Traffic Accident Risk Prediction for Multi-factor Spatio-temporal Networks

Qingrong Wang, Kai Zhang, Changfeng Zhu, Xiaohong Chen

Abstract—**Predicting the risk of traffic accidents has significant implications for emergency response and urban planning, and is critical in implementing intelligent transportation systems. However, there are some challenges in effectively capturing spatiotemporal correlations in accident-prone regions. Firstly, existing models focus on capturing spatio-temporal features of coarse-grained regions, and need to dynamically integrate predictions for fine-grained regions with those for coarse-grained regions. In addition to weather conditions, environmental factors such as points of interest (POI) and road attributes often influence traffic accidents. Therefore, it is crucial to extract the semantic relation between external factors and accident risk. Thirdly, since traffic accidents are rare events, the training of the model may encounter the problem of zero inflation. To tackle the difficulties above, we present a model for predicting traffic accident risk named Risk-CCNMAGU. Specifically, we design Spatial-channel CNNs and Multi-factor-Attention GCNs to catch spatial features in regions of different granularity. We also implement a dynamic fusion of coarse-grained and fine-grained regions through weighted aggregation. Meanwhile, we construct static, dynamic, and knowledge graph representation views to extract correlations between multiple external factors. We introduce the GRU-Attention module to capture nonlinear temporal correlations to learn temporal features. Additionally, we employ a sample-weighted MSE loss function to alleviate the data sparsity problem. Also, we add introduce to the raw data before feeding it into the model. Finally, we conducted comprehensive experiments on NYC and Chicago datasets, showing that Risk-CCNMAGU outperforms existing models on RMSE, MAE, MAP, and recall metrics.**

Index Terms—**Traffic accident risk, Fine-grained regions, Multi-factor, Zero inflation problem.**

I. INTRODUCTION

WITH the rising urban demand, the prominence of traffic accidents has also increased. According to statistics from the World Health Organization [1], nearly 1.35 million individuals succumb to traffic accidents yearly. Consequently, there is a growing importance in accurately predicting the risk of traffic accidents. Such predictive models can alert drivers in advance, allowing them to avoid areas or periods with a higher likelihood of accidents and enabling them to take proactive measures to mitigate the risk of collisions. This area of research has garnered significant interest among scholars.

Traffic accident risk prediction utilizes forecasting future accident risks based on historical spatiotemporal data. Compared to traditional time series prediction problems, traffic accident risk prediction faces challenges such as spatiotemporal heterogeneity and complex nonlinear relationships, making it difficult to establish high-performance prediction models. Initial research predominantly employed methodologies like machine learning to predict traffic accident risks [2][3][4]. However, these approaches fell short in accounting for both spatial and temporal correlations, leading to a decrease in the accuracy of predictions.

The recent advances in deep learning have significantly improved the prediction of spatio-temporal data problems [5][6]. Relevant scholars have applied Recurrent Neural Networks (RNNs) and their variants, such as Long Short-term Memory Networks (LSTMs) [7] and Gated Recurrent Units (GRUs), to model the temporal dimension. Meanwhile, Convolutional Neural Networks (CNNs) [8] have found wide applications in capturing spatial correlations. However, forecasting traffic collision probabilities is frequently characterized as a graph modeling problem. Consequently, to gain a more comprehensive understanding of the nonlinear spatial correlations within traffic accident data, some scholars have introduced GCNs [9][10] to investigate the non-Euclidean correlation of road networks and use attention mechanisms to capture dynamic features [11], thus obtaining better predictive performance. However, external environmental factors, including weather, distribution of points of interest (POI), holidays, and road properties, often influence traffic accidents [12]. These features are necessary for the predictive accuracy of these models to be satisfactory.

In general, there is a direct or indirect relationship between external environmental factors and traffic data, which can impact the traffic circumstances within a city. For example, related scholars [13][14][15] have effectively improved traffic forecasting by incorporating a limited set of external factors. However, these studies ignored the impact of correlations between external factors and traffic conditions. Therefore, some researchers [16][17] collected factors such as traffic volume, meteorological conditions, road infrastructure, POIs, and historical traffic accident information to investigate the prediction of traffic accident risks. Huang et al. [18] formulated a deep dynamic amalgamation network framework to examine the influence of intricate temporal associations and external variables on traffic accidents. Similarly, Liu et al. [19] employed a multi-task learning framework to integrate external variables to characterize traffic accidents. Although these prior studies

Manuscript received June 28, 2023; revised September 9, 2023. This work was supported in part by the National Natural Science Foundation of China (No. 71961016, 72161024), "Double-First Class" Major Research Programs, Educational Department of Gansu Province (No. GSSYLXM-04).

Qingrong Wang is a professor at School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: wangqr003@163. com).

Kai Zhang is a postgraduate student at School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China. (Corresponding author, e-mail: zhangk925@163.com).

Changfeng Zhu is a professor at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: cfzhu003@163. com)

Xiaohong Chen is a postgraduate student at School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: 2937325132@qq.com).

effectively integrated external factors with traffic accident data to enhance prediction accuracy, the key to further improving prediction performance lies in the adept fusion of diverse, multi-source heterogeneous data, particularly concerning accident-prone regions.

The development of Knowledge Graphs (KGs) has provided a broader way of looking at these issues. KGs exhibit remarkable efficacy in handling graph structures and facilitating information retrieval, particularly excelling in scenarios featuring complex topological arrangements. Therefore, researchers increasingly apply KGs in social networks [20] and intelligent recommendations [21]. Meanwhile, related scholars also use KGs in transportation, such as Xu et al. [22], who added temporal information to the knowledge base model, which performs well in analyzing and detecting emergency events. Muppalla [23] et al. pioneered constructing a KG utilizing dynamic image features, giving rise to the Image-based Traffic Sensing Knowledge Graph Model (ITSKG), significantly advancing traffic event detection. In order to construct semantic relationships between various environmental factors and traffic data, Zhu et al. [24] designed a KF-Cell to fuse the KGs embedding of environmental features and traffic flow features into a spatio-temporal graph convolutional network, which better captures the semantic relationships between spatio-temporal features and traffic information. Wang et al. [25] introduced a traffic flow forecasting approach called KGR-STGNN, which harnesses the complete information from external variables by KGs. Given this backdrop, a pressing need exists to explore the potential of KGs in capturing intricate correlations among multiple external factors and traffic information. This path shows substantial potential for improving the anticipation of traffic accident probability.

Furthermore, traffic collisions tend to happen in various city regions, each exhibiting distinct spatiotemporal attributes. For example, scholars [26][27] have effectively harnessed the segmentation of the study area into grids to capture the spatio-temporal relationships. Others [28] used the "global interaction + node query" approach to capture dynamic spatio-temporal correlations. Moreover, some studies [29][30] focus on the disparate distribution of traffic conditions across different spatial domains, aiming to capture spatio-temporal features. However, these methods tend to make predictions in a coarse-grained area, ignoring the impact of local vital features. Hence, there is an immediate requirement for a forecasting approach that incorporates fine-grained regional and multi-source heterogeneous spatio-temporal data.

This study addresses the challenges above and proposes a model called Risk-CCNMAGU, which adopts a multi-view and multi-factor approach and aims to forecast the likelihood of traffic accidents in a unit area over some time in the future. Specifically, the model divides the road network into unit areas of different granularity by the latitude and longitude of the city. We should consider the effects of static factors (e.g., POI, road attributes, etc.) and dynamic factors (e.g., weather, traffic flow, etc.) for different granularity areas. In the coarse-grained regions, We use the Spatial-channel Convolutional Neural Networks (Spatial-channel CNNs) to train the weights of various feature channels (e.g., weather, POI distribution.) to capture the space-time attributes of the region. Meanwhile, we access the correlation of static and dynamic features in the fine-grained region to construct the static, dynamic, and knowledge graph representation views. Then, we employ multi-factor-CCNs models to learn multi-view feature representations.

We introduce a GRU-Attention module dedicated to temporal feature acquisition to learn the non-linear time correlation. Meanwhile, this article employs a weighted MSE cost function, alleviating data sparsity issues. Finally, we dynamically fuse coarse-grained and fine-grained regions to understand the non-linear, evolving spatio-temporal correlations.

II. PRELIMINARIES AND PROBLEM DEFINITION

This section will introduce the close correlation between spatiotemporal attributes and traffic accident risks, such as time, weather conditions, road attributes, and other external environmental factors. We will construct multi-views through these external factors and define the problem studied in this article.

A. Spatio-temporal Features

The presence of complex topologies and traffic flows in urban networks is a potential factor for traffic accidents. Additionally, severe weather conditions, like heavy rain and snow, significantly reduce roadway capacity and visibility, increasing the risk of road accidents. Therefore, the occurrence of traffic collision incidents is influenced by various static and dynamic factors with strong spatial and temporal correlations.

A.1 Static Features

Static factors significantly influence traffic accidents, including road attributes, POI, and time attributes [31]. For example, a high traffic volume during peak commuting hours, especially on shorter road segments, can lead to widespread vehicle congestion. Additionally, congestion dissipates at a slow rate, making it more likely for traffic accidents to occur.

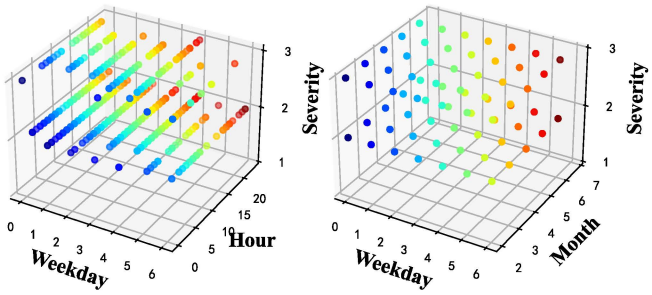

Fig.1 Temporal features.

Fig. 1 on the left shows that traffic incidents happen more frequently on weekdays during peak commuting hours, whereas the probability decreases during the evening or late at night. Moreover, the data in Fig. 1 on the right indicates an upsurge in traffic accident risk from May to July, which could be attributed to increased summer tourism or distracted driving due to higher temperatures. Hence, predicting traffic accident risk displays conspicuous temporal features that exhibit periodicity and trends.

A.2 Dynamic Features

In the spatial dimension, there are interactions between

traffic conditions on different roads, e.g., a congested road section may affect the traffic conditions on its neighboring roads, and the interactions are significantly dynamic. Furthermore, changes in weather conditions may also significantly impact road traffic conditions, thereby elevating the likelihood of accidents, as shown in Fig. 2.

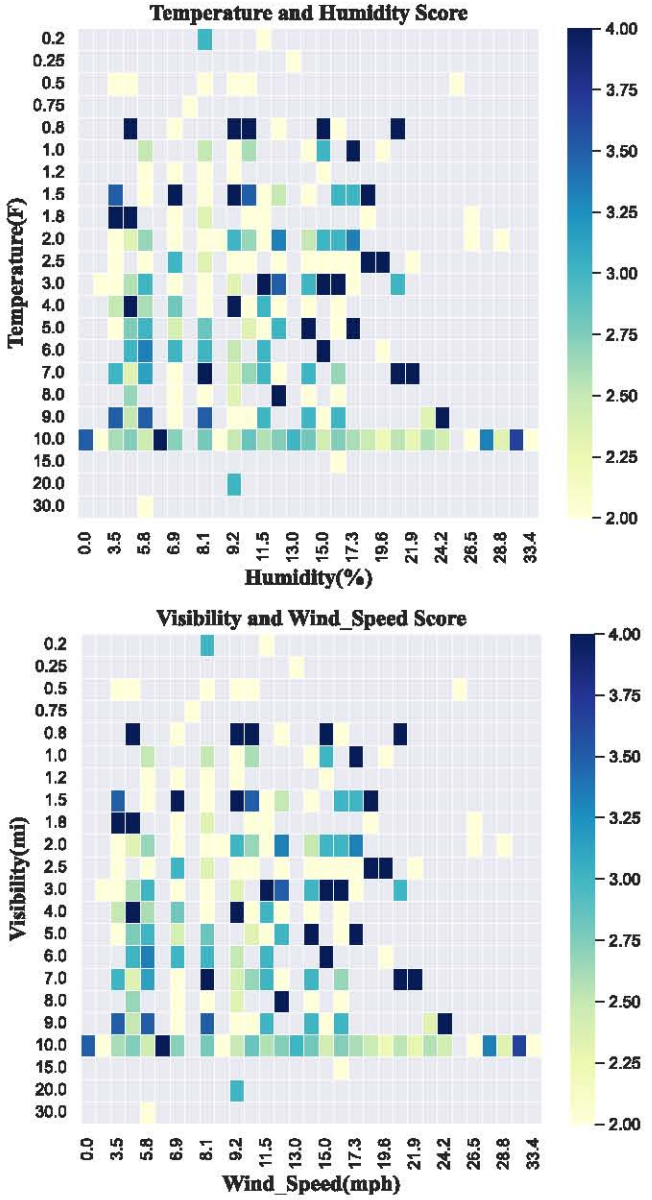

Fig. 2 Weather features.

B. Multi-view Construction

In the urban road network, semantic relationships exist between different unit areas, e.g., they may have similar road features, POI distributions, time characteristics, etc. As a result, these areas have similar traffic accident risk trends. In order to grasp the intricate spatiotemporal associations from diverse viewpoints, this article proposes a method to construct a static, dynamic, and knowledge graph representation view. Finally, it predicts the traffic accident risk in fine-grained areas through these views.

B.1 Static and Dynamic Views

We hold that the distribution of POIs, road grades, road types, and other factors within a region significantly impact the traffic conditions of that area. Hence, these features can reflect the semantic correlation of traffic accident risk [32].

This article considers each fine-grained area as a node in a graphical network, and the nodes contain the features in the unit region. Therefore, this section calculates the similarity between two nodes based on the Jensen-Shannon (JS), which calculates the similarity of road features, POI distribution, and accident risk. As an example, the procedure for calculating road characteristics is as follows.

$$Res_{R(i,j)} = 1 - JS(R_i(x), R_j(x)) \tag{1}$$

$$JS(R_i(x), R_j(x)) = \frac{1}{2}\left(\sum_{k=1}^{K} R_i^k(x) \log \frac{2R_i^k(x)}{R_i^k(x) + R_j^k(x)} + \sum_{k=1}^{K} R_i^k(x) \log \frac{2R_j^k(x)}{R_i^k(x) + R_j^k(x)} \right) \tag{2}$$

Where K is the dimension of road characteristics, $R_i(x)$ and $R_j(x)$ denote the road segment characteristics of node i and node j, respectively. Similarly, we calculate the similarity of the POI distribution $Res_{P(i,j)}$ simultaneously and use the risk of regions to calculate $Res_{S(i,j)}$. In addition, we will select the L most similar regions to obtain the adjacency matrix $A \in (A_{sr}, A_{sp}, A_{ds})$ of the topology map, which is set to 1 when there is an edge between two nodes; otherwise, it is set to 0. Therefore, we will construct the static view $G_s(V,E,A)$ and the dynamic view $G_d(V,E,A)$.

B.2 KG View Construction

KG, a semantic-based network structure, comprises multiple triplets (head entity, relationship, tail entity) and represents the semantic relationships between entities [33]. Fig. 3 shows that external environmental factors, such as weather conditions, peak periods, POI distribution, and road class, easily affect traffic accident risk. In particular, the risk of traffic accidents is higher under unfavorable conditions such as low visibility, peak hours, and superstores. Meanwhile, KG can integrate multiple heterogeneous data sources while retaining the original information. Hence, using KG for knowledge representation is more appropriate for forecasting the likelihood of traffic accidents.

In order to understand the knowledge structure and semantical context between accident segments and environmental features, this article employs a KG model KR-EAR [34] based on entity-attribute-relation. It divides KG relationships into attribute triplets and relation triplets. We use the triplet form of $G_{kg} = (R_e, S_a, Y_{aa})$ to denote the accident section, its attributes, and their relationships, where R_e represents the relation triplet between section v_i and section v_j, S_a represents the attribute triplet between the accident section and corresponding attribute, and Y_{aa} represents the correlation between various attributes.

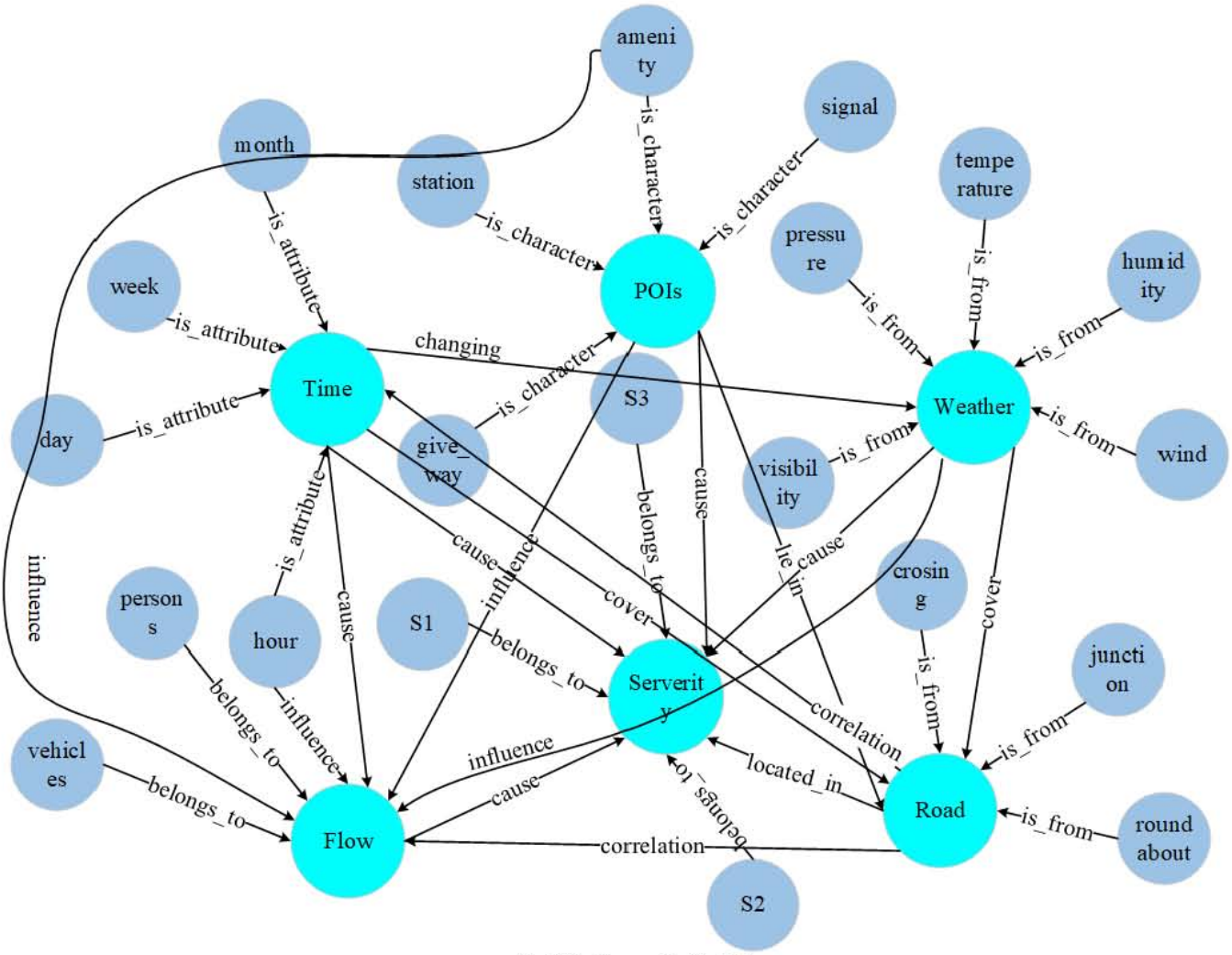

Fig.3 The frame of traffic KG.

Since the KR-EAR model can represent the connections between entities and their attributes, this model can be used to acquire embedding vectors X_e for entities, relationships, and properties. Given the low-dimensional vector X_e, the objective function can be formulated to maximize the combined probability of relationship and attribute triplets, effectively capturing the correlations between accident roads and attributes and between attributes. The objective function is defined as follows.

$$P(R_e, S_a|X_e) = \prod_{(v_i,r_e,v_j)\in R_e} P((v_i,r_e,v_j)|X_e) \cdot \prod_{(v_i,a_k,a_{k-v_i})\in S_a} P((v_i,a_k,a_{k-v_i})|X_e) \quad (3)$$

In Eq., the relationship triple (R_a) and the attribute triple (S_a) are mutually independent. $P((v_i,r_e,v_j)|X_e)$ represents the probability of the relationship triple of the accident road, given the embedded vector X_e. $P((v_i,a_k,a_{k-v_i})|X_a)$ is the conditional probability of attribute triple (v_i,a_k,a_{k-v_i}).

Moreover, TransR [35] generates the conditional probabilities for the relation triplet and the attribute triplet, which will be defined as follows.

$$P(v_i,r_e,v_j) = \frac{\exp(g(v_i,r_e,v_j))}{\sum_{\hat{v}_i\in V}\exp(g(\hat{v}_i,r_e,v_j))} \quad (4)$$

$$g(v_i,r_e,v_j) = -\|-v_iM_r + r_e - v_jM_r\|_{L1/L2} + b_1 \quad (5)$$

Where $g(\cdot)$ denotes the energy function portraying the correlation between road segments v_i and v_j. M_r stands for the mapping matrix, which is responsible for projecting entities from the entity space to the relationship space. Additionally, b_1 represents the bias value, while $L1$ and $L2$ correspond to the $L1$ norm and $L2$ norm, respectively.

We know that relationships and attributes will exhibit different characteristics. Therefore, we utilize a classification model to process data effectively to capture the relationship between entities and attributes.

$$P((v_i,a_k,a_{k-v_i})|X_e) = \frac{\exp(h(v_i,a_k,a_{k-v_i}))}{\sum_{\hat{a}_{k-v_i}\in a_{k-v}}\exp(h(v_i,a_k,\hat{a}_{k-v_i}))} \quad (6)$$

$$h(v_i,a_k,a_{k-v_i}) = -\|f(v_iW_a + b_a) - V_a\|_{L1/L2} + b_2 \quad (7)$$

Where $h(\cdot)$ denotes the scoring function. $f(\cdot)$ is a nonlinear function that projects the entity embeddings into the attribute

space via 1-NN and then computes the correlation between the transformed embeddings and the corresponding attribute value embeddings. In addition, V_a is the embedding vector of attribute value a_{k-v}, and b_a and b_2 are the bias values.

Traffic accidents in a particular region are often strongly correlated with various attributes of that region. Therefore, it is necessary to consider the correlation between these attributes.

Suppose $S(att)=\{(v_i,\hat{a}_k,\hat{a}_{k-v_i})|(v_i,\hat{a}_k,\hat{a}_{k-v_i})\in S_a\}$ are other attributes of v_i except (v_i,a_k,a_{k-v_i}). The conditional probability of the property triplet is defined as.

$$P((v_i,a_k,a_{k-v_i})|S(att))=\frac{\exp(z(v_i,a_k,a_{k-v_i},S(att)))}{\sum_{\hat{a}_{k-v_i}\in a_{k-v}}\exp(z(v_i,a_k,\hat{a}_{k-v_i},S(att)))} \quad (8)$$

$$z(v_i,a_k,a_{k-v_i},S(att))\propto \sum_{(v_i,\hat{a}_k,\hat{a}_{k-v_i})\in S(att)} P((a_k,a_{k-v_i})|(\hat{a}_k,\hat{a}_{k-v_i}))\,(A_a\cdot A_{\hat{a}}) \quad (9)$$

Where $z(\cdot)$ is the scoring function between the measured attributes. $(A_a\cdot A_{\hat{a}})$ denotes the dot product of attributes A_a and $A_{\hat{a}}$. $P((a_k,a_{k-v_i})|(\hat{a}_k,\hat{a}_{k-v_i}))$ notes the correlation between attribute values (a_k,a_{k-v_i}) and $(\hat{a}_k,\hat{a}_{k-v_i})$.

In summary, we construct the traffic knowledge graph representation $G_{kg}=(X_e,S(att),A)$ given the traffic network topology matrix A, the knowledge graph embedding vector X_e, and other attributes $S(att)$.

C. Problem Definition

Definition 1. Cell region. In order to collect spatio-temporal features that impact the incidence of road accidents, we divide the irregular urban road network into $a\times b$ grid regions, thereby constructing equally sized unit areas. Each coarse-grained region comprises multiple fine-grained regions, and a coarse-grained region contains $a/c\times b/c$ grid regions, with each coarse-grained unit area containing $c\times c$ fine-grained regions, where c is the coarsening factor.

Definition 2. Accident risks. This article sets the risk level of traffic accidents to 1, 2, and 3 based on the number of casualties [36]. For example, an area with three minor or one fatal accident has a risk level of 3.

Definition 3. Traffic accident data. In order to comprehensively leverage both Near-term and extended-term temporal dependencies, we employ a multi-scale approach informed by the findings of our time attribute analyses in Section A.1. The model's inputs comprise traffic accident data for the same periods in the previous k periods, the previous p weeks, and the previous np weeks, $t=k+p+np$.

Definition 4. Multiple views. We are inspired by the GSNet model, and in this article, we capture the nonlinear spatio-temporal correlation by constructing different views, specifically static, dynamic, and knowledge graph representation views. Among them, the static view is constructed based on the distribution of POIs and road characteristics, $G_{sp}=(V,E_{sp},A_{sp})$ and $G_{sr}=(V,E_{sr},A_{sr})$. The dynamic view then constructs $G_{ds}=(V,E_{ds},A_{ds})$ from the traffic accident risk values. Finally, the knowledge graph view constructs $G_{kg}=(X_e,S(att),A_{kg})$ by fusing traffic flow, weather conditions, and POI. In the above view, V, E, and A denote the nodes, edges, and graph adjacency matrix. See Section B for specific construction information.

This article's traffic accident risk problem is a nonlinear model f. With inputs including a traffic network topology matrix A, a multi-view feature X_G of the previous T periods, and a traffic knowledge graph G_{kg}, our objective is to forecast the probability of road incidents during the upcoming T+1 periods.

$$y=f(A,X_G,G_{kg}) \quad (10)$$

III. MODEL CONSTRUCTION

Fig.4 shows the model framework proposed in this article. In order to make the traffic data more in line with reality, we add noise to the raw data in coarse-grained and fine-grained regions. The model primarily comprises three central components: spatial feature module, temporal feature module, and multi-region spatiotemporal fusion module. Firstly, in the spatial feature module, the grid data with added noise is used as input in the coarse-grained region, and Spatial-channel CNNs are used to capture complex spatial correlations. After convolution processing, the residual structure enhances the network's learning ability. Meanwhile, in the fine-grained region, the fusion of dynamic and static features' multi-views and knowledge graph representations are used as GCN inputs, and we use the multi-factor self-attention mechanism for spatial feature learning.

Secondly, for the time feature module, this article uses information such as hours, days, and whether it is a holiday as data sources to capture time features. We employ the Attention-GRU module to capture insights into the temporal correlations within the input data. Next, in the multi-region spatio-temporal fusion module, this article integrates the spatio-temporal features of coarse-grained and fine-grained regions. Then, it inputs the fused features to the Fully Connected Layer to acquire the predicted outcomes. Finally, we also designed a sample-weighted MSE loss function to remission the sparsity problem of traffic data. This approach significantly enhances the model's efficacy in sparse data.

A. Spatial Features Module

A.1 Spatial Features of Coarse-grained Regions

Each coarse-grained region contains multiple interrelated unit regions, and the traffic conditions between these unit regions are highly correlated. Therefore, this article adopts a CNN-based method to capture the spatial correlation of coarse-grained regions. However, using traditional CNN also leads to some problems, such as neglecting the influence of external environmental variables on the risk of road accidents.

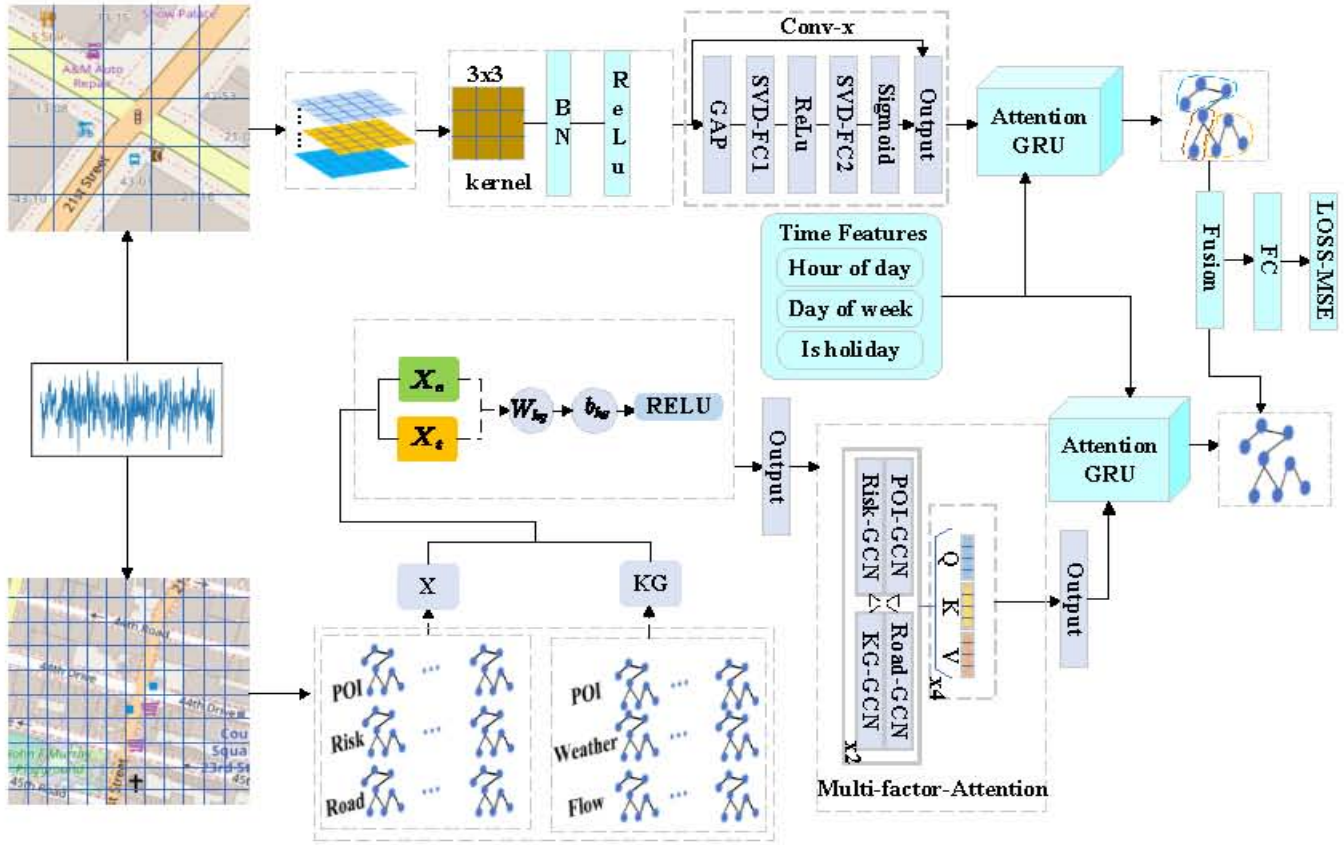

Fig. 4 The framework of the proposed Risk-CCNMAG U model.

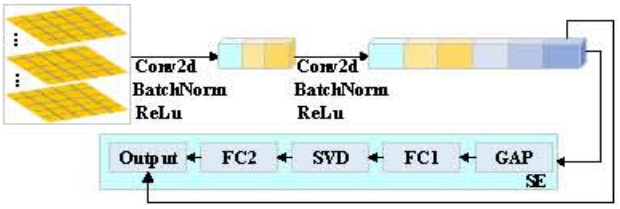

Fig.5 The framework of Spatial-channel CNNs.

As a result, this article further improves the CNN model. As shown in Fig. 5. We fully consider external environmental factors (e.g., weather, holidays, POI, etc.) to capture spatial correlation more effectively. In addition, this article introduces Global Average Pooling (GAP) and a Fully Connected Layer based on Singular Value Decomposition (SVD). We utilize these components to compute attention scores for each channel, enabling adaptive learning of the impacts of distinct features on the risk of traffic accidents.

Different unit areas often present similar spatial features within the same time frame. Therefore, this article uses multi-channel feature C (including external environmental factors, road attributes, traffic accident risks, etc.) as the input of Spatial-channel CNNs to obtain spatial features that affect the risk of traffic accidents.

$$C_t^k = \psi(W_t^k * C_t^{k-1} + b_t^k) \tag{11}$$

Where $*$ represents the convolution process, W_t^k and b_t^k are the trainable parameters of the convolution operation, respectively. ψ is the ReLU function, and C_t^k represents the output generated by the *k-th* convolutional layer within the given time interval t.

After the convolution operation, we obtain the output $C_t^k = [c_1, c_2, ..., c_m]$, where c_i denotes the feature embedding learned on each channel. Subsequently, we aim to reduce the number of parameters in the CNNs. We use GAP to compute the average of all elements within each feature map (channel) to mitigate overfitting.

$$GAP\_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} c(i,j) \tag{12}$$

Subsequently, we employ a two-layer, fully connected layer based on Singular Value Decomposition (SVD) to learn the non-linear correlations among different channels. This approach enables the capture of spatial correlations at a coarse-grained region.

$$svd\_fc' = GAP\_c * svd\_fc \tag{13}$$

$$C_f = Sigmoid(W_{c2} * ReLu(W_{c1} * svd\_fc')) \tag{14}$$

$$C_t^z = C_t^k * C_f \tag{15}$$

Where svd_fc is the singular value decomposition operation, C_f denotes the value of the channel weights obtained after two fully connected layers based on SVD, and C_t^z denotes the output after enhancing or suppressing certain channels on the feature map, i.e., spatial features in coarse-grained regions.

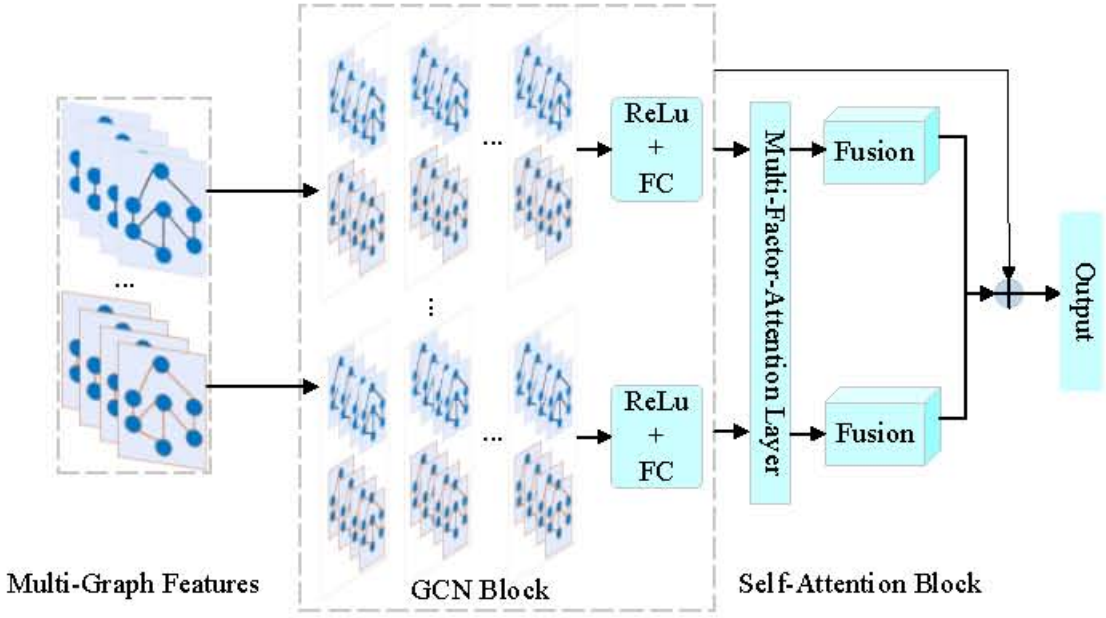

Fig.6 The framework of Multi-factor-Attention GCNs.

A.2 Spatial Features of Fine-grained Regions

Typically, factors such as road characteristics, traffic flow, weather conditions, and POI distribution are fused differently in each unit area, which affects the risk of traffic accidents to different degrees. However, as fine-grained regions can extract spatial variations on smaller scales, they help to capture spatial features within a localized area better. Therefore, in order to catch the spatial relationship of fine-grained regions, this study will model the spatial correlation of fine-grained regions from three perspectives: static view, dynamic view, and knowledge graph representation view. See Fig. 6 for an illustration.

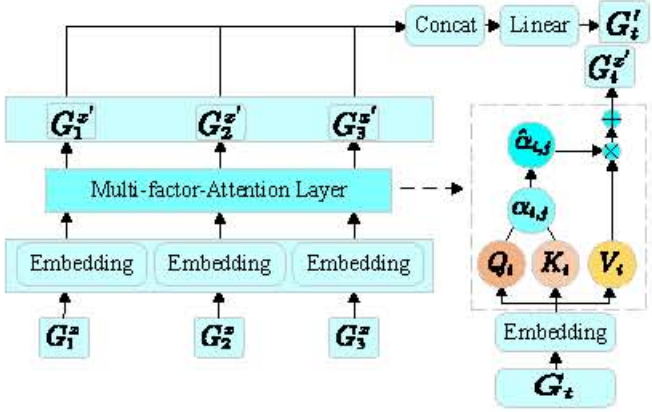

Fig.7 Self-Attention block.

With multiple views constructed by Definition 4, we will employ the Multi-factor GCNs module to learn about each view to capture the spatial correlation of fine-grained regions. The Multi-factor GCNs operation at time t proceeds as follows.

$$G_t^l = \sigma(A * G_t^{l-1} * W_g + b_j) \tag{16}$$

Where W_g and b_g denote the learnable parameters of the graph convolution process, respectively. σ is the ReLu activation function. G_t^l denotes the multi-factor convolution output at time t. We use $G_t = [G_t^s, G_t^z, G_t^f]$ to denote the graph embedding with three views.

However, we must perfect the features obtained through the Multi-factor GCNs to reflect the spatial correlation of fine-grained regions, as they may be affected by node sparsity and imbalanced neighbor nodes. Therefore, we employ a self-attention mechanism to allocate weights to distinct nodes flexibly, thereby better considering each node's importance to capture the different impacts of the three views on road accident risk forecasting. Inspired by self-attention, we map the output G_t of the multigraph convolution into Q, K, and V matrices using three matrices, namely W_q, W_k, and W_v, respectively, as shown in Fig. 7. The following is the calculation process of the self-attention mechanism.

$$\alpha = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{17}$$

In Eq., Q, K, and V represent query vectors, key vectors, and value vectors, respectively, and $\sqrt{d_k}$ is the dimension of the attention mechanism.

This section employs a multi-head attention mechanism to simultaneously attend to data from different spatial locations to extract more comprehensive and representative features from each view. We contact the h output matrices $\alpha_1, \alpha_2, \ldots, \alpha_h$ in Eq. (17) to acquire the final ultimate matrix $G_t' \in \mathbb{R}^{n \times d}$.

$$G_t' = \text{concat}(\alpha_1, \alpha_2, \ldots, \alpha_h) W_O \tag{18}$$

In Eq., $W_O \in \mathbb{R}^{d \times d}$ is a weight matrix that is independent of the number of heads.

After the multi-head attention layer, we use residuals to concatenate the final output G_t', thus avoiding the destruction and information loss of the output features by the multi-head attention layer. This approach expedites the model's convergence rate and improves training efficiency. A softmax layer is employed to gauge the significance of different

features.

$$G_t^x = softmax(\tau G_t + (1-\tau)G_t') \tag{19}$$

In Eq., G_t^x denotes the spatial feature captured at the fine-grained region.

B. Time Features Module

Predicting traffic accident risk is a classic problem in spatiotemporal data prediction, and unlike other time series prediction problems, it faces a more significant challenge in the time dimension. With the results of the analysis of temporal properties in Section A.1, we should consider the dynamic spatio-temporal properties of both neighboring and long-term periods, and then predict the risk of future periods more accurately.

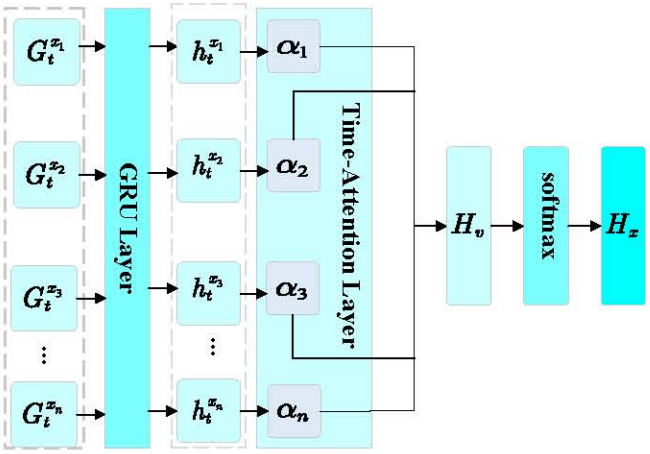

Fig.8 The framework of GRU-Attention.

In this article, we introduce the GRU-Attention module to catch short-term and long-term temporal characteristics to extract the temporal relevance of historical road crash risk, as shown in Fig.8. Specifically, taking fine-grained regions as an example, we extract temporal features based on the output of Multi-factor-Attention GCNs. Following Definition 3, we construct a data sequence $[G_1^x ..., G_p^x ..., G_{np}^x ..., G_t^x]$, using t-period historical traffic data, and then employ GRU to capture the potential temporal patterns that influence the risk of traffic accidents.

$$z_t = \sigma(W_z \cdot [f_{t-1}, G_t^x] + b_z) \tag{20}$$

$$r_t = \sigma(W_r \cdot [f_{t-1}, G_t^x] + b_r) \tag{21}$$

$$\tilde{f}_t = tanh(W \cdot [r_t \odot f_{t-1}, G_t^x] + b) \tag{22}$$

$$f_t = (1-z_t) \odot f_{t-1} + z_t \odot \tilde{f}_t \tag{23}$$

Where G_t^x denotes the current input, and f_{t-1} denotes the hidden state at the previous moment, W_z, W_r, W and b_z, b_r, b denote the weight matrix and the bias term, respectively. We use σ to denote the sigmoid function to limit the values of z_t and r_t to between 0 and 1. The hyperbolic tangent activation function $tanh$ calculates the candidate hidden state $\tilde{f}_t$. $\odot$ denotes the Hadamard product.

In addition, the GRU-Attention module effectively captures crucial information across different time steps through an attention mechanism when dealing with long-time series. Thus, we use this module to assign different weights to each time step, placing more weight on the most critical information for the current task. This approach provides richer temporal features in traffic accident risk prediction, which enhances prediction accuracy.

$$e_{t'} = ReLu(W_e[f_{t'}, f_t] + b_e) \tag{24}$$

$$\alpha_{t'} = \exp(e_{t'}) \Big/ \sum_{t'=1}^{t} \exp(e_{t'}) \tag{25}$$

$$Gf_t^* = \sum_{t'}^{t} \alpha_{t'} f_{t'} \tag{26}$$

Where t' denotes the time step in the past, $\alpha_{t'}$ denotes the weight value from the historical time to the target time. Gf_t^* denotes the weighted hidden state of the fine-grained region, emphasizing the historical information more critical for the prediction task. Similarly, we denote the spatio-temporal features of the coarse-grained region by Cf_t^*.

C. Multi-region Spatio-temporal Fusion Module

In traffic accident risk prediction, coarse-grained regions focus on global-level factors, such as overall traffic conditions and weather. In contrast, fine-grained regions focus on local details like road type, traffic flow, and POI distribution. Thus, we dynamically fuse the coarse-grained region with the fine-grained region, which helps to extract the multi-scale spatiotemporal characteristics. This approach enables the model to maintain excellent generalization while attending to local information, thereby enhancing the stability of the prediction results. We employ fully connected layers to perform a weighted fusion of spatio-temporal features from coarse-grained and fine-grained regions to achieve this objective. This results in prediction results considering features at all scales.

$$GCf = FC(W_1 * Gf_t^* + W_2 * Cf_t^*) \tag{27}$$

In Eq., $FC(\cdot)$ denotes the fully connected layer, W_1 and W_2 are learnable parameters, and GCf denotes the likelihood of traffic accidents at the subsequent time after dynamic fusion.

IV. LOSS FUNCTIOON

In general, road accidents are small probability occurrences, which may lead to a zero-inflated problem in the prediction process. In order to address this challenge, we employ the weighted Mean Square Error (MSE) loss function in this study. We utilize this loss function to assign weights to each sample, allowing the model to focus more on high-risk samples. Specifically, according to the actual accident risk, we categorize the samples into four levels corresponding to weight values of 0.05, 0.2, 0.25, and 0.5. The computation of the loss function proceeds as follows.

$$Loss(\hat{y}, y) = \frac{1}{2} \sum_{i=1} w_i (\hat{y}_i - y_i)^2 \tag{28}$$

Where $\hat{y}_i$ represents the predicted values, y_i represents the

actual values, and w_i represents the weight of the road crash risk class i.

V. EXPERIMENT

We experimentally assess the performance of Risk-CCNMAGU in this section. Firstly, we outline the experimental configuration, encompassing the dataset, evaluation metrics, hyperparameter configurations, and the baseline model. We then analyze the experimental results and compare Risk-CCNMAGU's performance with the baseline model using NYC and Chicago datasets. In addition, in this section, we perform an ablation study and a hyperparameter analysis to confirm the effectiveness of the model components introduced in this article and to assess how different hyperparameters affect the model's performance. Finally, we perform an experimental analysis of multi-step prediction and provide a visual presentation of the prediction results for the NYC dataset.

A. Datasets

A.1 Data Description

As shown in Table I, this study utilizes two traffic accident datasets from NYC and Chicago for experimental analysis. In addition, we use taxi trip data to accurately reflect the urban traffic flow as taxis frequently traverse urban areas, recording the number of taxi entrances and exits in each area. Therefore, we use this data as an essential parameter for measuring traffic volume, enabling better analysis and predicting traffic accident risk.

Table I Details of dataset

Data Types	Features
Accident Data	Latitude, longitude, times, Casualties, etc.
Traffic flow	Number of cab inflow/outflow.
POIs	Schools, recreation, residential areas, etc.
Weather	Temperature, humidity, visibility, etc.
Risk	Mild, moderate, severe
Road Features	Road type, road length and width, etc.
Peak time	8:00-13:00，16:00-21:00

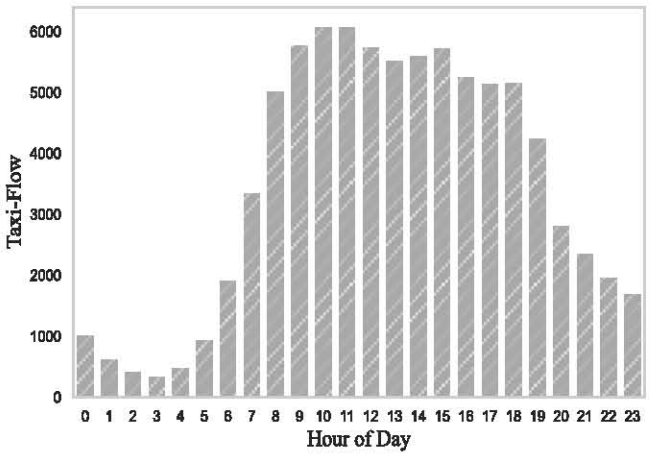

Fig. 9 Taxi flow

A.2 Data Preparation

(1) Taxi traffic data

This study reveals the dynamics of the urban traffic flow by analyzing taxi volume data. As shown in Fig. 9, taxi flows are usually at their maximum during the tidal hours of the day. This phenomenon indicates the quantity of vehicles on the road increases and the level of traffic congestion rises accordingly, thus raising the risk of traffic accidents.

(2) POI and accident risk data

Within each area of a city, the distribution and density of POIs greatly affect the probability of accidents. As shown in Fig. 10, traffic accidents are more likely to occur at intersections, junctions, and traffic signal locations, typically characterized by high traffic volumes and a tendency to violate traffic rules. Fig. 10 also illustrates that traffic accidents are rare, and accident data generally suffer from the zero-inflation problem. Therefore, it is necessary to consider the data imbalance when constructing models.

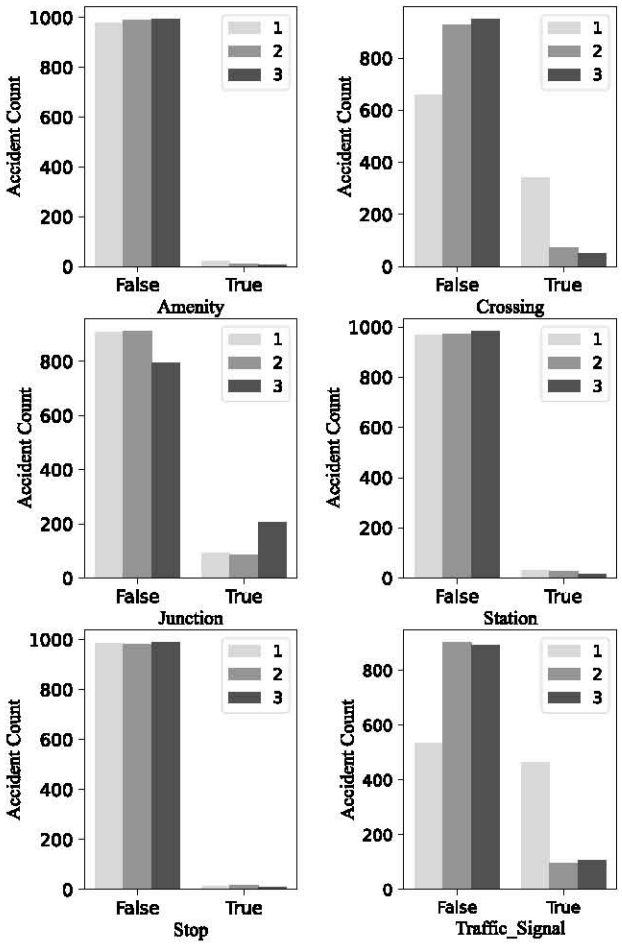

Fig. 10 Influence of POI on risk level.

(3) KG

In this study, we construct a knowledge graph using multiple sources of information, including taxi flows, POI counts, weather conditions, and road features. Inspired by the KST-GCN model, we use attribute triples to represent this information, e.g., (road 1, schools, 3) indicates that there are three schools around road section 1, and (road 2, taxi flow, accident risk) shows that there is a correlation between taxi mobility in the road section 2 and the accident risk in that road section. In addition, we consider the relationship between time and weather. For instance, (road 3, weather, time t) represents the weather conditions on Road 3 at time t, reflecting the impact of weather changes on traffic circumstances.

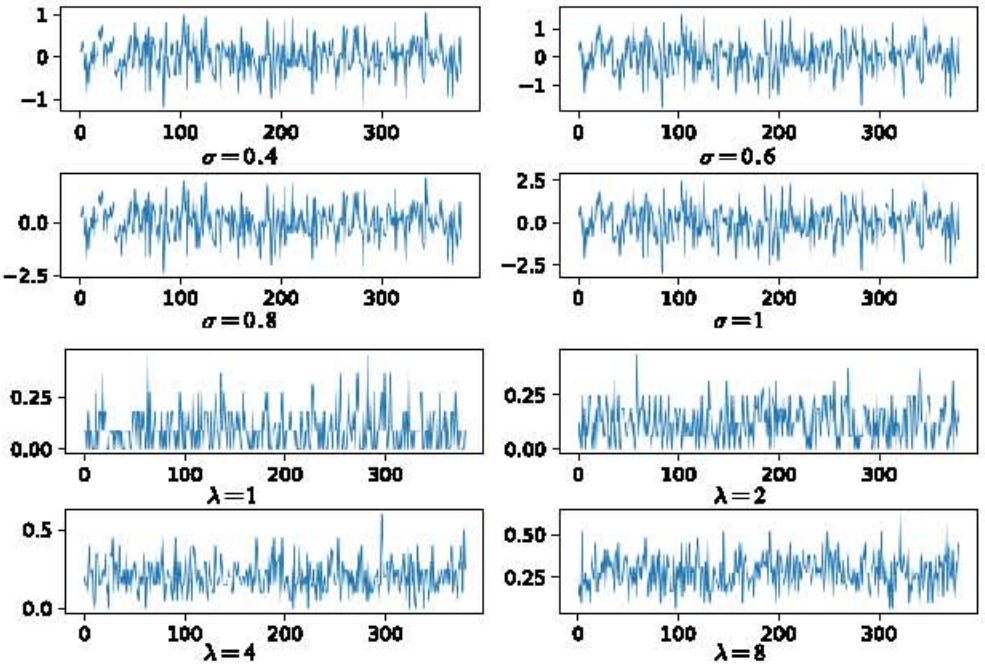

Fig.11 The data after adding noise.

A.3 Data Processing

We understand that actual traffic accident data contains rich information but also suffers from various types of noise. We introduce noise before feeding the raw data into the model to make the dataset used in this article closer to the actual situation. Therefore, we add noise following the Gaussian distribution $\boldsymbol{N \in (0, \sigma^2)\ (\sigma = 0.4, 0.6, 0.8, 1)}$, and Poisson distribution $\boldsymbol{P(\lambda)\ (\lambda \in 1, 2, 4, 8)}$. Fig. 11 illustrates the data after adding noise.

B. Evaluation Metrics

We employed four evaluation metrics to evaluate the model's performance: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Average Precision (MAP), and Recall. Lower RMSE and MAE values signify enhanced predictive model performance, indicating that the predicted risk closely aligns with the actual risk. Conversely, higher MAP and Recall values indicate superior model performance.

$$\boldsymbol{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\|y_i - \hat{y}_i\|^2} \tag{29}$$

$$\boldsymbol{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{30}$$

$$\boldsymbol{MAP} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{|X'_i|}P(j)\times R(j)\Big/|X'_i| \tag{31}$$

$$\boldsymbol{Recall} = \frac{1}{n}\sum_{i=1}^{n}X_i \cap X'_i \Big/ |X'_i| \tag{32}$$

Where $\boldsymbol{y_i}$ and $\boldsymbol{\hat{y}_i}$ denote the actual and forecasted values at time t, respectively. $\boldsymbol{P(j)}$ denotes the precision of the top j rank list, and $\boldsymbol{R(j)}$ denotes the occurrence of regional accidents. $\boldsymbol{X_i \cap X'_i}$ denotes the intersection of the set of regions with the highest predicted risk and the region of the actual occurrences of accidents, and $\boldsymbol{X'_i}$ is the region of the actual occurrences of accidents in the prediction results.

C. Hyperparameters

In this study, we implemented the Risk-CCNMAGU model using the PyTorch framework. We split the dataset into 60% as the training set, 20% as the validation set, and 20% as the test set. In order to enhance the efficiency of model training, we employ the Max-Min technique to normalize all the data within the range of [0, 1]. When setting the model parameters, we choose the adjacency time k = 3 and the long period np = 4. For the Spatial-channel CNNs module, we set the convolution kernels to 3x3 with a convolutional layer depth of 2. For the Multi-factor-Attention GCNs module, we have configured the graph convolutional network with a depth of 2 layers, each with 64 kernels. Moreover, within the GRU-Attention module, we configure the GRU network with 5 layers and hidden state sizes of 256. The training uses a learning rate of 0.00001 with a batch size of 32. The Adam optimizer is employed, and a prediction time step 1 is used.

D. Baselines

(1) LSTM: This model effectively captures the long-term dependencies in time series data of road accidents, enabling the estimation of traffic accident hazards in forthcoming times.

(2) SDCAE [37]: This model divides the city into regions. It captures spatial dependencies by stacking multiple denoising convolutional layers to forecast the risk of traffic accidents at the city scale.

(3) T-GCN [38]: This model combines GCN and GRU to consider spatial and temporal influences, resulting in more accurate traffic flow predictions.

(4) KST-GCN: This model introduces KF-Cell to effectively integrate traffic features and knowledge graph

representations, allowing KST-GCN to comprehend the connections between traffic conditions and external variables.

(5) ST-RiskNet: This model integrates local and global spatiotemporal features and considers multiple source factors such as time, weather, and traffic flow that influence accident occurrences, enabling the prediction of traffic accident risk.

(6) GSNet: This model learns spatiotemporal correlations from geographic and semantic perspectives to predict traffic accident risk.

E. Experiment Result

E.1 Performance Comparison

Tables II and III show the Risk-CCNMAGU and baseline models' prediction performance under the four evaluation metrics on both datasets. The LSTM does not perform as well as the other composite models because it only considers temporal correlations and ignores spatial ones. While SDCAE captures spatial correlations, it performs poorly in mining temporal features. T-GCN captures nonlinear spatio-temporal correlations through a combination of GCN and GRU, which improves the model's predictive performance. Still, its RMSE value is higher than that of the SDCAE model, possibly due to the influence of outliers or outliers in the data distribution. However, both SDCAE and T-GCN ignore the impact of multiple external variables on traffic data. The KST-GCN model considers both temporal and spatial correlations and the impact of multiple external variables. Thus, the KST-GCN model outperforms the T-GCN model, which validates the importance of multiple external factors for enhancing model performance. Unfortunately, the KST-GCN model has a higher MAE value, which may be due to the imbalanced dataset, which causes the model to be more inclined to predict common risk classes. However, the lower RMSE value of the KST-GCN model compensates for this drawback to some extent and has a minor impact on the model's predictive performance.

Compared to the results of KST-GCN, we find that ST-RiskNet and GSNet, which are models based on local and global regions, perform better in their prediction performance, which indicates that local and global semantic features help improve the model performance. However, these models ignore the influence of interconnections among various external variables on traffic accidents. Therefore, the proposed Risk-CCNMAGU model can adeptly encompass the influence of diverse external variables on traffic accident risk and fuse the temporal and spatial correlations of coarse- and fine-grained regions. Risk-CCNMAGU performs better on all four metrics than the baseline models mentioned above.

Table II Performance comparison of different models on the NYC.

Metric / Models	RMSE	MAE	MAP	Recall
LSTM	9.9143	7.2647	0.1216	0.2287
SDCAE	7.1782	6.8576	0.1418	0.2906
T-GCN	8.5873	6.0572	0.1461	0.2954
KST-GCN	7.0829	7.4691	0.1538	0.3022
ST-RiskNet	7.0056	6.9992	0.1674	0.3293
GSNet	6.6694	6.3488	0.1689	0.3300
Risk-CCNMAGU	**6.5729**	**6.3874**	**0.1809**	**0.3314**

Table III Performance comparison of different models on the Chicago.

Metric / Models	RMSE	MAE	MAP	Recall
LSTM	10.7582	8.4617	0.0624	0.1209
SDCAE	8.4871	7.9335	0.0845	0.1958
T-GCN	8.5913	6.0304	0.1462	0.1893
KST-GCN	8.4868	7.8015	0.0878	0.1931
ST-RiskNet	8.7527	7.4495	0.0906	0.1986
GSNet	8.7010	8.2616	0.1109	0.2277
Risk-CCNMAGU	**8.4168**	**7.8632**	**0.1293**	**0.2236**

E.2 Ablation Study

We assess the efficacy of all components of the Risk-CCNMAGU model through ablation experiments in this section. We design five models, MAG-HAU, CAC-HAU, CCMG-HAU, CCN-MAG-HAU, and CCN-MAG-AU, by removing coarse-grained regions, fine-grained regions, spatial attention mechanism, dynamic feature fusion, and peak hours, respectively.

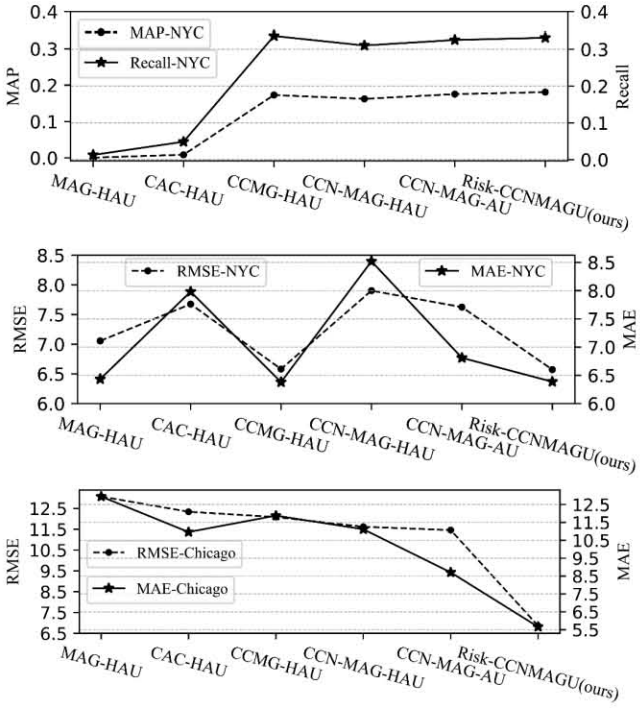

Fig. 12 Performance comparison of Risk-CCNMAGU with its variant model

Fig. 12 illustrates the comparison results between Risk-CCNMAGU and its five variant models on RMSE, MAE, MAP, and Recall. The results show that Risk-CCNMAGU outperforms the other variant models under all four evaluation metrics, affirming the efficacy of the introduced components in the model. The model prediction performance decreases when removing coarse-grained or fine-grained regions, especially on MAP and Recall. It is much lower than that of the other variant models, which indicates that it is vital to consider the characteristics of coarse-grained and fine-grained regions. When we exclude the spatial attention mechanism, CCMG-HAU displays slightly lower predictive accuracy than the Risk-CCNMAGU model on the NYC dataset. Still, the error value is much higher than that of Risk-CCNMAGU on the Chicago dataset, which suggests that

capturing the dynamic features can improve the prediction performance.

Additionally, our simple linking of two different granularity region features leads to performance degradation, demonstrating the effectiveness of dynamically fusing spatio-temporal features from two regions. Model performance is also impaired when we remove peak hour features, demonstrating the necessity of introducing peak hour data. As a result, the components designed in the Risk-CCNMAGU model effectively enhance the accuracy of the likelihood of traffic accident prediction.

E.3 Hyperparameter Study

In order to analyze the effects of various hyper-parameters on model performance, we perform experiments on the NYC dataset. This includes investigating parameters like the number of GCN filters, GRU layers, and the inclusion of Gaussian and Poisson noise.

(1) The effect of GCN filters on the model.

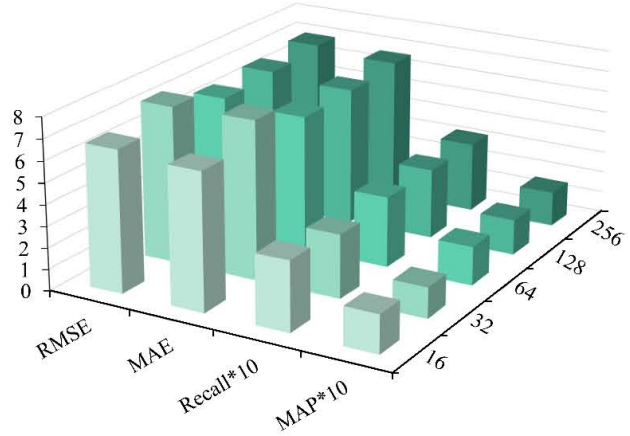

Fig. 13 Comparison of GCN-filters performance.

The results shown in Fig. 13 show that the model works well when the filter count is 64. However, both insufficient and excessive numbers of filters have an influence on the model's performance. This occurrence is attributed to the fact that an excessive number of filters augment the model's parameter count, potentially causing overfitting issues. On the other hand, insufficient filters can`t adequately capture the complex relationships and features in multiple views.

(2) The effect of GRU hidden layers on the model.

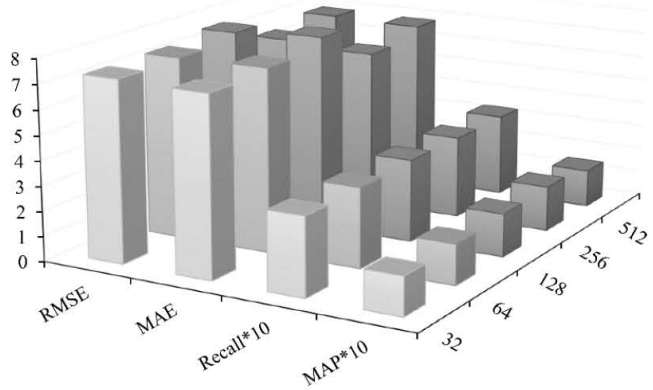

Fig. 14 Comparison of GRU hidden layers performance.

As presented in Fig. 14, the experiments in this study involve different quantities of GRU hidden layer units, precisely 32, 64, 128, 256, and 512. The results indicate that the model's performance improves considerably when the number of GRUs is 256. Conversely, the model's performance across the four evaluation metrics decreases significantly when the number of hidden units is reduced below 256.

(3) The effect of Gaussian/Poisson noise on the model.

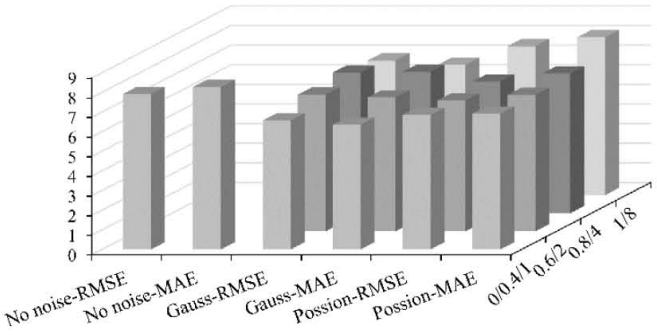

Fig. 15 Risk-CCNMAGU disturbance analysis.

Fig.15 presents the performance analysis outcomes of the Risk-CCNMAGU model using MAE and RMSE metrics for different noise conditions. The Risk-CCNMAGU model has the highest RMSE and MAE values without noise. This phenomenon is because actual traffic accident data commonly contain noise, and therefore, the spatio-temporal characteristics cannot be adequately captured using data with noise removed. However, when we add noise, the RMSE and MAE values of the model decrease significantly. In particular, we use Gaussian noise ($\sigma = 0.4$) when the error metric is the smallest.

In contrast, the RMSE and MAE values are higher than Gaussian noise when the parameters λ of the Poisson noise are 1, 2, 4, and 8. This phenomenon demonstrates the influence of added noise on the performance of the Risk-CCNMAGU model. Moderate Gaussian noise can improve the model's prediction accuracy, while higher-intensity Poisson noise may lead to performance degradation.

E.4 Multi-step Forecasting

This article also evaluates the performance of the Risk-CCNMAGU model under different prediction horizons, validating the model performance on the NYC dataset and analyzing its performance by setting the prediction steps to 1, 2, and 3. When a long period is 4, we compare the MAE and RMSE metrics between peak- and non-peak-hour data.

It can be seen by observing Figures 16, 17, and 18 that the Risk-CCNMAGU model exhibits better performance in MAE and RMSE values when using peak-hour data, compared to when not using peak-hour data when considering prediction horizons 1, 2, and 3. This phenomenon indicates that the peak-hour data favorably impacts the model's performance. In addition, the peak-hours MAE value in Figure 17 is marginally higher than the off-peak hours MAE value. Still, as the prediction step increases, the MAE value of step 1 becomes progressively lower than the off-peak-hour MAE value, indicating that Risk-CCNMAGU has an advantage in capturing long-term correlations.

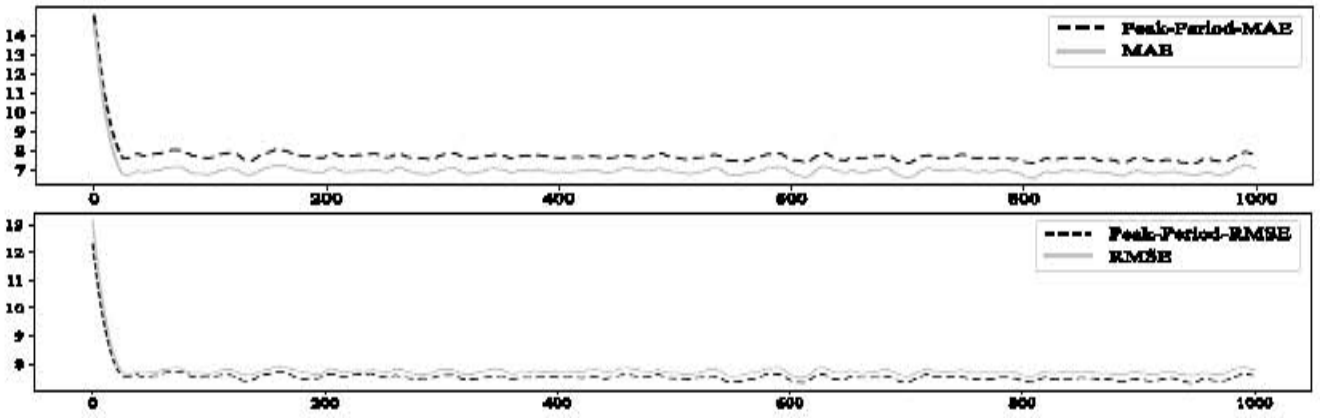

Fig. 16 Change of the error value for the prediction step size 1 .

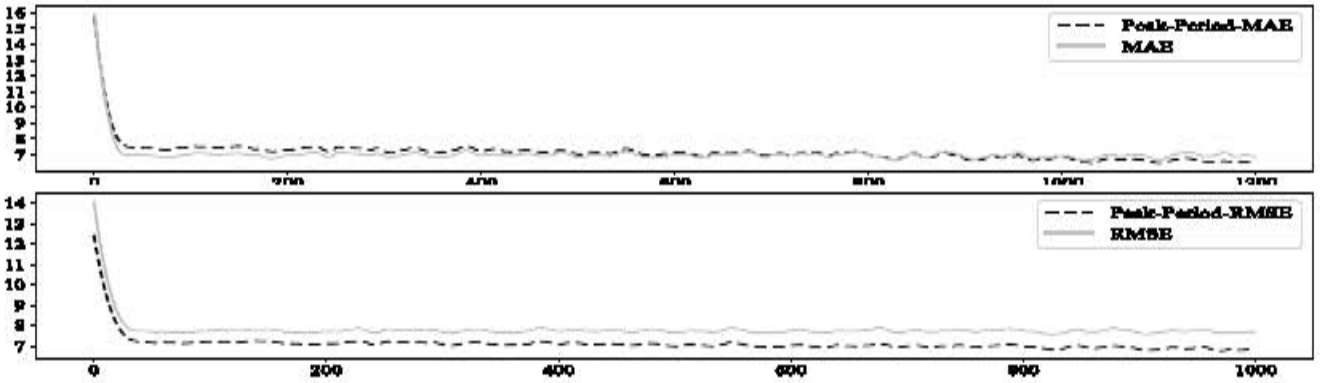

Fig. 17 Change of the error value for the prediction step size 2 .

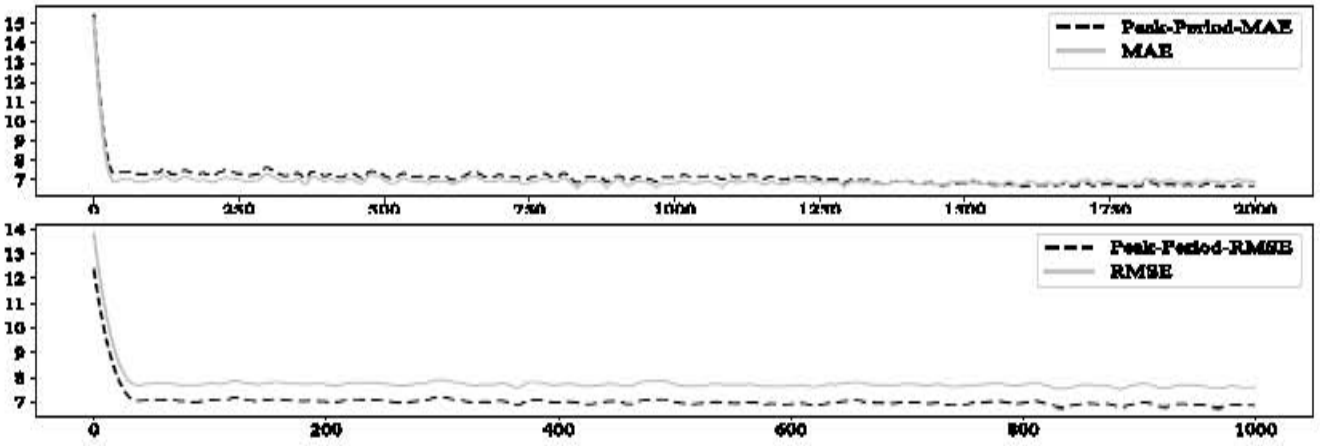

Fig. 18 Change of the error value for the prediction step size 3.

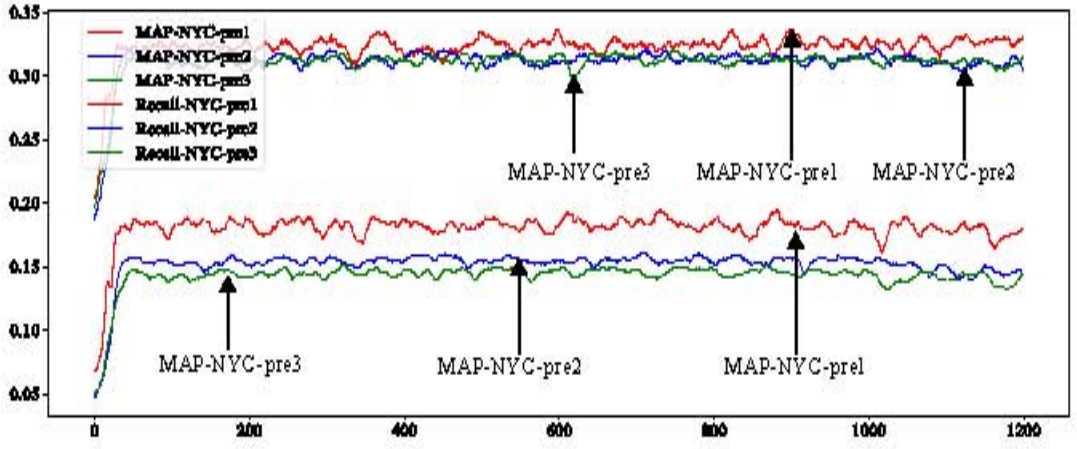

Fig. 19 Performance changes of NYC dataset.

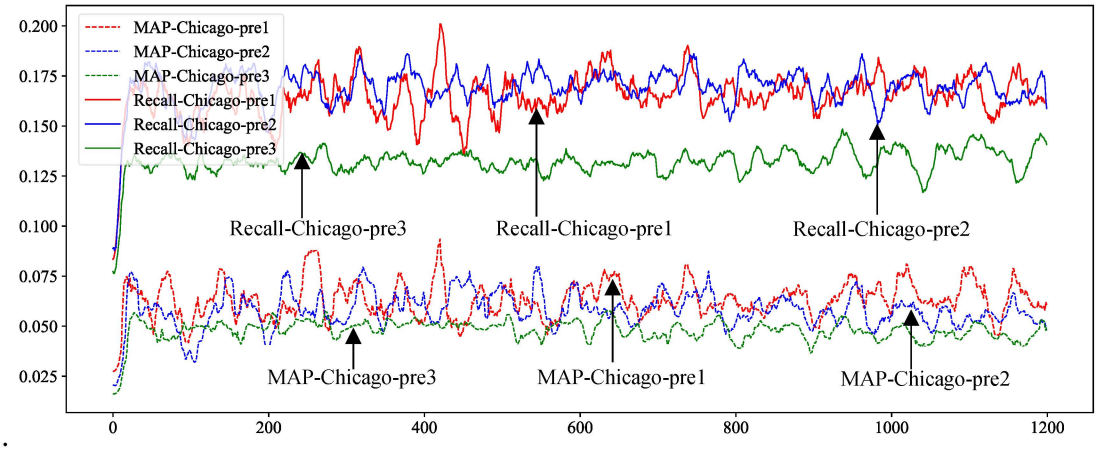

Fig. 20 Performance changes of the Chicago dataset.

We validate the Risk-CCNMAGU model on the NYC and Chicago datasets for prediction steps 1, 2, and 3 by adding high-peak period data, as shown in Figures 19 and 20. We observe that the Risk-CCNMAGU model's performance deteriorates with the prediction step size increase. In particular, the model has the lowest MAP and Recall values at a prediction step of 3. This phenomenon indicates that the model performs better in short-term prediction and can effectively exploit the correlation between static and dynamic factors. However, as the prediction step increases, the model requires longer time intervals, which may result in the loss of information and error accumulation, thus decreasing the prediction accuracy.

The MAE value gradually decreases with increasing forecast step, improving the performance of Risk-CCNMAGU. However, the MAP and Recall values are decreasing now, suggesting that they may sacrifice the precision of near-term forecasting to capture long-term spatiotemporal correlations. Therefore, this article chooses to set the forecast step size to 1 to balance capturing spatio-temporal short-term and long-term correlations better.

E.5 Research on Multi-factor Fusion

We validate the feature extraction performance of the multi-view approach introduced in our model. Therefore, we removed the knowledge graph representation view, static view, and dynamic view, respectively, thereby constructing three variant models of $G_{sp}-G_{sr}-G_{ds}-View$, $G_{ds}-G_{kg}-View$, and $G_{sp}-G_{sr}-G_{kg}-View$.

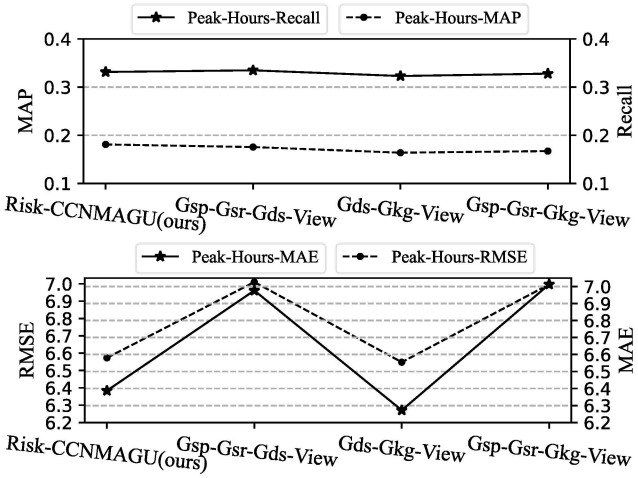

Fig. 21 Performance variation of various multi-views during peak hours.

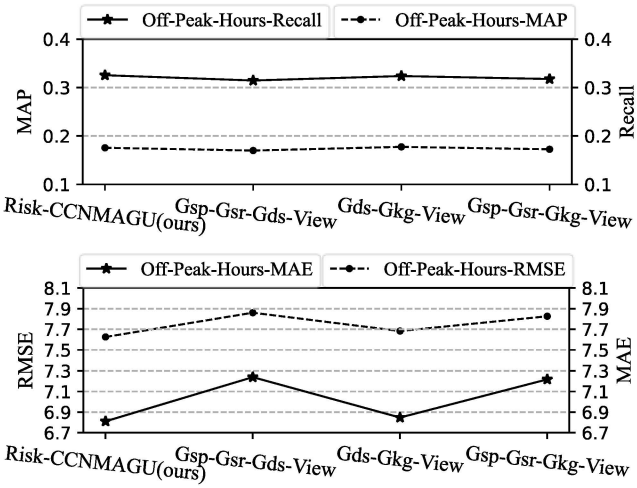

Fig. 22 Performance variation of various multi-views during off-peak hours.

According to the results in Figures 21 and 22, whether we introduce peak hours or not when we remove the knowledge graph representation, the MAP and Recall values of the Risk-CCNMAGU model have a smaller gap than other models. Still, the MAE and RMSE values are significantly higher. This phenomenon illustrates that capturing the correlation between static and dynamic factors is crucial to improving the model's predictive capability. In addition, when we remove only static or dynamic views, it will also impact the model's prediction accuracy. It's important to mention that during peak hours, the error values of the $G_{ds}-G_{kg}-View$ and $G_{sp}-G_{sr}-G_{ds}-View$ models are the smallest, further illustrating that introducing peak hour data can effectively predict traffic accident risks.

F. Model Interpretation

In section E.4, the Risk-CCNMAGU model utilizes historical data to predict traffic accident risk with step sizes of 1, 2, and 3. We visualize the prediction results in Figures 23 to 25 for a detailed explanation. Note that this section only showcases the results predicted through the NYC dataset in a limited space.

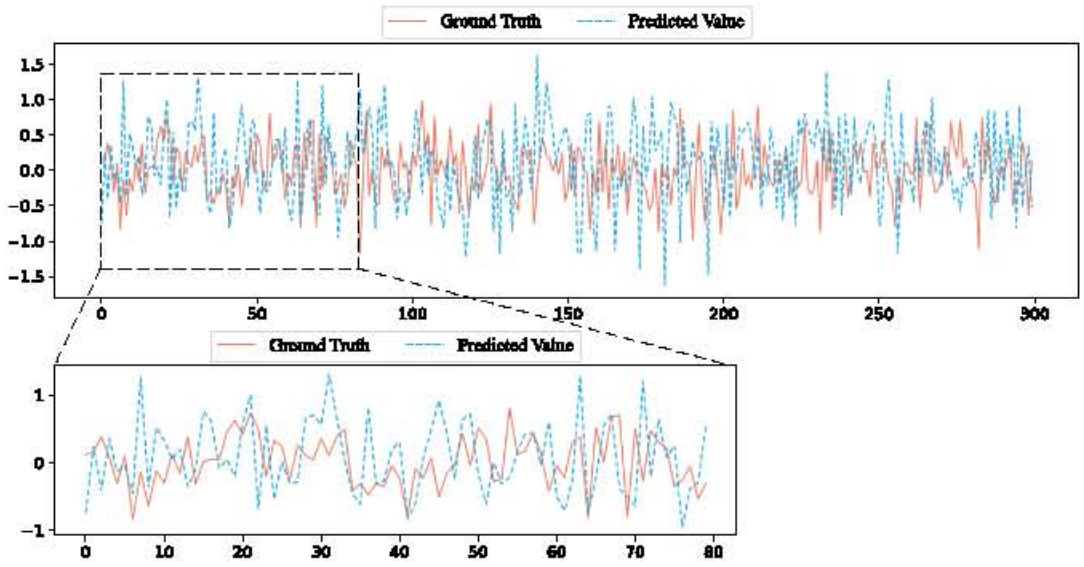

Fig. 23 Results for prediction step 1.

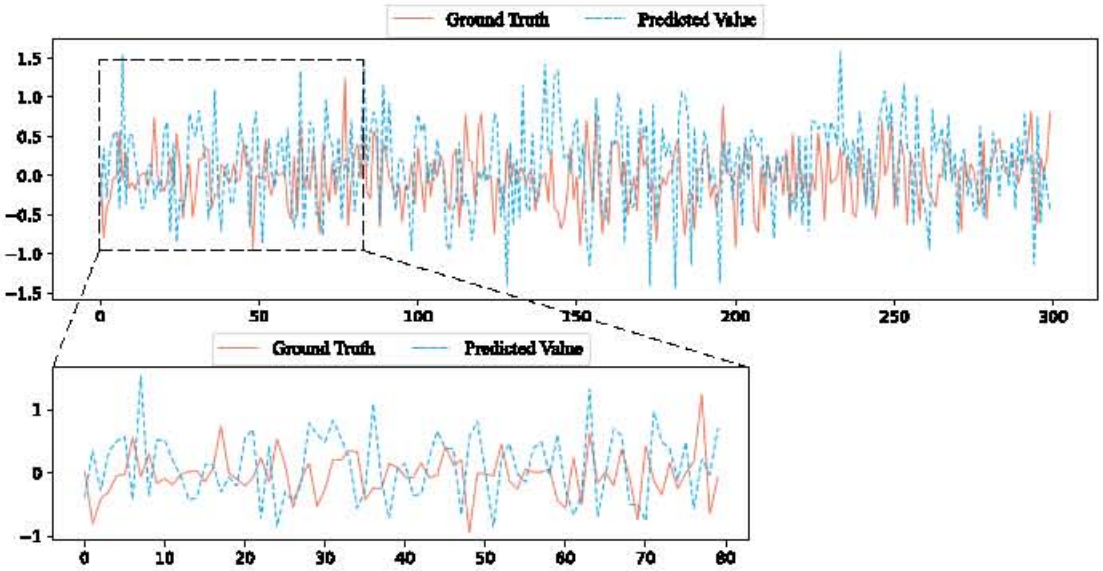

Fig. 24 Results for prediction step 2.

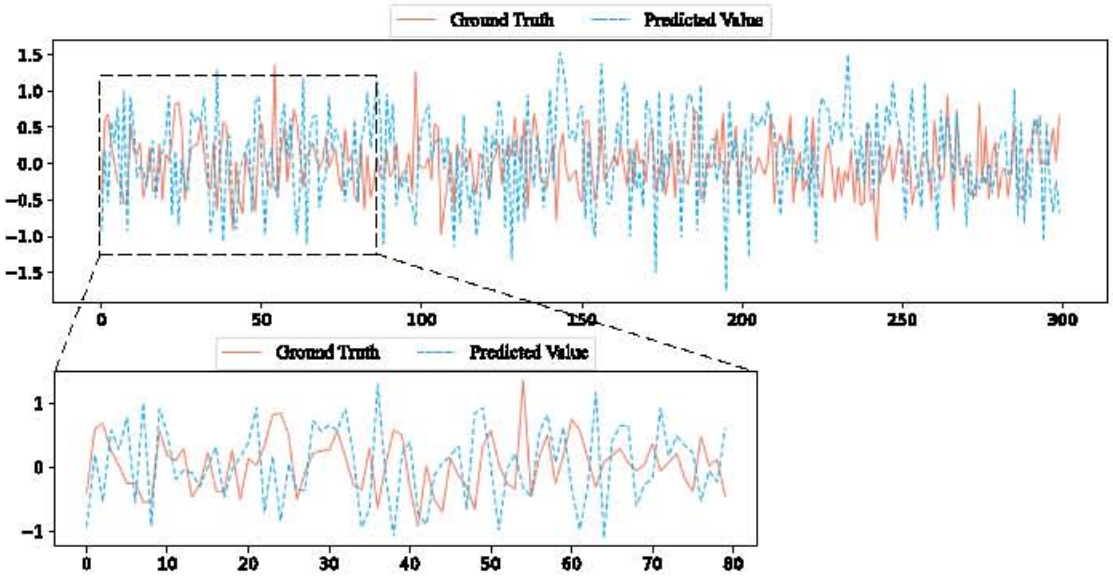

Fig. 25 Results for prediction step 3.

The results in Figures 23 to 25 demonstrate that the Risk-CCNMAGU model effectively captures the varying trends of traffic accident risk. In particular, the model aligns more closely with the ground truth for prediction step size 1, indicating higher accuracy. However, the model's predictive performance is relatively poor when the forecast step size is 2 and 3, which may be because long-term forecasts need to consider more factors to capture complex spatio-temporal dependencies. Moreover, we observe significant discrepancies between the model predictions and the actual observations at the inflection points in the predictions. This deviation may be due to extreme conditions such as heavy rainfall, fog, and other unexpected events. Such abrupt events may impact alterations in the probability of road accidents, influencing the model's performance.

VI CONCLUSION

This article proposes a spatio-temporal network model called Risk-CCNMAGU to enhance the prediction of traffic accident risk in coarse-grained and fine-grained regions. In order to catch the complex spatial and temporal dependencies in various granular regions, we designed the Spatial-channel CNNs module and the Multi-factor-Attention GCNs module. Meanwhile, we adopted a multi-view approach to extract correlations between static and dynamic factors to incurporate external factors into the model better. In addition, we employ a sample-weighted MSE loss function to mitigate the issue of data sparsity.

We assess the efficacy of the Risk-CCNMAGU model under different prediction step sizes through experimental results, and when the prediction step size is 1, the model performs better. In addition, further verified the prediction performance using ablation experiments and perturbation analysis with added noise. Compared to the baseline model, the model presented in this article exhibits superior performance across four key indicators, such as RMSE, MAE, MAP, and Recall, which fully proves the effectiveness of Risk-CCNMAGU in traffic accident risk prediction.

While the Risk-CCNMAGU model shows an improvement in prediction performance compared with other baseline models, we can still narrow its prediction horizon to the minute level, and the prediction granularity can be further refined to the road segment level, which represents its limitation. Therefore, we shall consider these factors in our forthcoming research to improve the generalization potential of the model.

REFERENCES

[1] Chinmoy Pal, Shigeru Hirayama, Sangolla Narahari, Manoharan Jeyabharath, Gopinath Prakash, and Vimalathithan Kulothungan, "An insight of world health organization (who) accident database by cluster analysis with self-organizing map (som)," *Traffic injury prevention*, 19. sup1. S15–S20, 2018.

[2] Hossain, et al, "A Bayesian network based framework for realtime crash prediction on the basic freeway segments of urban expressways," *Accident Analysis & Prevention*. vol. 45, pp. 373–381,2012.

[3] Bharti Sharma, Vinod Kumar Katiyar, and Kranti Kumar, "Traffic accident prediction model using support vector machines with gaussian kernel," *In Proceedings of fifth international conference on soft computing for problem solving*, vol. 2, pp. 1-10, 2016.

[4] Lei Lin, Qian Wang, and Adel W Sadek, "A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction," *Transportation Research Part C: Emerging Technologies*, vol. 55, pp. 444-459, 2015.

[5] Senzhang Wang, Jiannong Cao, and Philip Yu, "Deep learning for spatio-temporal data mining: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol.34, no.8, pp. 3681–3700, 2020.

[6] Wang, S., Zhang, M., Miao, H., & Yu, P. S, "MT-STNets: Multi-task spatial-temporal networks for multi-scale traffic prediction," I*n Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*, pp. 504–512, 2021, doi:10.1137/1.9781611976700.57.

[7] Huang Sen, "Time series prediction based on improved deep learning," *IAENG International Journal of Computer Science*, vol. 49, no. 4, pp. 1133-1138, 2022.

[8] Xingjian Shi, Zhourong Chen, Hao Wang, Dit Yan Yeung, Wai Kin Wong, and Wang Chun Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," *Advances in neural information processing systems*, vol. 28, 2015.

[9] P. Trirat, S. Yoon and J. -G. Lee, "MG-TAR: Multi-View Graph Convolutional Networks for Traffic Accident Risk Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 4, pp. 3779-3794, 2023, doi: 10.1109/TITS.2023.3237072.

[10] Wu M, Jia H, Luo D, et al, "A multi-attention dynamic graph convolution network with cost‐sensitive learning approach to road-level and minute‐level traffic accident prediction," *IET Intelligent Transport Systems*, vol. 17, no. 2, pp. 270-284, 2023.

[11] S. Wang, J. Zhang, J. Li, H. Miao and J. Cao, "Traffic Accident Risk Prediction via Multi-View Multi-Task Spatio-Temporal Networks," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1-1, 2021, doi: 10.1109/TKDE.2021.3135621.

[12] I. Lana, J. D. Ser, M. Velez, and E. I. Vlahogianni, "Road traffic forecasting: Recent advances and new challenges," *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 2, pp. 93-109, 2018.

[13] Wang, B., Lin, Y., Guo, S., & Wan, H, "GSNet: Learning spatial-temporal correlations from geographical and semantic aspects for traffic accident risk forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 5, pp. 4402-4409, 2021.

[14] Liao B, Zhang J, Wu C, et al., "Deep sequence learning with auxiliary information for traffic prediction," *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 537–546, 2018, doi:10.1145/3219819.3219895.

[15] D. Zhang and M. R. Kabuka, "Combining weather condition data to predict traffic flow: A GRU-based deep learning approach," *IET Intelligent Transport Systems*, vol. 12, no. 7, pp. 578–585, 2018.

[16] Yu, B. Du, et al., "Deep spatio-temporal graph convolutional network for traffic accident prediction," *Neurocomputing*, vol. 423, pp. 135–147, 2020.

[17] Moosavi, S., et al., "Accident risk prediction based on heterogeneous sparse data: New dataset and insights," *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, vol. abs/1909.09638, pp. 33–42, 2019.

[18] Huang, C., Zhang, C., Dai, P., & Bo, L, "Deep dynamic fusion network for traffic accident forecasting," *Proceedings of 28th ACM International Conference on Information and Knowledge Management*, pp. 2673–2681, 2019, doi:10.1145/3357384.3357829.

[19] Liu Z, Chen Y, Xia F, et al. TAP: Traffic Accident Profiling via Multi-Task Spatio-Temporal Graph Representation Learning[J]. ACM Transactions on Knowledge Discovery from Data, 2023, 17(4): 1-25.

[20] N. Noy, Y. Gao, A. Jain, A. Narayanan, A. Patterson, and J. Taylor, "Industry-scale knowledge graphs: Lessons and challenges," *Queue*, vol. 17, no. 2, pp. 48–75, 2019.

[21] X. Sha, Z. Sun, and J. Zhang, "Hierarchical attentive knowledge graph embedding for personalized recommendation," *Electronic Commerce Research and Applications*, vol. 48, pp. 101071, 2021.

[22] Z. Xu et al., "Building knowledge base of urban emergency events based on crowdsourcing of social media," *Concurrency and Computation: Practice and experience*, vol. 28, no. 15, pp. 4038–4052, 2016.

[23] Muppalla, RoopTeja, et al, "A knowledge graph framework for detecting traffic events using stationary cameras," *Proceedings of the 2017 ACM on Web Science Conference*, pp. 431–436, 2017.

[24] J. Zhu et al., "KST-GCN: A Knowledge-Driven Spatial-Temporal Graph Convolutional Network for Traffic Forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 15055-15065, 2022.

[25] Wang, S., Lv, Y., Peng, Y., Piao, X., & Zhang, Y, "Metro Traffic Flow Prediction via Knowledge Graph and Spatiotemporal Graph Neural Network". *Journal of Advanced Transportation*, vol. 2022, 2022.

[26] Zhou, Zhengyang, et al, "RiskOracle: a minute-level citywide traffic accident forecasting framework," *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 1, pp. 1258-1265, 2020.

[27] Bao, J., Liu, P., & Ukkusuri, S. V, "A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-

source data," *Accident Analysis & Prevention*, vol. 122, pp. 239–254, 2019.
[28] SUN, Bin, et al., "Modeling global spatial–temporal graph attention network for traffic prediction," *IEEE Access*, vol. 9, pp. 8581-8594, 2021.
[29] CHEN, Changlu, et al. Bidirectional Spatial-Temporal Adaptive Transformer for Urban Traffic Flow Forecasting. IEEE Transactions on Neural Networks and Learning Systems, 2022.
[30] Huang Y, Zhang F, Hu J., "Deep SpatialTemporal Graph Modeling of Urban Traffic Accident Prediction," *Lecture Notes in Electrical Engineering*, vol. 813, pp. 445 – 455, 2022.
[31] Chukwutoo C. Ihueze and Uchendu O. Onwurah, "Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria," *Accident Analysis & Prevention*, vol. 112, pp. 21–29, 2018.
[32] Yuxuan Liang, Zhongyuan Jiang, and Yu Zheng. Inferring traffic cascading patterns. *Proceedings of the 25th acm sigspatial international conference on advances in geographic information systems*, pp. 1-10, 2017, do:10.1145/3139958.3139960.
[33] X. Chen, S. Jia, and Y. Xiang, "A review: Knowledge reasoning over knowledge graph," *Expert Systems with Applications*, vol. 141, pp. 112948, 2020.
[34] Y. Lin, Z. Liu, and M. Sun, "Knowledge representation learning with entities, attributes and relations," *Ethnicity*, vol. 1, pp. 41–52, 2016.
[35] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," *Proceedings of the AAAI conference on artificial intelligence*, vol. 29, no. 1, pp. 2181–2187, 2015.
[36] Chen, Q., Song, X., Yamada, H., "Learning deep representation from big and heterogeneous data for traffic accident inference," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, pp. 338–344, 2016.
[37] C. Chen, X. Fan, C. Zheng, L. Xiao, M. Cheng, C. Wang, "SDCAE: Stack Denoising Convolutional Autoencoder Model for Accident Risk Prediction Via Traffic Big Data," *2018 Sixth International Conference on Advanced Cloud and Big Data (CBD)*, pp. 328-333, 2018, doi: 10.1109/CBD.2018.00065.
[38] Ling Zhao, Yujiao Song, Chao Zhang, Yu Liu, Pu Wang, Tao Lin, Min Deng, and Haifeng Li, "T-gcn: A temporal graph convolutional network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3848–3858, 2019.

Qingrong Wang, born in 1977, is a professor and master's supervisor, a teacher in the Department of Computer Science and Technology, whose main research interests are the application of big data and data mining in intelligent transportation.

Kai Zhang, the correspqonding author, was born in 1997 in Guizhou, China. He is currently a master student at Lanzhou Jiaotong University, majoring in computer technology. His main research interests are machine learning and intelligent transportation.