

Article

Crime Spatiotemporal Prediction Through Urban Region Representation by Using Building Footprints

Tao Wang, Peng Chen * and Miaoxuan Shan

School of Information and Cyber Security, People's Public Security University of China, Beijing 100038, China; wangtao466@stu.ppsuc.edu.cn (T.W.); 2023111021@stu.ppsuc.edu.cn (M.S.)

* Correspondence: chenpeng@ppsuc.edu.cn

Abstract

Current crime spatiotemporal prediction models are limited by the insufficient ability of POI data to represent the continuity and mixed-use nature of urban spatial functions. To address this, our study applies an urban region representation method based on building footprints and validates its effectiveness in improving the accuracy of crime spatiotemporal prediction. Specially, we first use the Region Dual Contrastive Learning algorithm to generate region representations as a region graph by integrating building footprints and POI data. Then, the region graph combined with crime data is input into crime prediction models to predict four crime types, including Burglary, Robbery, Felony Assault, and Grand Larceny. Finally, ablation experiments are conducted to quantify the contribution of building footprints to prediction improvement. The experimental results on New York City crime data indicate that (1) the region representations significantly improve deep learning model performance, with the most improved LSTM achieving average increases of 5.66% in Macro-F1 and 18.57% in Micro-F1, particularly benefiting baseline models with lower accuracy, and (2) the region representations yield more significant improvements for low-frequency crime categories and mitigates temporal memory decay in long-term predictions. These findings confirm that incorporating urban region representation based on building footprints effectively enhances crime spatiotemporal prediction performance, providing a more precise and efficient tool for urban security management to optimize police resource allocation and crime prevention strategies.



Academic Editor: Fabrizio Messina

Received: 8 September 2025

Revised: 11 November 2025

Accepted: 25 November 2025

Published: 27 November 2025

Citation: Wang, T.; Chen, P.; Shan, M. Crime Spatiotemporal Prediction Through Urban Region Representation by Using Building Footprints. *Big Data Cogn. Comput.* **2025**, *9*, 301. <https://doi.org/10.3390/bdcc9120301>

Copyright: © 2025 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/>).

Keywords: crime spatiotemporal prediction; building footprints; urban region representation; property crime; data fusion

1. Introduction

Crime poses a significant threat to social stability and public safety. According to statistics, more than seven individuals lose their lives to violent causes every hour in the United States [1], while crimes result in annual losses and associated costs exceeding \$4.9 trillion nationwide [2]. Confronted with the complex evolution of criminal patterns, traditional passive-reactive governance models urgently require transformation into data-driven proactive early warning mechanisms. Crime spatiotemporal prediction refers to the analysis and modeling of implicit spatiotemporal distribution patterns from historical crime and related data, thereby effectively forecasting the time and location of future criminal activities [3]. As a crucial tool for preventing and combating crime, the accuracy of such predictions is essential for enhancing the efficiency and effectiveness of urban security management [4].

The urban environment influences the spatiotemporal distribution of crime, a mechanism primarily explained by four major theoretical frameworks in environmental criminology: Routine Activity Theory emphasizes that the three conditions for crime occurrence (suitable target, motivated offender, and absence of guardian) are affected by the functional layout of the environment [5]; Broken Windows Theory focuses on the signal of neglect and lack of control conveyed by physical environmental decay [6]; Social Disorganization Theory highlights how neighborhood spatial structures shape informal social control [7]; and Crime Prevention through Environmental Design (CPTED) proposes proactive intervention strategies through dimensions such as spatial accessibility and natural surveillance [8].

The evolution of crime prediction models reflects the ongoing pursuit of accurately capturing these complex spatiotemporal dynamics. Early traditional statistical models primarily focused on the overall trends and distribution characteristics of crime phenomena. For instance, Polvi et al., based on the near-repeat victimization effect, found that the risk of burglary increased significantly in the short term but decayed rapidly over time [9]; Mohler et al. proposed self-exciting point process models, revealing the contagion and diffusion mechanisms of crime risk in space [10]; Kalinic et al. compared kernel density estimation and hotspot analysis methods, indicating that their combination could effectively improve crime hotspot identification [11]. Subsequently, machine learning methods demonstrated stronger capabilities in handling heterogeneous crime data, capturing non-linear spatiotemporal relationships, and enabling dynamic predictions. For example, Law et al. used Bayesian models to identify crime trend variations across different areas [12]; Yi et al. proposed a hybrid model integrating LSTM with autoencoders for crime prediction, maintaining accuracy while reducing computational complexity [13]. Currently, deep learning has become the primary technical choice for crime spatiotemporal prediction, where Graph Neural Networks (GNNs) are widely used to capture spatial dependencies, while Recurrent Neural Networks (RNNs) and their variants (e.g., LSTM) or Transformer architectures are employed for temporal modeling, continuously pushing the boundaries of predictive accuracy. Representative works include Huang et al.'s DeepCrime and MiST models, which capture complex dependencies in crime sequences through multi-modal encoding and attention mechanisms [14,15]; Sun et al.'s CrimeForecaster, which combines Graph Convolutional Networks (GCN) with gated recurrent units to effectively model spatiotemporal dynamics [16]; and addressing the limitations of pre-defined spatial relationships, Sun et al.'s AGL-STAN model, which introduces adaptive graph learning and employs Transformer architecture for more expressive power and parallel computation capability [17].

A critical enabler for these data-driven models is the incorporation of urban environmental data. In existing research, Point of Interest (POI) data, owing to its open accessibility and ability to quantify urban functional attractiveness, has become a crucial type of urban environmental data in crime spatiotemporal prediction models for characterizing spatial dependencies. By categorizing urban functions, POIs can reflect the three conditions emphasized in Routine Activity Theory, thereby influencing the spatiotemporal distribution of crime. For instance, commercial facilities (e.g., shopping malls, banks) are positively correlated with theft [18], while transportation hubs (e.g., bus stops) increase the risk of pickpocketing due to complex human flows [19]. This provides a theoretical foundation for utilizing POI data in crime spatiotemporal prediction. Early studies employed linear regression [20] and spatiotemporal association mining [21] to validate the enhancing effect of POI data on community crime rate prediction. With the advancement of spatiotemporal graph neural networks, the application of POIs in crime spatial modeling has become more widespread. For example, Wang et al. [22] proposed a homogeneity-aware graph neural network that innovatively introduced an adaptive regional graph learning mechanism,

using POI and administrative boundary data to generate a homogeneity-aware crime propagation topology. However, although POIs provide urban functional features for crime prediction, they lack spatial morphological information, making it difficult to represent the continuity and mixed-use nature of spatial functions [23]. As a result, POI data cannot fully capture the physical environmental decay characteristics central to Broken Windows Theory, the neighborhood spatial structural attributes highlighted by Social Disorganization Theory, or the spatial form control mechanisms relied upon in CPTED.

Buildings are a fundamental component of urban spaces. As open-source urban environmental data, building footprints provide detailed information on urban structure and spatial layout. Identifying and analyzing their morphology is of great significance for modeling and characterizing urban geography, semantically classifying social functions, and predicting economic activities [24]. Extracted from remote sensing imagery, building footprints capture the geometric forms and spatial distribution features of structures, offering a high-resolution continuous representation for analyzing the spatiotemporal distribution of crime. A separate line of research in urban analytics and criminology has leveraged building footprints to uncover correlations with crime, providing empirical support for the environmental criminology theories. For example, Pation et al. [25] extracted structural and textural features from remote sensing imagery in a study conducted in Medellín, Colombia, and found that areas with high homicide rates often exhibit higher local variability and lower overall homogeneity. This indicates more crowded and disordered urban layouts, which are associated with weaker social cohesion, consistent with Social Disorganization Theory. Meanwhile, Broken Windows Theory suggests that chaotic building layouts may signal physical disorder in an area, thereby attracting criminal activity. For example, Silva and Li [26] developed multiple metrics based on building footprints in Bissau, Guinea-Bissau, and conducted regression analyses between these metrics and crime rates. They found that a higher percentage of open space is correlated with lower crime rates, whereas older neighborhood age is associated with higher crime rates. According to CPTED, open spaces enhance visibility and reduce fear of crime, thereby improving safety. Broken Windows Theory, in turn, explains that older neighborhoods with more dilapidated and damaged buildings tend to experience higher crime rates. Ioannidis et al. [27] investigated the correlation between building density and crimes such as burglary and street theft using remote sensing imagery from Stockholm, Sweden. The results indicated that both types of crime are associated with building density: burglary occurs more frequently in areas with high building density, while the relationship between street theft and building density varies significantly across different planning zones and is easily influenced by other factors. Routine Activity Theory suggests that areas with higher building density offer more criminal opportunities, such as a greater number of valuable targets.

These studies demonstrate that building footprints can represent the continuous and mixed-use nature of urban spatial functions and provide theoretical support for explaining crime distribution. However, they have primarily focused on macro-level correlation analysis or static regression modeling, and have not been sufficiently explored as deep features for enhancing end-to-end spatiotemporal prediction models. This creates a gap between the proven explanatory power of building morphology and its underutilization in dynamic forecasting frameworks.

To address this gap, this paper applies building footprint data to characterize urban regions to construct spatial dependencies in crime spatiotemporal prediction models. Specially, we first fuse building footprints and POI data using Region Dual Contrastive Learning (RegionDCL) [28] to represent urban areas. Then, the learned region representations are incorporated as a region graph into crime prediction models based on different technical approaches. Finally, we use these prediction models to predict the occurrence of

four types of crimes (Burglary, Robbery, Felony Assault, and Grand Larceny) to validate the effectiveness of building footprints in crime spatiotemporal prediction. The experimental results on New York City crime data indicate that: (1) the region representations significantly improve deep learning model performance, with the most improved LSTM achieving average increases of 5.66% in Macro-F1 and 18.57% in Micro-F1, particularly benefiting baseline models with lower accuracy; (2) the region representations yields more significant improvements for low-frequency crime categories and mitigates temporal memory decay in long-term predictions.

The rest of this paper is organized as follows. Section 2 elaborates on the proposed methodology, including the construction of the region graph based on building footprints, the formation of the crime tensor, the fusion module, and the spatiotemporal prediction models. Section 3 introduces the study area and describes the datasets used in this research. Section 4 presents the experimental results and discussions, covering model comparisons, predictions for different crime types, and ablation studies. Finally, Section 5 concludes the paper and outlines potential directions for future work.

2. Methodology

This study adopts a progressive research framework of “data-representation-prediction”. For building footprints and POI data, the RegionDCL algorithm is employed to learn region representations, which serve as the region graph. Crime data are transformed into a crime tensor for representation. The region graph and the crime tensor are then fused and input into a set of crime spatiotemporal prediction models to perform forecasting. Figure 1 illustrates the specific steps implementing the above workflow.

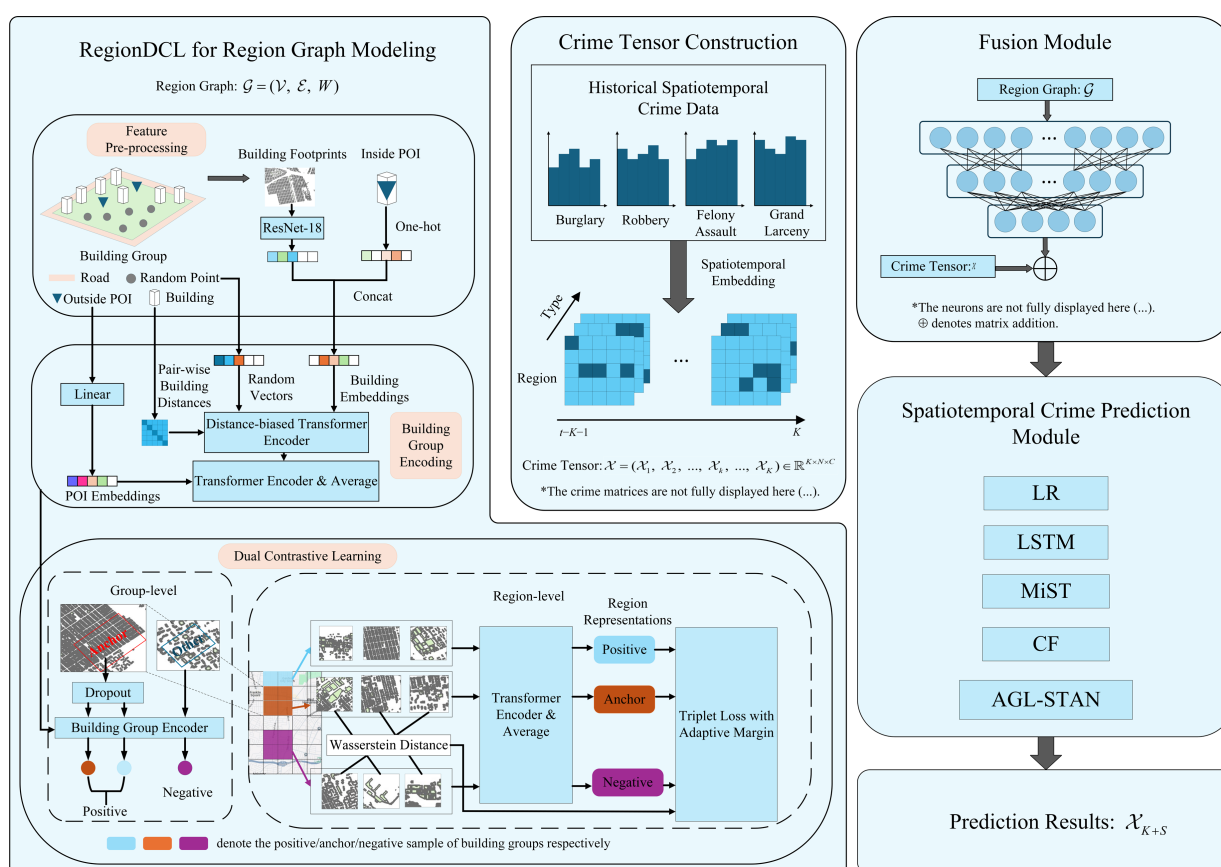


Figure 1. Research methodology framework.

2.1. RegionDCL for Region Graph Modeling

The study area is divided into N mutually non-overlapping geographic units, denoted as $R = \{r_1, r_2, \dots, r_i, \dots, r_N\}$, where r_i denotes the i -th unit. To model spatial dependencies among geographic units, a region graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ is defined over the geographic unit set R , where $\mathcal{V}$ is the node set, with each geographic unit corresponding to one node; $\mathcal{E}$ denotes the set of edges connecting the geographic units, which can be predefined or learned from external data sources; and $\mathbf{W} \in \mathbb{R}^{N \times N}$ is the graph weight matrix. We employ the RegionDCL algorithm to learn urban region representation based on building footprints [28], which constructs a region graph through three key steps: feature preprocessing, building group encoding, and dual contrastive learning. Critically, we frame this process not merely as a feature extraction technique, but as a method for implicitly operationalizing the environmental factors central to criminological theories.

(1) In the feature preprocessing stage, building footprints are partitioned into non-overlapping clusters based on high-resolution OpenStreetMap (OSM) road network data to form building group units. A pre-trained ResNet-18 model [29] extracts 64-dimensional deep visual features from rasterized building footprints. These are concatenated with a 64-dimensional one-hot encoded vector representing POI categories within the building (if any), resulting in a building embedding of dimension $d_{\text{building}} = 128$. The geometric features of building footprints provide direct signals about the level of physical order/disorder and the potential for natural surveillance, thereby operationalizing constructs from Broken Windows Theory and CPTED. For non-building urban areas (e.g., green spaces, plazas), Poisson Disk Sampling [30] with a radius $r = 100$ m is employed to generate uniformly distributed random point sets to represent the semantics of empty areas.

(2) During the building group encoding stage, building embeddings, random point features, and external POI features are fused in a hierarchical manner. First, a pair-wise distance matrix is computed between each pair of buildings and random points using the Haversine formula. This matrix, along with the building embeddings and random point feature vectors, is then fed into a distance-biased Transformer encoder [31] (utilizing 8 attention heads) to capture global interactions among all objects within the building group. The distance matrix is incorporated as a bias term in the self-attention mechanism (Equations (5) and (6) in [28]), ensuring that the model accounts for the spatial configuration of the environment. Meanwhile, external POI features are processed via a linear layer. The fusion of diverse POI categories with building data serves as an indicator of land-use mix, operationalizing concepts from Social Disorganization Theory. Conversely, the intensity and type of POIs directly signal the presence of human activity and “suitable targets”, a key element of Routine Activity Theory. The outputs from both the distance-biased encoder and the linear layer are subsequently combined and passed into a standard Transformer encoder equipped with average pooling, ultimately generating a unified building group embedding of dimension $d_{\text{group}} = 64$.

(3) The dual contrastive learning mechanism optimizes feature representations at two distinct scales: the building group level and the region level. At the building group level, the algorithm randomly selects a building group as an anchor sample (denoted as P_i). A positive sample (denoted as P_i^+) is generated by randomly dropping out 20% of the buildings within the anchor group to simulate structural perturbation. The remaining building groups within the same batch are treated as negative samples. The learning objective is to discriminate the similarity differences between the anchor and the positive/negative sam-

ples, training the encoder to capture the essential functional characteristics of the building groups. This is formalized using the InfoNCE loss:

$$\mathcal{L}_{group} = -\log \frac{\exp(\text{sim}(P_i, P_i^+) / \tau)}{\sum_{i=0}^n \exp(\text{sim}(P_i, P_j) / \tau)} \quad (1)$$

where $\text{sim}(\cdot, \cdot)$ is the KL-divergence similarity measure, and the temperature parameter $\tau = 0.05$.

At the region level, sliding windows are used as training units. A standard Transformer encoder with average pooling is applied to derive the region representation within each window, resulting in a final region representation of dimension $d_{region} = 64$. The similarity between regions is measured not only by geographic proximity but also by the distributional similarity of their building groups, quantified using the Wasserstein distance [32] between their embedded vectors. The contrastive learning of region representations employs a triplet loss [33]:

$$\mathcal{L}_{region} = \max(\|z_a - z_p\| - \|z_a - z_n\| + \lambda \cdot W, 0) \quad (2)$$

where the margin threshold is dynamically adjusted based on the aforementioned similarity metric. Here, $\|\cdot\|$ represents the L1 distance, z_a , z_p and z_n denote the anchor, positive and negative region representations, respectively. W is the Wasserstein distance, and the scaler λ controls the adaptive margin. This adaptive strategy operationalizes the theoretical premise that areas with similar urban morphology and function should share similar crime risk profiles even if they are geographically distant. It promotes the separation of region representations that are spatially adjacent yet exhibit divergent architectural patterns, while bringing together those that are geographically distant but share similar building configurations. Ultimately, the dual contrastive learning mechanism produces region representations that simultaneously encode both micro-level functional attributes and macro-level spatial regularities.

2.2. Crime Tensor Construction and Prediction

Crime data consist of the occurrence time, location, and crime type. Given consecutive and non-overlapping time slots $T = (t_1, t_2, \dots, t_k, \dots, t_K)$, where K is the length of the time series, a crime matrix $\mathcal{X}_k = (x_{1,1}^k, \dots, x_{i,c}^k, \dots, x_{N,C}^k) \in \mathbb{R}^{N \times C}$ is defined for each time slot t_k . Here, $x_{i,c}^k = 1$ indicates the occurrence of crime type c in geographic unit r_i during t_k , and 0 otherwise. The complete crime tensor is denoted as $\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_k, \dots, \mathcal{X}_K) \in \mathbb{R}^{K \times N \times C}$. Given the sparsity of crime data, the number of incidents for specific crime type in the vast majority of fine-grained spatiotemporal units is either 1 or 0. Therefore, crime spatiotemporal prediction can be regarded as a classification problem [22]. Figure 2 illustrates the detailed process of the crime tensor construction.

Given crime data from time t_1 to t_K across the study area R , a crime tensor $\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_k, \dots, \mathcal{X}_K)$ is constructed. Utilizing the region graph obtained in the previous step and a predictive model, i.e., the mapping function f , we aim to predict crimes over the entire study area for the next S time slots, denoted as $\mathcal{X}_{K+S}$. The entire process can be formally expressed as Equation (3):

$$\{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_k, \dots, \mathcal{X}_K; \mathcal{G}\} \xrightarrow{f} \mathcal{X}_{K+S} \quad (3)$$

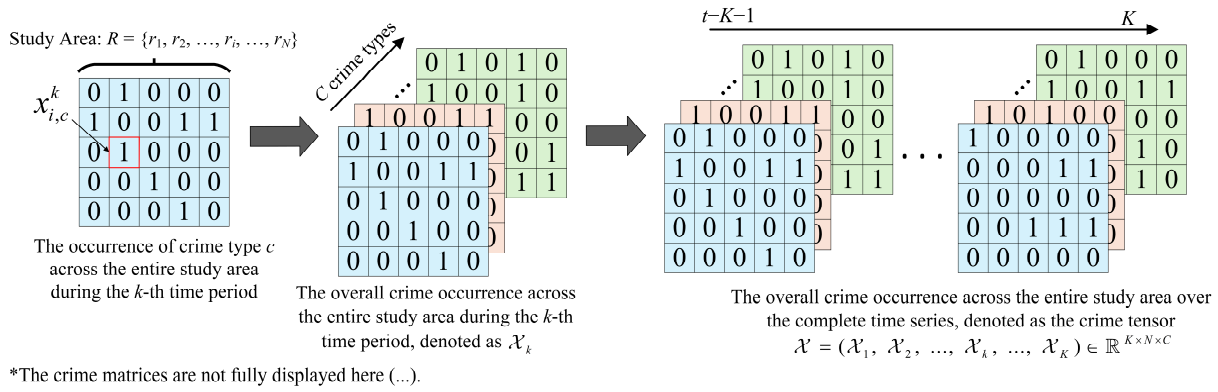


Figure 2. The specific construction process of crime tensor.

2.3. Fusion Module

To enable the aforementioned prediction process, the region graph must be fused with the crime tensor. Considering the high-dimensional density characteristics of the region representations, an MLP-based feature fusion module is introduced to align them with the sparser crime tensor.

Let the region representations be $\mathcal{G} \in \mathbb{R}^{N \times d_e}$. Equation (4) illustrates how the global perception capability of fully connected structures and the mapping mechanism of nonlinear activation functions (ReLU) progressively decouple complex spatial correlation patterns within high-dimensional features through layered processing. Here, L denotes the number of hidden layers, $W^{(l)}$ and $b^{(l)}$ represent the learnable weights and bias of the l -th hidden layer, respectively, and $H^{(l)}$ denotes the hidden state of the l -th layer with the initial hidden state given as $H^{(0)} = \mathcal{G}$. On this basis, the module achieves implicit spatial alignment between the region representations and the low-dimensional sparse crime tensor via a linear dimensionality reduction layer, as shown in Equation (5). Here, W_{align} and b_{align} denote the learnable weights and bias of the linear reduction layer respectively. The resulting output $\mathcal{G}_{output} \in \mathbb{R}^{N \times C}$ matches the sparsity structure of the crime tensor.

$$H^{(l)} = \text{ReLU}(W^{(l)}H^{(l-1)} + b^{(l)}), l = 1, \dots, L \quad (4)$$

$$\mathcal{G}_{output} = W_{align}H^{(L)} + b_{align} \quad (5)$$

For cross-modal fusion, $\mathcal{G}_{output}$ is first expanded along the temporal dimension into $\mathcal{G}_{output} \in \mathbb{R}^{K \times N \times C}$, which is then combined with the crime tensor via element-wise addition. The resulting tensor $\mathcal{Z}$ serves as the input to the crime spatiotemporal prediction model, as formalized in Equation (6).

$$\mathcal{Z} = \mathcal{G}_{output} \oplus \mathcal{X} \quad (6)$$

Compared to a CNN-based fusion module [34], the fully connected structure of the MLP differs from the fixed-size kernels used in convolutional operations. Its global feature interaction mechanism preserves cross-region dependency information in high-dimensional representations entirely, avoiding issues such as feature fragmentation and loss caused by the sliding of convolutional kernels.

2.4. Crime Spatiotemporal Prediction Module

To validate the universal enhancement effect of urban region representation based on building footprints on crime spatiotemporal prediction models, we employ five representative algorithmic categories for implementation: traditional machine learning, recurrent neural networks (RNNs), encoder–decoder architectures, graph neural networks (GNNs), and Transformer architectures with adaptive graph learning modules. This multi-paradigm

selection facilitates systematic verification of the compatibility and enhancement effects of the region representations across diverse technical pathways, ensuring comprehensive generalizability of methodological validation.

(1) Logistic Regression (LR) [35]: As a traditional machine learning algorithm, LR is implemented following the configuration by Sun et al. [16]. By incorporating the entire historical crime records, this approach enables the model to capture spatial patterns rather than relying solely on temporal features, which would otherwise lead to significant degradation in predictive capability.

(2) Long Short-Term Memory (LSTM) [36]: LSTM is a representative model of RNNs, extending traditional RNNs by enhancing the ability to capture long-term dependencies in time series data. It has been widely adopted in the domain of crime spatiotemporal prediction.

(3) Multi-View and Multi-Modal Spatial-Temporal learning framework (MiST) [15]: MiST employs an encoder–decoder architecture, with the encoder comprising LSTM layers and the decoder comprising RNN layers. An attention mechanism is positioned between the encoder and decoder to incorporate crime category relationships and geographical adjacency relationships.

(4) CrimeForecaster (CF) [16]: CF leverages a graph convolutional network (GCN) to extract spatial dependencies and utilizes a recurrent neural network with diffusion convolutional gated recurrent units (DCGRU) to capture temporal dynamics.

(5) Adaptive Graph Learning based Spatial-Temporal Attention Network (AGL-STAN) [17]: AGL-STAN learns spatial relationships through an adaptive graph learning module, rather than relying on predefined graphs between research units. In addition, its temporal-aware self-attention module, built upon the Transformer architecture, captures both local and global temporal dependencies more effectively than RNN-based architectures, while also enabling parallel computation.

2.5. Evaluation Metrics

The task focuses on predicting the occurrence of four crime types (Burglary, Robbery, Felony Assault, Grand Larceny) under specific spatiotemporal conditions, which is inherently framed as a multi-class classification problem. To comprehensively evaluate model performance, Macro-F1 [37] and Micro-F1 [38] are adopted as multi-class metrics to assess the overall prediction performance across all crime categories. The metrics are computed as follows:

$$\text{Macro-F1} = \frac{1}{C} \sum_{c=1}^C \frac{2TP_c}{2TP_c + FP_c + FN_c} \quad (7)$$

$$\text{Micro-F1} = \frac{2 \sum_{c=1}^C TP_c}{2 \sum_{c=1}^C TP_c + \sum_{c=1}^C FP_c + \sum_{c=1}^C FN_c} \quad (8)$$

These metrics comprehensively evaluate prediction accuracy through macro-averaging (emphasizing class balance) and micro-averaging (weighted by sample distribution). For evaluating crime type, since each type constitutes a binary classification problem, F1-score is used to measure discriminative performance.

3. Study Area and Date Description

3.1. Study Area

New York City is selected as the study area. Situated at 40°45′19.80″ N, 73°58′26.04″ W along the eastern U.S. coast where the Hudson River meets the Atlantic Ocean, New York City covers a total area of 778.2 km² [39]. As a global metropolis, it serves as an international hub for finance, commerce, and media. The city exhibits high population density and demographic complexity, with 8.8 million residents in 2020 comprising 35.9%

White, 22.7% Black/African American, 14.6% Asian, 10.5% Mixed Race, 0.7% Native American, 0.1% Pacific Islander, and 28.4% Hispanic/Latino [40]. These characteristics contribute to elevated crime rates and complex urban functional zoning, making it an ideal study sample. Figure 3 shows the community zoning map of New York City, with building footprints and crime hotspots distributions shown in illustrative areas.

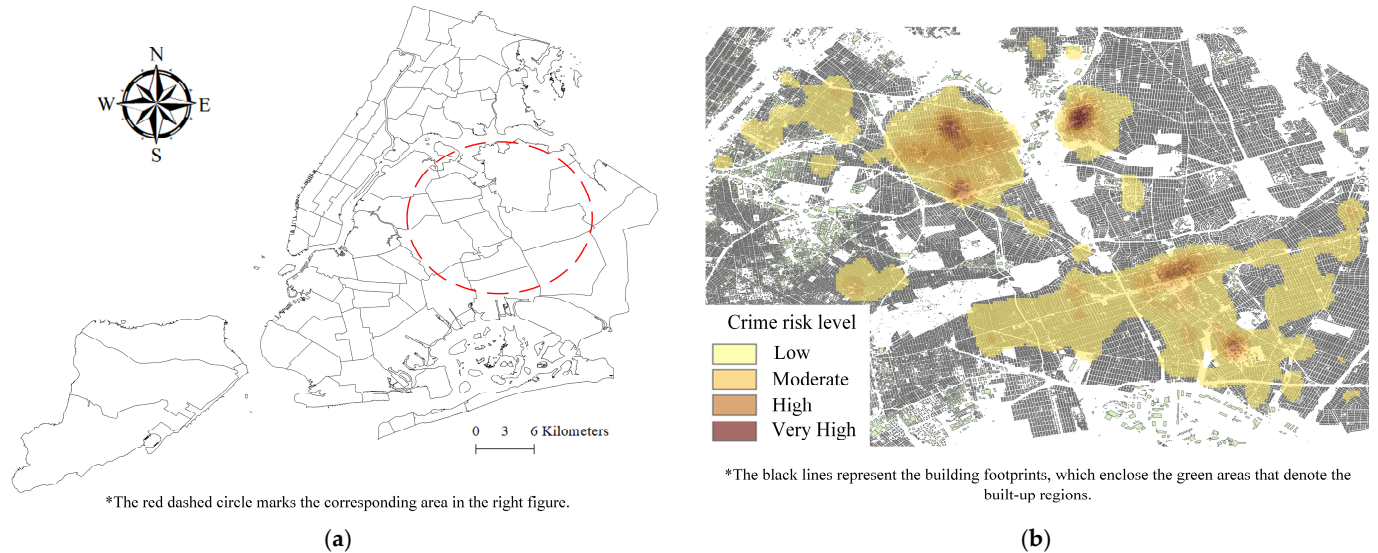


Figure 3. Study area and part of its crime hotspots with building footprints. (a) Study area. (b) Crime hotspots with building footprints.

3.2. Data Description

3.2.1. Urban Environmental Data

Open-source urban environmental data collected from OpenStreetMap (<https://www.openstreetmap.org/> (accessed on 7 September 2025)) include building footprints and POIs. To ensure temporal alignment with the 2019 crime data, the building footprints and POI data are retrieved as their 2019-version historical snapshots. Community boundary data are obtained from the NYC Open Data Portal (<https://opendata.cityofnewyork.us/> (accessed on 7 September 2025)). Statistical details are presented in Table 1.

Table 1. Statistics of urban environmental data.

Urban Environment	Building Footprints	POIs	Community Boundaries
Counts	1,081,256	41,963	71

Building footprints are polygon data delineating architectural outlines derived from remote sensing imagery. Among these, 141,595 footprints contain functional attributes directly inherited from OpenStreetMap's tagging schema, which categorizes buildings into 93 function types (e.g., residential, commercial, industrial, school, hospital). These categories are not manually defined but come from the original OSM data. It is worth noting that these fine-grained categories are not directly used as independent features; instead, they serve to indicate the semantic richness of the original dataset rather than to construct explicit model inputs. POI data, as key indicators of regional functional activities, provide fine-grained spatial semantics for urban analysis through their distribution patterns. A total of 41,963 POIs are integrated, covering essential categories such as daily services, recreation, healthcare, and education, effectively complementing the functional attributes of building footprints. Community boundaries divide the entire New York City into

71 non-overlapping communities, which also form the fundamental geographic units of this study.

3.2.2. Crime Data

Crime data for previous years are available through the NYC Open Data Portal. We collected a full year of crime data from 1 January to 31 December 2019, including information such as crime time, crime location (specific latitude/longitude coordinates), crime type, and offender characteristics. The original dataset contains 459,296 crime records. For this research, four representative categories are selected: Burglary, Robbery, Felony Assault, and Grand Larceny. These four crime types were selected primarily for their established theoretical relevance to environmental criminology [5,8,14,15], making them ideal for validating the impact of urban form representations. Additionally, their varying incidence rates provide a spectrum of data density conditions to thoroughly evaluate the method's robustness. Statistical details are presented in Table 2.

Table 2. Crime data statistics.

Crime Types	Burglary	Robbery	Felony Assault	Grand Larceny
Counts	10,886	13,434	20,860	43,116

4. Experiment

4.1. Experimental Setup

The learned region representations are tensors with a shape of 71×64 , meaning each of the 71 communities is characterized by a 64-dimensional vector. Following the dataset partitioning method used in prior studies [16,17], the crime data are divided chronologically into 6.5 months for training, 0.5 months for validation, and 5 months for testing. The training window length K is set to 10 days, and the prediction period S to 1 day. The geographic units of analysis are the 71 communities in New York City. For the fusion module, a three-layer MLP is used to map the region representations to the tensor space of the crime tensor, after which the mapped representations and the original crime tensor are combined via element-wise addition. The code is released in <https://github.com/Erdengxin/BF2Crime> (accessed on 7 September 2025).

4.2. Comparative Analysis of Prediction Performance Across Models

Different models are tested with and without the region representations for multi-class prediction of the four crime types mentioned above. Daily predictions are averaged monthly, and the results are shown in Table 3.

The “BF” column shows the baseline performance, while the “BF” column presents the absolute performance difference after integrating the region representations based on building footprints. A positive value indicates a performance gain. The rows labeled “Avg Macro↑” and “Avg Micro↑” display the average improvement ratios of Macro-F1 and Micro-F1. To statistically validate these improvements, we conducted one-tailed paired t -tests comparing the daily prediction performance with and without the region representations across the entire test period. The t -statistics and their significance levels are reported at the bottom of the table. Key observations include:

(1) All models except LR exhibit statistically significant improved prediction accuracy after integrating the region representations. As a machine learning model with only dozens of parameters and low complexity, LR is unable to utilize the high-dimensional region representations. Other models, whether based on RNN frameworks, Transformer frameworks, or encoder–decoder architectures, demonstrate consistent performance improvements following the integration of the region representations. These results indicate that deep learning

frameworks for crime spatiotemporal prediction can effectively leverage the rich urban spatial semantics encoded in the region representations to enhance prediction performance.

(2) Models lacking explicit spatial dependency modeling experience significant performance gains when incorporating the region representations. For example, LSTM with the region representations achieves substantial improvements in both Macro-F1 and Micro-F1, outperforming the original RNN-based models MiST and CF. This finding confirms the effectiveness of the region representations in strengthening spatial dependency modeling.

(3) The improved accuracy of AGL-STAN with the region representations suggests that the predefined graph derived from the region representations complements the adaptive graph learning module. While the adaptive graph learning module dynamically captures spatial relationships through data-driven approaches, it may overlook complex or implicit spatial correlations. The region representations provide complementary urban spatial semantics, offering a comprehensive and accurate initial spatial framework to refine adaptive graph learning.

(4) The region representations yield more substantial improvements in long-term prediction tasks. For instance, LSTM with the region representations achieves increases of 2.82% in Macro-F1 and 14.60% in Micro-F1 in August, while 9.48% and 24.35% in December, respectively. Similar trends observed across models indicate that temporal dependencies degrade in long-term forecasting [41], while the region representations mitigate this degradation by reinforcing the models' understanding of urban spatial structures.

To summarize, the region representations enhance prediction accuracy across all deep learning frameworks, particularly for models with lower baseline performance and long-term forecasting tasks, where the improvements are more pronounced.

Table 3. Multi-class crime prediction performance with/without the region representations.

Month	F1	LR		LSTM		MiST		CF		AGL-STAN	
		BF	BF	BF	BF	BF	BF	BF	BF	BF	BF
8	Macro	0.6430	+0	0.6764	+0.0191	0.6845	+0.0011	0.6862	+0.0318	0.7420	−0.0009
	Micro	0.6316	+0	0.5335	+0.0779	0.5679	+0.0383	0.5798	+0.0544	0.6947	+0.0126
9	Macro	0.6400	+0	0.6621	+0.0407	0.6805	−0.0048	0.6787	+0.0006	0.7375	−0.0014
	Micro	0.6318	+0	0.5331	+0.1000	0.5924	+0.0133	0.5911	+0.0132	0.6852	+0.0048
10	Macro	0.6567	+0	0.6621	+0.0424	0.6639	+0.0157	0.6897	+0.0084	0.7282	+0.0077
	Micro	0.6490	+0	0.5333	+0.0816	0.5543	+0.0394	0.5970	+0.0173	0.6794	+0.0242
11	Macro	0.5999	+0	0.6867	+0.0238	0.6960	+0.0002	0.6837	+0.0225	0.7274	+0.0073
	Micro	0.5829	+0	0.5166	+0.1026	0.5582	+0.0401	0.5570	+0.0395	0.6796	+0.0229
12	Macro	0.6097	+0	0.6476	+0.0614	0.6512	+0.0552	0.6784	+0.0298	0.7064	+0.0087
	Micro	0.5967	+0	0.5039	+0.1227	0.5134	+0.1161	0.5592	+0.0636	0.6578	+0.0245
Avg Macro ↑			0%		5.66%		2.07%		2.72%		0.60%
Avg Micro ↑			0%		18.57%		9.05%		6.60%		2.90%
t-statistic (Macro)			—		8.45 ***		2.85 **		5.12 ***		2.15 *
t-statistic (Micro)			—		16.23 ***		7.12 ***		10.35 ***		5.45 ***

Note: significance levels for the one-tailed paired *t*-test on daily performance differences are denoted as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. “↑” denotes increase and “—” indicates omission.

4.3. Comparative Analysis of Prediction Performance Across Crime Types

Following a comprehensive evaluation of multi-class crime prediction performance with and without the region representations, the impact of the region representations on the prediction performance for specific crime types is further examined. Using different models (excluding LR due to its lack of improvement), binary classification predictions are conducted for Burglary, Robbery, Felony Assault, and Grand Larceny under two conditions: with and without the region representations. This granular analysis aims to uncover differences in spatiotemporal prediction across crime types and elucidate how the region representations enhance the models' ability to capture these variations. Experimental results are shown in Figure 4, where the y-axis represents F1-scores and the x-axis denotes months. Key findings are as follows:

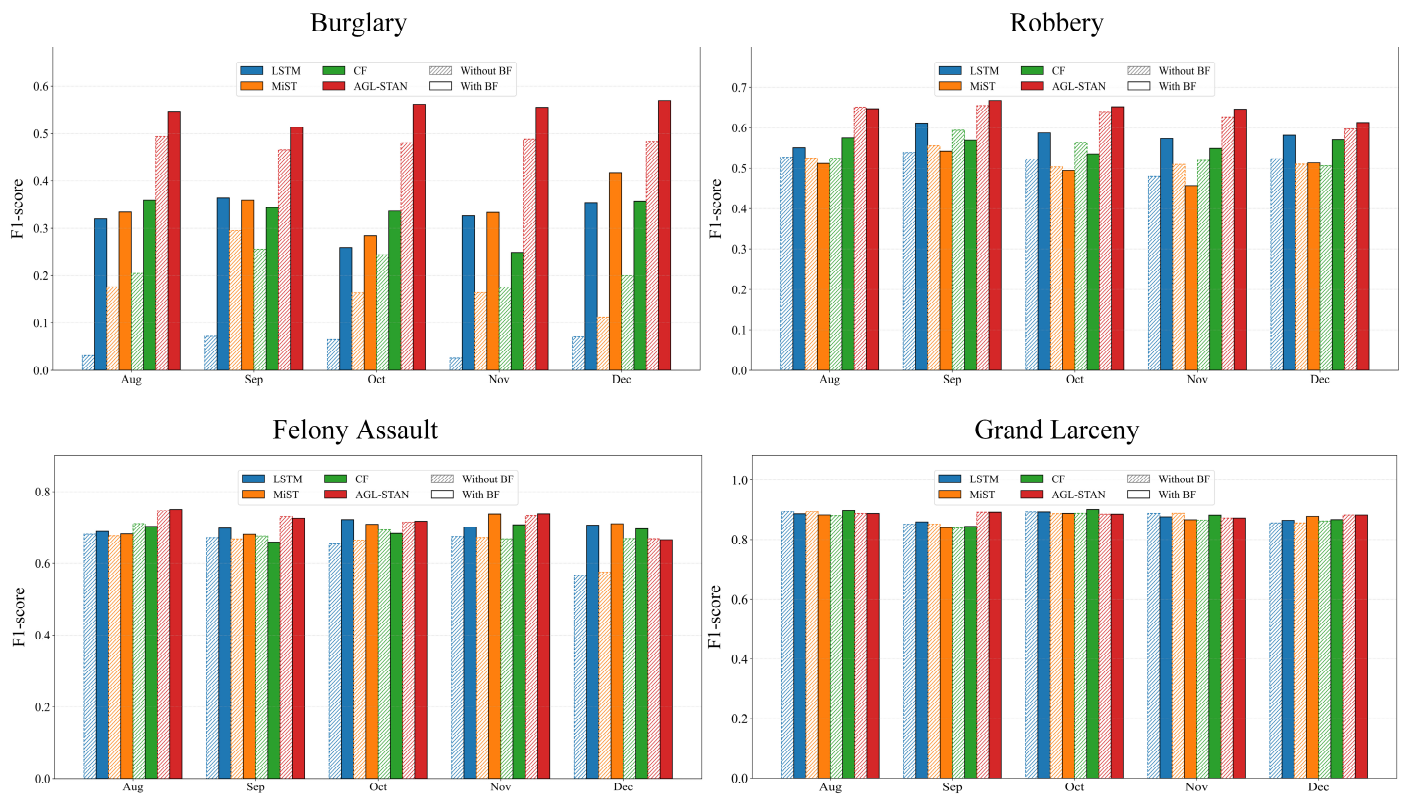


Figure 4. Prediction performance of models with/without the region representations for each crime type.

Burglary: As shown in Figure 4, LSTM, MiST, and CF exhibit limited predictive capability for Burglary without the region representations. It suggests that while these models can capture temporal dependencies, they lack sufficient spatial feature modeling. After integrating the region representations, LSTM achieves an average monthly improvement of 639.48%, MiST 112.03%, and CF 53.68%. Notably, AGL-STAN, which already demonstrates strong baseline performance without the region representations, still improves by 13.77% on average across months. These results indicate that even models explicitly modeling spatial dependencies can benefit from the additional spatial semantics provided by the region representations.

Robbery: After incorporating the region representations, LSTM shows consistent improvements in Robbery prediction. From August to December, F1-scores increase by 4.42%, 13.55%, 12.40%, 19.82%, and 11.18%, respectively, with a notable 19.82% improvement in November, validating the effectiveness of the region representations for long-term forecasting. The sustained 11.18% gain in December may be attributed to the region representations' ability to capture spatial heterogeneity in street crimes, enhancing cross-temporal prediction robustness. AGL-STAN also records long-term improvements. However, MiST and CF exhibit unstable results: MiST's attention mechanisms for spatial dependency modeling do not benefit from the region representations, while CF improves in August, November, and December but declines in September and October.

Felony Assault: All models exhibit strong baseline performance for Felony Assault but are affected by significant temporal performance decay. The region representations not only mitigate this decay but also improve accuracy in most months. For example, LSTM achieves a 24.80% improvement in December, substantially outperforming the baseline model's long-term decay trend.

Grand Larceny: As shown in Figure 4, all baseline models achieve F1-scores above 0.8 across all months for Grand Larceny, with minimal impact observed from the region representations. This crime type has the highest case count (43,116 cases), far exceeding the

others, reducing the impact of spatiotemporal data sparsity [3]. Models effectively capture its spatiotemporal patterns even in the absence of the region representations.

To further investigate the relationship between data volume, model performance, and the utility of our region representations, we conducted a down-sampling analysis on Grand Larceny. The results, illustrated in Figure 5, compellingly demonstrate that the benefit of the region representations is profoundly modulated by data sparsity. When the data is severely limited (e.g., at 10% and 20% sampling rates), the incorporation of BF provides a dramatic performance boost across all models. For instance, the F1-score of LSTM more than doubles at the 10% level with BF, transforming it from a weak predictor to a reasonably accurate one. This indicates that the rich spatial semantics from building footprints serve as a critical source of information, effectively compensating for the lack of training examples. As the volume of training data increases (to 40% and 80%), the relative advantage of BF gradually diminishes, though it continues to deliver consistent improvements. Finally, when the entire dataset (100%) is utilized, the models become sufficiently saturated with crime-specific data, and the marginal gain from the spatial prior provided by BF becomes negligible, as originally observed.

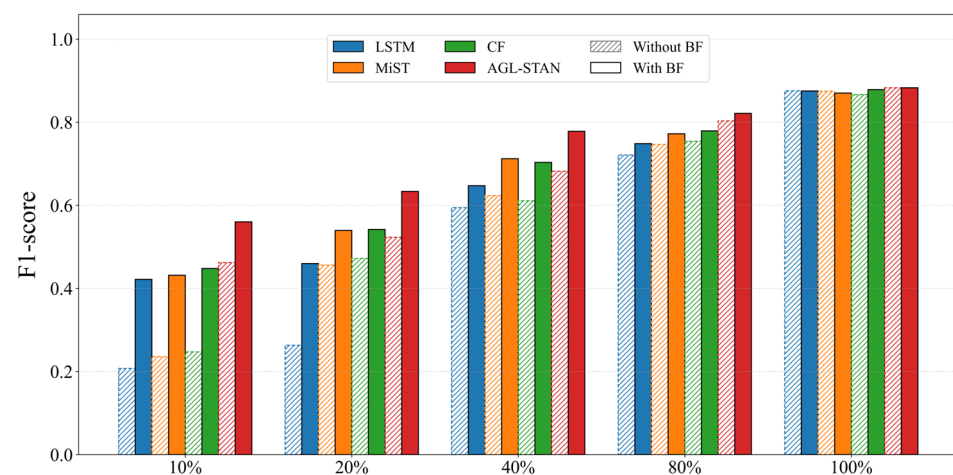


Figure 5. Impact of the region representations on Grand Larceny prediction under varying data availability.

Overall, the accuracy improvement provided by the region representations exhibits an inverse correlation with crime frequency, significantly alleviating spatiotemporal sparsity challenges for low-frequency crimes. In long-term forecasting tasks, the region representations effectively counteract performance degradation by capturing the persistent influence of spatial factors, thereby maintaining temporal stability in predictive efficacy.

4.4. Ablation Study

4.4.1. Ablation Analysis of Components in the Region Representations

To analyze the impact of different components on improving crime spatiotemporal prediction performance through the region representations, feature ablation experiments were conducted across various models. Each model was evaluated under the following four experimental settings for multi-class crime prediction:

- (1) Base model: without incorporating any region representations.
- (2) Base model + POI: the base model augmented with region representations constructed solely from POI data.
- (3) Base model + building footprints: the base model enhanced with region representations derived only from building footprints.

(4) Base model + POI + building footprints: the base model integrated with the complete region representations combining both POI and building footprints.

The experimental results, shown in Figure 6, demonstrate that the incorporation of the region representations generally improves the predictive performance of the base models, with the complete representation yielding the most significant gains, confirming the effectiveness of each component. Furthermore, the region representations based on building footprints contribute more notably to performance improvement than those based solely on POI, underscoring the dominant role of building footprints in constituting the region representations.

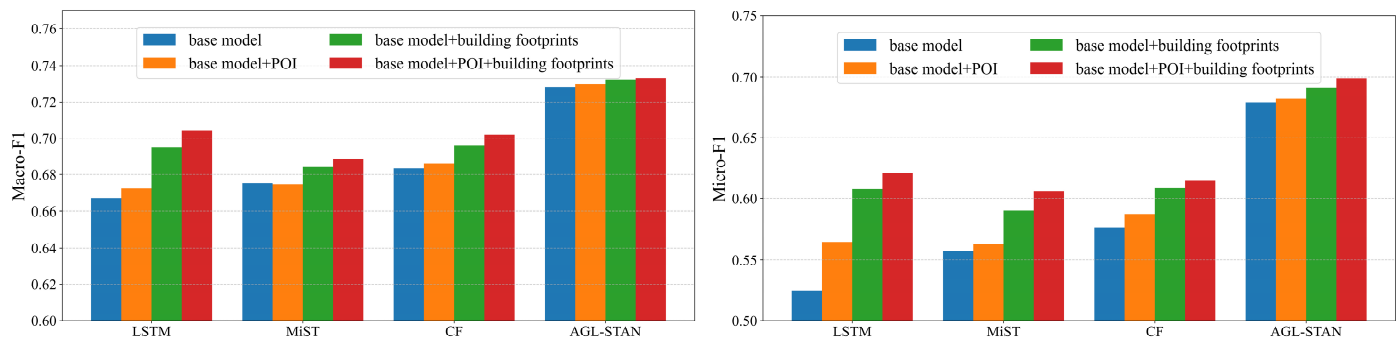


Figure 6. Ablation analysis of region representation components: prediction performance comparison.

4.4.2. Comparison of Two Fusion Modules

To effectively integrate the region representations with the crime tensor for enhancing prediction performance, two fusion modules are explored: a CNN-based module and an MLP-based module. The CNN-based module employs convolutional layers to extract local spatial features from the region representations, generating high-dimensional embeddings that are then fused with the crime tensor. In contrast, the MLP-based module applies non-linear transformations via multi-layer perceptrons to map the region representations into more expressive feature vectors before fusion. Table 4 compares the prediction performance of models using the two fusion strategies, with the last two rows indicating the average improvement of MLP-based fusion over CNN-based fusion.

Table 4. Prediction performance of models with CNN-based vs. MLP-based fusion modules.

Month	F1	LSTM		MiST		CF		AGL-STAN	
		BF_CNN	BF_MLP	BF_CNN	BF_MLP	BF_CNN	BF_MLP	BF_CNN	BF_MLP
8	Macro	0.6952	+0.0003	0.6903	−0.0047	0.6817	+0.0363	0.7409	+0.0002
	Micro	0.6004	+0.0110	0.5972	+0.0054	0.5700	+0.0642	0.7069	+0.0004
9	Macro	0.6996	+0.0032	0.6681	+0.0076	0.6617	+0.0176	0.7362	−0.0001
	Micro	0.6259	+0.0072	0.5743	+0.0314	0.5601	+0.0442	0.6993	−0.0003
10	Macro	0.6600	+0.0445	0.6701	+0.0095	0.6566	+0.0415	0.7334	+0.0025
	Micro	0.5597	+0.0552	0.5674	+0.0263	0.5275	+0.0868	0.7002	+0.0034
11	Macro	0.6941	+0.0164	0.6920	+0.0042	0.6896	+0.0166	0.7334	+0.0013
	Micro	0.5887	+0.0305	0.5687	+0.0296	0.5415	+0.0550	0.7013	+0.0012
12	Macro	0.6811	+0.0279	0.6808	+0.0256	0.6526	+0.0556	0.7134	+0.0017
	Micro	0.5852	+0.0414	0.5814	+0.0445	0.5257	+0.0971	0.6800	+0.0023
Avg Macro ↑		2.74%		1.25%		5.05%		1.54%	
Avg Micro ↑		5.09%		4.90%		12.85%		2.02%	
t-statistic (Macro)		3.35 **		2.15 *		6.91 ***		2.38 *	
t-statistic (Micro)		5.78 ***		4.72 **		12.43 ***		3.25 **	

Note: significance levels for the one-tailed paired *t*-test on daily performance differences are denoted as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. “↑” denotes increase.

Experimental results demonstrate that the MLP-based fusion module enables models to leverage the region representations more effectively, leading to greater performance gains compared to the CNN-based approach. For LSTM, the CNN-based module performs comparably to the MLP-based module in August and September but exhibits noticeable declines in October, November, and December. This can be attributed to the limitations of convolutional layers in processing high-dimensional features, which may cause information loss, whereas MLPs better preserve and utilize rich high-dimensional information [42,43]. These results suggest that information loss in the region representations degrades long-term prediction accuracy, further supporting the role of region representations in compensating for temporal dependency decay by reinforcing spatial structure understanding. For MiST, CF, and AGL-STAN, models using CNN-based fusion underperform compared to their counterparts without the region representations in most months. This implies that information loss in the region representations introduces additional noise, which is particularly detrimental to complex models.

Although the incorporation of building footprints based contextual information generally enhances model performance, slight declines are observed in some cases. These variations can be attributed to several factors. First, the added high-dimensional structural features may overlap with spatial dependencies already captured by crime or POI data, leading to feature redundancy and potential overfitting in models with limited regularization capacity, such as MiST and CF. Second, not all models are equally compatible with heterogeneous spatial features. CNN-based fusion modules, for instance, may suffer from information loss when processing dense embeddings. Third, for high-frequency crime types such as Grand Larceny, the spatiotemporal signals are already strong, and additional structural features contribute less marginal information, occasionally introducing minor instability. Lastly, inconsistencies in the completeness or labeling quality of OpenStreetMap footprint data across communities may introduce noise into model learning.

5. Conclusions

Currently, POI data applied in crime spatiotemporal prediction exhibit limitations in providing urban spatial semantic information, as they struggle to represent the continuity and mixed-use nature of urban spatial functions. To address this issue, this study introduces open-source building footprint data as a new influencing factor. By generating urban region representation based on building footprints and integrating them into various crime spatiotemporal prediction models, the effectiveness of building footprints in enhancing prediction accuracy is validated through multi-dimensional evaluations, including cross-model comparison, long-term forecasting analysis, and fine-grained prediction across different crime types.

Using New York City's building footprints, POIs and community boundaries, region representations were generated for NYC communities. Leveraging 2019 crime data, these representations were integrated into models such as LR, LSTM, MiST, CF, and AGL-STAN for multi-class predictions (covering Burglary, Robbery, Felony Assault, and Grand Larceny) as well as binary classification tasks for each of these four crime types.

The experimental results of multi-class predictions demonstrate that the region representations significantly enhance the prediction performance of LSTM, MiST, CF, and AGL-STAN models, proving their universal enhancement effect across deep learning frameworks. These representations exhibit cross-model generalizability and deliver greater improvements for models with lower baseline accuracy. Specifically, the region representations show more pronounced enhancements for models lacking spatial dependency modeling, such as LSTM, MiST, and CF, highlighting their capacity to encode rich urban spatial semantics. For AGL-STAN, the integration of the region representations also im-

proves prediction performance, indicating that the predefined graph derived from these representations effectively collaborates with adaptive graph learning. This underscores the complementary value of combining prior knowledge with data-driven approaches in spatial relationship modeling.

The binary classification results for the four crime types reveal that incorporating the region representations leads to massive accuracy improvements for Burglary; results in greater long-term accuracy gains than short-term improvements for Robbery and Felony Assault; and yields limited enhancements for Grand Larceny due to its large sample size, where baseline models already achieve high accuracy. From the perspective of case volume, crimes with fewer instances exhibit lower prediction accuracy, while the region representations provide more significant improvements for these crimes. These patterns reveal varying levels of sensitivity across crime types to the urban spatial semantics embedded in the region representations.

For both multi-class crime predictions and type-specific tasks, the region representations consistently yield greater accuracy improvements in long-term forecasting. This is further supported by the fusion module experiments: CNN-based fusion causes information loss in the region representations, resulting in inferior long-term improvements compared to MLP-based fusion. By providing stable urban spatial semantics, the region representations effectively alleviate the memory decay issue in long-term predictions for temporal models.

The region representations offer a practical, low-cost pathway to enhance policing efficiency and strategic planning. By enabling lightweight models like LSTM to achieve performance competitive with complex architectures, this approach makes effective crime prediction accessible even for resource-limited police departments, lowering the barrier to adopting data-driven strategies. The marked improvement in forecasting low-frequency crimes such as Burglary allows law enforcement to move beyond generic alerts and conduct precisely targeted interventions in specific spatiotemporal contexts, addressing the challenge of data sparsity for these crime types. Furthermore, the stability of these representations supports not only dynamic resource allocation across different time scales but also provides a foundation for long-term urban safety planning. Insights derived from the built environment can inform crime prevention through CPTED, guiding infrastructure investments and urban management decisions to proactively create safer communities.

Despite the promising results, this study has limitations that offer avenues for future work. The primary limitation is the external validity, as our experiments were conducted on data from New York City for the year 2019. Consequently, the generalizability of the findings to other urban contexts with different spatial structures (e.g., low-density cities, grid-based layouts) or to different time periods affected by unique socioeconomic factors or events (e.g., the COVID-19 pandemic) remains to be fully established. Therefore, a key direction for future research is to validate and potentially adapt the proposed framework across a diverse set of cities and over extended multi-year periods to rigorously assess its robustness and transferability. Future work also includes benchmarking against other representation methods and enhancing the interpretability of the learned features to better understand the specific urban factors driving the predictions. Additionally, the crime-influencing factors used in this study are all derived from open-source urban environmental data. In future research targeting specific cities, other types of influencing factors (such as population and weather data) could be integrated to further enhance the accuracy of crime spatiotemporal prediction.

Author Contributions: Conceptualization, T.W.; methodology, T.W. and P.C.; software, T.W.; validation, T.W. and M.S.; formal analysis, P.C.; investigation, T.W.; resources, T.W.; data curation, T.W.; writing—original draft preparation, T.W.; writing—review and editing, T.W., P.C. and M.S.; visualization, T.W.; supervision, P.C.; project administration, P.C.; funding acquisition, P.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Basic Scientific Research Business Expense Project of the People’s Public Security University of China, grant number 2024JKF04, and Innovative Talent Introduction Base for Disciplines in Higher Education Institutions, grant number B20087.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study.

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

References

1. Web-Based Injury Statistics Query and Reporting System. NVDRS Violent Deaths Report. 2024. Available online: <https://wisqars.cdc.gov/nvdrs/?rt=3&rt2=0&y=2021&g=00&i=0&m=20810&s=0&r=0&e=0&rl=0&pc=0&pr=0&h=0&ml=0&a=ALL&a1=0&a2=199&g1=0&g2=199&r1=NVDRS-INTENT&r2=NONE&r3=NONE&r4=NONE> (accessed on 21 March 2025).
2. Anderson, D.A. The aggregate cost of crime in the United States. *J. Law Econ.* **2021**, *64*, 857–885. [\[CrossRef\]](#)
3. Kang, H.W.; Kang, H.B. Prediction of crime occurrence from multi-modal data using deep learning. *PLoS ONE* **2017**, *12*, e0176244. [\[CrossRef\]](#)
4. Shan, M.; Ye, C.; Chen, P.; Peng, S. Ada-GCNLSTM: An adaptive urban crime spatio-temporal prediction model. *J. Saf. Sci. Resil.* **2025**, *6*, 226–236. [\[CrossRef\]](#)
5. Cohen, L.E.; Felson, M. Social change and crime rate trends: A routine activity approach. *Am. Sociol. Rev.* **1979**, *44*, 588–608. [\[CrossRef\]](#)
6. Wilson, J.Q.; Kelling, G.L. Broken windows. In *The City Reader*; Routledge: Abingdon, UK, 2015; pp. 303–313. [\[CrossRef\]](#)
7. Sampson, R.J.; Groves, W.B. Community structure and crime: Testing social-disorganization theory. *Am. J. Sociol.* **1989**, *94*, 774–802. [\[CrossRef\]](#)
8. Jeffery, C.R. Crime prevention through environmental design. *Am. Behav. Sci.* **1971**, *14*, 598. [\[CrossRef\]](#)
9. Polvi, N.; Looman, T.; Humphries, C.; Pease, K. The time course of repeat burglary victimization. *Br. J. Criminol.* **1991**, *31*, 411–414. [\[CrossRef\]](#)
10. 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. [\[CrossRef\]](#)
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 Association of Geographic Information Laboratories in Europe 2018 (AGILE), Lund, Sweden, 12–15 June 2018.
12. 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. [\[CrossRef\]](#)
13. Yi, F.; Yu, Z.; Zhuang, F.; Zhang, X.; Xiong, H. An integrated model for crime prediction using temporal and spatial factors. In Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 17–20 November 2018; IEEE: New York, NY, USA, 2018; pp. 1386–1391. [\[CrossRef\]](#)
14. Huang, C.; Zhang, J.; Zheng, Y.; Chawla, N.V. DeepCrime: Attentive hierarchical recurrent networks for crime prediction. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 22–26 October 2018; pp. 1423–1432. [\[CrossRef\]](#)
15. Huang, C.; Zhang, C.; Zhao, J.; Wu, X.; Yin, D.; Chawla, N. Mist: A multiview and multimodal spatial-temporal learning framework for citywide abnormal event forecasting. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 717–728. [\[CrossRef\]](#)
16. Sun, J.; Yue, M.; Lin, Z.; Yang, X.; Nocera, L.; Kahn, G.; Shahabi, C. Crimeforecaster: Crime prediction by exploiting the geographical neighborhoods’ spatio-temporal dependencies. In *Machine Learning and Knowledge Discovery in Databases, Proceedings of the Applied Data Science and Demo Track: European Conference, ECML PKDD 2020, Ghent, Belgium, 14–18 September 2020*; Proceedings, Part V; Springer International Publishing: Berlin/Heidelberg, Germany, 2021; pp. 52–67. [\[CrossRef\]](#)

17. Sun, M.; Zhou, P.; Tian, H.; Liao, Y.; Xie, H. Spatial-temporal attention network for crime prediction with adaptive graph learning. In Proceedings of the International Conference on Artificial Neural Networks, Bristol, UK, 6–9 September 2022; Springer Nature: Cham, Switzerland, 2022; pp. 656–669. [\[CrossRef\]](#)
18. Liu, K.; Zhang, L.; Tsou, S.; Wang, L.; Hu, Y.; Yang, K. Exploring the Complex Association Between Urban Built Environment, Sociodemographic Characteristics and Crime: Evidence from Washington, DC. *Land* **2024**, *13*, 1886. [\[CrossRef\]](#)
19. Ceccato, V.; Cats, O.; Wang, Q. The geography of pickpocketing at bus stops: An analysis of grid cells. In *Safety and Security in Transit Environments: An Interdisciplinary Approach*; Palgrave Macmillan: London, UK, 2015; pp. 76–98. [\[CrossRef\]](#)
20. Wang, H.; Kifer, D.; Graif, C.; Li, Z. Crime rate inference with big data. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 635–644. [\[CrossRef\]](#)
21. Zhao, X.; Tang, J. Modeling temporal-spatial correlations for crime prediction. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, Singapore, 6–10 November 2017; pp. 497–506. [\[CrossRef\]](#)
22. Wang, C.; Lin, Z.; Yang, X.; Sun, J.; Yue, M.; Shahabi, C. Hagen: Homophily-aware graph convolutional recurrent network for crime forecasting. In Proceedings of the 36th AAAI Conference on Artificial Intelligence, Virtual, 22 February–1 March 2022; Volume 36, pp. 4193–4200. [\[CrossRef\]](#)
23. Yao, Y.; Li, X.; Liu, X.; Liu, P.; Liang, Z.; Zhang, J.; Mai, K. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 825–848. [\[CrossRef\]](#)
24. Yan, X.; Ai, T.; Yang, M.; Yin, H. A graph convolutional neural network for classification of building patterns using spatial vector data. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 259–273. [\[CrossRef\]](#)
25. Patino, J.E.; Duque, J.C.; Pardo-Pascual, J.E.; Ruiz, L.A. Using remote sensing to assess the relationship between crime and the urban layout. *Appl. Geogr.* **2014**, *55*, 48–60. [\[CrossRef\]](#)
26. Silva, P.; Li, L. Urban crime occurrences in association with built environment characteristics: An African case with implications for urban design. *Sustainability* **2020**, *12*, 3056. [\[CrossRef\]](#)
27. Ioannidis, I.; Haining, R.P.; Ceccato, V.; Nascetti, A. Using remote sensing data to derive built-form indexes to analyze the geography of residential burglary and street thefts. *Cartogr. Geogr. Inf. Sci.* **2025**, *52*, 259–275. [\[CrossRef\]](#)
28. Li, Y.; Huang, W.; Cong, G.; Wang, H.; Wang, Z. Urban region representation learning with openstreetmap building footprints. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Long Beach, CA, USA, 6–10 August 2023; pp. 1363–1373. [\[CrossRef\]](#)
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [\[CrossRef\]](#)
30. Bridson, R. Fast Poisson disk sampling in arbitrary dimensions. In Proceedings of the ACM SIGGRAPH 2007 Sketches, San Diego, CA, USA, 5–9 August 2007. [\[CrossRef\]](#)
31. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762. [\[CrossRef\]](#)
32. Ni, K.; Bresson, X.; Chan, T.; Esedoglu, S. Local histogram based segmentation using the Wasserstein distance. *Int. J. Comput. Vis.* **2009**, *84*, 97–111. [\[CrossRef\]](#)
33. Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 815–823. [\[CrossRef\]](#)
34. Liu, C.; Yang, S.; Xu, Q.; Li, Z.; Long, C.; Li, Z.; Zhao, R. Spatial-temporal large language model for traffic prediction. In Proceedings of the 2024 25th IEEE International Conference on Mobile Data Management (MDM), Brussels, Belgium, 24–27 June 2024; IEEE: New York, NY, USA, 2024; pp. 31–40. [\[CrossRef\]](#)
35. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2013. [\[CrossRef\]](#)
36. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. *Neural Comput.* **2000**, *12*, 2451–2471. [\[CrossRef\]](#)
37. Lin, Z.; Lyu, S.; Cao, H.; Xu, F.; Wei, Y.; Samet, H.; Li, Y. Healthwalks: Sensing fine-grained individual health condition via mobility data. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*; Association for Computing Machinery: New York, NY, USA, 2020; Volume 4, pp. 1–26. [\[CrossRef\]](#)
38. Grover, A.; Leskovec, J. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 855–864. [\[CrossRef\]](#)
39. Where is New York City, NY, USA on Map Lat Long Coordinates. Lat Long Finder. Available online: <https://www.latlong.net/place/new-york-city-ny-usa-1848.html> (accessed on 22 March 2025).
40. United States Census Bureau QuickFacts. U.S. Census Bureau QuickFacts. Available online: <https://data.census.gov/table?q=new%20york%20city> (accessed on 7 September 2025).
41. Jing, L.; Gulcehre, C.; Peurifoy, J.; Shen, Y.; Tegmark, M.; Soljagic, M.; Bengio, Y. Gated orthogonal recurrent units: On learning to forget. *Neural Comput.* **2019**, *31*, 765–783. [\[CrossRef\]](#)

42. Liu, H.; Dai, Z.; So, D.; Le, Q.V. Pay attention to mlps. *arXiv* **2021**, arXiv:2105.08050. [[CrossRef](#)]
43. Tolstikhin, I.O.; Houlsby, N.; Kolesnikov, A.; Beyer, L.; Zhai, X.; Unterthiner, T.; Yung, J.; Steiner, A.; Keysers, D.; Uszkoreit, J.; et al. Mlp-mixer: An all-mlp architecture for vision. *arXiv* **2021**, arXiv:2105.01601. [[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.