of IoV data leads to a dispersed distribution of vehicle observations across the target region, thereby weakening the effective spatial correlations among neighboring traffic states. Moreover, the heterogeneity and structural complexity of urban road networks render traffic spatio-temporal correlations highly dynamic and nonstationary, further complicating their accurate modeling [20]. In light of these challenges, deep learning (DL) approaches, known for their strong representation learning capabilities, have emerged as a promising solution for TSE under sparse IoV sensing by effectively capturing complex spatio-temporal dependencies [21].

In this work, we propose a cost-effective TSE framework that leverages sparse IoV data and develop a novel spatio-temporal attention model to recover accurate traffic states from initial estimations obtained under sparse data conditions. To simulate sparse IoV data, a small subset of vehicle-level driving records is randomly sampled from the complete dataset. The urban area is partitioned into uniformly spaced grid regions, where the traffic state of each grid is defined as the average vehicle speed in four travel directions over a fixed time interval. We further observe that IoV data sparsification leads to degraded TSE accuracy, and the resulting estimation errors exhibit characteristics consistent with the central limit theorem. To address this challenge, we propose a novel spatio-temporal attention model, termed the convolutional retentive network (CRNet), which mitigates the adverse effects of data sparsification and enables accurate real-time TSE by effectively capturing spatio-temporal correlations inherent in urban traffic dynamics. For spatial correlation modeling, CRNet employs convolutional neural networks (CNNs) to learn local spatial traffic features. For temporal correlation modeling, a retentive network (RetNet), which is an attention-based architecture, is utilized to capture temporal dependencies in traffic states. Comprehensive experiments conducted on a large-scale real-world IoV dataset validate the feasibility of TSE with sparse IoV sensing and demonstrate the effectiveness of CRNet in improving estimation accuracy. The results show that the proposed approach achieves high accuracy at a relatively low cost and outperforms existing methods. The main contributions of this paper are summarized as follows:

- *Framework design for cost-effective TSE:* We present a cost-effective TSE framework that leverages sparse IoV data by uniformly selecting a small subset of probe vehicles across the entire urban area. This framework enables accurate TSE while significantly reducing data collection overhead.
- *Theoretical analysis of sparsification effects:* We analyze the impact of IoV data sparsification on average vehicle speed estimation and show that the resulting errors approximately follow a Gaussian distribution, in accordance with the central limit theorem. Based on this insight, we reformulate sparse IoV-based TSE as a sequential image denoising problem.
- Model design and experimental validation: We propose a novel spatio-temporal attention model, CRNet, which jointly captures spatial and temporal traffic correlations for robust estimation. Extensive experiments on a large-scale real-world IoV dataset demonstrate that CRNet achieves accurate and stable performance even under highly sparse mobile crowdsensing conditions.

The remainder of this paper is organized as follows. Section II reviews TSE and related DL approaches. Section III formulates the TSE problem and presents the corresponding data analysis. Section IV describes the architecture of CRNet. Section V reports extensive experimental results to evaluate the effectiveness and performance of the proposed method. Section VI concludes the paper.

II. Related Work

A. Traffic State Estimation With Sparse IoV Data

The task of TSE is paramount for ITS. The advent of connected vehicles with edge computing has ushered in a new era of IoV data, which encompasses extensive vehicular driving information, such as GPS coordinates and driving speeds across the traffic network [27], [28], [29], [30], [31], [32]. These rich IoV data provide the possibility to perform accurate TSE. Nonetheless, the granularity of vehicular data is frequently coarse, attributed to the disparate temporal and spatial distribution of vehicles, coupled with suboptimal data acquisition practices, culminating in less accurate TSE outcomes. To enhance TSE accuracy, numerous studies have focused on the imputation of missing traffic data, which are either passively or actively missing. A data completion approach based on the tensor decomposition method is proposed to recover missing values in inter-regional traffic data by dividing the target area into regions [33]. A Laplacian-enhanced low-rank tensor completion framework is constructed to recover reliable estimates from incomplete traffic data and achieve good performance under low observation rates [34]. Meanwhile, there has been growing emphasis on the development of DL methods for traffic data imputation in recent years [15]. As summarized in Table I, these works generally operate under the assumption that the missing values are confined to specific regions while the available data are accurate and can serve as reliable anchors.

However, when sparsity originates at the raw data source, for example when only a small portion of probe vehicles upload their information, even the observed data do not perfectly reflect the true traffic state. In such sparse IoV scenarios, existing imputation-oriented methods become insufficient because they rely on the premise of locally complete and accurate observations. This gap highlights the necessity of studying TSE directly under sparse IoV sensing conditions, where all regions may suffer from insufficient raw information and accurate reconstruction must be achieved without assuming perfect anchor states.

B. Deep Learning for Traffic Correlations

Traffic data exhibit rich spatial and temporal correlations due to the inherent characteristics of transportation systems. With the

TABLE I
THE SUMMARY OF TYPICAL TRAFFIC STATE ESTIMATION APPROACHES WITH SPARSE OBSERVATION

Data Source	Ref.	Scenario	Error Analysis	Object	Summary
Fixed Detector	[22]	Highway	✗	Data imputation: Estimate freeway traffic states from sparse fixed-detector observations.	Physics-informed constraints with S3 fundamental diagram exploit traffic flow consistency to reconstruct freeway states despite sparse fixed-detector coverage.
	[23]	Highway	✗	Data imputation: Fill missing values in highway detector time series under high missing ratios.	Global-local spatial-temporal dependencies are jointly exploited via LSTM, SVR and collaborative filtering to robustly impute high-ratio missing detector sequences.
	[24]	Urban	✗	Data imputation: Estimate incomplete urban traffic speed tensors from partially observed entries.	Low-rank spatial–temporal structure is extracted through SVD-initialized tensor decomposition to estimate incomplete urban speed tensors under structural missing.
Probe vehicle	[25]	Highway	✗	Data fitting: Reproduce detector-like traffic states using sparse probe vehicles.	Nonlinear inter-segment dependencies within sparse probe vehicle data are learned to reproduce detector-like speed states on highway networks.
	[26]	Urban	✗	Data imputation: Impute missing values in rasterized urban traffic grids.	Latent spatial-temporal embeddings with VAE approach are leveraged to interpolate rasterized urban traffic grids with missing entries.
	Our	Urban	✓	Data recovery: Reconstruct urban traffic states from initial estimation obtained by sparse IoV data.	The impact of data sparsification is analyzed, and then urban traffic states are reconstructed from the initial estimation by the CRNet.

rapid development of DL paradigms, deep neural networks have demonstrated strong capability in learning and extracting such correlations [35], [36], [37], [38], [39], [40], [41], [42]. For spatial feature extraction, CNNs are widely adopted and particularly effective for gridded traffic representations. In contrast, temporal correlations must be captured from sequential data. Recurrent neural networks (RNNs), especially long short-term memory (LSTM) networks, have been extensively used for modeling temporal dependencies by alleviating the vanishing gradient problem. More recently, transformer-based architectures leverage self-attention mechanisms to model long-range dependencies and have achieved notable success in traffic data mining tasks [43], [44], [45]. Building upon these advances, the RetNet has emerged as a promising successor to the transformer. By integrating the principles of recurrence and attention through the retention mechanism, RetNet enables parallel training, low-cost inference, and strong modeling capability [46]. Consequently, RetNet is well suited for capturing temporal correlations in traffic data with high accuracy and efficiency.

Moreover, some DL models are adept at capturing the spatial and temporal correlations present in input data. Convolutional LSTM network (ConvLSTM) extends LSTM with convolutional layers, enabling itself to capture spatial dependencies and long-term temporal relationships, particularly suited for sequential grid data [47]. PredRNN developed from RNN is used for spatial-temporal predictive learning, in which a pair of memory cells are explicitly decoupled and operate in nearly independent transition manners [48]. SimVP is a relatively simple video prediction model consisting of encoder, translator, and decoder which are completely built on CNN [49]. In SimVP, the encoder extracts spatial features, the translator learns temporal evolution by employing inception modules, and the decoder integrates spatial-temporal features.

Although existing methods can effectively capture spatial and temporal correlations of traffic states, their capacity to adeptly learn the distinct properties of initial TSE obtained from sparse IoV data is not assured. It is essential to develop novel solutions that enhance TSE accuracy by leveraging the unique features of the initial TSE results obtained from sparse IoV data, and address the instability and inaccuracy issues introduced by data sparsification.

TABLE II
SUMMARY OF NOTATIONS

G	Grid-structured city map
$v^{h,w}$	Driving information of vehicles in the (h,w) grid region
F_t	Entire IoV data
$y_t^{h,w}$	Ideal traffic state of the (h,w) grid at time slot t
$\boldsymbol{Y}_t$	Ideal traffic state image from entire IoV data
S_t	Sparse IoV data
$x_t^{h,w}$	Initial traffic state of the (h,w) grid at time slot t
$\boldsymbol{X}_t$	Initial traffic state image from sparse IoV data
$g(t,h,w)$	Gaussian noise of (h,w) grid at time slot t
$z(t,h,w)$	Data missing caused by data sparsification

III. FRAMEWORK AND PROBLEM ANALYSIS

This section presents the TSE framework with sparse IoV data, introduces the traffic state image representation, analyzes sparsification-induced errors, and formulates the recovery problem. For convenience, the major notations in this paper are shown in Table II.

A. Sparse IoV Data Enabled TSE Framework

The cost-effective TSE framework based on sparse IoV data is shown in Fig. 1. Instead of discarding data from specific regions, sparse IoV data are constructed by randomly sampling incoming vehicle records in real time, such that each record across the entire urban area has an equal probability of being selected. As a result, the participation of individual vehicles may vary over time. The retained probe records are transmitted to the backend

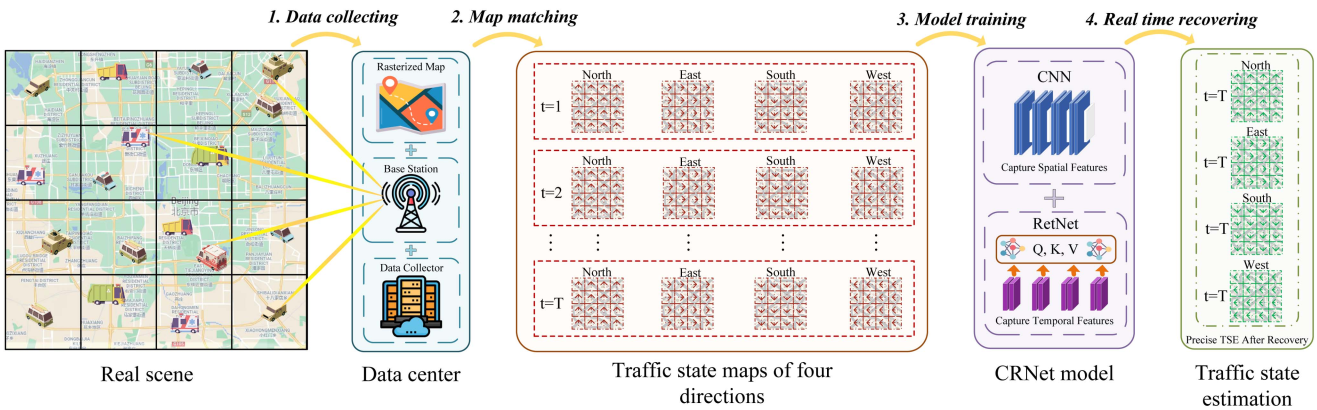


Fig. 1. The framework of sparse IoV data enabled traffic state estimation: The framework consists of sparse IoV data collection, map matching and grid-based representation, CRNet-based spatial–temporal modeling, and real-time traffic state recovery.

server via vehicular networks and contain mobility information, including vehicle speed, travel direction, and GPS coordinates.

The urban area is divided into uniform grid regions, each representing a small area of the city. For the collected sparse IoV data, each probe vehicle is assigned to a grid according to its GPS coordinates. Within each grid, vehicles are categorized into four groups based on their travel directions, namely East, South, West, and North. The initial TSE is obtained by computing the average speed of vehicles in each direction, resulting in four directional speed values per grid. This grid-based representation is analogous to a digital image, in which each grid corresponds to a pixel and the directional speeds serve as multiple channels associated with that pixel. We therefore refer to the grid-structured TSE results derived from sparse IoV data as traffic state images, which are subsequently refined by a recovery algorithm to obtain accurate TSE results. Note that after map matching and grid-level aggregation, minor GPS positioning errors generally have limited influence on grid-level traffic states, especially when the grid size is larger than the expected localization error. Moreover, since CRNet is trained on large-scale real-world IoV data rather than synthetically generated data with idealized noise, the learned model implicitly captures various real-world uncertainties, including localization noise and map-matching imperfections.

B. Traffic State Image

In this study, we construct traffic state images with a spatial resolution of $H \times W$ and four channels. The city map is evenly divided into uniform grids, analogous to pixels in a digital image. Each grid is associated with four directional traffic speeds (East, South, West, and North), which are treated as multi-channel attributes of a pixel. Specifically, for a $H \times W$ gridded city map, four sub-maps are constructed according to the four driving directions and subsequently merged to form a complete city map,

$$\mathrm{G}_c = \begin{bmatrix} \mathrm{G}_c^{1,1} & \cdots & \mathrm{G}_c^{1,W} \\ \vdots & \mathrm{G}_c^{\mathrm{h,w}} & \vdots \\ \mathrm{G}_c^{H,1} & \cdots & \mathrm{G}_c^{H,W} \end{bmatrix}, \tag{1}$$

$$\mathrm{G} = \mathrm{Concat}\left(\mathrm{G}_c | c = \{1, 2, 3, 4\}\right), \tag{2}$$

where $\mathrm{G}_c^{h,w}$ denotes the grid located at position (h, w) corresponding to driving direction c, G_c represents the directional sub-map, and G denotes the complete city map composed of four sub-maps.

To represent the traffic state of each grid, the average vehicle speeds in four directions are computed. Given the IoV data J_t collected at time slot t, which contains complete driving information of all probe vehicles, each data record is mapped to its corresponding grid on the predefined gridded city map $\mathbf{G}$. This mapping yields a grid-structured dataset defined as,

$$\mathbf{F}_t = \begin{bmatrix} F_t^{1,1} & \cdots & F_t^{1,W} \\ \vdots & F_t^{h,w} & \vdots \\ F_t^{H,1} & \cdots & F_t^{H,W} \end{bmatrix}, \tag{3}$$

where $F_t^{h,w} = \{v^{h,w} \mid v^{h,w} \in J_t\}$ denotes the set of vehicle speeds associated with grid (h, w) during time slot t, and $v^{h,w}$ represents the speed magnitude of a probe vehicle whose GPS coordinates fall within the spatial bounds of the grid.

To obtain a direction-aware representation of traffic states, each speed value $v^{h,w}$ is further grouped into directional categories $v_c^{h,w}$, where $c \in \{1, 2, 3, 4\}$ corresponds to the eastbound, southbound, westbound, and northbound driving directions, respectively. The average traffic speed in direction c for grid (h, w) at time slot t is then computed as,

$$y_{t,c}^{h,w} = \mathrm{Mean}\left(v_c^{h,w} \mid v_c^{h,w} \in F_{t,c}^{h,w}\right), \tag{4}$$

where $y_{t,c}^{h,w}$ represents the ideal traffic state obtained by averaging the speeds of all probe vehicles traveling in direction c within the grid. Accordingly, the complete traffic state of a grid located at (h, w) is given by,

$$y_t^{h,w} = \mathrm{Concat}\left(y_{t,c}^{h,w} | c = \{1, 2, 3, 4\}\right). \tag{5}$$

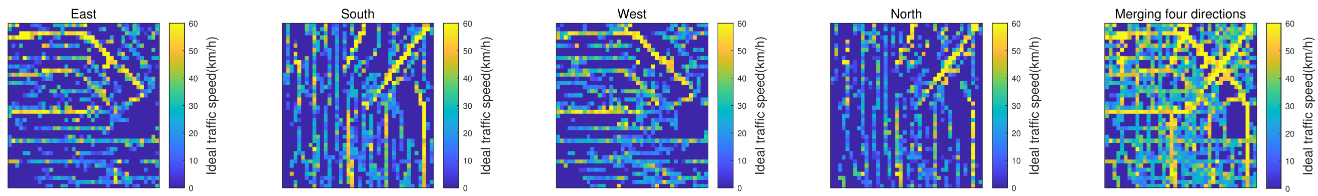


Fig. 2. An illustration of gridded traffic state images: Traffic state images for four travel directions (East, South, West, and North) are shown, along with the merged multi-channel representation used for spatial–temporal modeling.

For the entire city map, we integrate all of its grids and denote the traffic state image at time slot t as,

$$\boldsymbol{Y}_t = \begin{bmatrix} y_t^{1,1} & \dots & y_t^{1,W} \\ \vdots & y_t^{\mathrm{h,w}} & \vdots \\ y_t^{H,1} & \dots & y_t^{H,W} \end{bmatrix}. \tag{6}$$

In summary, Y_t represents the traffic state image at time slot t, where each grid corresponds to a pixel and each directional traffic speed serves as a channel. Therefore, Y_t can be viewed as a four-channel image with a spatial resolution of $H \times W$. An illustration of the gridded traffic state images is shown in Fig. 2.

C. Estimation Error Analysis

When sparse IoV data is utilized instead of the complete dataset, traffic states can be obtained through the same estimation procedure. However, although the average speed derived from sparse IoV data remains close to that obtained from the full data, it exhibits increased noise and instability. Through theoretical analysis based on the central limit theorem, together with empirical observations, we find that the estimation error at each time slot approximately follows a Gaussian distribution.

The procedure for constructing initial traffic state images from sparse IoV data is described as follows. The sparse IoV data collected at time slot t is denoted as $J_t^{'}$, which is obtained by randomly sampling from the complete IoV dataset J_t with equal selection probability for each vehicle record. Using the same data organization process, the sparse vehicle driving information collected over a time duration T up to time slot t is organized as,

$$S_t = \begin{bmatrix} S_t^{1,1} & \dots & S_t^{1,W} \\ \vdots & S_t^{\mathrm{h,w}} & \vdots \\ S_t^{H,1} & \dots & S_t^{H,W} \end{bmatrix}, \tag{7}$$

where $S_t^{h,w} = \{v^{h,w} | v^{h,w} \in J_t'\}$ represents the sparse driving information of probe vehicles matched to the grid located at (h, w). Since sparse IoV data is a subset of the complete dataset, it follows that $J_t' \subseteq J_t$ and $S_t^{h,w} \subseteq F_t^{h,w}$. Accordingly, the initial TSE at time slot t derived from sparse IoV data is expressed as,

$$\boldsymbol{X}_t = \begin{bmatrix} x_t^{1,1} & \dots & x_t^{1,W} \\ \vdots & x_t^{\mathrm{h,w}} & \vdots \\ x_t^{H,1} & \dots & x_t^{H,W} \end{bmatrix}, \tag{8}$$

$$x_{t,c}^{h,w} = \mathrm{Mean}(v_c^{h,w} | v_c^{h,w} \in S_{t,c}^{h,w}), \tag{9}$$

$$x_t^{h,w} = \mathrm{Concat}\,(x_{t,c}^{h,w} | c = \{1, 2, 3, 4\}), \tag{10}$$

where $x_t^{h,w}$ denotes the average vehicle speeds obtained from sparse IoV data at time slot t for the grid located at (h, w).

Theoretical analysis based on the central limit theorem indicates that $x_t^{h,w}$ can be modeled as the ideal traffic state $y_t^{h,w}$ corrupted by Gaussian noise and modulated by a random data-missing indicator. Specifically, the ideal average vehicle speed computed from the complete IoV dataset corresponds to the population mean, whereas the initial average speed derived from sparse IoV data can be regarded as a sample mean. When only a limited number of samples are randomly drawn, the resulting sample mean approximately follows a normal distribution centered at the population mean [50]. Consequently, the estimation error induced by data sparsification can be modeled as Gaussian noise. For grids with insufficient retained probe data, sparsification may lead to the absence of valid observations, resulting in missing initial estimates. Accordingly, the relationship between $x_t^{h,w}$ and $y_t^{h,w}$ can be expressed as

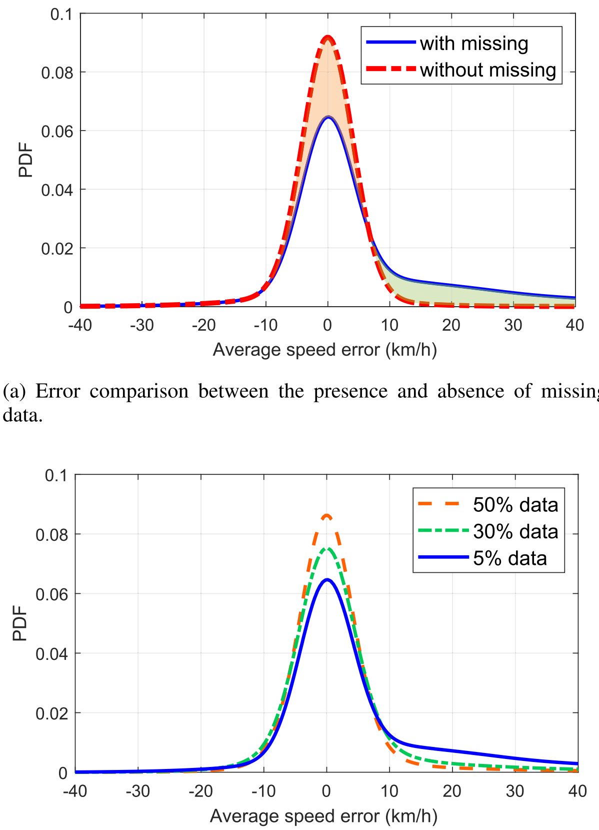
$$x_t^{h,w} = [y_t^{h,w} + g(t, h, w)] \cdot z(t, h, w), \tag{11}$$

where $g(t, h, w) \sim N(0, \sigma_{t,h,w}^2)$ denotes additive Gaussian noise and $z(t, h, w)$ is a binary indicator accounting for data missing caused by sparsification. The noise variance $\sigma_{t,h,w}^2$ varies across time and space, depending on the number of retained vehicle records in grid (h, w).

To validate the above analysis, Fig. 3 presents the error distributions induced by data sparsification. In Fig. 3(a), the overall estimation error exhibits an approximately Gaussian shape with distortion caused by missing values, while zeroing the errors of missing grids yields an almost ideal Gaussian distribution. Fig. 3(b) further shows that the error distributions under different sparsity levels remain bell-shaped, with reduced distortion as sparsity decreases. These observations are consistent with the Gaussian noise model predicted by the central limit theorem.

D. Traffic State Estimation Problem Formulation

The initial average vehicle speeds derived from sparse IoV data form the initial traffic state images; however, they are inaccurate compared with the ideal estimates obtained from complete data and therefore cannot be directly used as the final TSE results. Based on the above analysis, real-time TSE with sparse IoV data can be formulated as a recovery problem, which aims to reconstruct the grid-structured traffic state image by exploiting spatial and temporal correlations in time-series



(a) Error comparison between the presence and absence of missing data.

(b) Error comparison among data with different degrees of sparsity.

Fig. 3. The distribution of errors caused by data sparsification: Estimation error distributions are compared under missing data and varying sparsity levels, validating the statistical characteristics induced by sparse IoV sensing.

traffic state images. Specifically, the inputs to the TSE recovery algorithm consist of historical and current initial traffic state images obtained from sparse IoV data, denoted as $[\boldsymbol{X}t-T+1,\ldots,\boldsymbol{X}t-1,\boldsymbol{X}t]$. The objective is to generate an accurate traffic state image $\boldsymbol{Y}t$ for the current time slot that closely approximates the ideal estimation derived from the complete data. The key to this problem is to capture short-term spatial-temporal correlations in the input sequence to mitigate the effects of noise and data loss introduced by sparsification. Accordingly, the TSE recovery problem can be formulated as learning a mapping function $\mathcal{F}(\cdot)$,

$$\boldsymbol{Y}t=\mathcal{F}([\boldsymbol{X}t-T+1,\ldots,\boldsymbol{X}t-1,\boldsymbol{X}t]),\tag{12}$$

which transforms the initial estimations into an accurate TSE result for the current traffic state.

IV. Methodology

In this section, we present CRNet, which enhances initial TSE by leveraging spatial-temporal traffic correlations. We first introduce the CNN and RetNet modules for modeling spatial and temporal dependencies, and then describe the overall architecture of CRNet.

A. Spatial Correlation Modeling

In our framework, the traffic state is represented as a grid-structured image with a resolution of $H\times W$ and four channels. Spatial correlations of traffic states are naturally embedded in this image representation, where traffic conditions in a region are primarily influenced by those in its neighboring areas and the influence decays with increasing spatial distance. Such localized spatial dependencies can be effectively captured by CNNs, which are well suited for grid-structured data. Therefore, we employ CNNs to extract local spatial features from traffic state images using learnable convolutional filters.

Specifically, the traffic state images are first processed by CNNs in a channel-wise manner to capture independent spatial features corresponding to different travel directions. The convolution operation is formulated as,

$$X'_n=\mathrm{CNN}(X_n), n=t-T+1,\ldots,t-1,t\tag{13}$$

where $\mathrm{CNN}(\cdot)$ denotes the convolution operation, $X_n\in\mathbb{R}^{4\times H\times W}$ is the initial traffic state image derived from sparse IoV data, and $X'_n\in\mathbb{R}^{4\times H\times W}$ represents the extracted spatial features. To further enhance feature expressiveness and enable interaction across different directions, a fully connected (FC) layer is applied after the CNN to form the encoder block. This operation lifts the channel dimension from 4 to C and yields a high-dimensional latent representation, expressed as,

$$X''_n=\mathrm{FC}_e(X'_n), n=t-T+1,\ldots,t-1,t\tag{14}$$

where $\mathrm{FC}e(\cdot)$ denotes the encoder FC layer and $Xn''\in\mathbb{R}^{C\times H\times W}$ is the resulting spatial features. Subsequently, a reshape operation is applied to organize the spatial features into sequential inputs for temporal modeling, denoted as,

$$I^{h,w}=\mathrm{Concat}(x_n^{h,w}|n=t-T+1,\ldots,t-1,t),\tag{15}$$

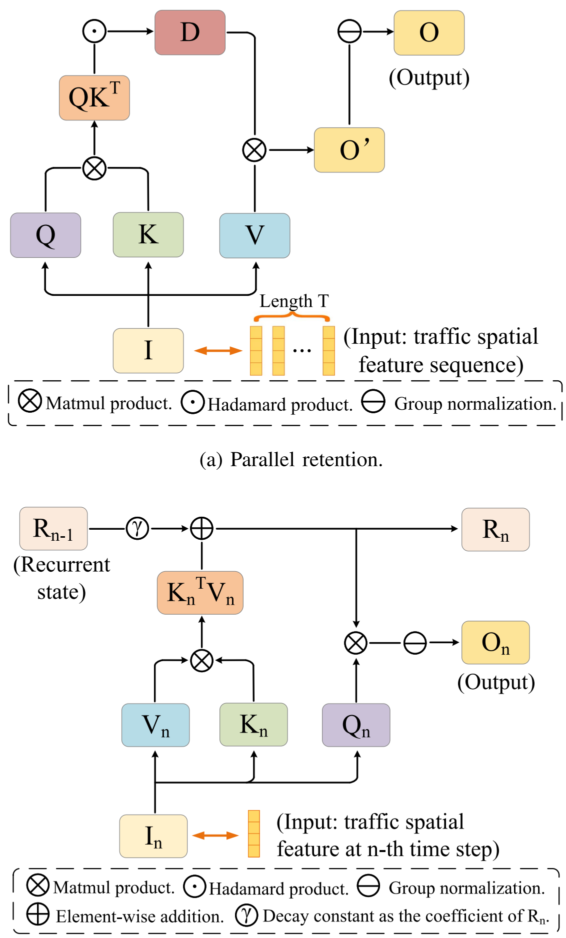
where $x_n^{h,w}$ denotes the spatial feature vector extracted from X''_n at (h,w). The resulting sequence $I^{h,w}\in\mathbb{R}^{T\times C}$ aggregates spatial features over T consecutive time slots and serves as the input for subsequent temporal correlation modeling.

B. Temporal Correlation Modeling

Temporal correlations of traffic states are embedded in the sequential structure of the processed traffic data. To capture such correlations, we employ RetNet, a self-attention–based architecture designed for efficient sequence modeling. The input to RetNet consists of traffic features from the previous $(T-1)$ steps and the current step t. By combining parallel computation for training and recurrent inference for deployment, with optional chunkwise recurrence for long sequences, RetNet enables fast training, memory-efficient inference, and low-cost deployment, which makes it suitable for temporal traffic modeling.

To model temporal correlations, we take $I\in\mathbb{R}^{C\times T}$ as the RetNet input and construct the standard self-attention mechanism as,

$$\begin{cases}Q=IW_Q,\\K=IW_K,\\V=IW_V,\end{cases}\tag{16}$$



(a) Parallel retention.

(b) Recurrent retention.

Fig. 4. The RetNet model for traffic temporal correlation: Illustrations of the parallel retention and recurrent retention mechanisms used to capture temporal dependencies in traffic spatial feature sequences.

where Q, K, V are all $\mathbb{R}^{T\times E}$ matrices and are short for Query, Key, and Value in turn, T is the number of tokens, E represents the embedding size, W_Q, W_K, W_V are learnable neural network weight matrices. The retention mechanism of RetNet is then applied to these embeddings to extract temporal dependencies across time.

We first introduce the traffic parallel retention mechanism, which enables efficient temporal correlation modeling by supporting fully parallel computation during training. By incorporating relative positional information into the query and key representations, the retention mechanism allows RetNet to aggregate information from all time steps. In addition, a predefined decay matrix D is employed to assign exponentially decaying importance to historical traffic states while enforcing causal masking to prevent information leakage from future time steps. As a result, temporal dependencies can be modeled in a parallel manner, formulated as,

$$Q = (IW_Q)\odot\Theta, K = (IW_K)\odot\bar{\Theta}, V = IW_V, \Theta = e^{i\theta}, \tag{17}$$

$$D_{nm} = \begin{cases}\gamma^{n-m} & n\geq m\\ 0 & n<m\end{cases}, \tag{18}$$

$$\text{Retention}_p(I) = (QK^{\mathrm{T}}\odot D)V, \tag{19}$$

where Θ and its complex conjugate $\bar{\Theta}$ encode relative positional information for Q and K, and $D\in\mathbb{R}^{T\times T}$ combines causal masking with exponential decay. The decay factor γ controls the contribution of historical time steps, assigning lower weights to traffic states further in the past.

We then introduce the traffic recurrent retention mechanism, which enables RetNet to perform temporal correlation modeling with low computational and memory complexity, making it suitable for efficient and low-latency inference. At each time step, the retention mechanism takes the current input $I_n\in\mathbb{R}^C$ together with a recurrent state R_n that propagates information from previous time steps, following a recurrent processing paradigm similar to recurrent neural networks. During recurrent inference, the recurrent state and output of the retention mechanism are computed as,

$$R_n = \gamma R_{n-1} + K_n^{\mathrm{T}}V_n, \tag{20}$$

$$\text{Retention}_r(I_n) = Q_nR_n, \tag{21}$$

where R_n denotes the recurrent state at the n-th time step, Q_n, K_n, and V_n are the n-th row vectors of Q, K, and V, respectively, and γ is identical to that defined in (18). Here, $I_n\in\mathbb{R}^C$ represents the input feature at the n-th time step and corresponds to the n-th element of the input sequence I.

For long input sequences, we adopt the chunkwise recurrent retention mechanism, which combines parallel computation within chunks and recurrent computation across chunks to improve efficiency and reduce memory overhead. Specifically, the input sequence I is divided into multiple chunks of equal length. Parallel retention is applied within each chunk, while recurrent retention is performed across consecutive chunks. The retention computation for the i-th chunk is given by,

$$Q_{[i]} = Q_{B_i:B_{i+1}}, K_{[i]} = K_{B_i:B_{i+1}}, V_{[i]} = V_{B_i:B_{i+1}}, \tag{22}$$

$$R_{[i]} = \gamma^B R_{[i-1]} + K_{[i]}^{\mathrm{T}}(V_{[i]}\odot\gamma^{B-i-1}), \tag{23}$$

$$\text{Retention}_k(I_{[i]}) = (Q_{[i]}K_{[i]}^{\mathrm{T}}\odot D)V_{[i]} + (Q_{[i]}R_{[i-1]})\odot\gamma^{i+1}, \tag{24}$$

where $[i]$ denotes the i-th chunk, B is the chunk length, and Q, K, V, and γ are defined as in (17). The recurrent state R is identical to that in (20), and $I_{[i]}\in\mathbb{R}^{C\times B}$ represents the input features of the i-th chunk.

Based on the retention mechanism described above, we further adopt the multi-scale retention (MSR) mechanism to enhance the performance, efficiency, and robustness of temporal correlation modeling. MSR assigns different fixed exponential decay factors γ to multiple retention heads, enabling the model to capture temporal dependencies at different scales and from diverse perspectives. The multi-scale retention mechanism is formulated as,

$$head_j = \text{Retention}(I), j = 1, 2, \ldots, u \tag{25}$$

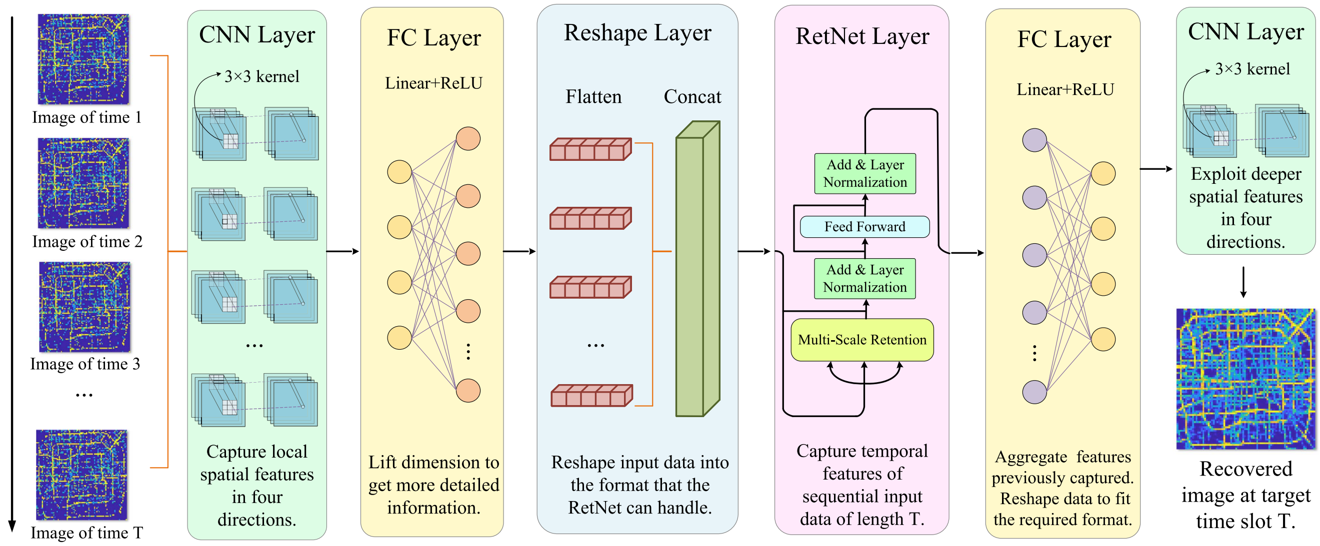


Fig. 5. The CRNet model for traffic spatial-temporal correlations: The model consists of an encoder for spatial feature extraction, a RetNet block for temporal correlation modeling, and a decoder for recovering the traffic state at the target time slot.

where Retention$(\cdot)$ can operate in parallel, recurrent, or chunkwise recurrent modes depending on the input sequence characteristics, u denotes the number of retention heads, and j indexes the heads. After obtaining the outputs of all heads, the MSR output is computed as,

$$N = \text{GN}(\text{Concat}(head_j | j = 1, 2, \ldots, u)), \tag{26}$$

$$\text{MSR}(I) = (\text{Swish}(IW_G) \odot N)W_O, \tag{27}$$

where GN$(\cdot)$ denotes group normalization applied to balance the variances across different heads, Swish$(\cdot)$ is a gating function that introduces non-linearity into the retention mechanism.

Finally, the RetNet architecture is constructed by stacking the MSR layer with a feed-forward network (FFN) layer. Given sparse IoV data $I \in \mathbb{R}^{C\times T}$ as input, the MSR module produces an intermediate representation M, which is further refined by the FFN module to generate the output $O \in \mathbb{R}^{C\times T}$. The overall computation of RetNet is formulated as,

$$M = \text{MSR}(\text{LN}(I)) + I, \tag{28}$$

$$O = \text{FFN}(\text{LN}(M)) + M, \tag{29}$$

where LN$(\cdot)$ denotes layer normalization.

C. Outcome Decoding

The TSE outcome is decoded from the RetNet output through a decoder block composed of a FC layer and a CNN layer. The decoder transforms the extracted spatial-temporal features into the target traffic state image. Specifically, the sequential traffic features O produced by RetNet are first reorganized across all grids to form a traffic state image with shape $\mathbb{R}^{(C\times T)\times H\times W}$. The FC layer then aggregates the learned spatial-temporal information and reduces the channel dimension to four, executed as,

$$U = \text{FC}_d(O), \tag{30}$$

Algorithm 1: CRNet Algorithm.

Input: Traffic state images $[\boldsymbol{X}_{t-T+1}, \ldots, \boldsymbol{X}_{t-1}, \boldsymbol{X}_t]$
Output: Recovered TSE outcome at time slot t
1: **for** each vehicle driving direction $c = 1, 2, 3, 4$ **do**
2: Capture spatial correlation in direction c by CNN.
3: **end for**
4: Exchange information of different directions and lift data dimension.
5: Concatenate spatial correlations at different time step to obtain sequential data for each grid.
6: Capture temporal correlation of each grid by RetNet.
7: Convert the processed sequential traffic data into grid-structured image format.
8: Integrate the spatial-temporal correlations previously obtained and reduce data dimension.
9: **for** each vehicle driving direction $c = 1, 2, 3, 4$ **do**
10: Decode the estimation in direction c.
11: **end for**
12: Export recovered TSE at time slot t as the final output.

where FC$d(\cdot)$ projects the feature dimension from $C \times T$ to 4, and $U \in \mathbb{R}^{4\times H\times W}$ represents an intermediate traffic state image. Subsequently, a CNN layer is applied to refine the spatial structure of the decoded traffic states and enforce spatial consistency among neighboring grids. The final output is obtained as,

$$Z = \text{CNN}(U), \tag{31}$$

where $Z \in \mathbb{R}^{4\times H\times W}$ denotes the recovered traffic state at time slot t.

D. CRNet Model

The architecture of CRNet is shown in Fig. 5, consisting of an encoder, a RetNet block, and a decoder. Given a sequence of

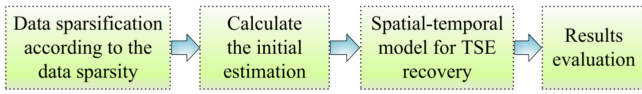


Fig. 6. The flowchart of the experiments: The experimental procedure includes data sparsification, initial traffic state estimation, CRNet-based spatial–temporal recovery, and performance evaluation.

traffic state images as input, the encoder applies CNN to extract spatial correlations from the four directional channels, followed by an FC layer that fuses directional features and lifts channel dimension. The resulting features are reshaped into temporal sequences and fed into RetNet to model temporal dependencies across T steps. The decoder then maps the processed sequences back to the grid format via an FC layer and uses CNN to obtain the final four-channel TSE at time slot t. Algorithm 1 summarizes the forward propagation process.

V. Experiments

In this section, we perform extensive simulations with real-world IoV data to validate the feasibility of the proposed TSE framework and evaluate the efficacy of the CRNet in achieving accurate TSE. Initially, we delineate the experimental setup, the benchmark methods, and the metrics employed for evaluation. We then analyse the experimental results in detail.

A. Experimental Setup

The experimental evaluation is conducted within the area enclosed by the fourth ring road of Beijing, covering longitudes from 116.2610°E to 116.4920°E and latitudes from 39.8235°N to 39.9930°N, using real-world IoV data collected over six days in 2012: Nov. 1, 5, 7, 9, 10, and 11, covering both weekdays and weekends. The dataset comprises approximately 5.9 to 7.8 million IoV records per day, spanning from 7:30 AM to 10:30 PM. Each record includes basic traffic attributes such as vehicle speed, travel direction, and GPS coordinates, essential for TSE. TSE is performed at one-minute intervals, resulting in 900 estimations per day. Traffic state images are of 80×80 resolution. Note that the proposed framework is not restricted to a specific grid resolution and remains applicable across different resolutions. The sparsity level is defined as the ratio of the sampled sparse IoV data to the total available IoV data, where the higher the sparsity is, the more IoV data is sampled. To simulate data sparsity, we randomly sample the original dataset to create sparse IoV data. For each record, a uniformly distributed random number between 0 and 1 is generated; if the number is smaller than the target sparsity level, that record is selected. This process ensures that sparse data are randomly and uniformly distributed, effectively reflecting real-world sensing uncertainty. The flowchart of the experiment is shown in Fig. 6. All experiments are implemented using MongoDB for data handling and Python/PyTorch for modeling on a Tesla V100-DGXS-32 GB GPU platform.

The hyper-parameters are as follows. For the encoder block, the FC layer comprises one FC network, lifting data dimensions from 4 to 16. Prior to the RetNet layer, a reshape layer is implemented, performing operations of flattening and concatenating. The RetNet layer has 2 heads and the dimension of its FFN is 32. The decoder block includes three FC networks within its FC layer. The batch size is 32 and the loss function is the mean squared error (MSE). The dataset from Nov. 1 is utilized for training while the datasets from the remaining five days serve for testing, aimed at evaluating the robustness and adaptability of CRNet over a comprehensive time scale including weekdays and weekends. The model undergoes training for over 200 epochs, starting with an initial learning rate of 0.001, which experiences a 10% decay after each epoch.

B. Baselines and Evaluation Metrics

Our study compares CRNet against several benchmark models, including classic and state-of-the-art spatial-temporal neural networks:

- Initial: The direct estimation from sparse IoV data.
- HA: An arithmetical operation averages the values of historical and current Initial estimations.
- CNN: A four-layer CNN designed to capture the spatial correlation of current Initial estimation.
- ConvLSTM: An integration of convolutional operation and LSTM framework, enabling it to capture both spatial and temporal correlations [47].
- E3DLSTM: A spatial-temporal model that integrates 3D convolutions into RNNs [54].
- PredRNN: A recurrent network in which a pair of memory cells are explicitly decoupled and finally form unified representations of the complex environment [48].
- SimVP: A simple video prediction model that is completely built upon CNN [49].
- TAU: A spatio-temporal model which decomposes temporal attention into intra-frame statical attention and inter-frame dynamical attention [51].
- WaST: A wavelet-based spatio-temporal framework which extracts and adaptively controls both low and high-frequency components via 3D discrete wavelet transform for faster processing [53].
- SimVPv2: A streamlined spatio-temporal model uses the Gated Spatial-temporal Attention mechanism to improve performance and reduce complexity [52].

In this context, the Initial refers to the unprocessed TSE, which also serves as the input for subsequent DL models. HA computes the mean of the past 5 minutes values of Initial estimations along the temporal dimension. CNN leverages the spatial correlation present in the traffic state image at the current time slot to perform TSE. Other models are adept at capturing both spatial and temporal correlations within the historical and current traffic state images, utilizing the past 20 minutes of Initial estimations as their input. Table III compares the parameter count and floating-point operations (FLOPs) of CRNet with some baseline models. We can see that the CRNet based on Retention employs fewer parameters and requires substantially less computational power.

TABLE III
COMPARISON OF CRNET AND BASELINE MODELS

Model	Parameters	FLOPs (G)
CRNet	102,292	0.7913
TAU [51]	2,279,624	2.0973
SimVPv2 [52]	2,375,912	2.1409
WaST [53]	13,526,232	34.7468

To quantitatively assess the TSE performance facilitated by DL models, five evaluation metrics are employed. Let $y_t^{h,w}$ represents the Ideal TSE and $y_t^{\hat{h,w}}$ denotes the estimated TSE of the image grid located at coordinates (h, w) at time slot t, the metrics are expressed as:

1) Root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{1}{T \times H \times W} \sum_{t=1}^{T} \sum_{h=1}^{H} \sum_{w=1}^{W} |y_t^{h,w} - y_t^{\hat{h,w}}|^2}. \tag{32}$$

RMSE measures the average magnitude of squared estimation error; smaller values imply higher accuracy.

2) Improved percentage (IPV):

$$IPV = \times \left(1 - \sqrt{\frac{\sum_{t=1}^{T} \sum_{h=1}^{H} \sum_{w=1}^{W} |y_t^{h,w} - y_t^{\hat{h,w}}|^2}{\sum_{t=1}^{T} \sum_{h=1}^{H} \sum_{w=1}^{W} |y_t^{h,w} - x_t^{\hat{h,w}}|^2}}\right) \times 100\%. \tag{33}$$

IPV quantifies the relative RMSE reduction over the Initial estimation; larger values indicate greater improvement.

3) Mean absolute error (MAE):

$$MAE = \frac{1}{T \times H \times W} \sum_{t=1}^{T} \sum_{h=1}^{H} \sum_{w=1}^{W} |y_t^{h,w} - y_t^{\hat{h,w}}|. \tag{34}$$

MAE gives the average absolute deviation; lower values denote better estimation accuracy.

4) Structural similarity index measure (SSIM):

$$SSIM = \frac{(2\mu_a \mu_b + C_1) \cdot (2\sigma_{ab} + C_2)}{(\mu_a^2 + \mu_b^2 + C_1) \cdot (\sigma_a^2 + \sigma_b^2 + C_2)}, \tag{35}$$

SSIM evaluates similarity between the ideal and estimated image; higher values preferred.

5) Peak signal-to-noise ratio (PSNR):

$$PSNR = 10 \cdot lg \left(\frac{60^2}{\frac{1}{H \times W} \sum_{h=1}^{H} \sum_{w=1}^{W} |y_t^{h,w} - y_t^{\hat{h,w}}|^2} \right). \tag{36}$$

PSNR measures reconstruction quality in decibels; higher values indicate better quality.

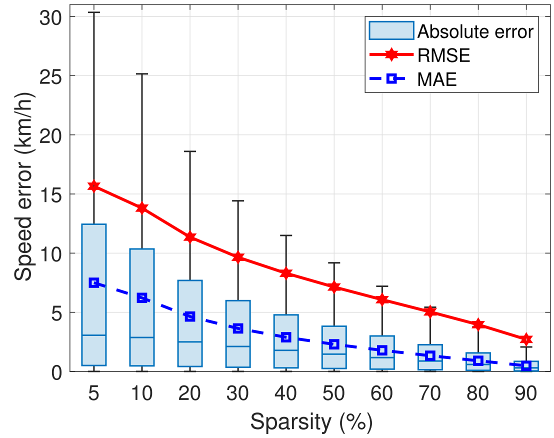


Fig. 7. The influence of sparsification: Estimation errors decrease as data sparsity is reduced, demonstrating the relationship between data density and TSE performance.

C. Performance Evaluation

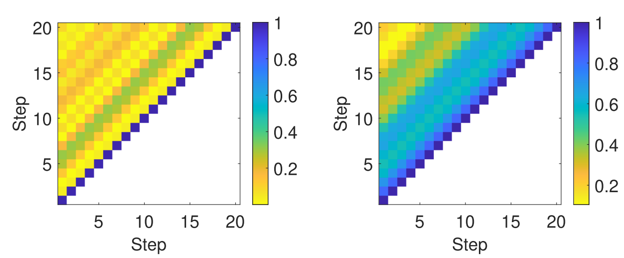
In order to illustrate the feasibility of the sparse IoV data enabled TSE framework and the effectiveness of the CRNet model, we conduct a comprehensive experiment including analyzing the influence of data sparsification, observing the improvement of the recovered estimation of CRNet compared to the Initial, and evaluating the performance of CRNet with the other data recovery methods.

1) Overall Performance: Fig. 7 shows the impact of data sparsification on TSE accuracy. The box plots show the distribution of the mean error for each grid in the map over the day for different levels of data sparsity. As the data becomes sparser, the median, range, and dispersion of the error values increase. As sparsity increases from 5% to 90%, RMSE decreases from 15.66 km/h to 2.71 km/h, and MAE decreases from 7.50 km/h to 0.48 km/h. The values and variations of the RMSE consistently exceed the MAE, suggesting that the sparser the data, the more abnormal errors tend to be generated, which negatively affects the accuracy of TSE.

Table IV presents a comparative evaluation of CRNet and other models for TSE recovery. Model performance is assessed using five metrics across six sparsity settings ranging from 5% to 50%. Each row corresponds to a fixed sparsity level, while the columns report the performance of different models. Compared with the baseline methods, CRNet consistently demonstrates superior recovery performance across all sparsity levels. Specifically, CRNet achieves the lowest RMSE and MAE, together with the highest SSIM and PSNR, indicating more accurate traffic state reconstructions. Notably, the estimation error of CRNet remains around 5 km/h even under highly sparse sensing, which is acceptable for practical TSE applications. Furthermore, under extremely sparse conditions (e.g., 3% data availability), CRNet reduces the RMSE by more than 40% compared with the initial estimation, demonstrating its robustness to severe data sparsification. These results validate the practicality and cost-effectiveness of CRNet, showing that accurate TSE can be achieved using sparse IoV data with substantially reduced sensing overhead.

TABLE IV
THE TSE ACCURACY OF THE CRNET AND BASELINE MODELS

Sparsity	Metric	Model										
		Initial	HA	CNN	ConvLSTM	E3DLSTM	PredRNN	SimVP	TAU	WaST	SimVPv2	CRNet
5%	RMSE (km/h)	15.750	14.681	11.422	9.185	9.132	8.928	8.956	9.410	9.009	9.479	**8.779**
	IPV (%)	*	6.790	27.479	41.685	42.020	43.317	43.137	40.258	42.800	39.818	**44.258**
	MAE (km/h)	7.532	7.140	7.231	5.364	5.375	5.031	5.290	5.644	5.238	5.728	**4.847**
	SSIM (-)	0.256	0.239	0.558	0.725	0.730	0.745	0.749	0.712	0.747	0.708	**0.759**
	PSNR (dB)	11.618	12.229	14.409	16.303	16.353	16.550	16.522	16.094	16.471	16.030	**16.695**
10%	RMSE (km/h)	13.906	12.602	10.518	7.835	8.992	8.069	8.216	8.682	8.537	9.063	**7.796**
	IPV (%)	*	9.376	24.361	43.656	35.331	41.974	40.921	37.571	38.606	34.825	**43.936**
	MAE (km/h)	6.254	5.810	6.307	4.193	5.266	4.360	4.671	5.030	4.876	5.414	**4.170**
	SSIM (-)	0.440	0.437	0.643	0.810	0.738	0.795	0.794	0.761	0.776	0.737	**0.814**
	PSNR (dB)	12.700	13.556	15.125	17.683	16.487	17.428	17.272	16.793	16.938	16.419	**17.727**
20%	RMSE (km/h)	11.445	10.149	9.508	6.726	8.704	7.162	6.984	7.354	7.734	7.653	**6.599**
	IPV (%)	*	11.328	16.924	41.233	23.945	37.424	38.978	35.746	32.426	33.131	**42.343**
	MAE (km/h)	4.672	4.335	5.567	3.334	5.223	3.694	3.703	4.001	4.245	4.219	**3.283**
	SSIM (-)	0.638	0.649	0.721	0.867	0.759	0.845	0.856	0.838	0.821	0.824	**0.873**
	PSNR (dB)	14.392	15.436	16.003	19.009	16.769	18.464	18.682	18.234	17.796	17.888	**19.175**
30%	RMSE (km/h)	9.731	8.618	8.509	5.780	7.983	6.538	6.108	6.465	7.014	6.632	**5.723**
	IPV (%)	*	11.445	12.556	40.600	17.960	32.805	37.230	33.564	27.922	31.849	**41.180**
	MAE (km/h)	3.657	3.494	4.501	2.733	4.701	3.187	3.025	3.401	3.772	3.575	**2.658**
	SSIM (-)	0.747	0.758	0.788	0.904	0.804	0.876	0.892	0.878	0.856	0.870	**0.908**
	PSNR (dB)	15.801	16.857	16.967	20.326	17.521	19.255	19.846	19.353	18.645	19.132	**20.411**
40%	RMSE (km/h)	8.370	7.506	7.654	5.069	7.014	6.055	5.345	5.816	6.333	5.911	**5.036**
	IPV (%)	*	10.333	8.551	39.438	16.203	27.660	36.137	30.513	24.331	29.381	**39.834**
	MAE (km/h)	2.909	2.930	3.933	2.206	3.933	2.817	2.474	2.873	3.243	2.928	**2.267**
	SSIM (-)	0.817	0.824	0.837	0.929	0.853	0.894	0.920	0.904	0.885	0.900	**0.931**
	PSNR (dB)	17.110	18.057	17.887	21.466	18.646	19.922	21.005	20.272	19.532	20.132	**21.523**
50%	RMSE (km/h)	7.198	6.631	6.842	4.599	6.338	5.646	4.709	5.140	5.525	5.200	**4.430**
	IPV (%)	*	7.872	4.936	36.106	11.938	21.561	34.570	28.580	23.231	27.753	**38.446**
	MAE (km/h)	2.307	2.514	3.324	1.892	3.456	2.614	2.160	2.421	2.748	2.490	**1.821**
	SSIM (-)	0.867	0.867	0.874	0.942	0.883	0.909	0.939	0.926	0.914	0.924	**0.948**
	PSNR (dB)	18.421	19.133	18.861	22.312	19.525	20.530	22.105	21.344	20.717	21.244	**22.636**



(a) Head 1. (b) Head 2.

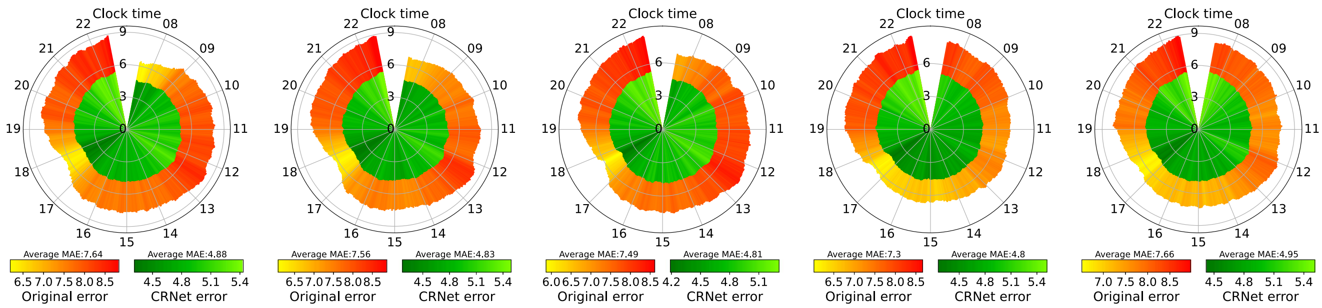
Fig. 8. Weights map of 2 heads: Multi-head mechanism enables effective multi-scale temporal modeling.

Fig. 8 shows the averaged normalized retention weight maps of two heads. Head 1 focuses on near-diagonal regions, indicating a preference for short-term temporal dependencies, whereas head 2 distributes its weights more broadly across historical steps, capturing longer-term dependencies. This complementary behavior confirms that the multi-head mechanism enables effective multi-scale temporal modeling.

2) Temporal Performance: Figs. 9 and 10 show CRNet's MAE per minute of TSE compared to the Initial estimation, spanning five days from 7:30 to 22:30, with sparsities of 5% and 30%, respectively. These figures are plotted in polar coordinates, with the X-axis representing time in minutes and the Y-axis representing MAE in km/h. In each polar plot, the orange area represents the MAE for Initial and the green area represents the MAE for CRNet. These simulations show the excellent stability of CRNet performance. The Initial estimated MAE values are quite high and show a lot of temporal variability, as evidenced by the irregular orange contours. In contrast, CRNet's MAE values show limited temporal variability, with smoother contours in the green region, suggesting that TSE accuracy improves throughout the day. CRNet is able to improve TSE accuracy at any time of the day with stable performance. In addition, CRNet shows good adaptability. Although there is a significant difference in traffic shapes between weekdays and weekends as seen in the raw results, CRNet consistently maintains similar MAE sizes and trends. For different sparsities from 5% to 30%, the improvement of CRNet over the Initial is considerable, and its MAE level is always kept low.

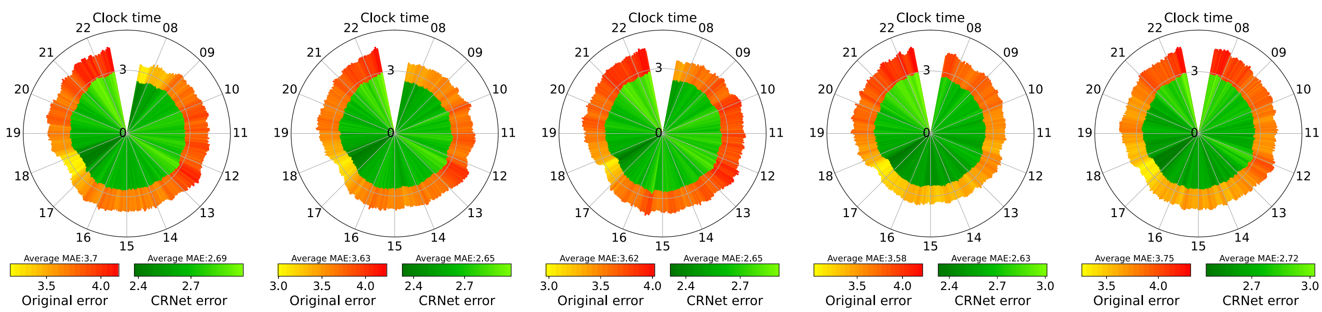
3) Spatial Performance: Fig. 11 shows the traffic states in Beijing at 1:00 PM on Monday under a sparsity level of 5% using sparse IoV data, where data sparsification leads to substantial information loss and consequently degrades TSE accuracy. Fig. 11(a) presents the ideal traffic state estimated from complete IoV data, in which road structures and overall traffic patterns are clearly visible. In contrast, Fig. 11(b) shows the initial estimation derived from 5% sparse IoV data, which contains a large number of missing grid values, resulting in a more incomplete and ambiguous representation. By comparison, the CRNet-recovered estimation in Fig. 11(c) closely matches the ideal traffic state. This demonstrates that CRNet effectively mitigates the adverse effects of data sparsification and substantially improves TSE accuracy under sparse IoV sensing conditions.

Fig. 12 shows the average estimation error for each grid over a 900-minute period on Monday, highlighting the spatial distribution of estimation accuracy. Before recovery, the initial estimation shown in Fig. 12(a) is dominated by orange–red regions, indicating large errors exceeding 20 km/h across most



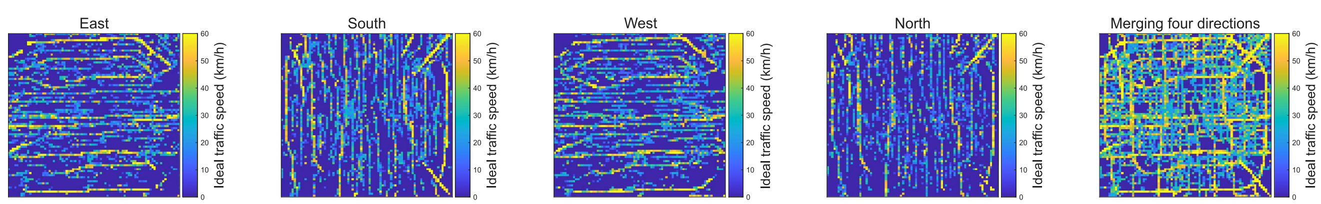
(a) Monday. (b) Wednesday. (c) Friday. (d) Saturday. (e) Sunday.

Fig. 9. Comparison of MAE (km/h) between CRNet and the initial at 5% sparsity: MAE distributions over different time periods across multiple days are shown, demonstrating that CRNet consistently reduces estimation errors compared with the initial estimation under highly sparse IoV data.

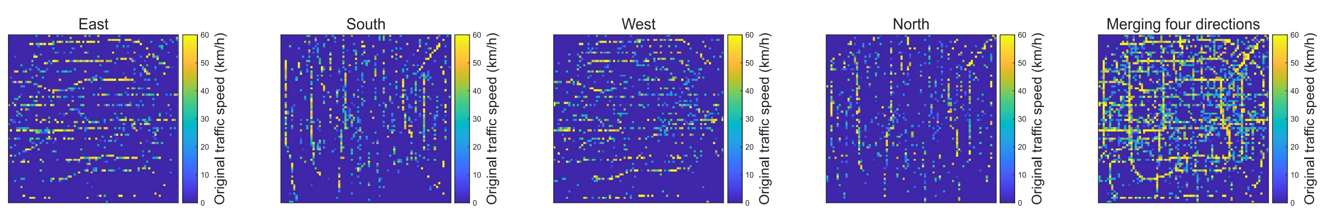


(a) Monday. (b) Wednesday. (c) Friday. (d) Saturday. (e) Sunday.

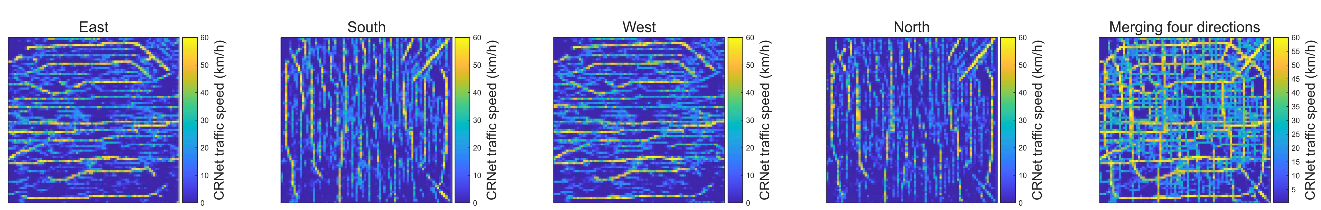
Fig. 10. Comparison of MAE (km/h) between CRNet and the initial at 30% sparsity: MAE distributions over different time periods for multiple days are shown, demonstrating that CRNet consistently outperforms the initial estimation under moderate data sparsity.



(a) Ideal estimation: TSE results when using complete data.

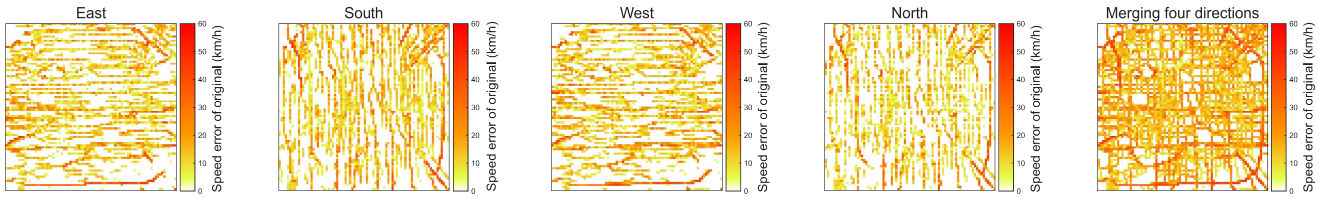


(b) Initial estimation: TSE results when using 5% data.

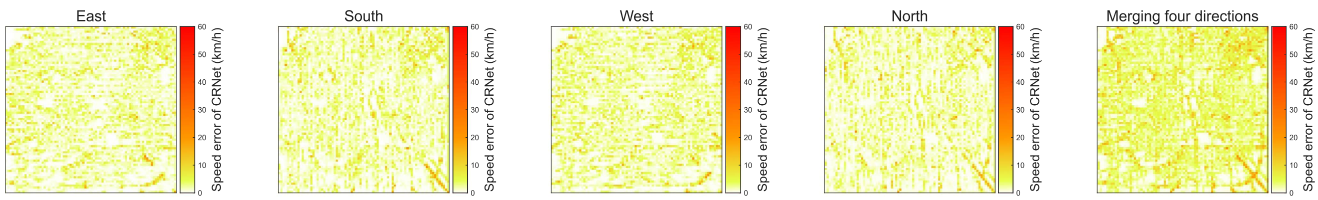


(c) CRNet estimation: TSE results with CRNet when using 5% data.

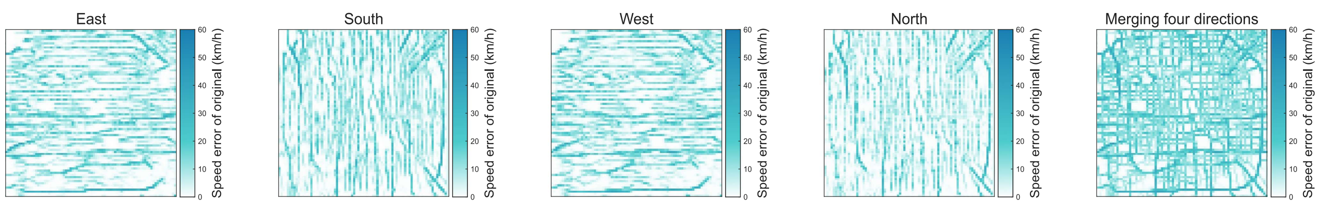
Fig. 11. The TSE outcomes visualization on Monday 1:00 PM at 5% sparsity: Traffic state maps obtained from complete data, sparse initial estimation, and CRNet recovery are compared, illustrating the effectiveness of CRNet in reconstructing spatial traffic patterns under sparse IoV sensing.



(a) Estimation error of Initial: there are large errors between Initial estimation and Ideal estimation.



(b) Estimation error of CRNet: the reconstructed estimation by the proposed CRNet is similar to the Ideal estimation.



(c) Reduced error by CRNet: the estimation errors are greatly reduce by the proposed CRNet.

Fig. 12. The accuracy improvement of CRNet on Monday 1:00 PM at 5% sparsity: Spatial distributions of estimation errors before and after CRNet recovery are shown, highlighting the significant error reduction achieved by the proposed method.

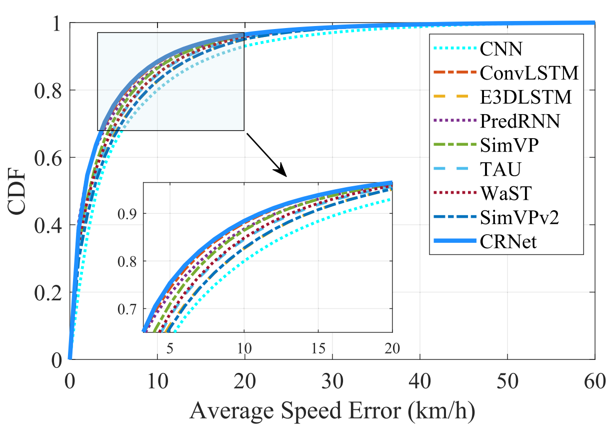
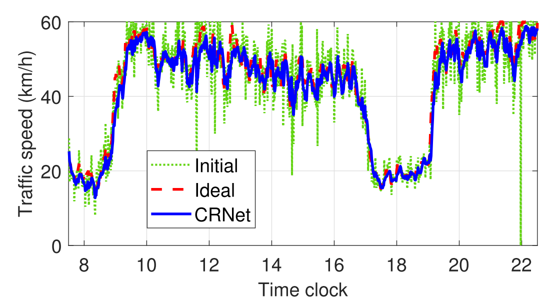
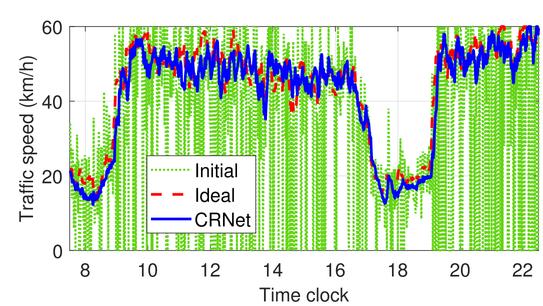


Fig. 13. The CDF of average error on Monday: Average speed error distributions for different models are compared, showing the advantage of CRNet.



(a) TSE results at 20% data sparsity.



(b) TSE results at 5% data sparsity.

Fig. 14. Visualization of estimated traffic speed comparison: CRNet improves estimation accuracy compared with the initial estimation, especially under highly sparse sensing.

grids. After CRNet recovery, Fig. 12(b) shows that the majority of grids shift toward the yellow–green range, with estimation errors largely reduced. Fig. 12(c) depicts the spatial distribution of average error reduction achieved by CRNet, with several regions showing reductions exceeding 50 km/h.

Furthermore, Fig. 13 presents the cumulative distribution function (CDF) of the average estimation error over a 900-minute period on Monday, facilitating a comparative evaluation of CRNet against other DL models. The CDF curves indicate that CRNet consistently yields lower estimation errors than all baseline methods. The magnified inset further confirms that the CDF curve of CRNet remains above those of the other models, demonstrating its clear advantage in achieving accurate TSE under sparse IoV sensing conditions.

4) Individual Grid Performance: Fig. 14 shows the recovered traffic speed under varying data sparsity levels, comparing

CRNet with both the Initial and Ideal estimations. In this figure, when no vehicle data are collected during a given time slot, the Initial speed is set to 0 km/h. At 20% data sparsity, the Initial estimation yields an RMSE of 5.309 km/h, whereas CRNet leverages short-term spatial-temporal correlations to reduce the RMSE to 3.025 km/h. The improvement is even more pronounced at 5% data sparsity: the frequent occurrence of missing data causes the Initial estimations to frequently default to 0 km/h, resulting in an RMSE of 23.321 km/h, while CRNet reduces this error to 4.389 km/h. Notably, although the Initial estimated speed fluctuates around the Ideal speed, these fluctuations become more severe as the available data decreases, thereby degrading estimation accuracy. In contrast, even when data are extremely sparse, CRNet effectively recovers missing information to produce estimates that closely approximate the Ideal speed. These findings demonstrate that CRNet significantly enhances TSE accuracy under limited data conditions, making it a robust solution for real-time TSE in sparse data scenarios.

VI. Conclusion and Future Work

In this paper, we have proposed a cost-effective TSE framework based on sparse IoV data. We have theoretically shown that sparsification errors approximately follow a Gaussian distribution and have reformulated the problem as a sequential data denoising task. To address this challenge, we have developed the CRNet, which jointly exploits spatial and temporal correlations for robust estimation. Extensive experiments on a large-scale real-world IoV dataset have confirmed the feasibility of sparse IoV-based TSE, showing that CRNet achieves reliable performance even with only 5% of data, where the MAE remains below 5 km/h. The main contributions of this study have included the establishment of a cost-effective framework, the provision of theoretical insights into sparsification errors, and the design of an accurate and robust spatial-temporal model. For future work, we will explore diffusion-based techniques as a complementary refinement mechanism to further enhance estimation accuracy and robustness, while carefully considering lightweight or selective designs consistent with the cost-effectiveness objective of this work, and extend the proposed framework to traffic flow prediction for forecasting future traffic dynamics.

References

[1] X. Guo, W. Xing, J. Fang, J. Chen, X. Chen, and R. Gao, "Noise-aware optimization for mobile crowdsensing-based travel time estimation," *IEEE Trans. Veh. Technol.*, vol. 73, no. 3, pp. 4067–4080, Mar. 2024.

[2] D. Yang et al., "Trajectory planning and tracking control for vehicles with tire blowout in complex traffic flows," *Sci. China Inf. Sci.*, vol. 68, no. 3, 2025, Art. no. 132202.

[3] H. Wu, H. Zhou, J. Zhao, Y. Xu, B. Qian, and X. Shen, "Deep learning enabled fine-grained path planning for connected vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 10, pp. 10303–10315, Oct. 2022.

[4] Y. Zhang, C. Li, K. Li, S. He, and J. Chen, "High-resolution large-scale urban traffic speed estimation with multi-source crowd sensing data," *IEEE Trans. Veh. Technol.*, vol. 73, no. 9, pp. 12345–12357, Sep. 2024.

[5] D. Yuan, J. Xue, J. Su, W. Xu, and H. Zhou, "ST-Mamba: Spatial-temporal mamba for traffic flow estimation recovery using limited data," in *Proc. IEEE/CIC Int. Conf. Commun. China*, 2024, pp. 1928–1933.

[6] M. C. Lucic, X. Wan, H. Ghazzai, and Y. Massoud, "Leveraging intelligent transportation systems and smart vehicles using crowdsourcing: An overview," *Smart Cities*, vol. 3, no. 2, pp. 341–361, 2020.

[7] J. Xue, K. Yu, T. Zhang, H. Zhou, L. Zhao, and X. Shen, "Cooperative deep reinforcement learning enabled power allocation for packet duplication URLLC in multi-connectivity vehicular networks," *IEEE Trans. Mobile Comput.*, vol. 23, no. 8, pp. 8143–8157, Aug. 2024.

[8] K. Moghaddasi, S. Rajabi, and F. S. Gharehchopogh, "An enhanced asynchronous advantage actor-critic-based algorithm for performance optimization in mobile edge computing-enabled internet of vehicles networks," *Peer-to-Peer Netw. Appl.*, vol. 17, no. 3, pp. 1169–1189, 2024.

[9] M. Zakarya, L. Gillam, A. A. Khan, O. Rana, and R. Buyya, "ApMove: A service migration technique for connected and autonomous vehicles," *IEEE Internet Things J.*, vol. 11, no. 17, pp. 28721–28733, Sep. 2024.

[10] A. M. Rahmani et al., "Self-learning adaptive power management scheme for energy-efficient IoT-MEC systems using soft actor-critic algorithm," *Internet Things*, vol. 31, 2025, Art. no. 101587.

[11] C. Tu et al., "Adaptive budgeting for collaborative multi-task data collection in online sparse crowdsensing," *IEEE Trans. Mobile Comput.*, vol. 23, no. 7, pp. 7983–7998, Jul. 2024.

[12] S. Taguchi and T. Yoshimura, "Online estimation and prediction of large-scale network traffic from sparse probe vehicle data," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 7233–7243, Jul. 2022.

[13] J. Xue et al., "Sparse mobile crowdsensing for cost-effective traffic state estimation with spatio–temporal transformer graph neural network," *IEEE Internet Things J.*, vol. 11, no. 9, pp. 16227–16242, May 2024.

[14] J. Liu, G. P. Ong, and X. Chen, "GraphSAGE-based traffic speed forecasting for segment network with sparse data," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 1755–1766, Mar. 2022.

[15] Y. Zhang, X. Kong, W. Zhou, J. Liu, Y. Fu, and G. Shen, "A comprehensive survey on traffic missing data imputation," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 12, pp. 19252–19275, Dec. 2024.

[16] L. Zhao et al., "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3848–3858, Sep. 2020.

[17] M. Saberi et al., "A simple contagion process describes spreading of traffic jams in urban networks," *Nature Commun.*, vol. 11, no. 1, pp. 1–9, 2020.

[18] E. Wang, W. Liu, W. Liu, Y. Yang, B. Yang, and J. Wu, "Spatiotemporal urban inference and prediction in sparse mobile crowdsensing: A graph neural network approach," *IEEE Trans. Mobile Comput.*, vol. 22, no. 11, pp. 6784–6799, Nov. 2023.

[19] A. Ali, Y. Zhu, and M. Zakarya, "A data aggregation based approach to exploit dynamic spatio-temporal correlations for citywide crowd flows prediction in fog computing," *Multimedia Tools Appl.*, vol. 80, no. 20, pp. 31401–31433, 2021.

[20] J. Xue, T. Zhang, W. Wu, H. Zhou, and X. Shen, "Sparse Big Data for vehicular network traffic flow estimation: A machine learning approach," in *Proc. IEEE Glob. Commun. Conf.*, 2022, pp. 4959–4963.

[21] S. Wang, J. Cao, and P. S. Yu, "Deep learning for spatio-temporal data mining: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 8, pp. 3681–3700, Aug. 2022.

[22] S. Chen, X. Ji, H. Shao, and Y. Zhang, "A physics-informed traffic state estimation model for freeways under sparse observation data," *Transportmetrica B: Transport Dyn.*, vol. 13, no. 1, 2025, Art. no. 2564701.

[23] L. Li, J. Zhang, Y. Wang, and B. Ran, "Missing value imputation for traffic-related time series data based on a multi-view learning method," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 2933–2943, Aug. 2019.

[24] X. Chen, Z. He, and J. Wang, "Spatial-temporal traffic speed patterns discovery and incomplete data recovery via SVD-combined tensor decomposition," *Transp. Res. Part C: Emerg. Technol.*, vol. 86, pp. 59–77, 2018.

[25] A. Abdelraouf, M. Abdel-Aty, and N. Mahmoud, "Sequence-to-sequence recurrent graph convolutional networks for traffic estimation and prediction using connected probe vehicle data," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 1, pp. 1395–1405, Jan. 2023.

[26] S. Zhang, X. Hu, W. Zhang, J. Chen, and H. Huang, "Learning traffic as videos: A spatio-temporal VAE approach to periodic traffic raster data imputation," *Intell. Data Anal.*, vol. 28, no. 5, pp. 1271–1292, 2024.

[27] O. Gkountouna, D. Pfoser, and A. Züfle, "Traffic flow estimation using probe vehicle data," in *Proc. IEEE Int. Conf. Data Sci. Adv. Analytics*, 2020, pp. 579–588.

[28] X. Huang, W. Wu, S. Hu, M. Li, C. Zhou, and X. Shen, "Digital twin based user-centric resource management for multicast short video streaming," *IEEE J. Sel. Topics Signal Process.*, vol. 18, no. 1, pp. 50–65, Jan. 2024.

[29] A. M. Rahmani, J. Tanveer, F. S. Gharehchopogh, S. Rajabi, and M. Hosseinzadeh, “A novel offloading strategy for multi-user optimization in blockchain-enabled mobile edge computing networks for improved Internet of Things performance,” *Comput. Elect. Eng.*, vol. 119, 2024, Art. no. 109514.
[30] H. Asgharzadeh, A. Ghaffari, M. Masdari, and F. S. Gharehchopogh, “An intrusion detection system on the Internet of Things using deep learning and multi-objective enhanced gorilla troops optimizer,” *J. Bionic Eng.*, vol. 21, no. 5, pp. 2658–2684, 2024.
[31] A. Naghib, F. S. Gharehchopogh, and A. Zamanifar, “A comprehensive and systematic literature review on intrusion detection systems in the internet of medical things: Current status, challenges, and opportunities,” *Artif. Intell. Rev.*, vol. 58, no. 4, 2025, Art. no. 114.
[32] A. Ali, Y. Zhu, and M. Zakarya, “Exploiting dynamic spatio-temporal graph convolutional neural networks for citywide traffic flows prediction,” *Neural Netw.*, vol. 145, pp. 233–247, 2022.
[33] A. B. Said and A. Erradi, “Spatiotemporal tensor completion for improved urban traffic imputation,” *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6836–6849, Jul. 2022.
[34] T. Nie, G. Qin, Y. Wang, and J. Sun, “Correlating sparse sensing for large-scale traffic speed estimation: A Laplacian-enhanced low-rank tensor Kriging approach,” *Transp. Res. Part C: Emerg. Technol.*, vol. 152, 2023, Art. no. 104190.
[35] C. Tan et al., “OpenSTL: A comprehensive benchmark of spatio-temporal predictive learning,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, vol. 36, pp. 69819–69831.
[36] J. Xue, T. Jiang, H. Huang, Z. Zhang, H. Zhou, and X. Shen, “Enabling data-driven OTFS modulation for 6G hyper reliable high-mobility communications,” *IEEE Wireless Commun.*, vol. 32, no. 3, pp. 196–203, Jun. 2025.
[37] Y. Yuan, J. Ding, J. Feng, D. Jin, and Y. Li, “A universal pre-training and prompting framework for general urban spatio-temporal prediction,” *IEEE Trans. Knowl. Data Eng.*, vol. 37, no. 5, pp. 2212–2225, May 2025.
[38] Y. Sun et al., “LOSEC: Local semantic capture empowered large time series model for IoT-enabled data centers,” *IEEE Internet Things J.*, vol. 12, no. 10, pp. 13144–13156, May 2025.
[39] A. Ali, Y. Zhu, and M. Zakarya, “Exploiting dynamic spatio-temporal correlations for citywide traffic flow prediction using attention based neural networks,” *Inf. Sci.*, vol. 577, pp. 852–870, 2021.
[40] J. Xue et al., “FD-RAN empowered multiple base stations cooperative ISAC for low-altitude networks,” *IEEE Trans. Netw. Sci. Eng.*, vol. 13, pp. 6927–6943, 2026.
[41] N. Awan et al., “Modeling dynamic spatio-temporal correlations for urban traffic flows prediction,” *IEEE Access*, vol. 9, pp. 26502–26511, 2021.
[42] J. Zhao et al., “Adaptive transparent cloaking tunnel enabled by meta-reinforcement-learning metasurfaces,” *PhotoniX*, vol. 7, no. 1, 2026, Art. no. 2.
[43] A. Vaswani et al., “Attention is all you need,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 6000–6010.
[44] T. Zhang et al., “Handover-free multi-connectivity mobility management for downlink FD-RAN: A hierarchical DRL-based approach,” *IEEE Trans. Cogn. Commun. Netw.*, vol. 11, no. 2, pp. 1281–1296, Apr. 2025.
[45] J. Jiang, C. Han, W. X. Zhao, and J. Wang, “PDformer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction,” in *Proc. AAAI Conf. Artif. Intell.*, 2023, pp. 4365–4373, vol. 37.
[46] Y. Sun et al., “Retentive network: A successor to transformer for large language models,” 2023, *arXiv:2307.08621*.
[47] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, “Convolutional LSTM network: A machine learning approach for precipitation nowcasting,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, vol. 28, pp. 802–810.
[48] Y. Wang et al., “PredRNN: A recurrent neural network for spatiotemporal predictive learning,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 2208–2225, Feb. 2023.
[49] Z. Gao, C. Tan, L. Wu, and S. Z. Li, “SimVP: Simpler yet better video prediction,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 3160–3170.
[50] J. A. Rice, *Mathematical Statistics and Data Analysis*. Noida, India: Cengage Learning, 2006.
[51] C. Tan et al., “Temporal attention unit: Towards efficient spatiotemporal predictive learning,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 18770–18782.
[52] C. Tan, Z. Gao, S. Li, and S. Z. Li, “SimVPv2: Towards simple yet powerful spatiotemporal predictive learning,” *IEEE Trans. Multimedia*, vol. 27, pp. 5170–5184, 2025.
[53] X. Nie et al., “Wavelet-driven spatiotemporal predictive learning: Bridging frequency and time variations,” in *Proc. AAAI Conf. Artif. Intell.*, 2024, vol. 38, pp. 4334–4342.
[54] Y. Wang, L. Jiang, M.-H. Yang, L.-J. Li, M. Long, and L. Fei-Fei, “Eidetic 3D LSTM: A model for video prediction and beyond,” in *Proc. Int. Conf. Learn. Representations*, 2019.

Jianzhe Xue (Graduate Student Member, IEEE) received the BS degree in communication engineering from Xidian University, Xi’an, China, in 2021. He is currently working toward the PhD degree with the School of Electronic Science and Engineering, Nanjing University, China. He was a joint PhD student with the Broadband Communications Research Group, Department of Electrical and Computer Engineering, University of Waterloo, Canada. His current research interests include internet of vehicles, intelligent signal processing, and machine learning for wireless communications. He was the recipient of the IEEE Daniel E. Noble Fellowship Award.

Dongcheng Yuan (Graduate Student Member, IEEE) received the BS degree in electronic information science and technology in 2024 from Nanjing University, Nanjing, China, where he is currently working toward the MS degree with the School of Electronic Science and Engineering. His current research interests include intelligent transportation system and Internet of Vehicles.

Yu Sun received the BS degree in communication engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2022. He is currently working toward the PhD degree with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. His current research interests include machine learning and optimization for data center, and resource allocation for wireless communications.

Tianqi Zhang received the BS degree in electronical information science and technology in 2021 from Nanjing University, Nanjing, China, where he is currently working toward the PhD degree in information and communication engineering. His current research interests include dynamic resource scheduling, mobility management, and machine learning in the field of emerging wireless networks.

Wenchao Xu (Member, IEEE) received the BE and ME degrees from Zhejiang University, Hangzhou, China, in 2008 and 2011, respectively, and the PhD degree with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, in 2018. In 2011, he joined Alcatel Lucent Shanghai Bell Company Ltd., where he was a software engineer for telecom virtualization. He is currently an assistant professor with the Division of Integrative Systems and Design, Hong Kong University of Science and Technology (HKUST), Hong Kong. He has authored or coauthored more than 100 research papers at top-tier journals including *Proceedings of IEEE*, *IEEE Transactions on Parallel and Distributed Systems*, *IEEE Transactions on Mobile Computing*, and prestigious conferences including IEEE/CVF CVPR, ICML, NeurIPS, IEEE INFOCOM, and IEEE ICDCS. His research interests include mobile computing, AI enabled networking, multimodal learning, and Internet of Vehicles.

Haibo Zhou (Fellow, IEEE) received the PhD degree in information and communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014. From 2014 to 2017, he was a postdoctoral fellow with the Broadband Communications Research Group, Department of Electrical and Computer Engineering, University of Waterloo. He is currently a full professor with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. His research interests include resource management and protocol design in B5G/6G networks, vehicular ad hoc networks, and space-air-ground integrated networks. He was the recipient of the 2019 IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award. He was a Track/Symposium co-chair of IEEE/CIC ICCC 2019, IEEE VTC-Fall 2020, IEEE VTC-Fall 2021, and IEEE GLOBECOM 2022. He is currently an associate editor for the *IEEE Transactions on Wireless Communications*, *IEEE Internet of Things Journal*, *IEEE Network Magazine*, and *IEEE Wireless Communications Letter*.

Xuemin Shen (Fellow, IEEE) received the PhD degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is currently a university professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include network resource management, wireless network security, Internet of Things, 5G and beyond, and vehicular ad hoc and sensor networks. He is a registered professional engineer of Ontario, Canada, an Engineering Institute of Canada fellow, a Canadian Academy of Engineering fellow, a Royal Society of Canada fellow, a Chinese Academy of Engineering foreign member, and a distinguished lecturer of the IEEE Vehicular Technology Society and Communications Society. Dr. Shen was the recipient of the Canadian Award for Telecommunications Research from the Canadian Society of Information Theory (CSIT) in 2021, R.A. Fessenden Award in 2019 from IEEE, Canada, Award of Merit from the Federation of Chinese Canadian Professionals (Ontario) in 2019, James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society, Joseph LoCicero Award in 2015 and Education Award in 2017 from the IEEE Communications Society, Technical Recognition Award from Wireless Communications Technical Committee (2019) and AHSN Technical Committee (2013), the Excellent Graduate Supervision Award in 2006 from the University of Waterloo, and the Premiers Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada. He was the technical program committee chair/cochair for IEEE Globecom 16, IEEE Infocom14, IEEE VTC10 Fall, IEEE Globecom07, and the chair for the IEEE Communications Society Technical Committee on Wireless Communications. He is also the president of the IEEE Communications Society. He was the vice president of Technical & Educational Activities, vice president of Publications, member-at-Large on the Board of Governors, chair of the Distinguished Lecturer Selection Committee, member of IEEE Fellow Selection Committee of the ComSoc. He was the editor-in-chief of the *IEEE Internet of Things Journal*, *IEEE Network*, and *IET Communications*.