

SIGN: A Scalable Incident Detection, Localization, and Severity Estimation using
Graph Neural Networks

by

Kaustubh Manoj Harapanahalli

A Thesis
Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved July 2025 by the
Graduate Supervisory Committee

Aviral Shrivastava, Chair
Nakul Gopalan
Ashif Iquebal

ARIZONA STATE UNIVERSITY
August 2025

ABSTRACT

Prior art has primarily focused on advancing incident detection for freeways. In contrast, incident detection on arterial roads typically occurs only at locations within the field of view of sensors or through the aggregation of sensor data. In all of these works, incident detection between sensors where there is no direct access to the occurrence of incidents is not a primary focus. Additionally, most incident detection-related implementations are either localized to the location where they are deployed or, if implemented on a city scale, are transductive, which challenges the real-world requirement of flexibility due to observed technical issues in the real world.

This thesis presents SIGN - a scalable deep-learning approach for classifying, localizing, and estimating the severity of traffic incidents occurring between intersections, using data captured from sparse sensor placements. By representing the road network as a dynamic graph through hidden embedding projections, it becomes inherently inductive, allowing a single model trained on one network to be deployed across varied topologies without requiring retraining. This approach explicitly models the spatial relationships between intersections, enabling a more comprehensive and rapid analysis of traffic patterns and addressing edge-case scenarios in real-world deployments, such as sensor failures and road closures.

Given the existing limitations in data availability, SIGN also presents a synthetic methodology for generating microscopic data. SIGN achieves a detection rate of 99%, with a false alarm rate of less than 11%, and a mean-time-to-detect incidents of 116 seconds on average in an urban environment with sensors at approximately 20% of traffic intersections. Additionally, the utilization of graph neural networks enables our network to be scalable, allowing for studies related to scalable incident detection, localization, and severity estimation. This paves the way for practical incident management systems in dynamic urban environments.

DEDICATION

To my friends and family, for their support and love.

ACKNOWLEDGMENTS

I would like to express my most sincere gratitude to my advisor, Dr. Aviral Shrivastava, for his active and patient support throughout my Thesis. I am also grateful to Edward Andert, whose guidance was invaluable through this journey.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES.....	vii
CHAPTER	
1 INTRODUCTION	1
2 RELATED WORKS	5
Macroscopic Data Bottleneck	5
Limitations Of Existing Incident Detection And Localization Methods	6
3 METHODOLOGY	8
Synthetic Data Simulations	8
Model Baseline Using TabNet	11
Graph Construction And Data Preparation	13
Definition Of The Graph Network	13
Feature Engineering Of The Dataset For Model Training	13
Incident Detection, Localization, And Severity Estimation Model	
Design	15
Node Feature Projection	15
Message Passing.....	17
Graph Pooling And Output Heads.....	18
Training Objective and Hybrid Loss Function	18
4 EXPERIMENTS	21
Dataset Simulation And Preprocessing	21
Model Training And Evaluation Considerations	21
5 RESULTS	23
SIGN Detects Traffic Incidents Better In Arterial Roads	23

SIGN Can Infer On Varying Road Network Topologies Without Requiring Retraining	23
6 CONCLUSION	26
REFERENCES	27

LIST OF TABLES

Table		Page
2.1	Summary of Incident Detection Literature and Its Presented Metrics. SIGN Has Shown the Best Detection and False Alarm Rates for Incident Detection in Arterial Roads.	6
4.1	Model Training Hyperparameters For TabNet.	21
4.2	Model Training Hyperparameters For SIGN.	22
4.3	Traffic Incident Detection Metrics and Their Definitions Based on Confusion Matrix, Where TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives.	22
5.1	Performance Comparison Between The Previous State-Of-The-Art Model And Our Baseline Using TabNet.....	23
5.2	Comparison Of Incident Detection, Top-3 Localization, And Severity Estimation Between Transductive And Inductive Approaches Obtained Using TabNet and GNN-based Models Respectively.	24

LIST OF FIGURES

Figure		Page
3.1	Comparison between the final curve (orange line) obtained through the optimized equation 3.1 and the original ground truth data, which is the aggregated vehicle count.	9
3.2	Shows a Tempe, AZ region selected as the test road network map for generating the simulations. All the points represent possible locations where sensors can be deployed, and the red dots indicate the intersections from where the sensor data was collected during simulation runs.	10
3.3	Represents the TabNet model encoder. The learnable mask is a trainable embedding that performs soft selection of crucial features from the data, the feature transformer processes the output from the learnable mask, and the attentive transformer learns to compute attention scores for each feature.	11
3.4	The block diagram for the baseline model training and inference. The raw data is pre-processed using the rolling window mechanism, as described in the section above, Synthetic Data Simulations.	12
3.5	Representation of the generated graph with all the necessary features. The data is generated to capture the node and edge features with a matching dimension of K . The connection between Junctions 4 and 6 is omitted as we are considering only major arterial roads.	14
3.6	High-level representation of the model architecture designed for model training. The goal of the model is to simultaneously detect if an incident is occurring, if so, localize the road segment where the incident is located, and estimate the severity of the incident.	16

3.7	Computation of Traffic Flow Conservation Loss during the training process, which takes the graph’s node-level predictions and computes the loss to fine-tune the network training to consider traffic flow conservation, allowing enhanced training.	19
-----	---	----

Chapter 1

INTRODUCTION

Traffic incidents are one of the most expensive sources of non-recurring traffic congestion, and they translate ordinary stretches of road into a high-risk environment. According to the 2023 study by the National Highway Traffic Safety Administration (NHTSA) Blincoe *et al.* (2022), traffic incidents account for a staggering \$1.4 trillion in expenses when accounting for lost productivity, medical costs, legal expenses, property damage, and other such socio-economic impacts. Studies of freeway service-patrol programs show that reducing the incident occurrence to clearance time by three to five minutes lowers the chances of a secondary incident by approximately 20% Salum *et al.* (2021). While the reverse scenario is that for every minute a traffic incident remains uncleared, it increases the probability of a secondary traffic incident by 2.8% per minute. Early detection is therefore imperative, and each one-minute reduction in emergency response time provides an improved survival chance by up to 6% Chen *et al.* (2023b). Furthermore, clearing traffic incidents more quickly reduces the probability of secondary traffic incidents by 21%.

Today, incident occurrence detection relies heavily on emergency calls, patrol sightings, or the use of perception sensors mounted at regular intervals on freeways or at intersections. According to the FHWA study Pecheux *et al.* (2023) conducted across ten states, it was found that only 30% of the reported traffic incidents were verified, indicating loopholes in the current system for detecting incidents. This also implies high false-alarm rates, slower confirmation of real incidents, and skewed performance metrics.

Hence, multiple research works have tried to tackle the problem of early traffic

incidents detection, such as Coursey *et al.* (2024), Sun *et al.* (2022), Yang *et al.* (2023), Han *et al.* (2020), Chen *et al.* (2023a), Zhu *et al.* (2018), to name a few. The research has been conducted primarily on two different types of data representations: macroscopic and microscopic. Macroscopic data is typically an aggregation of metrics collected through various sensor modalities, including loop detectors, Microwave Radar Sensors, and Passive Infrared Sensors. Microscopic data are fine-grained features that can be extracted from sensors such as cameras, GPS Ren *et al.* (2024), and Bluetooth, allowing for fine-grained modeling of driver behaviors, traffic interactions, and other related phenomena. Along with the type of sensor modality, incident detection is also dependent on the kind of road, specifically freeways, arterial roads, and internal urban streets.

Most incident detection work has focused on freeways using microscopic sensor-based modalities. While there have been works focused on arterial roads for incident detection, they primarily detect incidents within the field of view of the deployed sensors and utilize a macroscopic sensor modality, which limits the performance accuracy. In contrast, incidents that occur between intersections are usually not detected or identified by sensor modalities. The macroscopic datasets make it hard to locate incidents between intersections due to the aggregated representation of data features.

Given the vast improvements in computer vision technologies for traffic monitoring at intersections, the deployment of cameras and their utility for traffic use cases for urban roads and freeways have increased drastically. They also present the primary advantage of allowing for microscopic data to be derived from the captured video streams, such as individual vehicle speed, vehicle locations, vehicle occupancy on separate lanes, direction of vehicle movement, along with unique vehicle identifiers. Though cameras have great feature representation, deploying them to increase the road’s field of view coverage to 100% would be expensive. Hence, in this the-

sis, we aim to enhance the utility of data captured by microscopic sensors, such as cameras, at traffic intersections to detect traffic incidents that have occurred between intersections. We do so through the following contributions:

- A traffic simulation approach that generates realistic traffic flow, and allows capturing data features similar to the ones that could be derived through cameras. Our process utilizes readily available traffic count metrics provided by verified sources, such as the Department of Transportation. It generates traffic flow models that can be used to capture microscopic data through traffic simulators, such as SUMO.
- A novel approach that allows detection, localization, and severity estimation of traffic incidents between intersections without requiring the incident to be within the field of the sensors. The localization determines the exact road edge on which the incident occurs.
- An inductive graph neural network-based approach that incorporates the spatial features through a trainable projection, allowing for inductivity of the trained model. This allows for inference on road network topologies that contain a higher or lower number of intersections than the number on which the model is trained. This ensures the smooth execution of the model during scenarios such as sensor failures and road blockages.

We generate a dataset for a region of Tempe, Arizona, encompassing 12 urban arterial roads, classified as primary, secondary, and tertiary arterial road networks, which cover an area of approximately 4 square miles. We simulate the data for 31 days. We observed a detection rate of 99%, a mean-time-to-detect of 116 seconds, and a false alarm rate of 11.94%. The top-3 localization accuracy is 85.57%, and the severity estimation accuracy is 87.77%. The model demonstrates a tested capability

to infer with a sensor sparsity ratio ranging from 82% to 93%.

Chapter 2

RELATED WORKS

Macroscopic Data Bottleneck

Most data-driven incident-detection studies utilize macroscopic datasets, such as PEMS-Bay Cuturi (2011), I-880 Skabardonis *et al.* (1996), and METR-LA Li *et al.* (2018). These datasets are generated using inductive-loop detectors embedded in high-speed roadways and provide only aggregated metrics, such as mean speed, occupancy, and volume, at intervals of every five to ten minutes. Such coarse temporal granularity smooths out the rapid fluctuations that often precede an incident. Such aggregation also strips away vehicle-level cues (e.g., headway variance, abrupt deceleration, or lane-change bursts) that are highly discriminative in urban settings. Moreover, because the sensors are almost exclusively installed on freeways, the resulting data fail to represent arterial road networks where many urban crashes occur. Consequently, models trained on these benchmarks tend to exhibit inflated false-alarm rates and poor transferability to city streets.

To overcome these limitations, we capture data through a microscopic simulation that provides representations at both the individual vehicle level and the individual lane level, while allowing for generic macroscopic data representations for comparison through possible aggregations in the simulated data. This approach to data simulation enables us to inject the spatial and temporal details necessary for robust detection of node and edge-level features in graph networks, while avoiding the prohibitive cost of large-scale urban sensor deployment. This allows us to generate more realistic road topology features during model training.

Limitations Of Existing Incident Detection And Localization Methods

Table 2.1: Summary of Incident Detection Literature and Its Presented Metrics. SIGN Has Shown the Best Detection and False Alarm Rates for Incident Detection in Arterial Roads.

Work	Region	Dataset	DR	FAR	MTTD
Liang <i>et al.</i> (2022)	Highway	Macroscopic	88.09 %	2.80 %	26.80 sec
Chen <i>et al.</i> (2023a)	Highway	Macroscopic	99.33 %	6.50 %	NA
Han <i>et al.</i> (2020)	Urban	Microscopic	86.40 %	8.69 %	61 sec
Zhu <i>et al.</i> (2018)	Urban	Macroscopic	86.6 %	5.12 %	NA
Yang <i>et al.</i> (2023)	Both	Macroscopic	80 %	4.68 %	450 sec
Atilgan <i>et al.</i> (2023)	Highway	Macroscopic	74 %	7.6 %	300 sec
Ours	Both	Microscopic	99 %	11.94 %	116 sec

Several studies have addressed incident detection on arterial roads. The Table 2.1 presents a high-level comparison between the previous works and our study. However, upon closer analysis, we observe their limitations in terms of scalability and spatial awareness.

Many early approaches have often relied on macroscopic data. Methods such as Yu *et al.* (2015); Han *et al.* (2020) employed threshold-based pattern matching, while Zhu *et al.* (2018) utilized a convolution-based model architecture to train on the macroscopic data. While these studies have advanced the field, they often face challenges such as high false alarm rates in dynamic conditions, long detection intervals due to 5-minute data aggregation cycles, or a focus on freeways that does not translate well to complex arterial networks.

Although inductive loops are deployed at intersections to capture macroscopic data, they present the same limitations for urban incident detection. However, the

deployment of camera modules at intersections facilitated traffic analysis Yu *et al.* (2021); Shah *et al.* (2018). However, this is limited to the field of view of the cameras. According to the report Pecheux *et al.* (2023), more than 50% of traffic incidents occur in metropolitan regions. Given that between two intersections, the cameras at the intersections capture only about 10% to 20% of the area, incidents that occur outside this region are not captured. Along with these, the existing research implementations Shah *et al.* (2018); Yu *et al.* (2021); Zhu *et al.* (2018) focus on a fixed network dimension, which increases the challenge of deployment as they are transductive.

Chapter 3

METHODOLOGY

Synthetic Data Simulations

Given the requirement to simulate microscopic data, we begin by utilizing a 24-hour real-world vehicle count on the roads between the selected intersections, as provided by the Tempe Department of Transportation, and then generate an aggregation of these sparse vehicle counts. We then apply Fast Fourier Transforms Cooley and Tukey (1965) to this aggregated value to obtain a frequency-based relation to the traffic density on the roads. To simplify the simulation process, we consider the top two frequency components from the decomposed FFT signal and formulate a non-linear equation that can approximately model the average traffic behavior over time, as represented in Equation 3.1.

$$f(t; \theta) = A_1 \sin(\omega_1 t + \phi_1) + A_2 \sin(\omega_2 t + \phi_2) + A_3 + \alpha \quad (3.1)$$

To estimate the parameters for the Equation 3.1, we employ the Levenberg-Marquardt algorithm Levenberg (1944); Marquardt (1963). The goal is to find the particular parameter vector $\theta = [A_1, \omega_1, \phi_1, A_2, \omega_2, \phi_2, A_3]$, that minimizes the squared difference between the formulated traffic model equation 3.1 represented by $f(t; \theta)$ and the aggregated traffic vehicle counts $y(t)$:

$$\min_{\theta} \sum_{i=1}^n [y(t_i) - f(t_i; \theta)]^2 \quad (3.2)$$

The Levenberg-Marquardt algorithm 3.4 updates the parameters by solving the lin-

earized equation:

$$(J^T J + \lambda I) \delta = J^T \mathbf{r} \quad (3.3)$$

where δ is the parameter update, λ is the damping factor, $\mathbf{r}$ is the residual vector which is the difference between the ground truth and the predicted values, and J is the Jacobian matrix of the partial derivatives of the residuals with respect to each parameter, which can be represented as:

$$J_{ij} = \frac{\partial r_i(\theta)}{\partial \theta_j} \forall J \in \mathbb{R}^{(n,m)} \quad (3.4)$$

where m is the total number of parameters and n is the total number of data points.

This allows the Levenberg-Marquardt algorithm to optimize for the best approximations for the parameters that partially model the true traffic values. Figure 3.1 indicates the comparison between the final curve (orange line) obtained through the optimized equation 3.1 and the original ground truth data.

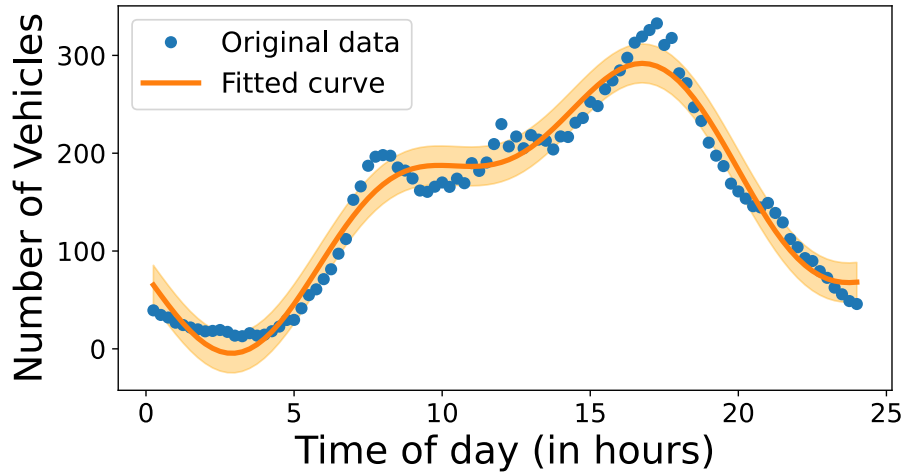


Figure 3.1: Comparison between the final curve (orange line) obtained through the optimized equation 3.1 and the original ground truth data, which is the aggregated vehicle count.

With the simulation running using the optimized 3.1 equation, traffic incidents are injected into the simulation process according to a probability score on a pre-defined scale. Whenever the relevant probability value is generated, a random vehicle

is selected in the running simulation, and either a stalled vehicle or a multi-vehicle crash incident is injected. To simulate realistic traffic behavior, a slowdown radius is enforced depending on the type of incident. Throughout the simulation execution, data is continuously collected from select intersections, as indicated in Figure 3.2, which presents inherent sensor sparsity in the collected data.

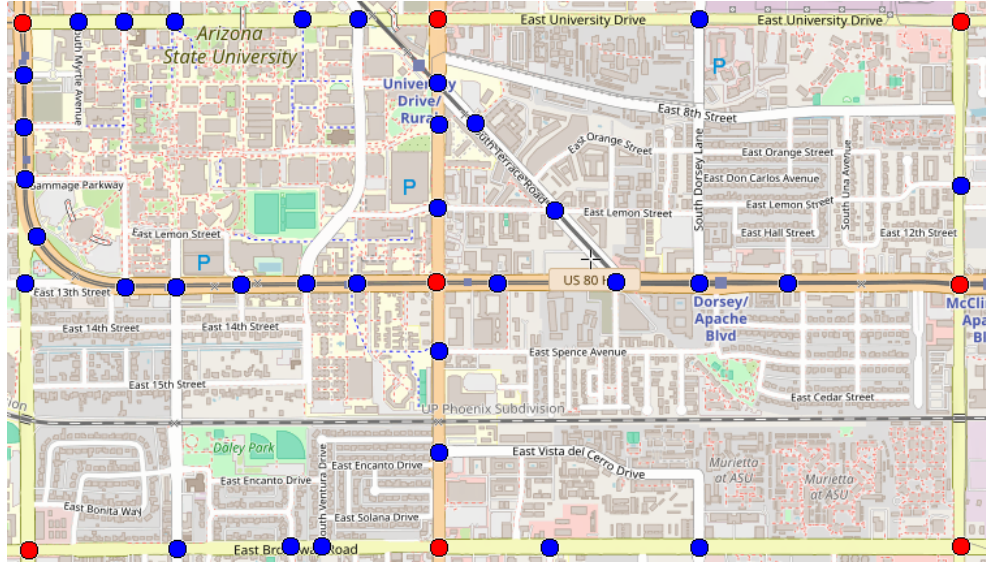


Figure 3.2: Shows a Tempe, AZ region selected as the test road network map for generating the simulations. All the points represent possible locations where sensors can be deployed, and the red dots indicate the intersections from where the sensor data was collected during simulation runs.

As there is sparsity in terms of collected incident data, we pre-process the data using a multitude of methods involving a rolling window mechanism to reduce impact of unscheduled vehicle stops, using vehicle re-identification to compute travel time of vehicles between all possible combinations of two contiguous intersections, depending on the considered intersections, resulting in additional features based on junction mean speed, vehicle count and vehicle occupancy enriching the data for model training, completing the entire data simulation process.

Model Baseline Using TabNet

To start with the model training, we consider the TabnetArik and Pfister (2019) model architecture as our baseline model, given its performance on temporal time-series data, which partially aligns with our data. The main reason for selecting TabNet as the baseline is due to the long-range sequences in the data (up to 2592000 seconds, equivalent to 30 days), as well as the history of performance of transformer models. During the training, the data is processed by the transformer encoder of TabNet, which uses an attentive transformer block, a learnable mask, and a feature transformer block. The learnable mask is a trainable embedding that performs soft selection of crucial features from the data, the feature transformer processes the output from the learnable mask, and the attentive transformer learns to compute attention scores for each feature. The overall output feature aggregation is used to predict the results. Figure 3.3 shows the architecture of the TabNet encoder, which is derived from its original paper Arik and Pfister (2019).

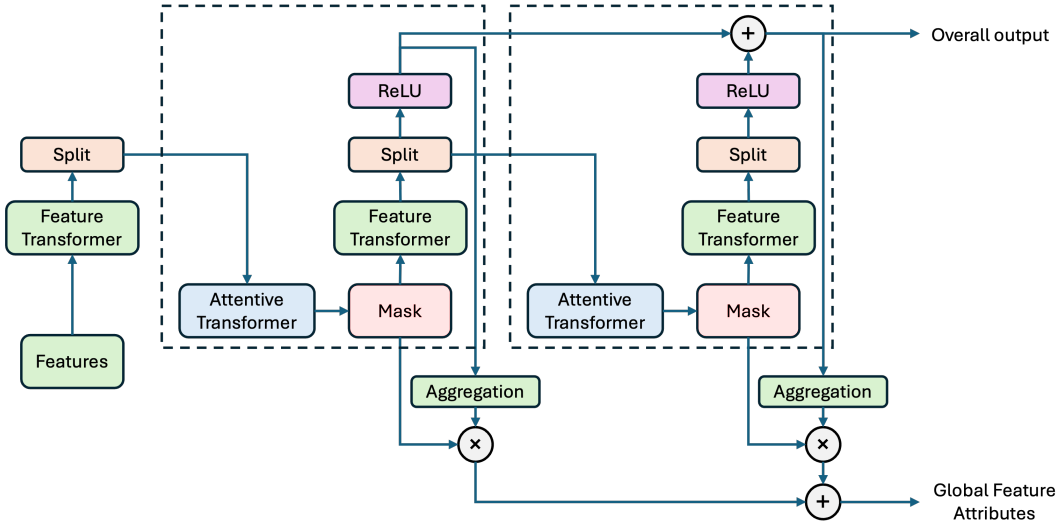


Figure 3.3: Represents the TabNet model encoder. The learnable mask is a trainable embedding that performs soft selection of crucial features from the data, the feature transformer processes the output from the learnable mask, and the attentive transformer learns to compute attention scores for each feature.

For our use case of the three tasks - incident detection, localization, and severity estimation, we employ an ensemble of three different TabNet models. Figure 3.4 shows the implementation of the training of the baseline. We begin with the processed data, as described in the Synthetic Data Simulations section, and use it to train three individual models for each task. However, during inference, the incident localization and incident severity models are triggered only when the incident detection model predicts the presence of an incident in the selected road topology network.

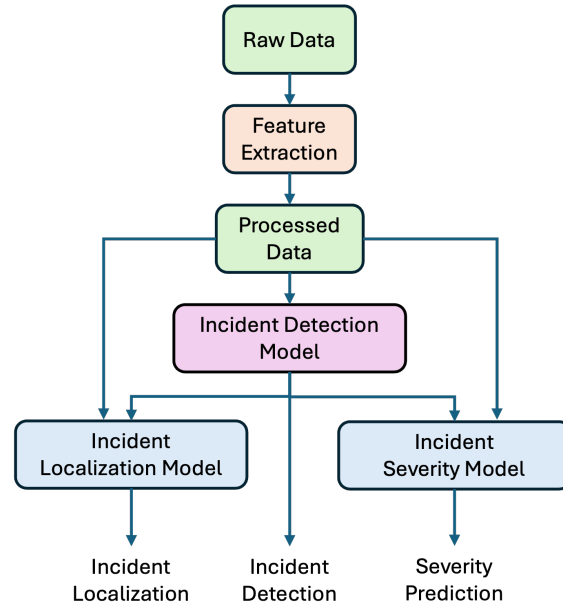


Figure 3.4: The block diagram for the baseline model training and inference. The raw data is pre-processed using the rolling window mechanism, as described in the section above, Synthetic Data Simulations.

With the baseline established, we observed a few limitations: (i) the trained model is not scalable, and (ii) the model works in a temporal aspect without considering any spatial representational aspect of the road network topology. To address these limitations, we develop a graph-based deep learning approach using graph neural networks, which enables the final model to be inductive, consider the spatial representation of the road network topology, and also facilitate multi-headed task prediction.

Graph Construction And Data Preparation

The road network's topology naturally lends itself to a graph network-based representation. Utilizing this as the foundation of our approach, we construct a graph network as follows.

Definition Of The Graph Network

The selected road topology network is modeled as a directed graph $\mathbb{G} = (\mathbb{V}, \mathbb{E})$, where $\mathbb{V}$ represents all the intersections as nodes with $|\mathbb{V}| = N$, N representing the total number of nodes and $\mathbb{E}$ represents all directed roads between intersections as directed edges with $|\mathbb{E}| = M$, M representing the total number of edges. An edge $e_{uv} \in E$ exists from node $u \in \mathbb{U}$ to node $v \in \mathbb{V}$ if vehicles can move from intersection u to v . Given the directed nature of the graph, the reverse direction representation is not automatically considered unless there is a corresponding edge e_{vu} present.

Feature Engineering Of The Dataset For Model Training

As the problem of incident detection is spatio-temporal, for each time step t , the spatio-temporal attributes are assigned as node features, edge features, and the respective set of task-specific labels as follows:

- **Edge Features:** A feature vector $a_{uv}^t \in \mathbb{R}^K$ is composed from the tabulated dataset, and associated with the edge e_{uv} . The feature vector comprises the rolling mean speed, traffic occupancy, and the vehicle counts. To ensure the data allows for inductive modeling, the feature vectors are padded to maintain a uniform dimension $K = \max(a_{uv}, x_v)$. Ensuring uniform feature dimensionality allows the model to handle edges inductively.
- **Inductive Node Features:** To enable inductive behavior for the trained

- **Multi-Task Labels:** For each time step t , the following three labels are generated to assist in collaborative training of the three tasks - incident detection, localization, and severity estimation.
 - **Detection label:** $y_{det}^{(t)} \in \{0, 1\}$, which indicates the absence or occurrence of an incident in the entire road network.
 - **Localization label:** $y_{loc}^{(t)} \in \{0, 1, \dots, C_{loc} - 1\} \cup \{-1\}$, where C_{loc} represents the total number of roads in the road network. Additionally, to allow for ignoring localization when there is no incident, we add the label of -1 .
 - **Severity label:** $y_{sev}^{(t)} \in \{0, 1, \dots, C_{sev} - 1\} \cup \{-1\}$, where C_{sev} represents the total number of severity classes. Additionally, to allow for ignoring severity estimation when there is no incident, we add the label of -1 .

Incident Detection, Localization, And Severity Estimation Model Design

Having constructed the graph-based representation of the traffic network, we designed the SIGN model as a multi-task graph neural network (GNN) for training on spatio-temporal graph data. Figure 3.6 presents the high-level overview of the model architecture that is designed, and below, we will deep dive into the mathematical flow of the model architecture.

Node Feature Projection

To ensure the network is inductive, it is crucial to learn the feature representation of the nodes $x_v \in \mathbb{V}$. So, we begin by projecting the raw features into a learnable embedding.

Given a single graph data object, $\mathbb{G}^{(t)}$, which represents the state of the traffic on the selected road network topology at time t . This single graph data object

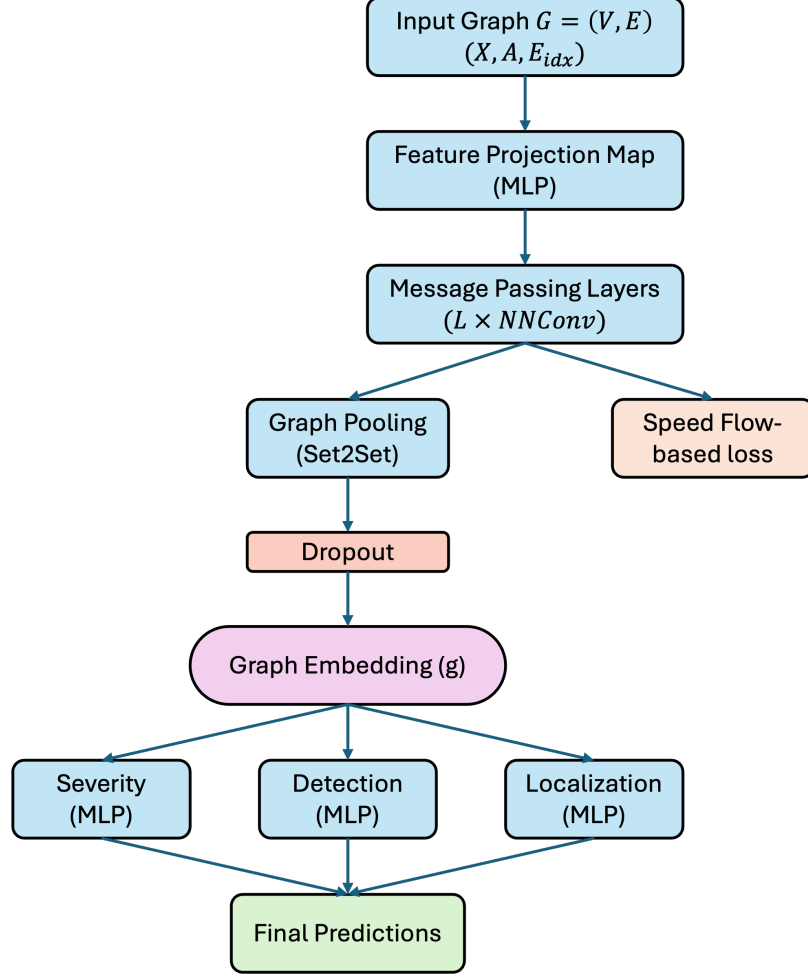


Figure 3.6: High-level representation of the model architecture designed for model training. The goal of the model is to simultaneously detect if an incident is occurring, if so, localize the road segment where the incident is located, and estimate the severity of the incident.

contains mainly node feature matrix $X \in \mathbb{R}^{N \times K}$ and edge feature matrix $A \in \mathbb{R}^{M \times K}$. The input node feature matrix $x_v \in \mathbb{R}^K$ at each node v is transformed into a hidden embedding representation as a learnable embedding. This is carried out with a simple MLP block, resulting in a hidden embedding $h^{(0)}$:

$$h_v^{(0)} = \text{ReLU}(W_{init}x_v + b_{init}) \in \mathbb{R}^{hidden_dim} \quad (3.6)$$

As the same projection can be applied to new graphs with the same number of features, this process makes the model inductive. Also, it enables it to learn a richer

representation of the traffic state at each intersection.

Message Passing

As per our implemented road graph network definition (3), the graph edges have essential information required to understand the traffic patterns, making it crucial to make the message passing mechanism the primary feature. Hence, we employ L message-passing layers (or graph convolutional layers) to transform node information based on the node and edge attributes. The node embedding features are updated from layer $h^{(l)}$ to layer $h^{(l+1)}$ as follows:

- **Message Creation:** The message creation occurs through a dedicated MLP block that processes the edge features a_{uv} from node u to node v , resulting in a unique transformation matrix $M_{uv}^{(l)}$ for each edge:

$$M_{uv}^{(l)} = \text{reshape} \left(MLP_{edge}^{(l)}(a_{uv}) \right) \quad (3.7)$$

- **Message Aggregation:** Once the message matrices are computed for all edges, each node v aggregates the messages coming from its neighbors as follows. For all the immediate incoming neighbors of node u in $\mathcal{N}(v)$, we take u 's current embedding $h_u^{(l)}$ and transform it by the message matrix: $M_{uv}^{(l)} \cdot h_u^{(l)}$. This yields the contribution of neighbor u to v 's new state. We then aggregate all such contributions by taking the mean over $u \in \mathcal{N}(v)$.
- **Node State Updation:** After the node's aggregates their embeddings, the updated nodes are propagated to the next layer through:

$$h_v^{l+1} = ReLU \left(W_{root}^{(l)} h_v^{(l)} + \text{mean}_{u \in \mathcal{N}(v)} (M_{uv}^{(l)} \cdot h_u^{(l)}) \right) \quad (3.8)$$

Graph Pooling And Output Heads

Relation between individual node embeddings is carried out by converting the final message passing layer embedding $H \in \mathbb{R}^{N \times \text{hidden_dim}}$ into an aggregated single graph-level embedding, g .

$$g = \text{Set2Set}(H) \in \mathbb{R}^{\text{hidden_dim} \times 2} \quad (3.9)$$

The *Set2Set* pooling operation Liu *et al.* (2022) aggregates information from all nodes into a fixed-length vector through a learnable attention mechanism, allowing the model to adaptively focus on the most relevant nodes when generating the graph-level embedding.

The final graph embedding g , which is a learned representation of the entire traffic graph at time t , is fed into three independent parallel prediction heads to determine the incident detection, localization, and severity estimation outputs, utilizing MLP blocks as shown below.

- Detection prediction: The detection prediction head $z_{det} = \text{MLP}_{det}(g) \in \mathbb{R}^1$ predicts whether an incident is occurring using
- Localization prediction: The localization prediction head $z_{loc} = \text{MLP}_{loc}(g) \in \mathbb{R}^{C_{loc}}$ is trained to determine the particular road segment where the incident could be occurring.
- Severity prediction: The severity prediction head $z_{sev} = \text{MLP}_{sev}(g) \in \mathbb{R}^{C_{sev}}$ predicts the severity of the incident that has occurred.

Training Objective and Hybrid Loss Function

Given the objective of multi-task classification, we defined a loss function for the three prediction tasks: incident detection, localization, and severity estimation.

Additionally, we introduce an auxiliary cost function, a traffic flow conservation regularization term, during training. This regularization term operates on the principle of traffic flow conservation, inspired by Greenshield’s Fundamental Diagram Kesting and Treiber (2013) from traffic flow theory. The Greenshield’s fundamental diagram states that the mean speed of vehicles reduces with an increase in the density of cars.

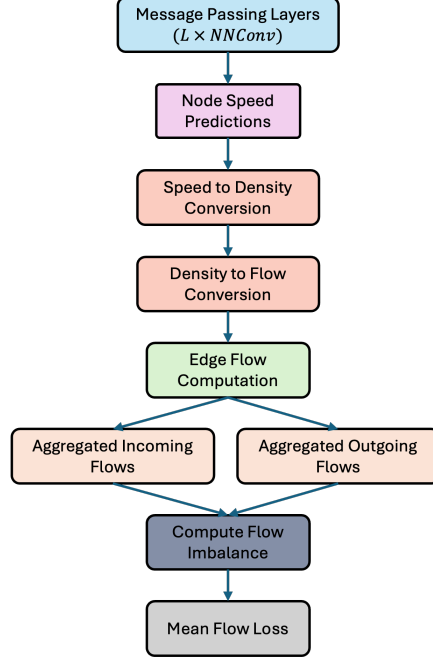


Figure 3.7: Computation of Traffic Flow Conservation Loss during the training process, which takes the graph’s node-level predictions and computes the loss to fine-tune the network training to consider traffic flow conservation, allowing enhanced training.

Figure 3.7 illustrates the computation flow for considering the traffic flow conservation loss. We formulate the Fundamental Flow Diagram, derived from Greenshield’s algorithm, as $q(\rho) = s(\rho) \cdot \rho$. The speed at a given occupancy is defined as:

$$s(\rho) = s_{max} \left(1 - \frac{\rho}{\rho_{max}} \right) \quad (3.10)$$

where s is the speed, ρ is occupancy, and q is the traffic flow. We can derive the vehicle occupancy as $\rho = \rho_{max} \left(1 - \frac{s}{s_{max}} \right)$. From the computed occupancy, we determine the

edge-level traffic flow. Then, for each node, we calculate the difference between the total incoming and outgoing flow, as well as the flow between intersections, based on the configured intersection sensor pairs using the equation 3.11.

$$\mathcal{L}_{flow} = \frac{1}{N} \sum_{i \in V} \left| \sum_{j \in \mathcal{N}_{in}(i)} q_j - \sum_{k \in \mathcal{N}_{out}(i)} q_k \right| \quad (3.11)$$

We use this traffic flow conservation loss function as a regularization parameter, as illustrated in Equation 3.12, with a weight, allowing for control over the impact of this loss computation during model training.

$$\mathcal{L}_{total} = \mathcal{L}_{supervised} + \lambda_{flow} \mathcal{L}_{flow} \quad (3.12)$$

where $\mathcal{L}_{supervised} = \mathcal{L}_{det} + \lambda_{loc} \mathcal{L}_{loc} + \lambda_{sev} \mathcal{L}_{sev}$.

EXPERIMENTS

Dataset Simulation And Preprocessing

For the dataset simulation, we generate an OpenStreetMap simulation configuration for the road network of interest and simulate the dataset for thirty consecutive days. We preprocess the dataset with three rolling window variations: 300, 600, and 900 seconds. The observed baseline F1-scores for the model developed using TabNet are 93.12%, 96%, and 96%, respectively. Given the observed scores, we choose a window size of 600 seconds for all our experiments.

Model Training And Evaluation Considerations

We trained the TabNet and the custom GNN architecture, SIGN, on an NVIDIA A6000 GPU, and the hyperparameters set for the experiments are listed in Tables 4.1 and 4.2.

Table 4.1: Model Training Hyperparameters For TabNet.

Hyper-parameters	Value
Prediction Layer Dimension	64
Attention Embedding Dimension	64
Optimizer Momentum	0.3
Optimizer	Adam
Learning Rate	0.02
Epochs	80
Loss Function	Cross Entropy

Table 4.2: Model Training Hyperparameters For SIGN.

Hyper-parameters	Value
Number of message passing layers	5
Hidden embedding dimension	32
Initial number of junctions	8
Optimizer	Adam
Scheduler	StepLR
Learning Rate	0.001
Epochs	100
Loss Function	Hybrid Loss
Detection Loss Weight	1.0
Localization Loss Weight	0.5
Severity Loss Weight	0.25
Traffic Flow Regularization Weight	0.1

The Table 4.3 presents the different metrics used to verify the quality of the model training, focusing on three standard metrics - Detection Rate (DR), Mean Time to Detect (MTTD), and False Alarm Rate (FAR) - to evaluate the model performance on incident-related tasks.

Table 4.3: Traffic Incident Detection Metrics and Their Definitions Based on Confusion Matrix, Where TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives.

Metrics	Definition
DR (Detection Rate)	$\frac{TP}{TP+FN}$
FAR (False Alarm Rate)	$\frac{FP}{FP+TN}$

RESULTS

SIGN Detects Traffic Incidents Better In Arterial Roads

We train the best-performing model, as per the related work Zhu *et al.* (2018), to compare how their approach performs against our approach of data simulation and model inference. Our evaluation results are presented in Table 5.1.

Table 5.1: Performance Comparison Between The Previous State-Of-The-Art Model And Our Baseline Using TabNet.

Algorithm	DR	MTTD	FAR
Zhu <i>et al.</i> (2018)	51%	471 secs	35.42%
TabNet (Baseline)	98%	197 secs	6.26%

We observe that our baseline approach, implemented using TabNet, is more accurate than the previous state-of-the-art approach proposed by Zhu *et al.* (2018) for incident detection on arterial roads. We observe that our detection rate is almost $1.9\times$ better, the mean-time-to-detect is almost $2.4\times$ better, and the false alarm rate is reduced by nearly 85%.

SIGN Can Infer On Varying Road Network Topologies Without Requiring Retraining

Table 5.2 illustrates the performance comparison between the trained inductive GNN, which is trained only once for a road topology of 8 junctions, and the transductive TabNet baseline model that is trained from scratch for all the road network topologies ranging from 3 junctions to 8 junctions.

Table 5.2: Comparison Of Incident Detection, Top-3 Localization, And Severity Estimation Between Transductive And Inductive Approaches Obtained Using TabNet and GNN-based Models Respectively.

Nodes	Model	Re-training?	DR	Top-3	Severity	FAR	MTTD
8 Node	GNN	Yes (Once)	100%	85.57%	87.77%	11.94%	116 secs
	TabNet	Yes	98%	78.15%	79.69%	6.26%	197 secs
7 Node	GNN	No	99%	68.45%	85.23%	15.15%	143 secs
	TabNet	Yes	97%	77.07%	77.19%	7.12%	199 secs
6 Node	GNN	No	96%	58.31%	84.07%	14.12%	168 secs
	TabNet	Yes	97%	75.27%	76.34%	8.24%	200 secs
5 Node	GNN	No	95%	45.26%	82.24%	17.85%	173 secs
	TabNet	Yes	95%	73.65%	74.51%	8.95%	236 secs
4 Node	GNN	No	96%	48.34%	83.53%	18.08%	175 secs
	TabNet	Yes	92%	67.65%	72.18%	11.28%	286 secs
3 Node	GNN	No	99%	32.43%	73.67%	36.59%	178 secs
	TabNet	Yes	90%	62.12%	72.15%	11.57%	314 secs

The performance of the inductive GNN model is observed to be superior with inevitable trade-offs. In terms of incident detection accuracy, the inductive model achieved a near-perfect detection rate for all nodes, ranging from 95% to 99%. Additionally, considering the spatial representation of the graph network, the mean-time-to-detect improved by approximately 50%, promoting faster first responder response times. However, a dip in localization accuracy is observed due to the behavioral difference of the model architecture. TabNet combines all road segments that were separate in an 8-intersection road network. When the number of intersections to consider is reduced, the inductive model maintains the total number of nodes constant, irrespective of the reduction in the number of intersections.

These results also prove the sparse sensor handling capability of our approach.

This provides an added benefit when considering real-world deployment scenarios. Our inductive model will not fail in scenarios such as sensor failures or road closures.

CONCLUSION

This thesis set out to address three main limitations in existing work related to incident detection management. First, the research community lacked access to either a real-world or simulated microscopic dataset for detecting incidents at intersections. We addressed this by building a SUMO-based simulation framework. The resulting simulation framework now provides an open-source, reproducible source of a microscopic data simulator, explicitly tailored to the intersection incident analysis.

Second, no approach previously allowed us to detect, localize, and estimate the severity of traffic incidents between intersections. Our results now demonstrate the capability of our approach to do this. We have shown that our approach enables our trained model to operate in multiple network modes within a single model. This also highlights the inductive nature of our graph neural network model. This was achieved by representing the road network as a dynamic graph and generating inductive node features, which enabled SIGN to decouple the model from the network’s structure successfully.

This thesis represents a significant step toward developing truly scalable and efficient incident management systems for complex urban environments. We plan to focus our future work on scaling up the dataset to city-scale from sub-city scales by exploring more advanced model architectures and physics-informed machine learning approaches. Future work should focus on validating this framework on larger, real-world datasets and exploring more advanced GNN architectures and physics constraints.

REFERENCES

- Arik, S. O. and T. Pfister, “Tabnet: Attentive interpretable tabular learning”, CoRR **abs/1908.07442**, URL <http://arxiv.org/abs/1908.07442>, arXiv: 1908.07442 (2019).
- Atilgan, I. *et al.*, “Towards accurate traffic accident detection: Developing deep learning strategies with distant past, recent past, and adjacency features”, in “2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)”, pp. 6120–6125 (IEEE, 2023).
- Blincoe, L., T. R. Miller, J.-S. Wang, D. Swedler, T. Coughlin, B. Lawrence, F. Guo, S. Klauser and T. Dingus, “The economic and societal impact of motor vehicle crashes, 2019”, Tech. rep., NHTSA (2022).
- Chen, J. *et al.*, “More robust and better: Automatic traffic incident detection based on XGBoost”, in “Advances in Traffic Transportation and Civil Architecture”, pp. 267–274 (CRC Press, 2023a), URL <https://www.taylorfrancis.com/chapters/edit/10.1201/9781003402220-31/robust-better-automatic-traffic-incident-detection-based-xgboost-junyu-chen-pan-wu-jinlong-li-lunhui-xu>.
- Chen, K. L., B.-W. Tsai, G. Fortin and J. F. Cooper, “2023 SafeTREC Traffic Safety Facts: Emergency Medical Services — Safe Transportation Research and Education Center — safetrec.berkeley.edu”, URL <https://safetrec.berkeley.edu/2023-safetrec-traffic-safety-facts-emergency-medical-services> (2023b).
- Cooley, J. W. and J. W. Tukey, “An algorithm for the machine calculation of complex Fourier series”, *Mathematics of computation* **19**, 90, 297–301 (1965).
- Coursey, A., J. Ji, M. Quinones Grueiro, W. Barbour, Y. Zhang, T. Derr, G. Biswas and D. Work, “Ft-aed: Benchmark dataset for early freeway traffic anomalous event detection”, *Advances in Neural Information Processing Systems* **37**, 15526–15549 (2024).
- Cuturi, M., “PEMS-SF”, UCI Machine Learning Repository, DOI: <https://doi.org/10.24432/C52G70> (2011).
- Han, X. *et al.*, “Traffic incident detection: A trajectory-based approach”, in “2020 IEEE 36th international conference on data engineering (ICDE)”, pp. 1866–1869 (IEEE, 2020), URL <https://ieeexplore.ieee.org/abstract/document/9101794/>.
- Kesting, A. and M. Treiber, “Traffic flow dynamics: data, models and simulation”, No. Book, Whole(Springer Berlin Heidelberg, Berlin, Heidelberg (2013).
- Levenberg, K., “A method for the solution of certain non-linear problems in least squares”, *Quarterly of applied mathematics* **2**, 2, 164–168 (1944).

- Li, Y. *et al.*, “Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting”, URL <http://arxiv.org/abs/1707.01926>, arXiv:1707.01926 [cs, stat] (2018).
- Liang, H. *et al.*, “Traffic incident detection based on a global trajectory spatiotemporal map”, *Complex & Intelligent Systems* **8**, 2, 1389–1408, URL <https://link.springer.com/10.1007/s40747-021-00602-8> (2022).
- Liu, C., Y. Zhan, J. Wu, C. Li, B. Du, W. Hu, T. Liu and D. Tao, “Graph pooling for graph neural networks: Progress, challenges, and opportunities”, arXiv preprint arXiv:2204.07321 (2022).
- Marquardt, D. W., “An algorithm for least-squares estimation of nonlinear parameters”, *Journal of the society for Industrial and Applied Mathematics* **11**, 2, 431–441 (1963).
- Pechoux, K. K., G. Carrick and B. B. Pechoux, “Secondary crash research: A multistate analysis”, Tech. rep., Federal Highway Administration, U.S. Department of Transportation (2023).
- Ren, Z., X. Li, J. Peng, K. Chen, Q. Tan, X. Wu and C. Shi, “Graph autoencoder with mirror temporal convolutional networks for traffic anomaly detection”, *Scientific reports* **14**, 1, 1247 (2024).
- Salum, J. H., A. E. Kitale, T. Sando and P. Alluri, “Evaluating the impact of road rangers in preventing secondary crashes”, *Accident Analysis & Prevention* **156**, 106129 (2021).
- Shah, A. *et al.*, “Cadp: A novel dataset for cctv traffic camera based accident analysis”, arXiv preprint arXiv:1809.05782 First three authors share the first authorship. (2018).
- Skabardonis, A. *et al.*, “I-880 field experiment: Data-base development and incident delay estimation procedures”, *Transportation Research Record* **1554**, 1, 204–212 (1996).
- Sun, Y., T. Mallick, P. Balaprakash and J. Macfarlane, “A data-centric weak supervised learning for highway traffic incident detection”, *Accident Analysis & Prevention* **176**, 106779 (2022).
- Yang, H. *et al.*, “Traffic Incident Generation And Supervised Learning Based Detection Via A Microscopic Simulation Platform”, in “2023 IEEE International Conference on Cybernetics and Intelligent Systems (CIS)”, pp. 146–151 (2023), URL <https://ieeexplore.ieee.org/abstract/document/10370973>, iSSN: 2326-8239.
- Yu, W., S. Park, D. S. Kim and S.-S. Ko, “Arterial Road Incident Detection Based on Time-Moving Average Method in Bluetooth-Based Wireless Vehicle Reidentification System”, *Journal of Transportation Engineering* **141**, 3, 04014084, URL <https://ascelibrary.org/doi/10.1061/%28ASCE%29TE.1943-5436.0000748> (2015).

- Yu, Y. *et al.*, “Citywide traffic volume inference with surveillance camera records”, IEEE Transactions on Big Data **7**, 6, 900–912 (2021).
- Zhu, L. *et al.*, “A deep learning approach for traffic incident detection in urban networks”, in “2018 21st international conference on intelligent transportation systems (ITSC)”, pp. 1011–1016 (IEEE, 2018), URL <https://ieeexplore.ieee.org/abstract/document/8569402/>.