Real-Time Traffic Incident Detection with Sparse Observations: A Masked Spatiotemporal Graph Learning Framework

1st Xiaolei Zhu
Department of Civil and Environmental Engineering
National University of Singapore
Singapore
xiaolei.z@u.nus.edu

2nd Qiaowen Bai
Department of Civil and Environmental Engineering
National University of Singapore
Singapore
isebq@nus.edu.sg

3rd Ghim Ping Ong
Department of Civil and Environmental Engineering
National University of Singapore
Singapore
ceeongr@nus.edu.sg

4th Biplab Sikdar
Department of Electrical and Computer Engineering
National University of Singapore
Singapore
bsikdar@nus.edu.sg

***Abstract*—Timely and accurate incident detection is essential for maintaining the efficiency and resilience of transportation systems, especially under sparse sensing conditions. This paper proposes a Masked Spatiotemporal Graph Learning (MSTGL) framework for robust incident detection using sparse traffic observations. The framework integrates a spatial graph attention encoder to model localized traffic interactions and a BERT-based temporal encoder to capture evolving temporal dynamics. A masked supervision strategy is introduced during training to enhance the predictive capability of the traffic state decoder, which forecasts future traffic conditions from spatiotemporal embeddings. On top of this, we propose a dual-window anomaly detection module, which utilizes both short-term and long-term prediction residuals to identify incident-induced deviations in real time. The proposed method is evaluated on a campus-scale traffic network in Singapore with sparse fixed-point camera coverage. Experimental results demonstrate that the proposed framework significantly improves detection performance, achieving 17.76% higher F1-score compared to state-of-the-art baselines.**

Index Terms*—intelligent transportation systems, incident detection, sparse observations, spatiotemporal graph learning, masked supervision.

I. Introduction

Stable road traffic operation is essential for maintaining the overall efficiency of transportation networks. However, abnormal incidents—such as vehicle breakdowns, roadway obstructions, or traffic accidents—inevitably occur, occupying limited road resources and causing serious disruptions. These incidents undermine network stability and lead to significant delays. Although it is impossible to completely prevent such events, rapid and effective responses can significantly mitigate their negative impacts [1]. Consequently, the ability to detect incidents in real time is critically important. Automatic Incident Detection (AID), as a key component of intelligent transportation systems, plays a vital role in enhancing road safety and improving operational efficiency.

Automatic incident detection methods can be broadly categorized into direct and indirect approaches [2]. Direct detection methods identify abnormal events by applying image processing techniques to determine whether incidents have occurred within a localized detection area. While this approach is straightforward, it is impractical to install and maintain camera coverage across an entire road network. In contrast, indirect detection methods infer the occurrence of incidents by capturing abnormal changes in traffic flow parameters, such as speed, volume, and occupancy, resulting from an incident. Given the widespread deployment of traffic monitoring sensors (e.g., cameras and ground loop detectors) across the road network in modern cities, traffic flow data can be easily and continuously collected [3]. Taking advantage of existing infrastructure, indirect detection methods provide a cost-effective solution for achieving real-time automatic incident detection [4]. However, a notable characteristic of indirect detection is that traffic sensors are often sparsely distributed, particularly in large-scale traffic networks [5]. This results in significant spatial coverage gaps, introducing major challenges for accurate and timely incident identification.

A lot of studies have explored indirect detection methods under sparse observation conditions, which can be broadly categorized into three main types: traffic parameter threshold-based methods, traffic feature-based classification methods, and time-series prediction-based methods.

For traffic parameter threshold-based methods, incidents are detected by monitoring traffic flow parameters (e.g., volume, speed) and identifying deviations beyond predefined thresholds. Typical examples include the California Algorithm and dynamic threshold set [6], [7]. However, these methods require extensive site-specific threshold calibration, limiting their adaptability across different networks, and also exhibit low robustness to noise, leading to a high false detection rate [8]. With the rise of machine learning techniques, traffic feature-based classification methods have gained increasing attention. Traffic features under normal and incident condi-

The authors express their gratitude to the financial support of A*STAR, CISCO Systems (USA) Pte. Ltd and National University of Singapore under its Cisco-NUS Accelerated Digital Economy Corporate Laboratory (Award I21001E0002). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of grantors.

tions are distinguished by learning from historical traffic flow data through classifier training such as Bayesian probability classification [6], probabilistic neural networks [10], Random Forests [19], and Support Vector Machines [12]. Although these methods offer improved performance, they suffer from challenges such as heavy dependence on feature selection, class imbalance due to the scarcity of incident data, and the requirement for large amounts of labeled historical data. To address these limitations, prediction-based methods have been proposed, which take time-series traffic parameters as input to predictive models, generate expected values under normal conditions, and identify incidents by analyzing deviations between predicted and observed values. The effectiveness of prediction-based methods relies on maintaining stable forecasting of normal traffic patterns, even when unexpected incidents occur. Techniques such as Kalman filtering [13], wavelet analysis [14], and neural network-based [15] forecasting are commonly used to achieve robust short-term traffic prediction. Additionally, robust criteria for incident detection are essential, often involving the development of spatiotemporal discriminators to assess whether observed deviations indicate abnormal events [16], [17]. Despite the above achievements, these methods often fail to capture the complex spatiotemporal dependencies in real-world traffic networks, particularly under conditions of sparse observations.

To address this issue, we develop a Masked Spatiotemporal Graph Learning (MSTGL) framework to enable robust incident detection under the challenges of sparse sensor deployments and limited traffic observability. The proposed method consists of two core modules: traffic state prediction and incident identification. During traffic state prediction, a spatial graph attention encoder is employed to capture localized spatial dependencies, and a BERT-inspired temporal transformer encoder is used to model dynamic temporal patterns from partially observed traffic sequences. The resulting spatiotemporal contextualized embeddings are then passed to a prediction decoder, which maps them to the expected traffic states at the next time window. To enhance the model's ability to learn generalized spatiotemporal correlations under sparse observations, a masked supervision strategy is integrated into the training process. The outputs of the traffic states are then passed to the dual-window incident identification module, which performs real-time detection by modeling the distribution of relative errors over short-term and long-term windows and measuring their divergence from a historical reference using Jensen–Shannon divergence. At last, numerical experiments are conducted based on a campus-scale traffic network in Singapore to evaluate the performance of the proposed method.

II. Problem Definition

We consider a traffic monitoring system deployed over a road network within a defined area. Traffic incidents—such as accidents, lane blockages, or vehicle breakdowns—may occur at any road segment, leading to abnormal traffic patterns that propagate across the network. Due to infrastructure limitations, only a subset of mid-segment locations is equipped with fixed-point cameras that continuously collect real-time traffic state data. The objective of this system is to infer, in real-time, potential traffic incidents occurring throughout the entire road network, based on sparse observations from mid-segment cameras.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{A})$ denote a road network, where $\mathcal{V} = \{1, \ldots, v, \ldots, N\}$ denotes the set of N road segments in the studied traffic network, and $\mathcal{E}$ represents the set of edges which indicates the pairwise connectivity among road segments. $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ denotes the spatial adjacency matrix, where $[\boldsymbol{A}]_{i,j}$ is set to 1 if road segments v_i and v_j are physically connected, and 0 otherwise. Let $\mathcal{O} \subseteq \mathcal{V}$ denote the set of observed segments, i.e., the road segments equipped with mid-segment cameras, and $\mathcal{U} = \mathcal{V}\backslash\mathcal{O}$ denote the set of unobserved segments.

At each time step t, the system collects sparse traffic information on observed links and performs incident detection. We define the observation matrix $\boldsymbol{X}_t$ to represent traffic states across observed segments in $\mathcal{O}$: $\boldsymbol{X}_t = \{\boldsymbol{x}_t^v | v \in \mathcal{O}\} \in \mathbb{R}^{F\times|\mathcal{O}|}$, where $\boldsymbol{x}_t^v = \{x_t^{v,c} | c = 1, \ldots, F\} \in \mathbb{R}^F$ denotes the feature vector (e.g., traffic volume, speed, occupancy) for road segment v at time t. A sequence of past observations is represented by $\boldsymbol{X}_{1:t} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_t\}$. Now we formulate the incident detection task over the road-network graph $\mathcal{G}$, which is decomposed into two tightly coupled subproblems:

- **Traffic State Prediction.** Given historical traffic observations $\boldsymbol{X}_{1:t}$ from the observed segments $\mathcal{O}$, the first subproblem aims to predict the future state of each segment $v \in \mathcal{V}$ at the next timestep $t + 1$. Formally, we seek to learn a prediction function $f_{\boldsymbol{A}}$:
$$\hat{\boldsymbol{x}}_{t+1}^v = f_{\boldsymbol{A}}\left(\boldsymbol{X}_{1:t}\right), \qquad \forall v \in \mathcal{V}, \tag{1}$$
where $\hat{\boldsymbol{x}}_{t+1}^v \in \mathbb{R}^F$ denotes the predicted traffic state vector at segment v. For unobserved segments ($v \in \mathcal{U}$), predictions must rely solely on learned spatiotemporal correlations and historical context from observed segments $\mathcal{O}$.
- **Incident Identification.** Given the predicted and observed traffic states at time $t+1$, the second subproblem aims to detect abnormal traffic incidents on each segment $v \in \mathcal{V}$. The short-term and long-term error distributions are compared with a historical reference via Jensen–Shannon Divergence (JSD). Formally, we define a JSD-based anomaly score $\hat{\mathcal{D}}_{t+1}^v \in [0, 1]$. An incident is flagged if:
$$\hat{\mathcal{D}}_{t+1}^v \geq \theta, \tag{2}$$
where $\theta \in [0, 1]$ is the incident detection threshold.

III. Methodology

A. Masked Spatiotemporal Prediction Framework

To address the traffic state prediction problem outlined above, we propose a Masked Spatiotemporal Graph Modeling framework. The framework integrates Graph Attention Networks (GAT) and BERT-style Transformer encoders to

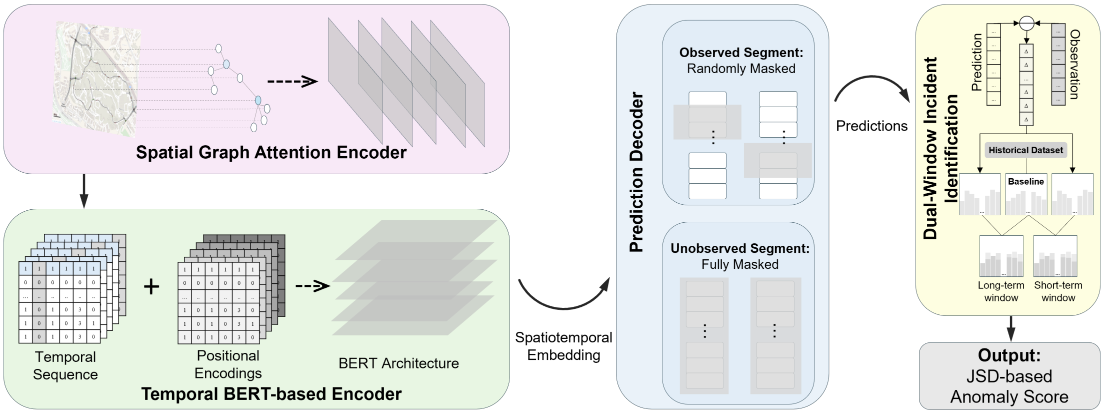

Fig. 1: The Proposed Masked Spatiotemporal Graph Learning Framework.

capture complex spatial correlations and dynamic temporal dependencies from sparse traffic observations. Specifically, the model predicts traffic states for all segments using only inputs on observed segments and is optimized via a masked loss that supervises only the masked outputs. The methodology consists of three major components: (1) Spatial Graph Attention Encoder, (2) Masked Temporal Transformer Encoder, and (3) Prediction Decoder.

1) Spatial Graph Attention Encoder: The road network $\mathcal{G}$ is encoded spatially using a GAT to aggregate spatially correlated traffic states. For each timestep t, we first encode the observed traffic features over all segments (with artificial masking as detailed in subsequent sections). A GAT layer transforms the feature vector of each segment v by attending over its spatial neighbors $\mathcal{N}(v)$, as defined by the adjacency matrix $\boldsymbol{A}$, to a spatial embedding $\boldsymbol{h}_t^v \in \mathbb{R}^d$:

$$\boldsymbol{h}_t^v = \sigma\left(\sum_{u \in \mathcal{N}(v)\cup\{v\}} \alpha_{v,u} \boldsymbol{W}\boldsymbol{x}_t^u\right), \quad \forall v \in \mathcal{V}, \tag{3}$$

where $\mathbf{W} \in \mathbb{R}^{d\times F}$ is a learnable projection matrix, $\sigma(\cdot)$ is a non-linear activation, and $\alpha_{v,u}$ denotes the attention weight between segment v and neighbor u, computed as:

$$\alpha_{v,u} = \frac{\exp\left(\mathrm{LeakyReLU}(\mathbf{a}^\top[\boldsymbol{W}\boldsymbol{x}_t^v \| \boldsymbol{W}\boldsymbol{x}_t^u])\right)}{\sum_{k\in\mathcal{N}v\cup\{v\}} \exp\left(\mathrm{LeakyReLU}(\mathbf{a}^\top[\boldsymbol{W}\boldsymbol{x}_t^v \| \boldsymbol{W}\boldsymbol{x}_t^k])\right)}, \tag{4}$$

with $\mathbf{a} \in \mathbb{R}^{2d}$ being a learnable attention vector and $\|$ denoting vector concatenation. The aggregated embeddings for all segments at time t are denoted as $\mathbf{H}_t = [\boldsymbol{h}_t^v]_{v\in\mathcal{V}} \in \mathbb{R}^{N\times d}$, which compactly summarize spatial contexts.

2) Temporal BERT-based Encoder: To model the temporal dependencies across traffic states, we adopt a Transformer encoder inspired by the BERT architecture, which is particularly well-suited for masked modeling tasks with irregular and sparse inputs. For each road segment $v \in \mathcal{V}$, we collect its spatial embeddings a historical window of length T: $\boldsymbol{H}_t^v = [\boldsymbol{h}_\tau^v]_{\tau=t-T+1:t} \in \mathbb{R}^{T\times d}$, where $\boldsymbol{h}_\tau^v \in \mathbb{R}^d$ is the output of the spatial encoder at time τ for segment v. We enrich this sequence with learnable positional encodings $\mathbf{P} \in \mathbb{R}^{T\times d}$ added element-wise to encode time-step indices, obtaining: $\bar{\mathbf{H}}_t^v = \mathbf{H}_t^v + \mathbf{P} \in \mathbb{R}^{T\times d}$. The BERT encoder with L transformer layers, including self-attention and feedforward networks, processes $\bar{\mathbf{H}}_t^v$ and outputs contextualized embeddings:

$$\widetilde{\mathbf{H}}_t^v = \mathrm{BERT}_\theta(\bar{\mathbf{H}}_t^v) \in \mathbb{R}^{T\times d}, \tag{5}$$

where $\widetilde{\mathbf{H}}_t^v = [\widetilde{h}_\tau^v]_{\tau=t-T+1:t}$. We extract the final embedding $\widetilde{\boldsymbol{h}}_t^v$ as the spatiotemporally contextualized representation for segment v at time t.

3) Prediction Decoder: The prediction decoder maps the contextualized representation $\widetilde{\boldsymbol{h}}_t^v \in \mathbb{R}^d$ to the predicted traffic feature vector $\hat{\boldsymbol{x}}_{t+1}^v \in \mathbb{R}^F$:

$$\hat{\boldsymbol{x}}_{t+1}^v = \mathcal{H}_\theta(\widetilde{\boldsymbol{h}}_t^v), \quad \forall v \in \mathcal{V}. \tag{6}$$

By stacking the outputs across all segments, we obtain the predicted traffic state matrix $\hat{\mathbf{X}}_{t+1} \in \mathbb{R}^{N\times F}$.

B. Training with Masked Supervision

To enable robust learning under partial observability, we adopt a masked supervision strategy that simulates real-world sensor sparsity and encourages the model to learn generalized spatiotemporal patterns. During training, we assume access to simulated ground-truth traffic states for all segments $v \in \mathcal{V}$, allowing the model to learn to predict both observed and unobserved segments.

At each timestep $t+1$, we construct a binary training mask vector $\boldsymbol{m}_{t+1}^{\text{train}} \in \{0,1\}^N$ for the prediction at time $t+1$, where $\boldsymbol{m}_{t+1}^{\text{train}}[v]$ is set to 1 if segment v is masked and should be predicted, and 0 otherwise. The training mask is constructed such that: (1) All unobserved segments $v \in \mathcal{U}$ are always masked, reflecting the absence of real-time data in deployment. (2) A random subset of observed segments $v \in \mathcal{O}$ is also masked to improve generalization.

Given the prediction produced by the masked spatiotemporal prediction framework $\hat{\boldsymbol{x}}_{t+1}^v$ and the ground-truth traffic feature vector $\boldsymbol{x}_{t+1}^v \in \mathbb{R}^F$ collected at time $t+1$, the model is trained by minimizing a masked reconstruction loss:

$$\mathcal{L} = \frac{1}{\sum_v \mathbf{m}_{t+1}^{\text{train}}[v]} \sum_{v\in\mathcal{V}} \mathbf{m}_{t+1}^{\text{train}}[v] \cdot \left\|\hat{\boldsymbol{x}}_{t+1}^v - \boldsymbol{x}_{t+1}^v\right\|^2. \tag{7}$$

C. Dual-Window Incident Identification

Given the traffic state prediction outputs, we propose a dual-window detection framework that identifies abnormal traffic conditions based on the deviation between predicted and observed values. For each segment $v \in \mathcal{V}$ and traffic feature $c \in \{1, \ldots, C\}$, the element-wise relative error at time t+1 is defined as:

$$\mathcal{E}_{t+1}^{v,c} = \frac{|\hat{x}_{t+1}^{v,c} - x_{t+1}^{v,c}|}{x_{t+1}^{v,c} + \epsilon}, \tag{8}$$

where $\hat{x}_{t+1}^{v,c}$ and $x_{t+1}^{v,c}$ denote the predicted and observed traffic states, respectively, and ϵ is a small constant to avoid division by zero. The matrix $\mathcal{E}_{t+1} \in \mathbb{R}^{|\mathcal{V}| \times F}$ collects relative errors for all segment-feature pairs at time $t+1$.

To balance sensitivity to short-term disruptions and robustness against noise, we construct two temporal windows: a short-term window $\mathbb{S}$ and a long-term window $\mathbb{L}$, both aligned up to time t. A short-term sliding window $\mathbb{S}$ captures rapid fluctuations by including the three most recent time steps, while a long-term sliding window $\mathbb{L}$ provides more stable error evaluation by including the six most recent time steps. Each sliding window $\mathbb{W} \in \{\mathbb{S}, \mathbb{L}\}$ consists of a sequence of relative error matrices from consecutive time steps, so we have $\mathbb{W} = \{\mathcal{E}_{t-|\mathbb{W}|+1}, \ldots, \mathcal{E}_t\}, \quad \mathbb{W} \in \mathbb{R}^{|\mathbb{W}| \times \mathcal{V} \times C}$. For each segment-feature pair (v, c), the corresponding sequence of errors is extracted as: $\mathbb{W}^{v,c} = [\mathcal{E}_\tau^{v,c}]_{\tau \in t-|\mathbb{W}|+1:t}$.

To characterize the distributional behavior of prediction errors, we apply Gaussian kernel density estimation (KDE) to each error sequence:

$$\hat{p}_{\mathbb{W}}^{v,c}(x) = \frac{1}{|\mathbb{W}| \cdot h} \sum_{i=1}^{|\mathbb{W}|} \kappa\left(\frac{x - \mathbb{W}_i^{v,c}}{h}\right), \tag{9}$$

where h is the bandwidth parameter, $\kappa(\cdot)$ is the Gaussian kernel, and $\hat{p}_{\mathbb{W}}^{v,c}(x)$ denotes the estimated probability density of relative errors in window $\mathbb{W}$.

To assess the abnormality of current observations, each window's KDE is compared to a reference distribution $\hat{p}_{\mathbb{N}}^{v,c}(x)$, constructed offline from historical data under normal traffic conditions. Notably, the reference KDE $\hat{p}_{\mathbb{N}}^{v,c}$ serves as a prior and remains fixed during detection. However, during real deployment, only a subset of segments $\mathcal{O}$ are observed. For unobserved segments $\mathcal{U} = \mathcal{V} \setminus \mathcal{O}$, their input features $x_{t+1}^{v,c}$ is unavailable. To overcome this, we impute these unavailable observations using predictions from a separate model trained under full supervision. At each time $t+1$, this auxiliary model takes the ground-truth observations from observed segments as input and outputs predictions for the unobserved segments. These predicted values are then used as pseudo-observations $\tilde{x}_{t+1}^{v,c}$ to compute relative errors.

We compare the divergence of the reference distribution with that obtained from both the short-term and long-term windows. If both windows exhibit significantly different distributions from the reference, this may indicate an abnormal traffic state and suggest the potential occurrence of an incident. To quantify the discrepancy between the error distribution in the current window and the reference distribution, we employ the Jensen–Shannon divergence (JSD). For each segment-feature pair (v, c) and sliding window $\mathbb{W} \in \{\mathbb{S}, \mathbb{L}\}$, the JSD-based anomaly score is computed as:

$$\begin{aligned} \mathcal{D}_{\mathbb{W}}^{v,c} = &\frac{1}{2} \int \hat{p}_{\mathbb{W}}^{v,c}(x) \log \frac{2\hat{p}_{\mathbb{W}}^{v,c}(x)}{\hat{p}_{\mathbb{W}}^{v,c}(x) + \hat{p}_{\mathbb{N}}^{v,c}(x)} dx \\ &+ \frac{1}{2} \int \hat{p}_{\mathbb{N}}^{v,c}(x) \log \frac{2\hat{p}_{\mathbb{N}}^{v,c}(x)}{\hat{p}_{\mathbb{W}}^{v,c}(x) + \hat{p}_{\mathbb{N}}^{v,c}(x)} dx. \end{aligned} \tag{10}$$

The feature-level divergences are averaged to obtain a segment-level score: $\hat{\mathcal{D}}_{\mathbb{W}}^v = \frac{1}{C} \sum_{c=1}^{C} \delta(\mathcal{D}_{\mathbb{W}}^{v,c})$. A final composite JSD-based anomaly score $\hat{\mathcal{D}}^v$ is calculated via a weighted fusion of short-term and long-term divergence:

$$\hat{\mathcal{D}}^v = \alpha \cdot \hat{\mathcal{D}}_{\mathbb{S}}^v + (1 - \alpha) \cdot \hat{\mathcal{D}}_{\mathbb{L}}^v, \tag{11}$$

where α is a weighting coefficient. Finally, an incident is identified on segment v if: $\hat{\mathcal{D}}^v \geq \theta$, where θ is a predefined detection threshold.

IV. Numerical Experiments

In this section, we utilize the traffic network of a large campus area in Singapore as an example to experiment with and evaluate the performance of the proposed method. A living lab is currently under development at this campus, with a lot of cameras being installed along bus stops, providing a low-risk, high-impact opportunity to pilot next-generation transportation technologies with direct applicability to national mobility strategies. Fig. 2 shows the layout of the testbed traffic network and the locations of the deployed cameras. Although the video-based data collection allows direct observation of real-time conditions at the sensor locations, the spatial distance between bus stops, combined with the campus's hilly and winding terrain, results in many blind spots where camera coverage is unavailable. Moreover, due to the frequent heavy rain and thunderstorms in Singapore, there have been past incidents where fallen trees blocked roads without timely detection, leading to prolonged disruptions. The proposed method can be used to address such challenges by enabling the timely detection of abnormalities through traffic data, allowing management authorities to respond promptly and effectively.

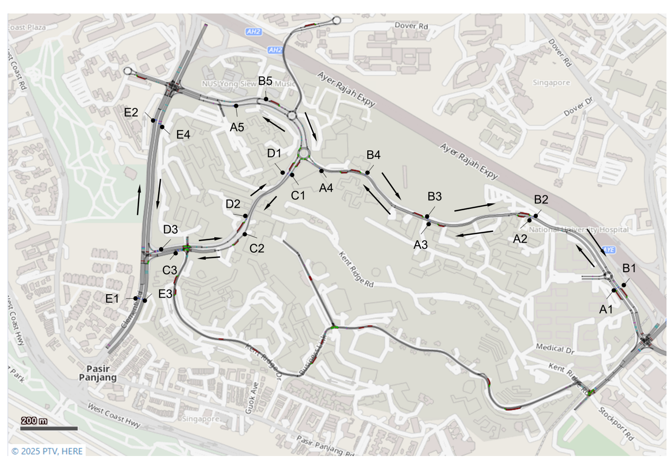

Fig. 2: Testbed Traffic Network with Camera Deployment.

A. Data Collection

Since the proposed method requires historical data covering both normal and abnormal traffic conditions for effective modeling and testing, and the sensing infrastructure has only recently been deployed, large-scale data collection is still ongoing. To address this limitation, for the normal dataset, we calibrated a range of traffic volumes based on the available collected data and synthetically generated 600 hours of traffic data by randomly sampling within this calibrated range. The synthetic data covers a variety of normal traffic conditions and was simulated using VISSIM on the campus network. Traffic parameters at each data collection point were recorded every five minutes. As a result, approximately 3,000 data points were obtained for each location, forming the foundation for subsequent model training and evaluation. The dataset was split into training, validation, and test sets during the training process with a ratio of 8:1:1.

We simulated incident scenarios by creating multiple incident blocks at different locations along various links in the network and triggering them at varying times over a 30-hour simulation period, using the first hour as a warm-up. To capture a wide range of traffic conditions, traffic volumes entering from A1, B5, and D3 were randomly varied between 200 veh/h and 1500 veh/h for each simulation hour. For each simulation run, a total of four incidents were generated, and the process was repeated across 30 random seeds. For each scenario, traffic parameters were recorded at designated data collection points, and the trained model was employed to detect the occurrence of incidents.

B. Result and Discussion

After training the model with 600 hours of historical data, we obtained a reasonably accurate prediction model with an MAPE under 13.47%. Using this trained model, we applied it to detect the simulated incident scenarios. We selected a weighting coefficient $\alpha = 0.7$ and a global threshold $\theta = 0.68$ by grid search. The KDE bandwidth parameter h was calculated by Scott's rule. The proposed framework is capable of maintaining second-level inference speeds for each run, which is sufficient for practical applications. Fig. 3 illustrates the performance of the prediction module under one such scenario. As shown in the figure, several error peaks emerge around the time of the incident. The model maintains high prediction accuracy under normal traffic conditions, while errors become noticeable during incident periods. Importantly, the model continues to follow the underlying normal traffic trend, making the anomaly caused by the incident more distinguishable through the prediction errors. This validates the feasibility of using the distribution of prediction errors as an indirect indicator for incident detection, supporting the effectiveness of our proposed method.

Fig. 4 illustrates an example of the JSD-based anomaly scores across different simulated incident scenarios, which occurred at the 3rd, 10th, 21st, and 25th hour of a 30-hour simulation period, located at segments A2–A3, A1–A2, A3, and A4–A5, respectively. The window length used in this study is 5 minutes [18]. The short-term and long-term horizons include 3 and 6 windows, respectively, which are commonly used in traffic prediction. As shown in the heatmap, four distinct regions with elevated JSD-based anomaly scores emerge, indicating abnormal shifts in the prediction error distribution that correspond to potential incidents at specific spatiotemporal locations.

The predicted incident matrix clearly highlights the affected regions across both time and space. Notably, these extended abnormal regions are largely due to the lagging impact of traffic disruptions, and thus, early detection, particularly at the onset of an incident, is more critical. From the incident matrix, we observe that each incident block begins at a specific segment: the second incident block begins at A1–A2, the third at A3–A4, and the fourth at A4–A5. These starting points, located at the top–left of each spatiotemporal block, indicate the initial locations affected by the incidents and align well

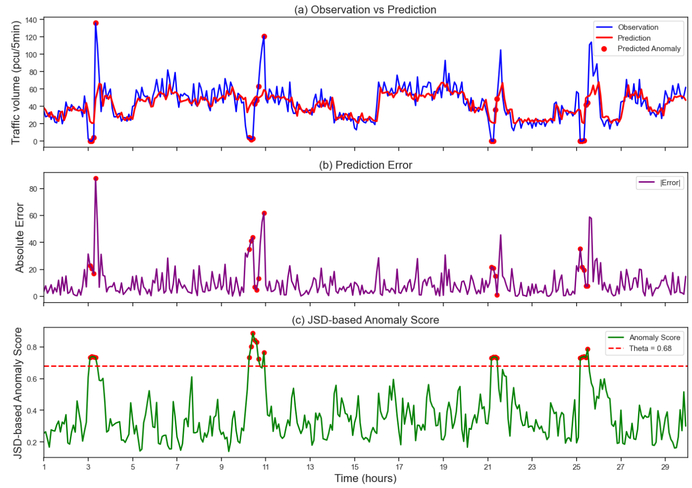

Fig. 3: Traffic State Prediction vs. Observation under an Incident Scenario Simulation.

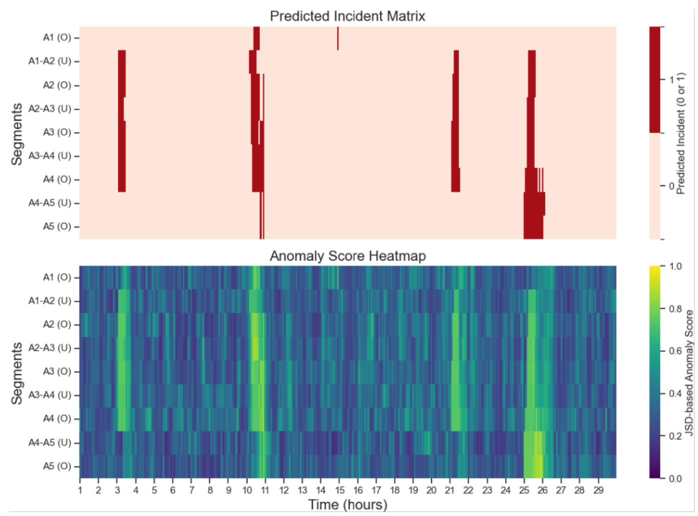

Fig. 4: Incident Identifications under an Incident Scenario Simulation.

with the simulation settings. For the first incident, although the detected region spans a broader range from A1 to A5, the true incident location at A2–A3 is still correctly covered within the identified block. Additionally, a false positive is detected at segment A1, but it only lasts for a single time window.

TABLE I: Performance comparison of different detection methods

Model	True Positive Rate	False Alarm Rate	F1 Score
Random Forest	0.7925	0.0250	0.6804
Kalman Filter	0.8926	0.0480	0.6103
LSTM + Anomaly Eval.	0.9134	0.0393	0.6624
GSTPL Network	0.9413	0.0339	0.7048
Poposed MSTGL Framework	**0.9653**	**0.0161**	**0.8348**

We further benchmarked the proposed method against several existing approaches adapted to our incident detection setting. Table I reports the true positive rate, false alarm rate, and F1 score for representative baselines, including Random Forest [19], Kalman Filter [20], Long-short-term-memory network + anomaly evaluation (LSTM+AE) [21], Graph Spatio-Temporal Pattern Learning (GSTPL) Network [2]. While most methods, except for Random Forest, achieve relatively high detection rates, they also suffer from high false alarm rates. In contrast, the method proposed in this study achieves both a high detection rate and a low false alarm rate, resulting in the highest F1 score among all compared approaches—an improvement of 17.76% over STGAD, the best-performing existing method.

V. Conclusions

In this paper, we proposed an MSTGL framework for real-time incident detection under sparse traffic observations. The framework integrates a spatial graph attention encoder and a BERT-inspired temporal transformer encoder to extract spatiotemporal features from partially observed traffic data, and employs a prediction decoder to map these embeddings to expected traffic states under normal conditions, with a masked supervision strategy incorporated during training to enhance generalization under sensor sparsity. A dual-window incident identification module then detects anomalies in real time by estimating the distribution of relative errors over both short-term and long-term windows, and measuring their divergence from a reference distribution constructed from historical normal traffic conditions. The proposed framework is further validated on a campus-scale traffic network in Singapore with limited camera coverage, achieving satisfactory performance by identifying the incident segment with high detection accuracy and low false alarm rates. Beyond traffic operations, the method has broader applicability for supporting timely incident detection by transportation authorities and emergency response agencies.

Future work may focus on three key directions: integrating vehicle-mounted sensor data to enhance detection robustness and applying the framework to real-world incident datasets; extending the approach to urban traffic networks with more complex patterns; and improving the method to characterize and classify incidents such as accidents and roadblocks by their unique spatiotemporal characteristics.

References

[1] W. Zhang, H. Zha, L. Gan, H. Wang, and T. Wang, "Incident detection for freeways based on a dual-state traffic factor state network," *Transportmetrica A: Transport Science*, pp. 1–29, 202.

[2] H. Li, Y. Ma, X. Wang, and Z. Li, "Graph spatiotemporal pattern learning network for real-time road network traffic abnormal incident detection," *Transportation Research Record*, vol. 2677, no. 12, pp. 815–829, 2023.

[3] T. Nie, G. Qin, L. Sun, W. Ma, Y. Mei, and J. Sun, "Contextualizing MLP-mixers spatiotemporally for urban traffic data forecast at scale," *IEEE Transactions on Intelligent Transportation Systems*, early access, 2024, doi: 10.1109/TITS.2024.

[4] J. Evans, B. Waterson, and A. Hamilton, "Evolution and future of urban road incident detection algorithms," *Journal of Transportation Engineering, Part A: Systems*, vol. 146, no. 6, p. 03120001, 2020.

[5] Q. Wang and K. Yang, "Privacy-preserving data fusion for traffic state estimation: A vertical federated learning approach," *Transportation Research Part C: Emerging Technologies*, vol. 168, p. 104743, 2024.

[6] Z. Li, J. Sun, and J. Jia, "A dynamic threshold-based traffic incident detection method considering traffic flow fluctuation characteristics," *Transportation Research Record*, vol. 2675, no. 9, pp. 528–538, 2021.

[7] Y. Zhang, J. Hu, and Y. Yang, "Real-time freeway incident detection based on adaptive threshold and traffic states," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 5674–5684, 2022.

[8] H. Li, "Research on traffic flow operation risk evaluation based on deep learning," Ph.D. dissertation, Transportation College, Jilin University, Changchun, China, 2021.

[9] Q. C. Liu, J. Lu, S. Y. Chen, *et al.*, "Multiple Naive Bayes classifiers ensemble for traffic incident detection," *Mathematical Problems in Engineering*, vol. 2014, p. 383671, 2014.

[10] G. W. Liu, H. L. Jin, J. Z. Li, *et al.*, "A Bayesian deep learning method for freeway incident detection with uncertainty quantification," *Accident Analysis & Prevention*, vol. 176, p. 106796, 2022.

[11] N. Dogru and A. Subasi, "Traffic accident detection using random forest classifier," in *Proc. 15th Learning and Technology Conf.*, New York, USA, 2018, pp. 40–45.

[12] J. L. Xiao, "SVM and KNN ensemble learning for traffic incident detection," *Physica A: Statistical Mechanics and its Applications*, vol. 517, pp. 29–35, 2019.

[13] M. Saeedmanesh, A. Kouvelas, and N. Geroliminis, "An extended Kalman filter approach for real-time state estimation in multi-region MFD urban networks," *Transportation Research Part C: Emerging Technologies*, vol. 132, p. 103384, 2021.

[14] Y. S. Jeong, M. Castro-Neto, M. K. Jeong, and L. D. Han, "A wavelet-based freeway incident detection algorithm with adapting threshold parameters," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 1, pp. 1–19, 2011.

[15] H. Ghahremannezhad, H. Shi, and C. Liu, "Real-time accident detection in traffic surveillance using deep learning," in *Proc. IEEE Int. Conf. Imaging Systems and Techniques (IST)*, Kaohsiung, Taiwan, 2022, pp. 1–6, doi: 10.1109/IST55454.2022.9827736.

[16] L. Deng, D. Lian, Z. Huang, and E. Chen, "Graph convolutional adversarial networks for spatiotemporal anomaly detection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 6, pp. 2416–2428, Jun. 2022, doi: 10.1109/TNNLS.2021.3136171.

[17] M. D. Stojanović, M. S. Stanković, and M. S. Stanković, "Spatiotemporal road traffic anomaly detection: A tensor-based approach," *Applied Sciences*, vol. 11, no. 24, pp. 12017, Dec. 2021, doi: 10.3390/app112412017.

[18] K. Zhao, D. Guo, M. Sun, C. Zhao, H. Shuai, and C. Shao, "Short-term traffic flow prediction based on hybrid decomposition optimization and deep extreme learning machine," *Physica A: Statistical Mechanics and its Applications*, vol. 647, Art. no. 129870, Aug. 2024.

[19] N. Dogru and A. Subasi, "Traffic accident detection using random forest classifier," in *Proc. 15th Learning and Technology Conf.*, Jeddah, Saudi Arabia, IEEE, 2018.

[20] R. Wang, D. B. Work, and R. Sowers, "Multiple model particle filter for traffic estimation and incident detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 12, pp. 3461–3470, 2016.

[21] Z. Zhao, W. Chen, X. Wu, P. C. Chen, and J. Liu, "LSTM network: A deep learning approach for short-term traffic forecast," *IET Intell. Transp. Syst.*, vol. 11, no. 2, pp. 68–75, 2017.