

A DEEP NEURAL NETWORK FOR SPATIOTEMPORAL PREDICTION OF THEFT CRIMES

Xinxin Lv¹, Changfeng Jing^{2*}, Yi Wang³, Shiyuan Jin¹

¹School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 100044, China – (X.L.) 2108570020068@stu.bucea.edu.cn, (S.J.) 2108160320003@stu.bucea.edu.cn

²School of Information Engineering, China University of Geosciences, Beijing 10083, China – (C.J.) jingcf@cugb.edu.cn

³School of Earth and Space Sciences, Peking University, Beijing 100871, China – (Y.W.) yiwanggis@gmail.com

Commission IV, WG IV/8

KEY WORDS: Spatiotemporal Prediction, Chicago Theft Crime, Grid Unit Division, Deep Neural Network, Spatiotemporal Dependence, Deep Learning.

ABSTRACT:

Accurate crime prediction plays an important role in public safety, providing technical guidance and decision support for the police and government departments. Due to the dynamics and imbalance of crime distribution, it is difficult to build predictive models for it. Specifically, the fine-grained and non-linear spatiotemporal dependencies of crime data cannot be captured accurately. In this paper, a neural network model ST-ACLCrime based on ConvLSTM and SE block was proposed to predict the number of theft crimes in hotspot areas. By overlaying ConvLSTM layers, fine-grained spatiotemporal dependencies are captured while preserving spatial location information. To further enhance the global channel feature representation, SE block is used to recalibrate the channel features and enhance the channel inter-dependencies. In addition, the closeness and the period components are set to dynamically capture the dependence of different time trends. We choose the city of Chicago as the study case, and use a multi-level spatial grid to divide the whole city area. The experimental results show that the proposed model exceeds all baseline model, such as HA, CNN, LSTM, CNN-LSTM and ConvLSTM. It was effectively capturing spatiotemporal dependence and improving prediction accuracy.

1. INTRODUCTION

Crime prediction has been a hot topic in modern research and business (Wang et al., 2019). Timely crime prediction can prevent the loss of physical life and property, while providing guarantees for urban operations, public security, and sustainable development. Theft crimes are related to the victims' property loss and order of life, presenting a high frequency, and the police department will always maintain a strict attitude toward such property crime (Chen et al., 2015; Ye et al., 2021). Therefore, it is great importance to deeply explore the spatiotemporal patterns for specific types of crime.

Crimes as a geography-based spatiotemporal event data, which has distinct information about the occurrence time and geographical location (Lin et al., 2018). Interestingly, hotspot areas are constantly changing over time. Some theoretical foundations, such as the routine activity theory and crime pattern theory, provide a foundation for the distribution pattern and hotspot patterns of criminal activity (Bernasco et al., 2012; Cohen and Felson, 1979). However, it is difficult to reflect the intrinsic mechanisms of crime in the short term. Tapping into the periodic patterns and spatiotemporal dependence in the short term for crime number prediction becomes a focus of attention.

Existing predictive methods are mainly based on historical data to predict the conditions at future moments or locations. Early time series models, such as autoregressive integrated moving average (ARIMA), tended not to accurately depict spatiotemporal patterns (Chen et al., 2008). The reason was that ARIMA models can easily fit linear and smooth time series, while non-linear and dynamic features could not be accurately

depicted. The vector autoregressive (VAR) required to set parametric variables along with extensive sample support, resulting in difficult parameter estimation (Zivot and Wang, 2006). With the rapid development of deep learning, prediction models such as LSTM and GRU have led to a significant improvement in predictive capability (Cortez et al., 2018; Su and Jiang, 2020). However, this focuses only on the time dependence of events, where these models ignore the impact of the spatial dependence. In another research direction, forecasting using spatiotemporal sequence also becomes of increasing interest. Convolution-based models captures spatial dependence at different distances for crowd flow prediction (Zhang et al., 2016). A deep residual network architecture called ST-ResNet, was implemented on crime datasets based grid division (Wang et al., 2017). However, they model temporal dependence through timestamps of different properties, which cannot capture continuous time-stamps within a property. In addition, some models use large spatial units in order to avoid problems caused by data sparsity.

Considering the deficiencies of existing studies, a neural network called ST-ACLCrime was proposed for crime prediction. It combines a Convolutional LSTM (ConvLSTM) layer and a Squeeze-and-Excitation (SE) block with channel attention to obtain fine-grained and non-linear spatiotemporal dependencies. In this paper, it is explored on a multi-level spatial division while being able to identify crime hotspot areas and quantities. The dependencies of different time properties are also dynamically fused in considering various properties. The contributions are as follows.

* Corresponding author, Changfeng Jing jingcf@cugb.edu.cn

(1) A model called ST-ACLCrime has been proposed with the aim of improving the prediction accuracy and hotspot hit rate. It takes into account spatiotemporal pattern interaction, global channel information and temporal periodicity.

(2) The ConvLSTM captures the spatiotemporal dependencies of fine-grained and non-linear features. The SE block that follows ConvLSTM immediately provides effective control to the global information.

(3) The Chicago theft dataset was selected and validated in two-level of spatial resolution, 2000 and 1000 meters, respectively. The experimental results show that proposed model outperforms the baseline models.

The article is organized as follows. The related work on crime prediction is reviewed in Section 2. The details of the proposed model are described in Section 3. We show the study area and crime dataset in Section 4, and the training procedure in Section 5. The experimental results are discussed in Section 6, and Section 7 summarizes the full work.

2. RELATED WORK

Prediction of the spatial and temporal patterns of crime occurrence is essential. According to the existing research based on, we classify crime prediction models into three categories: statistical approaches, machine learning approaches and deep learning approaches.

2.1 Statistical Approaches

ARIMA was used for short-term prediction of crime time series, and its results outperformed the two methods of Exponential Smoothing (Chen et al., 2008). Different crime types are considered as predictable, such as robbery, murder, burglary and total 8 crime types, as well as Linear Regression model approach was utilized in the crime data provided by Bangladesh police department (Awal et al., 2016). ARIMA and Linear Regression would be more suitable for linear relations, and the prediction performance is deficient for the crime of imbalanced feature distribution. To explore the spatial patterns of crime, Kernel Density Estimation (KDE) has been investigated to identify hot crime areas and explore their parameter settings (Hart and Zandbergen, 2014). Twitter data incorporated into the crime prediction model through thematic modeling, the KDE method for prediction and decision support for the city of Chicago (Gerber, 2014). KDE does not require a priori knowledge, but the fixed bandwidth will lead to a large gap between the probability density and the reality. Most importantly, the time-dependent relationships are ignored. To address the temporal deficiency, Spatio-Temporal Kernel Density Estimation (ST-KDE) was proposed (Brunsdon et al., 2007), where temporal weights and spatial kernels are combined to express three-dimensional data feature. It was employed for spatiotemporal analysis and prediction tasks, such as crime hotspot and medical service demand. For instance, a ST-KDE approach was applied to predictive hotspot maps of crime data, which spatiotemporal features are integrated (Hu et al., 2018b). Self-exciting point processes (SEPP) also help to understand crime patterns by spatial covariate information and parameter inference, improving the robustness by isotropic triggering (Reinhart and Greenhouse, 2017; Rosser and Cheng, 2019). Summarize the above, statistical-based methods require a lot of debugging to find the optimal parameters, which also results in a huge workload.

2.2 Machine Learning Approaches

With the enhancement of computer hardware and software devices, machine learning methods are gradually used to handle prediction tasks for big data. Common approaches include: Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (KNN), Decision Tree and Bayesian. SVM have been proposed to predict crime hotspots, and identify locations with high levels of crime hotspot classes by binary classification (Kianmehr and Alhaji, 2008). In a cyber-crime, where eight machine learning methods were studied, SVM was considered the most productive method for predicting the type of cyber-attacks and achieving high prediction accuracy (Bilen and Özer, 2021). This general framework can also be used for other spatial data tasks. KNN was used to observe crime rates and even possible locations and times of occurrence (Kumar et al., 2020). Spatiotemporal Bayesian approach were established to model the trend of property crime over time at small areas, providing new insights into prediction of hot and cold spots (Law et al., 2014). Decision Tree and Naive Bayesian classifier were compared to predict future crimes at specific times, and the results showed that Naive Bayesian has better accuracy (Almanic et al., 2015). In addition, demographic data were combined to capture future factors that may affect the occurrence of crimes. But in another Chicago crime study, decision trees were superior to the Naive Bayesian algorithm (Aldossari et al., 2020).

2.3 Deep Learning Approaches

Deep learning simplifies the workflow of machine learning, relying on its powerful temporal and spatial feature extraction capabilities to become the mainstream approach. Long Short-Term Memory (LSTM) was established for emergency event prediction that provided a response for unexpected events (Cortez et al., 2018). Its performance exceeds that of time series models such as Moving Average and ARIMA. Gated Recurrent Unit (GRU) was used for the prediction of important news events and incorporated the effect of prior knowledge (Su and Jiang, 2020). However, these models treat historical data as time series that ignore the effect of spatial dependence.

In the European space, the temporal and spatial factors are also captured at the same time. Crime data were mapped to the grid, and CNN and LSTM were combined to predict crime number, with prediction accuracy exceeding that of independent CNN and LSTM (Stec and Klabjan, 2018). It is suggested that the models have different levels of impact for different crime types, and CNN having a higher impact on theft crimes than narcotic drug crimes. Convolution-based deep neural network (DNN) models were designed to predict the occurrence of crimes, and the feature-level fusion approach preserves the non-linear dependencies from different domains (Kang and Kang, 2017). However, temporal and spatial are independent components ignoring their interactivity, and more importantly are not adapted to areas where data are not sufficient. Considering that crime data was sparsely distributed, ST-ResNet was implemented to predict when and where crimes would occur (Wang et al., 2019). However, more attention is paid to fine-grained temporal units rather than fine-grained spatial units. Three deep learning architectures, such as ST-ResNet, DMVST-Net and STD-Net, were compared for the prediction of Chicago crime (Matereke et al., 2021). Hetero-ConvLSTM framework performed the prediction of traffic accidents and solved the problem of spatial heterogeneity (Yuan et al., 2018). However, it is still a coarse-grained spatial division unit.

Attentional mechanisms assign weights based on the importance of elements, which is also a major direction of prediction research. A model called DeepCrime was proposed by adding attention to the residual neural network (Huang et al., 2018). The importance of across time was learned through the attention mechanism while dynamic spatiotemporal dependencies were obtained. The dual self-attention network (DSANet) for multivariate time series prediction (Huang et al., 2019). Multi-level attention networks, named GeoMAN, for time Series Prediction (Liang et al., 2018). In addition, there are applications of graph-based structured data under non-Euclidean spaces. A gated localised diffusion network (GLDNet) for sparse data prediction in network-level hotspot mappings, validated on a small region of the crime dataset (Zhang and Cheng, 2020). Temporal Graph Convolutional Neural Network (T-GCN) captured the dynamics of crime in temporal and spatial terms, and likewise proved to be effective for prediction tasks (Jin et al., 2020).

3. METHODOLOGY

3.1 Problem Definition

Encouraged by existing studies (Wang et al., 2019), the study area was divided into $H \times W$ grid with equal spatial intervals, where H is the number of rows and W is the number of columns. Grid location was defined as $L_{h,w} = [l_{0,0}, l_{0,1}, \dots, l_{H-1,W-1}]$. All crime information is divided according to time intervals, and represent as $T = [t_1, t_2, \dots, t_n]$. Crime incidents falling into each grid was counted, and the data distribution is given by $X_t = [[t_1, x_1], [t_2, x_2], \dots, [t_n, x_n]]$. Learning from historical multi-day crime data $X_m = \{m = 0, 1, \dots, n-1\}$ and predicting crime data for the coming day X_n .

3.2 Model Architecture

This paper is an exploration of the spatiotemporal distribution patterns of theft crimes, predicting the number of crimes within each grid unit for a city scale. A deep neural network model with attention was defined to extract the spatiotemporal dependence features of crime incidents. The general architecture of the ST-ACLCrime model is shown in Figure 1. It consists of three sections: dataset construction, spatiotemporal dependency modeling and parameter matrix fusion. The details of the methodology are as follows.

Considering the periodicity of theft crimes, there were two temporal components with the same structure constructed: daily closeness and weekly periods. In each temporal component, the spatiotemporal dependencies are extracted by Convolutional layers, ConvLSTM layer, Squeeze-and-Excitation (SE) block. Convolutional layer for transforming feature dimensions. ConvLSTM network uses convolutional structure instead of the fully-connected structure of LSTM (FC-LSTM) (Shi et al., 2015), which can establish both temporal sequence relationships and retain spatial information. As a lightweight CNN architecture, SE block can be adaptively recalibrated for channel features and obtained global information interdependencies (Hu et al., 2018a). Spatiotemporal dependencies are captured simultaneously more fine-grained and non-linear, and improving the accuracy of crime prediction. The parameter matrix approaches were used for the dynamic fusion of closeness and period components (Zhang et al., 2016).

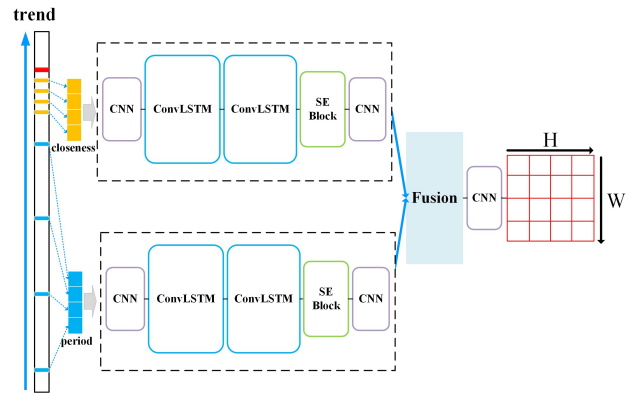


Figure 1. The framework of ST-ACLCrime.

3.3 Time Components

In the case where the input sequence is too long, the training time is too long and more difficult. To address this gap, the model chooses highly correlated timestamp sequences. This is to ensure the best prediction performance on the one hand and to reduce the input dataset on the other hand. Some timestamps are highly dependent compared to others and play a decisive role in crime incidents prediction. Therefore, the closeness dependence and period dependence of crime were considered to establish two different time trend components. They were denoted as X_c and X_d , respectively. The specific formulas are as follows.

$$X_c = [X_{t-l_c}, X_{t-(l_c-1)}, \dots, X_{t-1}] \quad (1)$$

$$X_d = [X_{t-l_d-p}, X_{t-(l_d-1)-p}, \dots, X_{t-p}] \quad (2)$$

Where l_c and l_d denote the length of the closeness component and the periodic component, respectively, and p is the time interval of the periodic component.

3.4 Convolutional LSTM

ConvLSTM was proposed to solve the prediction problem of spatiotemporal sequences. The fully connected operation of FC-LSTM is replaced by the convolution operation in the state transition process, due to the FC-LSTM is insufficient for spatial information. We consider stacking multiple layers of ConvLSTM to form a prediction structure. Unlike focusing only on the temporal or spatial dimensions, simultaneously acquiring spatiotemporal dependencies has better performance in prediction task. The formula of ConvLSTM is as follows.

$$\begin{aligned} i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i); \\ f_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f); \\ C_t &= f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c); \\ o_t &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o); \\ H_t &= o_t \circ \tanh(C_t) \end{aligned} \quad (3)$$

Where X_t denotes input, H_1, H_2, \dots, H_t denotes cell output, C_1, C_2, \dots, C_t denotes the hidden states, i_t, f_t, o_t the gates, ‘*’ indicates convolution operator and ‘ \circ ’ indicates Hadamard product.

3.5 Squeeze-and-Excitation Block

CNNs have proven to be useful in capturing spatial dependencies, with success in the field of vision and event prediction. However, its receptive field is limited and can only capture local spatial information. Therefore, the channel attention mechanism is incorporated, which improves the perception of global information by modeling the interdependencies between channels. The SE block is made up of two parts: squeeze and excitation. Squeeze portion gets the global information of the feature map, while the excitation portion learns the feature weights of each channel. Details are shown in Figure 2.

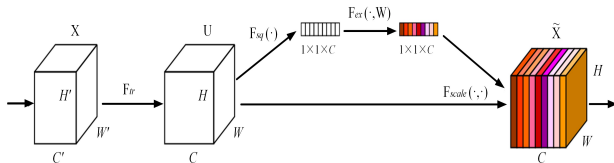


Figure 2. The structure of SE block.

Where W and H indicate the width and height of the input feature, respectively. C denotes the channel number of features, F_{sq} and F_{ex} denote the squeeze and excitation operations, respectively. F_{tr} denote the convolutional operator. F_{scale} denote the channel-wise multiplication.

3.6 Spatiotemporal Feature Fusion

To better capture the temporal dependence, the closeness and periodicity of temporal trends are considered. They are highly correlated with the occurrence of crime and are more likely to influence the prediction results. However, the degree of influence varies significantly across components. A parametric-matrix approach is employed to dynamically fuse the temporal dependence of different temporal trends. Unlike the manual parameter setting, dynamic learning achieves better effects.

$$X_o = W_c \circ X_c^L + W_d \circ X_d^L \quad (4)$$

Where X_o are the output, and the X_c and X_d are the closeness and period sequence, respectively; W_c and W_d represents the learnable parameters; and ‘ \circ ’ represents Hadamard product.

4. STUDY AREA AND DATASET

The city of Chicago, USA, was chosen as the study area. As the third largest city in the United States, Chicago is highly prosperous in transportation, economy, and education. However, behind the high prosperity of Chicago, it breeds a large number of urban crimes. It is much higher than the national average. Theft crimes, which directly involve the security of citizens' property, have a high frequency of occurrence. In addition, due to the fact that crime data is officially collected, managed and made public, it is of great investigative value and practical meaning to consider it as a case study. The study area and theft distribution are shown in the Figure 3.

The acquired crime dataset was downloaded through the Open Data Portal to be developed by the Chicago government (<https://data.cityofchicago.org/>). The dataset spans a total of 731 days and contains a total of about 124,620 theft records from January 1, 2016 to December 31, 2017. This collects complete

crime information, including incident number, crime type, time of reporting, geographic location, brief background description and so on. The daily number of crimes is displayed in Figure 4.

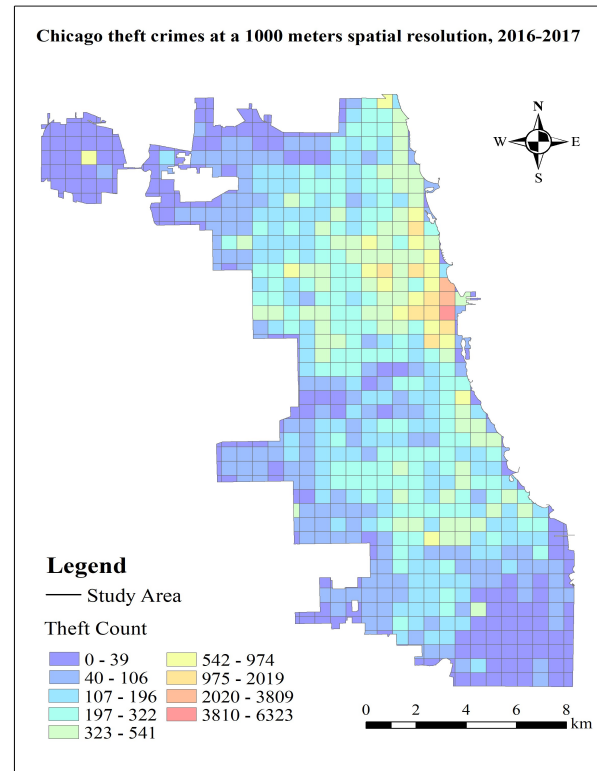


Figure 3. The study area and the distribution of theft crimes in 1000 meters.

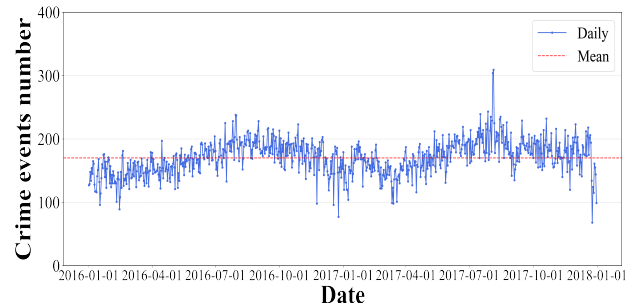


Figure 4. The daily number of theft crimes from January 1, 2016 to December 31, 2017.

5. EXPERIMENT

5.1 Grid Division

The grid can combine the attribute characteristics of crime (e.g., numbers) and geographic location which compensates for the coarse granularity of existing administrative areas. Based on the existing studies, 2000 meters and 1000 meters were considered as division units (Wang et al., 2021; Gerber, 2014). At each temporal granularity T , the two-dimensional tensor of response spatial location was constructed. The rasterized dataset (T, H, W) was used for the training of the model, where T is denoted as the timestamp, H and W are the number of rows and columns of dividing grid.

Gird Size	H (Rows)	W (Columns)
2000 meter	21	18
1000 meter	42	35

Table 1. Two-level spatial unit division.

5.2 Model Training

All models follow the PyTorch framework for training. To avoid the influence of objective factors, they use the same training parameters, as detailed below. The number of incidents falling into the grid is counted on a daily basis. The length of the closeness sequence and period sequence are 3 and 4, and time interval of the periodic is 7. The training and test sets are divided according to 9:1, i.e., the last 70 days as test sets. MSE is used as the loss function with Adam optimizer. The learning rate of 0.0001 and Batch size is 32. All models are terminated after 100 iterations. The number of convolution channels is 64.

5.3 Baseline Models

Considering the learning ability of the models in temporal and spatial terms, a total of five baseline models were selected for comparison and analysis in the experiment.

HA (Smith and Demetsky, 1997): Historical Average Model. The average of historical moments is used as the result of the prediction.

CNN (Fukushima, 1980): Convolutional Neural Networks. Spatial dependence of events captured by relying on convolution operations.

LSTM (Hochreiter and Schmidhuber, 1997): Long Short-Term Memory network. It is used for prediction of long sequences, avoiding gradient disappearance and explosion.

CNN-LSTM: CNN is used to capture spatial dependence and LSTM to capture temporal dependence.

ConvLSTM (Shi et al., 2015): Replacing the fully-connected operation of LSTM with convolution and having better performance in capturing spatiotemporal dependencies.

5.4 Model Evaluation

In order to evaluate the error between the predicted and true values, three evaluation metrics were chosen for our model and the baseline model: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Hit Rate. The smaller the result of the evaluation metric indicates the stronger the predictive ability of the model for MAE and RMSE. As shown in Eq. 5, Eq. 6 and Eq. 7 below.

$$MAE = \frac{1}{c} \sum_{i=1}^c |x - \hat{x}| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{c} \sum_{i=1}^c (x - \hat{x})^2} \quad (6)$$

$$Mean\ Hit\ Rate = \frac{1}{c} \sum_{i=1}^c \left(\frac{n_i}{N_i} \right) * 100\% \quad (7)$$

Where x and \hat{x} denote the true values and the predicted values, respectively, and c denotes the number of values. n_i and N_i denote the number of events in the hotspot area and the total study area, respectively.

6. RESULTS AND DISCUSSIONS

Three evaluation metrics were used to evaluate the models. In addition, we show the mean hit rate at 10% and 20% coverage levels. The mean hit rate ranks the predicted results. The results of the predictive metrics for all models are shown in Table 2 and Table 3. Figure 5 shows the trend of the mean hit rate with different coverage levels.

Model	MAE	RMSE	10%	20%
HA	0.5840	1.5121	41.73%	57.23%
CNN	0.3783	0.8191	55.01%	76.65%
LSTM	0.4795	0.8768	55.88%	78.53%
CNN-LSTM	0.3687	0.8116	56.55%	80.16%
ConvLSTM	0.3544	0.7774	58.51%	81.44%
ST-ACLCrime	0.3449	0.7658	58.52%	82.01%

Table 2. Comparisons the results of different models at 2000 meters spatial resolution.

Model	MAE	RMSE	10%	20%
HA	0.2017	0.8822	42.51%	52.23%
CNN	0.1438	0.4009	54.21%	72.20%
LSTM	0.1977	0.4184	59.14%	77.38%
CNN-LSTM	0.1455	0.3991	54.16%	72.16%
ConvLSTM	0.1432	0.3838	59.55%	78.96%
ST-ACLCrime	0.1409	0.3798	59.41%	80.57%

Table 3. Comparisons the results of different models at 1000 meters spatial resolution.

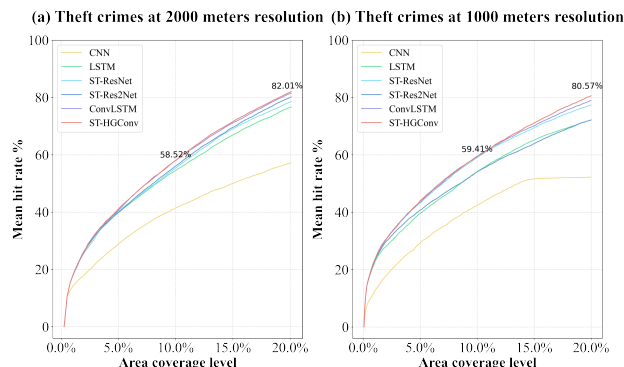


Figure 5. Trend of mean hit rate at different coverage levels.

The experimental results are as follows. At 2000 meters spatial resolution, the MSE and RMSE of ST-ACLCrime reached 0.3449 and 0.7658, respectively, and the mean hit rate reached 82.01% at 20% coverage level. The average hit rate is improved with a maximum of 24.78% compared to HA model. At 1000 meters spatial resolution, the MSE and RMSE of ST-ACLCrime reached 0.1409 and 0.3798, respectively, and the mean hit rate reached 80.57% at 20% coverage level. Compare to ConvLSTM, ST-ACLCrime models have a 1.61% improvement of hit rate in at 20% level.

HA ignored the temporal fluctuation and dynamic change of the events, and the prediction performance has significant shortcomings. In comparison, deep learning models have better performance. The CNN and LSTM were affected by the spatial unit division, so their performance was less stable with much room for improvement. Better prediction accuracy was obtained using the CNN-LSTM combination rather than focusing only, surpassing that of a single CNN and LSTM. ConvLSTM was superior to CNN-LSTM in terms of model architecture and thus

performed better in spatiotemporal prediction. ST-ACLCrime combines ConvLSTM and SE block, focusing on global channel features and fitting complex dependencies, with significantly improved fine-grained and non-linear capabilities. Its spatiotemporal performance exceeds all baseline models, achieving the best prediction accuracy.

To further explore the performance of predictive performance, a specific day was chosen to visualize the crime data. Figure 6 shows the predicted number of ST-ACLCrime on December 31, 2017. The performance of the true and predicted values at different resolutions are compared respectively. In addition, the following conclusions are drawn from the visualized results by comparing them at different units. The proposed model obtained good prediction performance at 2000 meters and was able to significantly predict the area where the cases occurred. At 1000 meters the performance is slightly lower than 2000 meters due to the sparse grid and low incidents values. Therefore, it is important to select the suitable grid units in order to obtain better performance.

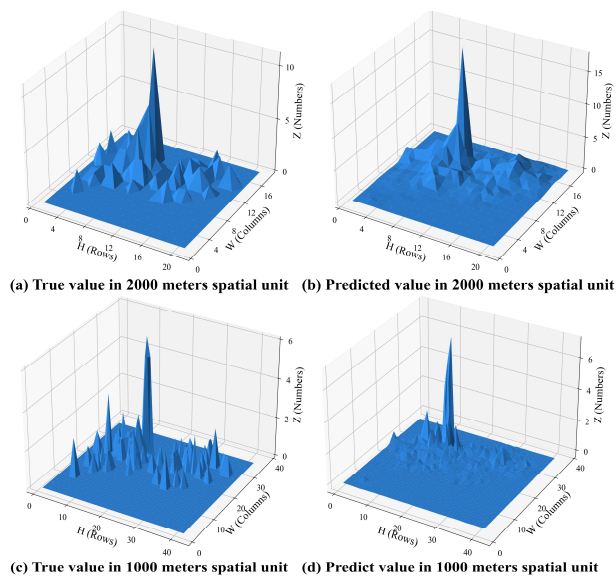


Figure 6. ST-ACLCrime predictions for December 31, 2017 at two-level spatial resolution, respectively.

7. CONCLUSION

In the paper, we propose an end-to-end neural network model, called ST-ACLCrime, to predict theft crimes at the urban scale. It combines ConvLSTM layer with SE block. ConvLSTM simultaneously captures spatial and temporal dependencies. The global channel-wise features are calibrated by the SE block immediately after the ConvLSTM. Theft crime in Chicago was selected as the experimental data for the study. ST-ACLCrime obtains more accurate prediction results through periodicity, spatiotemporal dependence and channel attention. Validation at multiple levels of spatial units to facilitate research on fine-grained spatial units. It is hoped that patrol routes can be arranged for police departments and governments. In the future, external influences on crime and network-level expansion can be taken into account.

ACKNOWLEDGMENTS

This research was funded by the China Scholarship Council (Grant #20221007), the Beijing Natural Science Foundation (Grant #8222009), and the BUCEA Post Graduate Innovation Project (Grant #03081022001).

REFERENCES

- Aldossari, B. S., Alqahtani, F. M., et al., 2020: A comparative study of decision tree and naive bayes machine learning model for crime category prediction in Chicago. *Proceedings of 2020 the 6th international conference on computing and data engineering*, 34-38.
- Almanie, T., Mirza, R., et al., 2015: Crime prediction based on crime types and using spatial and temporal criminal hotspots. *arXiv preprint arXiv:1508.02050*.
- Awal, M. A., Rabbi, J., et al., 2016: Using linear regression to forecast future trends in crime of Bangladesh. *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, 333-338.
- Bernasco, W., Block, R., et al., 2012: Go where the money is: Modeling street robbers' location choices. *Journal of Economic Geography* 13, 119-143.
- Bilen, A. and Özer, A. B., 2021: Cyber-attack method and perpetrator prediction using machine learning algorithms. *PeerJ Computer Science* 7.
- Brunsdon, C., Corcoran, J., et al., 2007: Visualising space and time in crime patterns: A comparison of methods. *Computers, Environment and Urban Systems* 31(1), 52-75.
- Chen, P., Yuan, H. Y., et al., 2008: Forecasting Crime Using the ARIMA Model. *2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery* 5, 627-630.
- Chen, X., Cho, Y., et al., 2015: Crime prediction using Twitter sentiment and weather. *2015 Systems and Information Engineering Design Symposium*, 63-68.
- Cohen, L. and Felson, M., 1979: Social Change and Crime Rate Trends: A Routine Activity Approach. *American Sociological Review* 44, 588-608.
- Cortez, B., Carrera, B., et al., 2018: An architecture for emergency event prediction using LSTM recurrent neural networks. *Expert Systems with Applications* 97, 315-324.
- Gerber, M. S., 2014: Predicting crime using Twitter and kernel density estimation. *Decision Support Systems* 61, 115-125.
- Hart, T. and Zandbergen, P., 2014: Kernel density estimation and hotspot mapping: Examining the influence of interpolation method, grid cell size, and bandwidth on crime forecasting. *Policing: An International Journal of Police Strategies and Management* 37(2), 305-323.
- Hochreiter, S. and Schmidhuber, J., 1997: Long Short-term Memory. *Neural computation* 9(8), 1735-1780.

- Hu, J., Shen, L., et al., 2018a: Squeeze-and-Excitation Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7132-7141.
- Hu, Y., Wang, F., et al., 2018b: A spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation. *Applied geography* 99, 89-97.
- Huang, C., Zhang, J., et al., 2018: DeepCrime: Attentive hierarchical recurrent networks for crime prediction. *Proceedings of the 27th ACM international conference on information and knowledge management*, 1423-1432.
- Huang, S., Wang, D., et al., 2019: Dsanet: Dual self-attention network for multivariate time series forecasting. *Proceedings of the 28th ACM international conference on information and knowledge management*, 2129-2132.
- Jin, G., Wang, Q., et al., 2020: Addressing Crime Situation Forecasting Task with Temporal Graph Convolutional Neural Network Approach. *2020 12th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, 474-478.
- Kang, H.-W. and Kang, H.-B., 2017: Prediction of crime occurrence from multi-modal data using deep learning. *PLoS one* 12(4), e0176244.
- Kianmehr, K. and Alhaji, R., 2008: EFFECTIVENESS OF SUPPORT VECTOR MACHINE FOR CRIME HOT-SPOTS PREDICTION. *Appl. Artif. Intell.* 22(5), 433-458.
- Kumar, A., Verma, A., et al., 2020: Crime Prediction Using K-Nearest Neighboring Algorithm. *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, 1-4.
- Law, J., Quick, M., et al., 2014: Bayesian Spatio-Temporal Modeling for Analysing Local Patterns of Crime Over Time at the Small-Area Level. *Journal of Quantitative Criminology* 30(1), 57-78.
- Liang, Y., Ke, S., et al., 2018: Geoman: Multi-level attention networks for geo-sensory time series prediction. *IJCAI* 2018, 3428-3434.
- Lin, Y.-L., Yen, M.-F., et al., 2018: Grid-Based Crime Prediction Using Geographical Features. *ISPRS International Journal of Geo-Information* 7(8), 298.
- Matereke, T., Nyirenda, C. N., et al., 2021: A Comparative Evaluation of Spatio Temporal Deep Learning Techniques for Crime Prediction. *2021 IEEE AFRICON*, 1-6.
- Reinhart, A. and Greenhouse, J., 2017: Self-exciting point processes with spatial covariates: Modelling the dynamics of crime. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 67(5), 1305-1329.
- Rosser, G. and Cheng, T., 2019: Improving the Robustness and Accuracy of Crime Prediction with the Self-Exciting Point Process Through Isotropic Triggering. *Applied Spatial Analysis and Policy* 12(1), 5-25.
- Shi, X., Chen, Z., et al., 2015: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems* 28.
- Smith, B. L. and Demetsky, M. J., 1997: TRAFFIC FLOW FORECASTING: COMPARISON OF MODELING APPROACHES. *Journal of Transportation Engineering-asce* 123(4), 261-266.
- Stec, A. and Klabjan, D., 2018: Forecasting Crime with Deep Learning. *ArXiv* 5.
- Su, Z. and Jiang, J., 2020: Hierarchical gated recurrent unit with semantic attention for event prediction. *Future Internet* 12(2), 39.
- Wang, B., Lin, Y., et al., 2021: GSNet: Learning spatial-temporal correlations from geographical and semantic aspects for traffic accident risk forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* 35(5), 4402-4409.
- Wang, B., Zhang, D., et al., 2017: Deep learning for real time crime forecasting. *arXiv preprint arXiv:1707.03340*.
- Wang, B., Yin, P., et al., 2019: Deep Learning for Real-Time Crime Forecasting and Its Ternarization. *Chinese Annals of Mathematics, Series B* 40(6), 949-966.
- Ye, X., Duan, L., et al., 2021: Spatiotemporal Prediction of Theft Risk with Deep Inception-Residual Networks. *Smart Cities* 4(1), 204-216.
- Yuan, Z., Zhou, X., et al., 2018: Hetero-ConvLSTM: A Deep Learning Approach to Traffic Accident Prediction on Heterogeneous Spatio-Temporal Data, *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* London, United Kingdom, 984-992.
- Zhang, J., Zheng, Y., et al., 2016: DNN-based prediction model for spatio-temporal data. *Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems*, 1-4.
- Zhang, Y. and Cheng, T., 2020: Graph deep learning model for network-based predictive hotspot mapping of sparse spatio-temporal events. *Computers, Environment and Urban Systems* 79, 101403.
- Zivot, E. and Wang, J., 2006: Vector autoregressive models for multivariate time series. *Modeling financial time series with S-PLUS®*, 385-429.