International Journal of Geo-Information

MDPI

Article

Graph Representation Learning for Street-Level Crime Prediction

Haishuo Gu [1], Jinguang Sui [1,2,*] and Peng Chen [1]

[1] School of Information Network Security, People's Public Security University of China, Beijing 100038, China; 2021111024@stu.ppsuc.edu.cn (H.G.); chenpeng@ppsuc.edu.cn (P.C.)
[2] School of Criminology, People's Public Security University of China, Beijing 100038, China
* Correspondence: suijinguang@ppsuc.edu.cn

Abstract: In contemporary research, the street network emerges as a prominent and recurring theme in crime prediction studies. Meanwhile, graph representation learning shows considerable success, which motivates us to apply the methodology to crime prediction research. In this article, a graph representation learning approach is utilized to derive topological structure embeddings within the street network. Subsequently, a heterogeneous information network that incorporates both the street network and urban facilities is constructed, and embeddings through link prediction tasks are obtained. Finally, the two types of high-order embeddings, along with other spatio-temporal features, are fed into a deep neural network for street-level crime prediction. The proposed framework is tested using data from Beijing, and the outcomes demonstrate that both types of embeddings have a positive impact on crime prediction, with the second embedding showing a more significant contribution. Comparative experiments indicate that the proposed deep neural network offers superior efficiency in crime prediction.

Keywords: crime prediction; graph representation learning; street network; deep learning

check for updates

Citation: Gu, H.; Sui, J.; Chen, P. Graph Representation Learning for Street-Level Crime Prediction. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 229. https://doi.org/10.3390/ijgi13070229

Academic Editors: Wolfgang Kainz, Hartwig H. Hochmair, Hao Li and Levente Juhász

Received: 29 April 2024
Revised: 26 June 2024
Accepted: 26 June 2024
Published: 1 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The spatio-temporal prediction of crime can detect or identify the future crime risk within a specific spatial unit. The prediction results can be used to guide law enforcement agencies to deploy prevention, control, and intervention strategies in a scientific way to combat crime [1]. So far, many spatio-temporal prediction models have been proposed, and they primarily employ statistical and machine learning algorithms to integrate the data of historical crime incidents, social and physical environments, and ambient populations to derive the estimated crime risk in a targeted area. Though the proposed models achieved satisfactory prediction results, the input variables (crime incidents and other related features) are mostly aggregated or organized by specifically designated areal units (e.g., uniform grids, administrative districts, and census tracts). Compared to fixed areal units, however, the street network is a unique spatial topology that encodes the connections between street segments and determines the physical links along which crime risk spreads [2]. Subsequently, the street network-based predictive crime mapping tools NTKDE and GLDNet were proposed, and they outperformed grid-based alternatives in property and assault crime forecasting [3,4]. Although these models showed better prediction performance, they did not include the potential spatial variables at multiple scales, which were thought to have a significant impact on the occurrence of crime [5], such as the raw attributes of street segments and the influence of adjacent urban functional facilities.

Representation learning methods have gained significant achievements in recent years, particularly in Natural Language Processing, Computer Vision, and Speech Recognition. The advancements also offer new perspectives for crime prediction research. The primary objective of representation learning is to uncover latent higher-order dependencies from big data. These feature vectors, also known as embeddings, are then generated and subsequently used as inputs for machine learning algorithms in downstream tasks, such

as classification and prediction. Considering the disadvantages of street network-based crime prediction research, we propose an approach which incorporates several techniques to learn temporal and spatial representation vectors potentially correlated with crime. The effectiveness of these vectors is then evaluated.

Specifically, our work comprised five sequential steps. In the first step, the issue of sparse crime data was addressed. A bi-exponential decay smoothing technique was employed to smooth the data establishing temporal dependencies in historical crime incidents. In the second step, a vector representing the intrinsic attributes of the street segment was generated. In the third step, the embedding of the street network was generated by employing Deepwalk learning. In the fourth step, a heterogeneous information network (HIN) was constructed by integrating the street network with urban functional facilities. After that, the street embedding of the HIN was obtained through link prediction tasks. Lastly, all vectors were fed into our deep neural network to realize crime prediction.

The contributions of this article are as follows:

It provides a flexible framework for crime prediction. Compared to the current state-of-the-art GLDNet, our framework treats each street segment as an independent unit. The general spatio-temporal feature representation vectors for each street segment are captured and subsequently aggregated to enable crime prediction via deep neural networks. The framework has the advantage of streamlining the integration of newly discovered features.

It utilizes graph representation learning. To capture the topological structure embedding of the street network (Street Network to Embedding, SN2V), the street network was converted into a dual graph, and the embedding learning process was conducted using the Deepwalk algorithm. Furthermore, to obtain the functional embedding (heterogeneous information network, HIN2V) of street segments within an urban built environment, a heterogeneous information network was established.

2. Related Work

2.1. Crime Prediction

Prior research has indicated that the distribution of crime is not uniform [6] but displays a phenomenon of repeat victimization in space and time [7]. It inspired the following crime concentration analysis and crime prediction research. The initial models proposed for crime prediction treated historical crime recordings as the sole input variables and borrowed the principles from other fields to predict future crime. Typical models include Prospective Mapping [8], the Contagion Model [9], the Kernel Density Estimation Model [10,11], the Self-Exciting Point Process Model [12], the Neural Network Model [13], etc. With the abundance of data, crime prediction models have started to incorporate environmental variables. These variables help capture the impacts of factors, such as weather and land use, on crime occurrences. Typical models include the Risk Terrain Model [14], Bayesian Spatio-temporal Model [15], Discrete Choice Framework [16], Tensor Decomposition [17], Graph Learning [18], and Flexible Search Window [19]. Moreover, some models include the movement pattern of the population. By analyzing patterns of human movement, such as commuting patterns or the influx of people during specific events, crime hotspots could be better understood and predicted [20–22]. However, all of the models are common in that they aggregate crime incidents into a specific temporal and areal unit and then predict the future crime risk within the unit. Most of the spatial units used are grids, such as arbitrarily defined grids, census tracts, and communities, but they also cause some problems for crime analysis and prediction. First, grid-based units have non-overlapping boundaries, which might interrupt the continuity of certain characteristics in adjacent units [23]. Second, these methods are applied under the assumption that the crime risk is uniformly distributed within a grid cell, which may be particularly problematic in cases where street segments that are co-located within a grid experience very different risks [3]. Third, the prediction accuracy varies across different spatial scales, which is known as the modified areal unit problem (MAUP) [24,25].

Facing the problem of fixed grids, researchers have been seeking to predict crime on street networks. There is growing evidence that crime variability can be attributed to street segments [26], because the shape and structure of the street network are thought to play a crucial role in influencing the distribution of crime [27,28]. More importantly, predicting crime on the street is practical and effective for police patrolling in urban cities [29,30]. The typical network-based predictive crime mapping, named NTKDE, was proposed by Rosser et al. They translated the kernel density estimation (KDE) used in Prospective Crime Mapping (ProMap) into a network space and achieved better prediction performance in property crime forecasting than grid-based alternatives. Subsequently, a deep learning model for network-based predictive mapping, GLDNet, was developed. GLDNet leveraged a graph-based representation of network-structured data and introduced a localized diffusion network to model spatial propagation. The experiment showed that GLDNet outperformed NTKDE [4]. However, there are still some limitations in network-based studies. Firstly, it is important to note that the above-mentioned studies did not fully consider the correlation between the attributes of each street segment and the occurrence of crime [31–34]. Secondly, previous research input the entire street network into deep learning models to learn spatial dependencies and try to predict the probability of crime occurrence across all street segments, which requires a huge amount of computational resources. Moreover, the role of urban functional facilities in impacting crime occurrence was proven by previous studies, for example, crime generators and crime attractors [35,36]. However, the above-mentioned studies primarily focused on learning the propagation patterns of crime risk within the network but ignored the correlation between the streets and these functional facilities in the built environment.

2.2. Street Network Modeling and Graph Representation Learning

In recent years, the availability of open-source projects like OpenStreetMap (OSM) has facilitated easy access to street network data. Various methods have been employed to model street networks, with the graph structure being the most prevalent due to its ability to represent both the geometric and topological structures of real-world street networks [37]. Two types of graph models are commonly utilized, namely primal graphs and dual graphs. In primal graphs, the nodes denote intersections, and edges denote street segments. In dual graphs, the nodes denote street segments, and edges denote intersections connecting street segments [38].

Representation learning based on street networks is a burgeoning topic in the Intelligent Transportation Systems (ITS) area. Applications of this approach span a range of tasks, including street classification, traffic flow prediction, and travel time estimation. Wang proposed a model called Road Network to Vector (RN2Vec) to jointly learn embeddings of intersections and road segments and evaluated the learned embeddings for node/edge classification and travel time estimation [39]. Zhang proposed a dual graph-based approach that encompasses both a simple graph and a hypergraph, and it is capable of capturing the low-order, high-order, and long-range relationships among roads simultaneously [40]. Zhang proposed a spatial-temporal generative adversarial network (ST-GAN) to disclose the relationship between underlying patterns and citywide traffic dynamics, which improved the prediction accuracy, and characterized the structural properties of the traffic evolution process [41]. Gharaee utilized a line graph transformation to learn highly representative road embeddings and proposed a Graph Attention Isomorphism Network to achieve road type classification [42].

Given the remarkable achievements of representation learning in ITS, it is worthwhile to further study its applicability in crime prediction. Analogous to the evaluation of traffic flow and travel time, crime incidents are also constrained by the spatio-temporal elements present in the urban built environment. This realization motivates us to employ representation learning techniques to capture higher-order spatial dependencies within street networks and subsequently examine their impact on crime prediction.

3. Problem Statement and Method

The objective of this study is to forecast the likelihood of crime occurrence across all street segments in the following week. This prediction is based on the analysis of historical crime risk patterns over time in conjunction with the attributes of street segments, spatial dependencies within the street network, and the influence of various functional facilities in the built environment.

The training data comprises instances (x, y), where x represents a four-field representation originating from various datasets related to a street segment, and y indicates whether a crime occurs on that street segment. If $y = 1$, it signifies that a crime occurs, and if $y = 0$, it signifies the absence of a crime. The x variable encompasses various spatio-temporal fields, including time-varying risk, street attributes, street network embedding, and heterogeneous information network embedding. x can be denoted as $[x_{tvr}, x_{sa2v}, x_{sn2v}, x_{hin2v}]$. Therefore, the objective of this prediction task is to construct a model, $y = CP_model(x)$, that estimates the probability of crime occurrence on a specific street in a specific time window. Essentially, this task can be framed as a binary classification problem, as demonstrated in Figure 1.

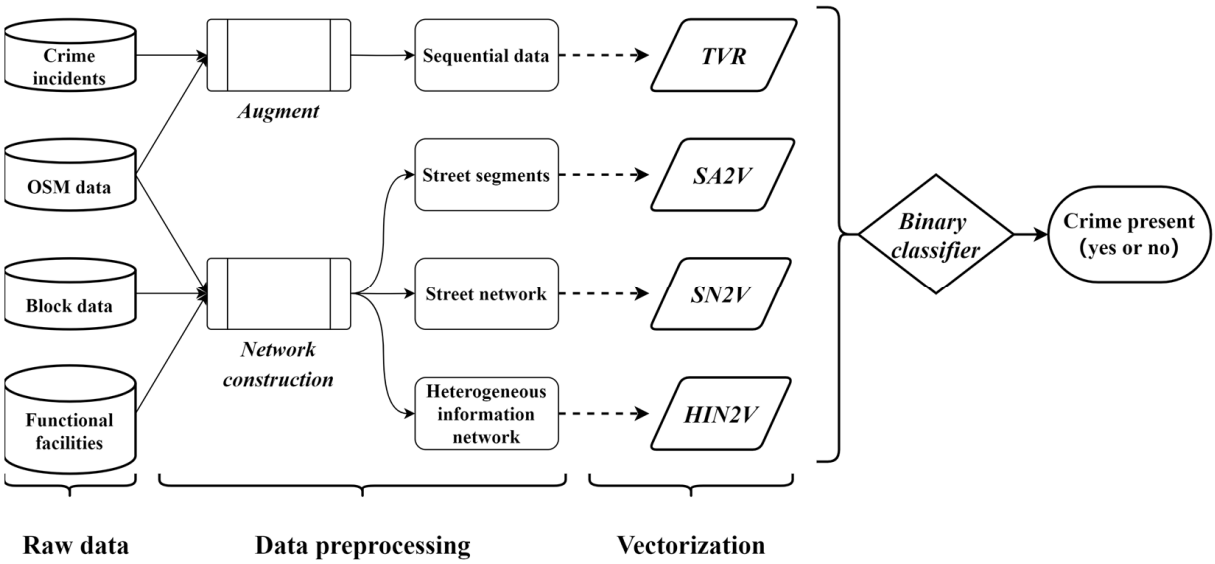


Figure 1. Crime prediction framework based on representation learning.

Initially, historical crime data were mapped onto street segments using the shortest distance method, and sequential risk data were generated for each street segment through data augmentation techniques. The risk data served as features along the temporal dimension, denoted as time-varying risk (TVR). Then, the fundamental attributes of the street segments were accessed from OpenStreetMap (OSM), and block characteristics were extracted from the Beijing Laboratory. These attributes were normalized and transformed into feature vectors to enhance their compatibility with deep learning models, referred to as Street Attributes to Vector (SA2V and SA2V_B). Next, the street network from OSM was converted into a graph structure, and the Deepwalk algorithm was applied to derive vectors encapsulating the network's structural characteristics. The vectors were named as Street Network to Vector (SN2V). Subsequently, a heterogeneous information network was constructed, incorporating the street network and functional facility entities obtained from Baidu. By performing link prediction tasks, advanced semantic features were derived and named as Heterogeneous Information Network to Vector (HIN2V). Given the advantages of the DeepFM model, such as its end-to-end processing capability and its ability to capture both low-order and high-order feature interactions [43], it was adopted as the binary classifier to fulfill the prediction task.

3.1. Time-Varying Risk (TVR)

Historical crime records have to be linked to specific street segments, a process that involves mapping each crime to its corresponding street segment. In this study, crime incidents were mapped to the nearest street segment using the Euclidean distance between their spatial coordinates. However, an issue that might be encountered is the highly sparse distribution of crime incidents on some street segments. Previous studies have demonstrated that crime risk follows an exponentially decaying trend over time [4,19]. To reflect this phenomenon, a bi-exponential decay model was utilized to construct an augmented data sequence that represents the time-varying risk. The principle of this method is illustrated below [44]:

$$S(t) = \frac{N}{p+r-q}\left[(p-q)e^{-(p+r)t} + re^{-qt}\right] \quad (1)$$

where t denotes time (hour), N denotes the number of crime incidents, $p + r$ denotes the communicative memory decay rate, q denotes the cultural memory decay rate, r denotes the rate of information flows from communicative memory to cultural memory, and p, q, and r can be estimated by data fitting.

This manipulation serves two purposes: firstly, it maintains the consistency of the processed data with the temporal trends observed in the historical data; secondly, it reduces the interference caused by a large number of zero values in subsequent learning tasks. The resulting output captures the temporal changes in crime risk associated with each street segment, and it is then used as an input for the subsequent prediction models, as shown in Figure 2.

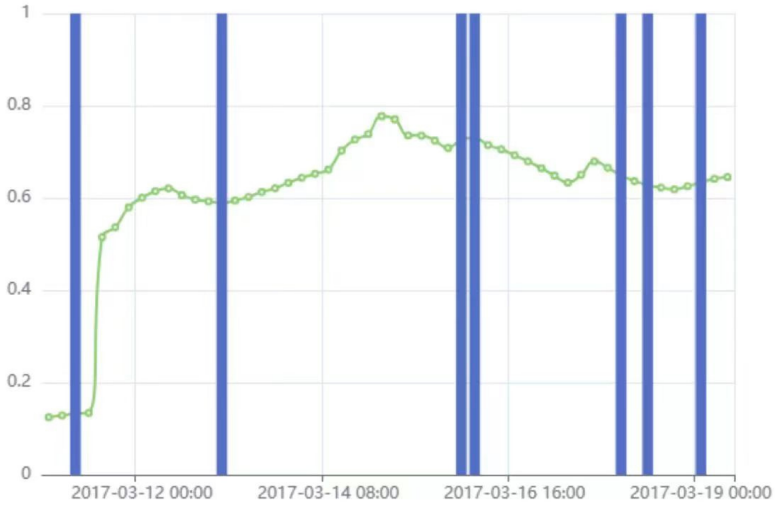


Figure 2. The crime risk following the application of bi-exponential decay smoothing on a street segment (the blue bars represent the "time of occurrence of the incidents").

3.2. Street Attributes to Vector (SA2V)

Statistical analyses have proven that a close relationship exists between street segment attributes and the occurrence of crime [45–47]. The attributes encompass street type (e.g., arterial or local and pedestrian friendliness), topological configuration, directional alignment, speed limit, length, and other relevant characteristics. However, the raw attributes are not directly suitable for model training. To integrate them into a unified representational space for training, preprocessing steps including feature selection, scaling, and normalization were performed (see Appendix A). These attributes, categorized as low-order features, were concatenated to form a 12-dimensional vector representative of each street segment.

Following the phenomenon that crime occurrence is correlated with street blocks [48,49], the block attributes were integrated as a distinctive feature set for intersecting streets within

our analysis. The one-hot encoding method was used to construct the feature vectors (see Appendix A). To distinguish this feature set from the basic street attributes, denoted as SA2V, the block attributes were referred to as SA2V_B.

3.3. Street Network to Vector (SN2V)

The street network can be conceptualized as a graph in which the nodes denote street segments and edges denote the intersections between every pair of street segments. This dual graph approach, introduced by Porta, differs from the conventional primal graph [38]. Figure 3 illustrates the process of converting a fictive urban street network into a dual graph. In this scenario, a local street network is composed of interconnected street segments (labeled by numbers) and intersections. Each street segment is assigned as a node, and the edges represent the possible routes or travel paths between pairs of nodes, which correspond to streets within the graph.

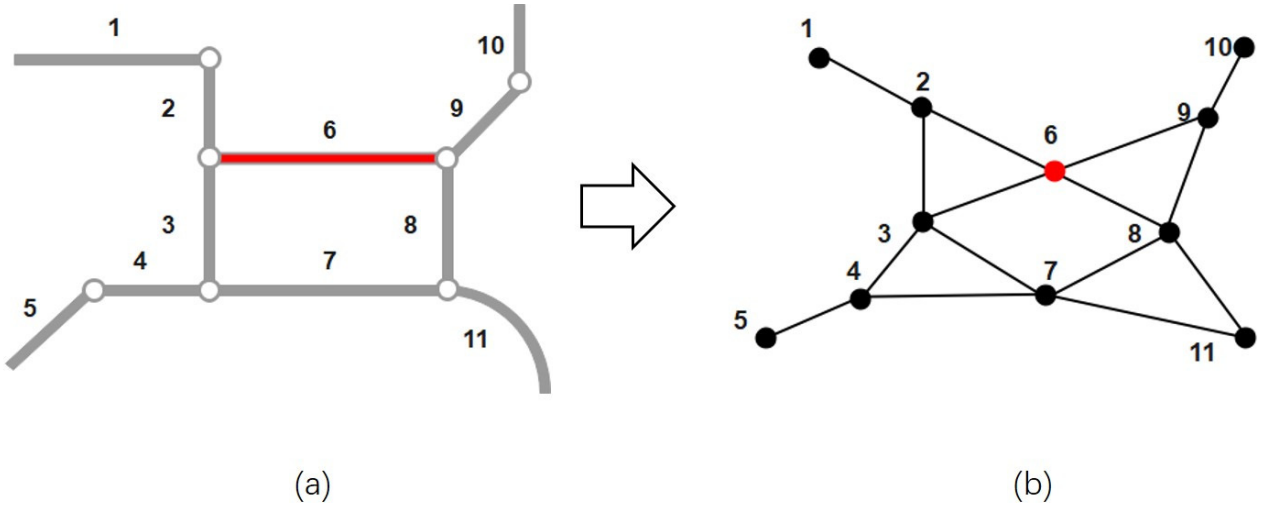


Figure 3. The dual graph transformation of a street network: (**a**) a fictive street network; (**b**) a dual graph after transformation.

As a general graph representation method, Deepwalk was utilized to derive embeddings for each street segment. It operated by simulating short-range random walks to learn node embeddings, which were then used to represent the similarity of neighborhoods and the social interaction identity of nodes [50]. The initial step involved converting the street segments into graph nodes. A node representation matrix ϕ was initialized, and a binary tree t was constructed from the node set v. Subsequently, a random walk generator was applied to each node, followed by a Skip-Gram procedure [51] that updated the node representations.

Following the experimental comparison, the optimal parameters for our model are as follows: window size (w = 5), embedding size (d = 128), walks per node (γ = 10), walk length (t = 80). Additionally, five negative samples and 22,789 iterations were utilized, respectively. Upon training the model to convergence, a 128-dimension vector was obtained for application in downstream tasks.

3.4. Heterogeneous Information Network to Vector (HIN2V)

Prior research has indicated that the urban environment significantly influences crime. However, considering that the urban environment is characterized by heterogeneous data from multiple sources, the data organization and feature extraction directly impact the model's predictive accuracy. In order to effectively model the influence of the urban environment on crime, a heterogeneous information network (HIN) was developed to represent the interconnections of fundamental urban entities.

HIN includes the administrative districts, business districts, Points of Interest (POIs)/Areas of Interest (AOIs), and streets. Figure 4 illustrates the process of transforming the layout structure of street networks and urban functional facilities into an HIN. The transformation involves converting the physical infrastructure of a city, encompassing streets and various urban amenities, into a unified network that integrates diverse types of information. The POIs were bi-directionally linked with their nearest street segments, utilizing OSRM's nearest point API service. The AOIs were processed in the same manner. The street segments were first divided at segmentation points, and then the directed relationships between them were maintained. This operation was facilitated by the NetworkX's line_graph function. For business districts, bi-directional links between the business districts and the street segments located within these areas were established. Lastly, the administrative districts were bi-directionally linked with the intersecting business districts.

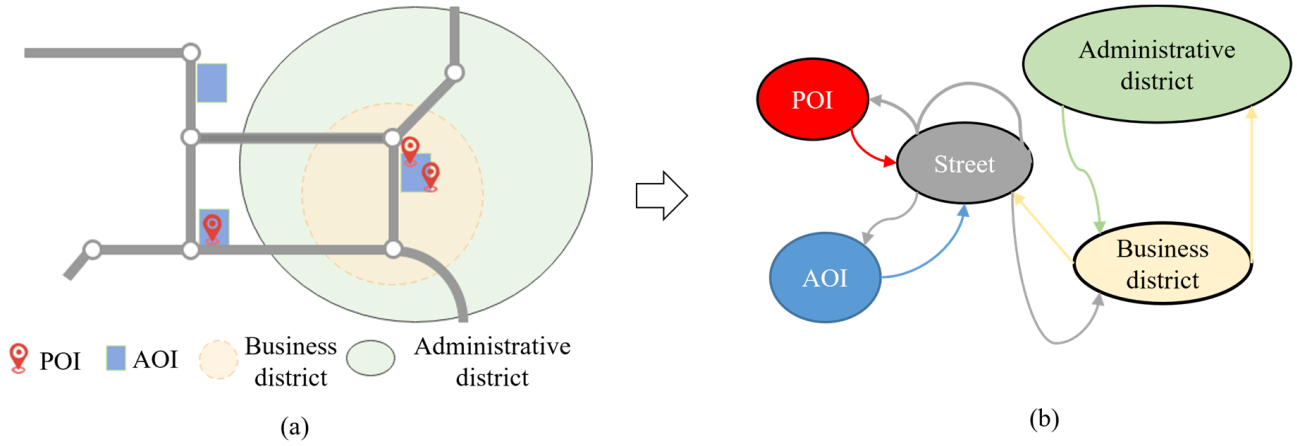


Figure 4. The HIN schema transformation: (**a**) a fictive urban built environment including an administrative district, business district, AOI, POI, and street network; (**b**) the schema of the HIN after transformation.

In an HIN, each node represents a different entity, and there are various relationships between these entities. For instance, self-loops on street nodes denote the interconnection relationship between streets themselves. Bi-directional edges between a street and POI (AOI, business district) nodes indicate their proximity within a specific distance range. Administrative districts are more macro-level entities and are represented by nodes that connect to business district nodes to capture more comprehensive semantic information.

The HIN can be simplistically viewed as a knowledge graph; thus, the link prediction techniques could be involved. The Relational Graph Convolutional Network (RGCN) algorithm was employed to predict potential connections between nodes in the network and obtain the embeddings for streets. The RGCN is designed to improve the model's sensitivity to various relationships by assigning distinct weight matrices to each type of relationship [52]. It assigns scores to possible edges (s, r, o) based on the function f(subject, relation, object). For any node pair s and o in the graph, a presentation pair ($h_s^{(L)}$ and $h_o^{(L)}$) is assigned to each. The prediction score is then calculated by directly taking the inner product of these representations. A four-layer RGCN model was constructed and trained, achieving convergence with an accuracy of 64%.

$$h_i^{(l+1)} = \sigma\left(\sum_{r\in R}\sum_{j\in N_i^r}\frac{1}{c_{i,r}}W_r^{(l)}h_j^{(l)} + W_o^{(l)}h_i^{(l)}\right) \quad (2)$$

where N_i^r denotes the set of neighbor indices of node i under relation $r \in R$, and $c_{i,r}$ is a normalization constant that can either be learned or predefined.

4. Data and Experimental Preparation

4.1. Data and Representations

Study area. The study area is Beijing, the capital city of China. In the year 2017, Beijing was home to over twenty million residents. The city is administratively divided into sixteen districts, encompassing a total area of 16,410 square kilometers.

Crime data. To evaluate the predictive performance of our proposed framework, an experiment focusing on two types of property crime in Beijing was conducted: residential burglary and pickpocketing. The recordings of residential burglaries spanned from January to December 2017 with a total of 22,478 incidents. Similarly, the recordings of pickpocketing incidents also spanned from January to December 2017 with a total of 11,274 incidents. Both datasets provided reliable temporal and spatial information. The distribution of crime incidents on the street segments is shown in Figures 5 and 6.

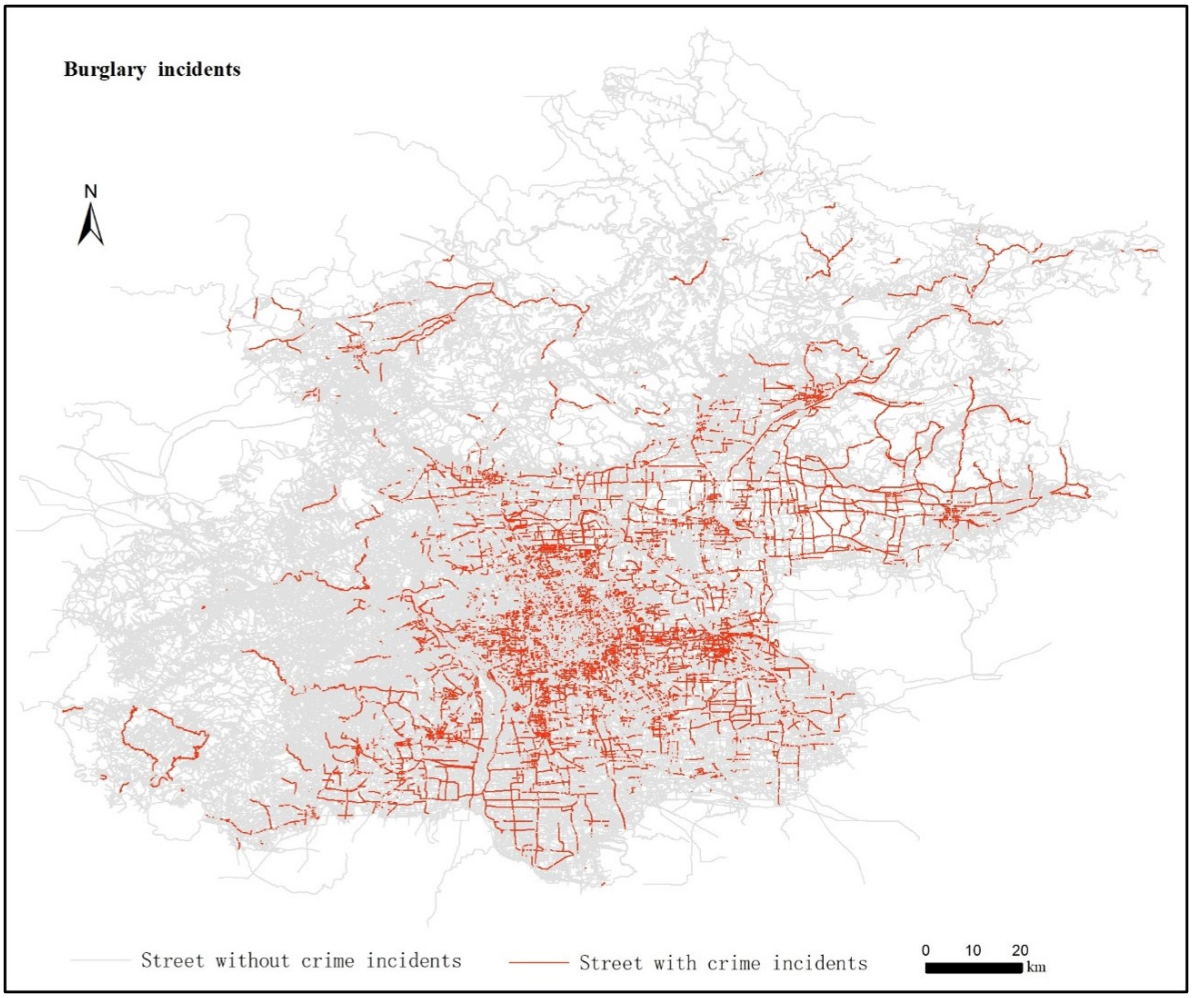


Figure 5. The distribution of burglary incidents on the street network.

Street network. The street network data of Beijing serves as the foundational layer due to the significant correlation between crime and the urban street structure [3,4,19,44,53]. A total of 292,468 instances of Beijing urban street information were acquired from OSM. The primary attributes of street segments encompass length, free_speed, allow_uses, lanes, capacity, link_type, geometry, and other relevant characteristics. Besides that, block forms, which play a significant role in shaping urban functionality and human mobility patterns, were also considered. Finally, 19,860 instances of block forms from the Beijing Urban Laboratory were collected.

Urban functional facilities. The facilities encompass POIs, AOIs, and business districts in Beijing. POIs can be used to represent the land use characteristics of specific locations. AOIs offer insights into the geographical entities that function as landmarks in local areas. Business districts reflect the spheres of influence of service centers. The data were accessed through the API interface provided by the Baidu Map Service. In total, 311 business districts, 789,245 POIs, and 31,900 AOIs were received.



Figure 6. The distribution of pickpocketing incidents on the street network.

Representations. The spatio-temporal vectors generated from all representation learning methods are displayed in Table 1.

Table 1. Input vectors list.

Factorization	Vector Name	Data Type	Dimension	Order
Temporal features	TVR	continuous	1	high
Street attributes	SA2V	discrete and continuous	12	low
Block attributes	SA2V_B	discrete	26	low
Street network	SN2V	continuous	128	high
HIN	HIN2V	continuous	16	high

4.2. Baseline Models

GLDNet. A baseline model was constructed following the previous studies [4,54]. The model is composed of two primary components. The temporal component is a two-layer Gated Recurrent Unit (GRU) [55] engineered to capture the temporal dynamics of event propagation. The spatial component integrates three distinct types of Graph Neural

Network (GNN) layers: Graph Attention Network (GAT) [56], Graph Gaussian Neural Network (GGNN) [57], and EdgeConv [58]. The layers are tailored to extract features of event propagations across the spatial dimension, which is confined by the street network. The model's architecture is shown in Figure 7. It is noted that the original model architecture of GLDNet, as presented in the literature, only accounts for two features: the time-varying risk (TVR) and Street Network to Vector (SN2V).

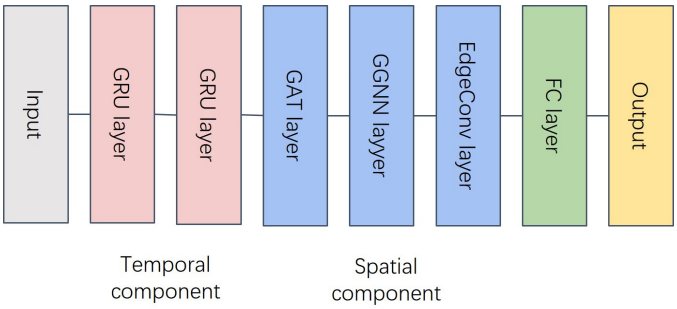


Figure 7. The structure of the baseline model.

Other machine learning algorithms. To substantiate the efficacy of our proposed prediction framework, a suite of conventional machine learning algorithms was incorporated for a comparative analysis. These algorithms include Logistic Regression (LR) [59], Decision Tree (DT) [60], Random Forest (RF) [61], XGBoost [62], and Support Vector Machines (SVMs) [63]. All of the algorithms utilized the spatio-temporal feature vectors introduced in this study.

The Stochastic Average Gradient (SAG) optimization algorithm was chosen for its efficiency with LR. Both Gini impurity and entropy were evaluated as splitting criteria with DT and RF. Gini impurity was selected for its simplicity and effectiveness. The performance of linear and RBF kernel types with the SVM was compared. Due to the frequent convergence issues encountered with the RBF kernel, we opted for the linear kernel as it provided more reliable results. Both 'gbtree' and 'gblinear' were tested as booster types with XGBoost, and 'gbtree' was chosen for its performance with tree-based models. In our model, a two-layer deep neural network (DNN) architecture was selected to test the configurations of (128,64), (256,128), and (512,256). The configuration of (256,128) was found to be the most effective one. The ReLU activation function was employed for the DNN layers, and the regularization coefficient was set to 1×10^{-5} to prevent overfitting.

4.3. Evaluation Metrics

Three standard evaluation metrics were utilized for classification tasks: the Area Under the Curve (AUC), Mean Squared Error (MSE), and accuracy (ACC). The AUC, representing the area beneath the receiver operating characteristic curve, is advantageous due to its insensitivity to class imbalance, a significant consideration given the infrequency of crime incidents. A higher AUC value indicates superior predictive model performance. Conversely, the MSE was utilized to assess the regression accuracy of the prediction models. A lower MSE value signifies enhanced prediction model performance. The ACC measures the overall correctness of the predictions made by a model, which is calculated as the ratio of the number of correct predictions to the total predictions made. A higher ACC suggests better overall performance. However, it is crucial to consider a comprehensive set of metrics, particularly in the crime prediction domain where datasets are often imbalanced.

4.4. Dataset Composition and Division

Due to the disparity between the number of streets where crimes have occurred and those that have not, it is essential to balance the dataset to prevent model bias towards the majority class. Therefore, a down-sampling technique was employed to construct a balanced dataset. Firstly, all street segments with recorded crimes were included as positive

samples in the dataset. An equivalent number of street segments with no crime records were randomly selected to serve as negative samples. The selected negative samples were then integrated into the dataset, ensuring an equal representation of both classes. The combined dataset was thoroughly mixed and randomized to eliminate any ordering bias. From the balanced dataset, a stratified random sampling method was used to allocate 60% of the data for training, 20% for validation, and 20% for testing purposes.

5. Experimental Results

5.1. Performance Comparison

The prediction results of different models are shown in Table 2 and Figure 8. As shown in the figures, the baseline model, GLDNet, does not demonstrate superior predictive performance. As discussed in Section 4.2, due to the constraints imposed by its model architecture, GLDNet is unable to incorporate all four feature vectors presented in this paper. Instead, it only takes into account TVR and SN2V. This limitation is evident and does not fully showcase the potential advantages of the GLDNet model. Consequently, comparative analyses using the identical input features were conducted. The results are presented in Table 3.

Table 2. Experimental results for burglary and pickpocketing.

Model	Burglary			Pickpocketing		
	AUC	**MSE**	**ACC**	**AUC**	**MSE**	**ACC**
GLDNet	0.506	0.341	0.572	0.505	0.431	0.429
LR	0.649	0.525	0.475	0.500	0.664	0.336
DT	0.688	0.308	0.692	0.692	0.331	0.669
RF	0.708	0.291	<u>0.709</u>	0.718	0.308	<u>0.692</u>
XGBoost	<u>0.760</u>	<u>0.233</u>	0.495	<u>0.765</u>	<u>0.249</u>	0.560
SVM	0.501	0.754	0.246	0.500	0.665	0.335
Ours	**0.808**	**0.208**	**0.738**	**0.805**	**0.226**	**0.721**

Bold indicates the best result, and underline indicates the second-best result.

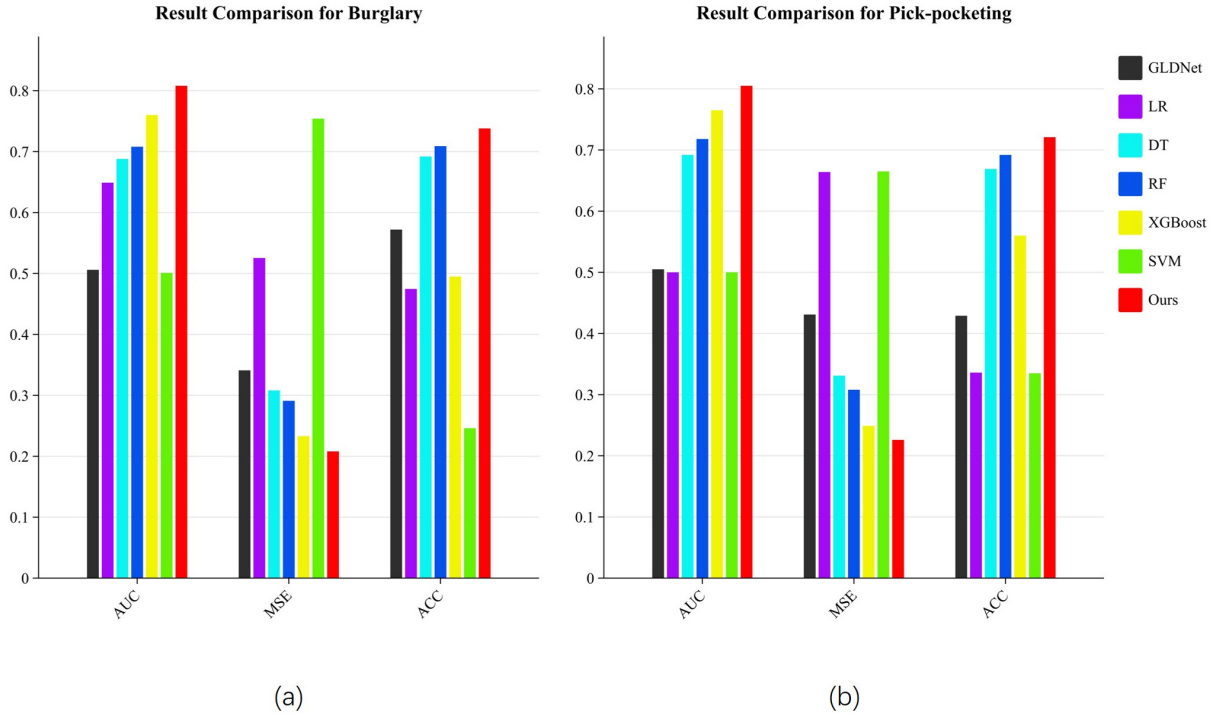


Figure 8. (**a**) The result comparison for burglary; (**b**) the result comparison for pickpocketing.

Table 3. Result comparison under identical input feature conditions.

Model	Burglary			Pickpocketing		
	AUC	MSE	ACC	AUC	MSE	ACC
GLDNet	0.506	0.341	0.572	0.505	0.431	0.429
LR	0.550	0.607	0.393	0.507	0.652	0.348
DT	0.647	0.416	0.584	0.638	0.447	0.553
RF	0.658	0.408	**0.592**	0.651	0.433	**0.567**
XGBoost	0.644	0.267	0.431	0.642	0.302	0.395
SVM	0.500	0.754	0.246	0.500	0.664	0.336
Ours	**0.717**	**0.221**	0.588	**0.716**	**0.257**	0.560

Bold indicates the best result, and underline indicates the second-best result.

Under conditions of equivalent input features, our model still demonstrates better performance. Although GLDNet shows some improvement, the enhancement is not significant. Thus, it can be inferred that the organization of the data structure has a noticeable impact on the model architecture. Subsequently, the result indicates that in addition to the topological structure (SN2V), the incorporation of high-order (HIN2V) and low-order (SA2V and SA2V_B) spatial features enables most models (except the SVM model) to achieve good performance. Notably, our model outperforms the others, realizing a 6.30% enhancement in the AUC and a 4.12% increase in ACC. Furthermore, our model demonstrates a 10.59% reduction in the MSE.

5.2. Ablation Experiment

Following the foundational experiments, ablation studies were conducted to thoroughly investigate the influence of different features on crime prediction capabilities. Four primary representation vectors were included, namely the SA2V, SA2V_B, SN2V, and HIN2V. All four vectors were sequentially eliminated to evaluate their contributions to the predictive model. The results of this assessment are presented in Table 4 and Figure 9. It needs to be clarified that the GLDNet model was not included in the ablation study due to its inherent structural constraints.

Table 4. Results of ablation experiments.

Model	Experiment 1 (All Vectors Input)		Experiment 2 (HIN2V Removed)		Experiment 3 (SN2V Removed)		Experiment 4 (SA2V_B Removed)	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
Burglary								
LR	0.500	0.475	0.523	0.421	0.515	0.407	0.501	0.389
DT	0.692	0.692	0.691	0.654	0.687	0.650	0.611	0.561
RF	0.718	0.709	0.710	**0.667**	0.702	**0.657**	0.616	**0.564**
XGBoost	0.765	0.495	0.653	0.428	0.640	0.420	0.626	0.430
SVM	0.500	0.246	0.500	0.245	0.474	0.245	0.479	0.298
Ours	**0.808**	**0.738**	**0.765**	0.645	**0.744**	0.600	**0.715**	0.510
Pickpocketing								
LR	0.500	0.336	0.523	0.373	0.515	0.360	0.501	0.337
DT	0.692	0.669	0.691	0.633	0.687	0.627	0.611	0.513
RF	0.718	0.692	0.710	**0.653**	0.702	**0.642**	0.616	0.517
XGBoost	0.765	0.560	0.653	0.420	0.640	0.408	0.626	0.382
SVM	0.500	0.335	0.500	0.336	0.474	0.440	0.479	0.413
Ours	**0.805**	**0.721**	**0.762**	0.623	**0.734**	0.589	**0.711**	**0.525**

Bold indicates the best result, and underline indicates the second-best result.

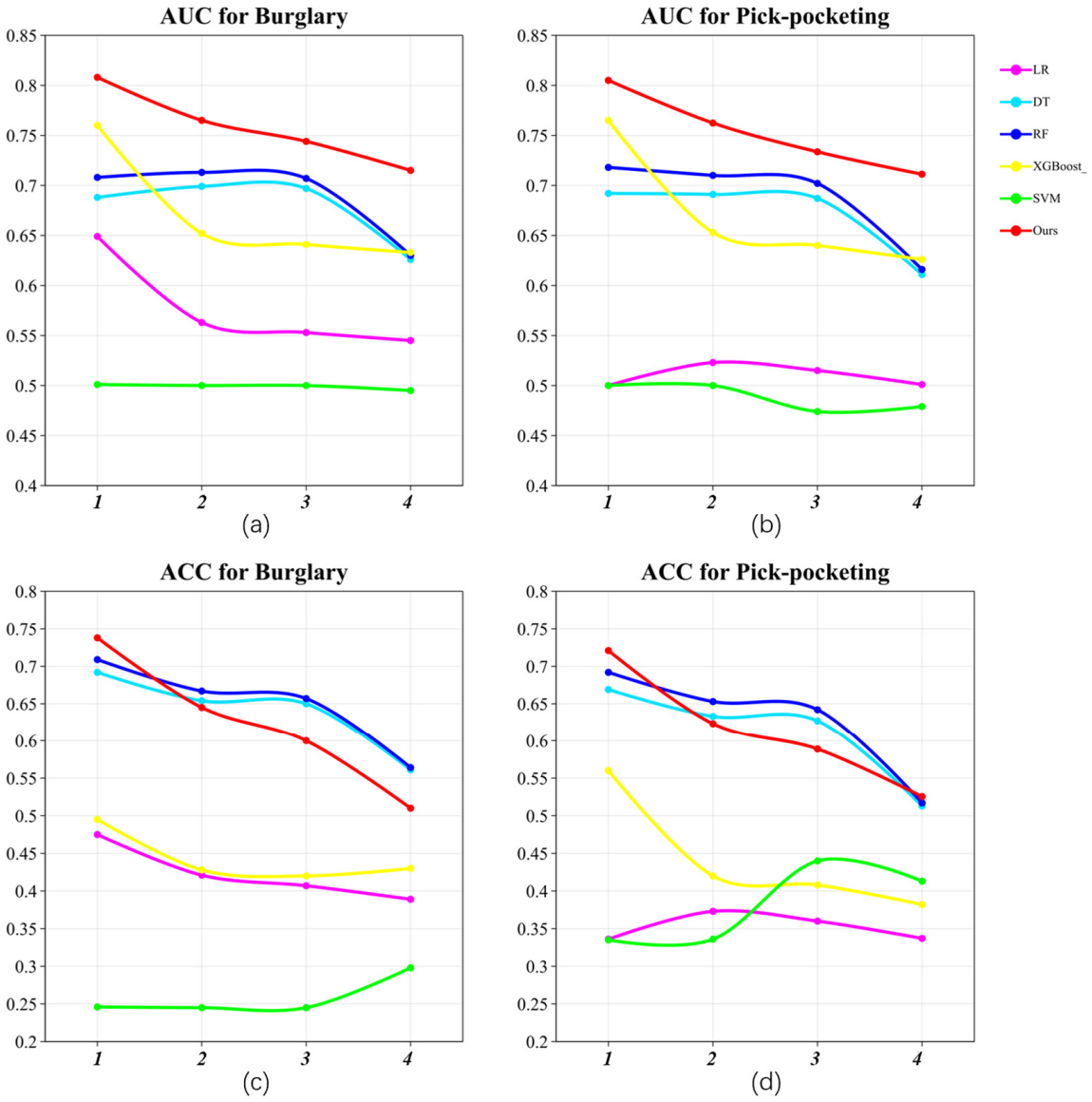


Figure 9. (**a**) AUC comparison for burglary; (**b**) AUC comparison for pickpocketing; (**c**) ACC comparison for burglary; (**d**) ACC comparison for pick-pocketing.

The ablation experiments revealed that the introduced representation vectors collectively enhanced predictive performance. Despite distinct variations in the SVM and LR models, their subpar predictive capabilities suggest that our proposed feature learning method may not be suitable for these models. Furthermore, the findings suggest that the HIN2V and SA2V_B vectors might significantly influence model performance, while SN2V generates moderate results. However, this ablation experiment only provides an intuitive insight into the contribution of feature vectors, while the exploration of potential interactive effects among multiple features is lacking. Consequently, a Shapley analysis was employed to investigate the contribution of the newly proposed features to the prediction model.

5.3. Feature Vector Interpretation

The effectiveness of machine learning models depends on the quality of feature engineering, as different representations can either obscure or reveal various explanatory factors of the underlying data [64]. The experimental results can provide a rough estimate of whether individual features or feature combinations contribute to the prediction, but it is necessary to have an accurate description of each feature's specific contribution. To

address this issue, the Shapley value was used to complete the task [65,66]. Equation (3) illustrates the working principle of the Shapley value.

$$\varnothing_j(val) = \sum_{S \subseteq \{x_i,\ldots,x_p\} \setminus \{x_j\}} \frac{|S|!(p-|S|-1)!}{p!} \left(val\left(S \cup \{x_j\}\right) - val(S)\right) \quad (3)$$

where S is a subset of the features used in the model, p represents the number of features, x_j denotes the j-th feature, and $val(S)$ represents the predicted value of the subset S, while $\varnothing_j(val)$ represents the contribution of the j-th feature to the model. The contribution of each feature to the model is derived from its marginal contribution relative to other features, measured in Shapley values.

The Shapley values of the AUC and ACC metrics were computed for two crime categories, as illustrated in Figure 10.

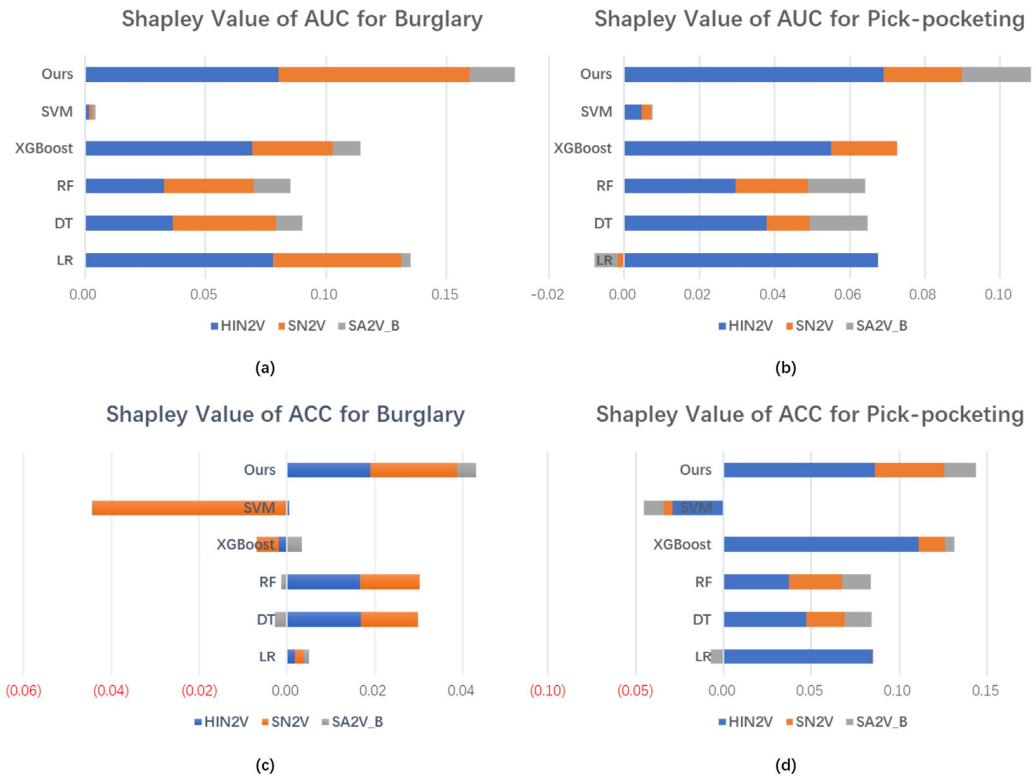


Figure 10. (**a**) Shapley value of AUCs for burglary; (**b**) Shapley value of AUCs for pickpocketing; (**c**) Shapley value of ACCs for burglary; (**d**) Shapley value of ACCs for pickpocketing.

The results indicate that HIN2V significantly enhances the prediction performance across nearly all models, as evidenced by the noteworthy improvements in both the AUC and ACC metrics for the two crime categories. This enhancement is likely attributable to the inclusion of a heterogeneous information network that encompasses more sophisticated semantic information. Such information represents nuances that conventional spatial feature representations fail to capture.

Moreover, SN2V, representing the topological structure of the street network, consistently demonstrates a notable positive impact across the majority of predictive models. This result further validates the significant influence of street network structural traits in various domains, including Intelligent Transportation Systems, as previously discussed. Thus, its favorable involvement in crime prediction is not unexpected.

Next, the block features exhibit variable performance across different models, predominantly positive effects, though occasionally negative. Within our proposed framework, they maintain a positive influence. This finding reaffirms earlier research, which indicates that

neighborhood characteristics can affect individuals' mobility and daily routines, potentially contributing to the occurrence of crime.

6. Conclusions

In this article, we initially applied a general graph representation learning approach to derive the topological structure embedding of the street network. Subsequently, we constructed a heterogeneous information network incorporating both street network and urban functional facilities. Through a link prediction task, we obtained the embeddings of street segments within the urban built environment. Subsequently, these two high-order embeddings, combined with other spatio-temporal features, were fed into a deep neural network to enable street-level crime prediction. The predictive outcomes demonstrate the positive impact of both embeddings, with a particular emphasis on the significant contribution of HIN2V. When employing our model to predict burglary crimes, the Shapley value calculations reveal that HIN2V contributes 45.06% and 44.23% to the AUC and ACC metrics, respectively. For pickpocketing crime prediction, HIN2V's performance is even more pronounced, with contributions to the AUC and ACC metrics reaching 63.8% and 59.9%, respectively. Comparatively, SN2V performs better in predicting burglary than pickpocketing crimes, contributing 44.29% and 45.76% to the AUC and ACC for the former, and 19.17% and 27.51% for the latter.

Comparative experiments have conclusively shown that our neural network outperforms other baseline models in terms of efficacy, exhibiting a 6.3% enhancement in the AUC and a 4.12% increase in ACC compared to the next best model.

There remains room for improvement in our research. Firstly, while there are various graph representation learning methods based on street networks, our study only utilized Deepwalk. In subsequent work, we intend to explore additional methods to uncover more effective embeddings. Secondly, the heterogeneous information network (HIN) we constructed is, in theory, still a static network and does not yet consider the impact of dynamic factors like population movement on crime prediction. Addressing this limitation is also a focus of our future research endeavors.

Author Contributions: Conceptualization, Haishuo Gu and Jinguang Sui; methodology, Peng Chen; software, Haishuo Gu; validation, Haishuo Gu and Jinguang Sui; formal analysis, Peng Chen; investigation, Jinguang Sui and Peng Chen; data curation, Haishuo Gu; writing—original draft preparation, Jinguang Sui; writing—review and editing, Peng Chen; visualization, Haishuo Gu and Jinguang Sui; funding acquisition, Jinguang Sui. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Fundamental Research Funds for the Central Universities under grant number 2023JKF01ZK12, the National Science and Technology Support Program Project [number 2023YFC3321604], and the Research and Innovation Project of Graduate Students Supported by Top-notch Innovative Talents Training Funds of the People's Public Security University of China under grant number 2022yjsky012.

Data Availability Statement: The dataset and codes used in this paper can be accessed at the following link: https://1drv.ms/f/s!AhzbAH_bYjqHgdwwqZ9bZUKy1QgV8Q?e=H4LDYu (accessed on 8 June 2023).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Pre-Processing Techniques for Street Attribute

The number of lanes, if empty, is taken as 0; if it is less than or equal to 2, it is taken as 1; and all other quantities are taken as 2.

The length of the street segment is discretized into 11 sections from 0 to 10. The first 10 sections are all 100 m long, and the length of the street segment over 1000 m is truncated to 10.

The bearing capacity of a street segment is converted into a three-dimensional multi-hot vector.

The speed of the street segment (free_speed) has been standardized according to the category of the street and its speed limit.

The geometry of the street segment is extracted using the h3 library developed by Uber for feature extraction.

The attributes of allow_uses and link_type are encoded with one-hot.

The basic attributes of block form, such as form_type and function, are encoded using one-hot.

The attributes of density, bld_ bldden, and mix are normalized.

The other attributes of block form, such as bld_far, bld_count, bld_floorn, shp_len, and shp_area, are discretized by a truncation value of 3, an interval value of 10, an interval value of 10, an interval value of 1000, and an interval value of 106, respectively.

The attribute of geometry is kept as default.

References

1. Weisburd, D.; Braga, A.A. *Police Innovation: Contrasting Perspectives*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2019.
2. Davies, T.P.; Bishop, S.R. Modelling Patterns of Burglary on Street Networks. *Crime Sci.* **2013**, *2*, 10. [CrossRef]
3. Rosser, G.; Davies, T.; Bowers, K.J.; Johnson, S.D.; Cheng, T. Predictive Crime Mapping: Arbitrary Grids or Street Networks? *J. Quant. Criminol.* **2017**, *33*, 569–594. [CrossRef]
4. Zhang, Y.; Cheng, T. Graph Deep Learning Model for Network-Based Predictive Hotspot Mapping of Sparse Spatio-Temporal Events. *Comput. Environ. Urban Syst.* **2020**, *79*, 101403. [CrossRef]
5. Hipp, J.R.; Williams, S.A. Advances in Spatial Criminology: The Spatial Scale of Crime. *Annu. Rev. Criminol.* **2020**, *3*, 75–95. [CrossRef]
6. Sherman, L.W.; Gartin, P.R.; Buerger, M.E. Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place. *Criminology* **1989**, *27*, 27–56. [CrossRef]
7. Farrell, G.; Pease, K. *Once Bitten, Twice Bitten: Repeat Victimisation and Its Implications for Crime Prevention*; Police Research Group Crime Prevention Unit Paper: London, UK, 1993.
8. Bowers, K.J. Prospective Hot-Spotting: The Future of Crime Mapping? *Br. J. Criminol.* **2004**, *44*, 641–658. [CrossRef]
9. Townsley, M.K.; Homel, R.; Chaseling, J. Infectious Burglaries. A Test of the Near Repeat Hypothesis. *Br. J. Criminol.* **2003**, *43*, 615–633. [CrossRef]
10. Chainey, S.; Ratcliffe, J. *Identifying Crime Hotspots*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2013.
11. Kalinic, M.; Krisp, J.M. Kernel Density Estimation (KDE) vs. Hot-Spot Analysis–Detecting Criminal Hot Spots in the City of San Francisco. In Proceedings of the 21st AGILE Conference on Geographic Information Science, Lund, Sweden, 12–15 June 2018.
12. Mohler, G.O.; Short, M.B.; Brantingham, P.J.; Schoenberg, F.P.; Tita, G.E. Self-Exciting Point Process Modeling of Crime. *J. Am. Stat. Assoc.* **2011**, *106*, 100–108. [CrossRef]
13. Corcoran, J.J.; Wilson, I.D.; Ware, J.A. Predicting the Geo-Temporal Variations of Crime and Disorder. *Int. J. Forecast.* **2003**, *19*, 623–634. [CrossRef]
14. Kennedy, L.W.; Caplan, J.M.; Piza, E. Risk Clusters, Hotspots, and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies. *J. Quant. Criminol.* **2011**, *27*, 339–362. [CrossRef]
15. Law, J.; Quick, M.; Chan, P. Bayesian Spatio-Temporal Modeling for Analysing Local Patterns of Crime Over Time at the Small-Area Level. *J. Quant. Criminol.* **2014**, *30*, 57–78. [CrossRef]
16. Bernasco, W. Modeling Micro-Level Crime Location Choice: Application of the Discrete Choice Framework to Crime at Places. *J. Quant. Criminol.* **2010**, *26*, 113–138. [CrossRef]
17. Ge, L.; Liu, J.; Zhou, A.; Li, H. Crime Rate Inference Using Tensor Decomposition. In Proceedings of the 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Guangzhou, China, 8–12 October 2018; pp. 713–717.
18. Wang, B.; Luo, X.; Zhang, F.; Yuan, B.; Bertozzi, A.L.; Brantingham, P.J. Graph-Based Deep Modeling and Real Time Forecasting of Sparse Spatio-Temporal Data. *arXiv* **2018**, arXiv:1804.00684.
19. Shiode, S.; Shiode, N. A Network-Based Scan Statistic for Detecting the Exact Location and Extent of Hotspots along Urban Streets. *Comput. Environ. Urban Syst.* **2020**, *83*, 101500. [CrossRef]
20. Kadar, C.; Pletikosa, I. Mining Large-Scale Human Mobility Data for Long-Term Crime Prediction. *EPJ Data Sci.* **2018**, *7*, 26. [CrossRef]
21. Song, G.; Liu, L.; Bernasco, W.; Xiao, L.; Zhou, S.; Liao, W. Testing Indicators of Risk Populations for Theft from the Person across Space and Time: The Significance of Mobility and Outdoor Activity. *Ann. Am. Assoc. Geogr.* **2018**, *108*, 1370–1388. [CrossRef]
22. Song, G.; Bernasco, W.; Liu, L.; Xiao, L.; Zhou, S.; Liao, W. Crime Feeds on Legal Activities: Daily Mobility Flows Help to Explain Thieves' Target Location Choices. *J. Quant. Criminol.* **2019**, *35*, 831–854. [CrossRef]

23. Hipp, J.R.; Boessen, A. Egohoods as Waves Washing across the City: A New Measure of "Neighborhoods". *Criminology* **2013**, *51*, 287–327. [CrossRef]
24. Dark, S.J.; Bram, D. The Modifiable Areal Unit Problem (MAUP) in Physical Geography. *Prog. Phys. Geogr. Earth Environ.* **2007**, *31*, 471–479. [CrossRef]
25. Openshaw, S. A Geographical Solution to Scale and Aggregation Problems in Region-Building, Partitioning and Spatial Modelling. *Trans. Inst. Br. Geogr.* **1977**, *2*, 459. [CrossRef]
26. Steenbeek, W.; Weisburd, D. Where the Action Is in Crime? An Examination of Variability of Crime Across Different Spatial Units in The Hague, 2001–2009. *J. Quant. Criminol.* **2016**, *32*, 449–469. [CrossRef]
27. Braga, A.A.; Papachristos, A.V.; Hureau, D.M. The Concentration and Stability of Gun Violence at Micro Places in Boston, 1980–2008. *J. Quant. Criminol.* **2010**, *26*, 33–53. [CrossRef]
28. Di Bella, E.; Corsi, M.; Leporatti, L.; Persico, L. The Spatial Configuration of Urban Crime Environments and Statistical Modeling. *Environ. Plan. B Urban Anal. City Sci.* **2017**, *44*, 647–667. [CrossRef]
29. Chen, H.; Cheng, T.; Wise, S. Developing an Online Cooperative Police Patrol Routing Strategy. *Comput. Environ. Urban Syst.* **2017**, *62*, 19–29. [CrossRef]
30. Chen, H.; Cheng, T.; Ye, X. Designing Efficient and Balanced Police Patrol Districts on an Urban Street Network. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 269–290. [CrossRef]
31. Kim, Y.-A.; Hipp, J.R. Pathways: Examining Street Network Configurations, Structural Characteristics and Spatial Crime Patterns in Street Segments. *J. Quant. Criminol.* **2020**, *36*, 725–752. [CrossRef]
32. Kim, Y.-A. Examining the Relationship Between the Structural Characteristics of Place and Crime by Imputing Census Block Data in Street Segments: Is the Pain Worth the Gain? *J. Quant. Criminol.* **2018**, *34*, 67–110. [CrossRef]
33. Kim, Y.-A.; Hipp, J.R. Physical Boundaries and City Boundaries: Consequences for Crime Patterns on Street Segments? *Crime Delinq.* **2018**, *64*, 227–254. [CrossRef]
34. Weisburd, D.; White, C.; Wooditch, A. Does Collective Efficacy Matter at the Micro Geographic Level?: Findings from a Study of Street Segments. *Br. J. Criminol.* **2020**, *60*, 873–891. [CrossRef]
35. Brantingham, P.; Brantingham, P. Criminality of Place: Crime Generators and Crime Attractors. *Eur. J. Crim. Policy Res.* **1995**, *3*, 5–26. [CrossRef]
36. Bernasco, W.; Block, R. Robberies in Chicago: A Block-Level Analysis of the Influence of Crime Generators, Crime Attractors, and Offender Anchor Points. *J. Res. Crime Delinq.* **2011**, *48*, 33–57. [CrossRef]
37. Boeing, G. Street Network Models and Indicators for Every Urban Area in the World. *Geogr. Anal.* **2022**, *54*, 519–535. [CrossRef]
38. Porta, S.; Crucitti, P.; Latora, V. The Network Analysis of Urban Streets: A Dual Approach. *Phys. A Stat. Mech. Appl.* **2006**, *369*, 853–866. [CrossRef]
39. Wang, M.-X.; Lee, W.-C.; Fu, T.-Y.; Yu, G. On Representation Learning for Road Networks. *ACM Trans. Intell. Syst. Technol.* **2021**, *12*, 1–27. [CrossRef]
40. Zhang, L.; Long, C. Road Network Representation Learning: A Dual Graph-Based Approach. *ACM Trans. Knowl. Discov. Data* **2023**, *17*, 1–25. [CrossRef]
41. Zhang, H.; Wu, Y.; Tan, H.; Dong, H.; Ding, F.; Ran, B. Understanding and Modeling Urban Mobility Dynamics via Disentangled Representation Learning. *IEEE Trans. Intell. Transport. Syst.* **2022**, *23*, 2010–2020. [CrossRef]
42. Gharaee, Z.; Kowshik, S.; Stromann, O.; Felsberg, M. Graph Representation Learning for Road Type Classification. *Pattern Recognit.* **2021**, *120*, 108174. [CrossRef]
43. Guo, H.; Tang, R.; Ye, Y.; Li, Z.; He, X. DeepFM: A Factorization-Machine Based Neural Network for CTR Prediction. *arXiv* **2017**, arXiv:1703.04247.
44. Candia, C.; Jara-Figueroa, C.; Rodriguez-Sickert, C.; Barabási, A.-L.; Hidalgo, C.A. The Universal Decay of Collective Memory and Attention. *Nat. Hum. Behav.* **2019**, *3*, 82–91. [CrossRef]
45. Frith, M.J.; Johnson, S.D.; Fry, H.M. Role of the Street Network in Burglars' Spatial Decision-Making. *Criminology* **2017**, *55*, 344–376. [CrossRef]
46. Zeng, M.; Mao, Y.; Wang, C. The Relationship between Street Environment and Street Crime: A Case Study of Pudong New Area, Shanghai, China. *Cities* **2021**, *112*, 103143. [CrossRef]
47. Zhou, B.; Chen, L.; Zhou, F.; Li, S.; Zhao, S.; Das, S.K.; Pan, G. ESCORT: Fine-Grained Urban Crime Risk Inference Leveraging Heterogeneous Open Data. *IEEE Syst. J.* **2021**, *15*, 4656–4667. [CrossRef]
48. He, Z.; Wang, Z.; Xie, Z.; Wu, L.; Chen, Z. Multiscale Analysis of the Influence of Street Built Environment on Crime Occurrence Using Street-View Images. *Comput. Environ. Urban Syst.* **2022**, *97*, 101865. [CrossRef]
49. Long, Y.; Li, P.; Hou, J. xuan Three-Dimensional Urban Form at the Street Block Level for Major Cities in China. *Shanghai Urban Plan. Rev.* **2019**, *3*, 10–15.
50. Perozzi, B.; Al-Rfou, R.; Skiena, S. DeepWalk: Online Learning of Social Representations. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 701–710.
51. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv* **2013**, arXiv:1301.3781.

52. Schlichtkrull, M.; Kipf, T.N.; Bloem, P.; van den Berg, R.; Titov, I.; Welling, M. Modeling Relational Data with Graph Convolutional Networks. In *The Semantic Web*; Gangemi, A., Navigli, R., Vidal, M.-E., Hitzler, P., Troncy, R., Hollink, L., Tordai, A., Alam, M., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2018; Volume 10843, pp. 593–607, ISBN 978-3-319-93416-7.
53. Summers, L.; Johnson, S.D. Does the Configuration of the Street Network Influence Where Outdoor Serious Violence Takes Place? Using Space Syntax to Test Crime Pattern Theory. *J. Quant. Criminol.* **2017**, *33*, 397–420. [CrossRef]
54. Yan, S.; Xiong, Y.; Lin, D. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 27 April 2018.
55. Cho, K.; van Merrienboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv* **2014**, arXiv:1406.1078.
56. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *arXiv* **2018**, arXiv:1710.10903.
57. Li, Y.; Tarlow, D.; Brockschmidt, M.; Zemel, R. Gated Graph Sequence Neural Networks. *arXiv* **2017**, arXiv:1511.05493.
58. Wang, S.; Cao, J.; Yu, P.S. Deep Learning for Spatio-Temporal Data Mining: A Survey. *IEEE Trans. Knowl. Data Eng.* **2019**, *34*, 3681–3700. [CrossRef]
59. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2013; ISBN 978-0-470-58247-3.
60. Salzberg, S.L. C4.5: Programs for Machine Learning by J. Ross Quinlan. *Mach. Learn.* **1994**, *16*, 235–240. [CrossRef]
61. Breiman, L. Random Forests. In *Machine Learning*; MIT Press: Cambridge, MA, USA, 2001; pp. 5–32.
62. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA USA, 13 August 2016; pp. 785–794.
63. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
64. Bengio, Y.; Courville, A.; Vincent, P. Representation Learning: A Review and New Perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *35*, 1798–1828. [CrossRef]
65. Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4 December 2017.
66. Zhang, X.; Liu, L.; Lan, M.; Song, G.; Xiao, L.; Chen, J. Interpretable Machine Learning Models for Crime Prediction. *Comput. Environ. Urban Syst.* **2022**, *94*, 101789. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.