Spatiotemporal Crime Prediction in New York City Using Convolutional LSTM Networks

Salem Fradi
Sup'Com
Tunis, Tunisia
salemaziz.fradi@supcom.tn

Ghassen Zrigua
Sup'Com
Tunis, Tunisia
ghassen.zrigua@supcom.tn

Fatma Ben Helal
Sup'Com
Tunis, Tunisia
fatmaezzahra.benhelal@supcom.tn

Ahmed Essouaied
Sup'Com
Tunis, Tunisia
ahmed.essouaied@supcom.tn

Abstract—Crime prediction is a critical challenge for urban safety and law enforcement resource allocation. This paper presents a deep learning framework for spatiotemporal crime prediction in New York City using Convolutional LSTM (ConvLSTM) networks. We transform historical NYPD crime complaint data into a grid-based representation and employ a dual-stack ConvLSTM architecture to capture complex spatial and temporal dependencies. Our approach incorporates geographic masking to focus predictions within NYC borough boundaries, ensuring realistic and actionable forecasts. We compare the performance of our deep learning model against two baseline Historical Average (HA) models: global average and weekday-specific average. Experimental results demonstrate that the ConvLSTM architecture effectively captures spatiotemporal crime patterns, achieving masked MSE of 0.7413 and MAE of 0.3997, significantly outperforming the HA global baseline (MSE: 1.1083, RMSE: 1.0527) and HA weekday baseline (MSE: 1.1391, RMSE: 1.0673). Our framework provides law enforcement agencies with data-driven insights for proactive crime prevention strategies.

Index Terms—Crime prediction, Convolutional LSTM, spatiotemporal analysis, deep learning, urban computing, New York City

I. INTRODUCTION

Urban crime prediction has emerged as a critical application of machine learning in smart city initiatives. Accurate forecasting of crime patterns enables law enforcement agencies to optimize patrol routes, allocate resources efficiently, and implement proactive intervention strategies. Traditional statistical methods often fail to capture the complex spatiotemporal dependencies inherent in urban crime dynamics, motivating the adoption of deep learning approaches.

New York City, with a population exceeding 8 million and spanning five boroughs (Manhattan, Brooklyn, Queens, The Bronx, and Staten Island), presents unique challenges for crime prediction. The city exhibits heterogeneous crime patterns influenced by diverse socioeconomic factors, population density variations, and complex urban geography. The New York Police Department (NYPD) maintains comprehensive crime complaint databases, providing rich historical data for predictive modeling.

Recent advances in deep learning, particularly Convolutional LSTM (ConvLSTM) networks [1], have demonstrated remarkable success in spatiotemporal sequence forecasting tasks. ConvLSTM extends traditional LSTM networks by incorporating convolutional operations, enabling the model to capture both spatial correlations and temporal dependencies simultaneously. This architecture has proven effective in various domains including weather prediction, traffic forecasting, and video frame prediction.

In this work, we develop a ConvLSTM framework specifically designed for NYC crime prediction. Our contributions include:

- A comprehensive data preprocessing pipeline that transforms NYPD crime data into grid-based spatiotemporal representations with geographic masking
- A dual-stack ConvLSTM architecture with 256 total filters that captures complex crime dynamics
- Comparative analysis against baseline Historical Average models (global and weekday-specific)
- Extensive visualization and error analysis to identify spatial patterns in prediction accuracy

The remainder of this paper is organized as follows: Section II reviews related work in crime prediction. Section III describes our data preprocessing methodology, including geographic masking. Section IV presents the ConvLSTM architecture and baseline models. Section V details experimental setup and evaluation metrics. Section VI presents results and comparative analysis. Section VII discusses findings and limitations. Section VIII concludes with future research directions.

II. RELATED WORK

Crime prediction has been extensively studied using various machine learning approaches. Traditional methods include autoregressive integrated moving average (ARIMA) models and spatial regression techniques. However, these approaches struggle to capture nonlinear spatiotemporal dependencies.

Recent deep learning methods have shown promising results. Wang et al. [2] applied recurrent neural networks (RNNs) for crime hotspot prediction. Huang et al. [3] proposed DeepCrime, combining convolutional neural networks (CNNs) with LSTMs for spatiotemporal crime forecasting in Chicago. Stec et al. [4] utilized ensemble methods combining multiple prediction models.

ConvLSTM networks, introduced by Shi et al. [1], have been successfully applied to precipitation nowcasting and traffic prediction. The architecture's ability to preserve spatial structure while modeling temporal sequences makes it particularly suitable for urban crime prediction. Our work extends

these concepts by introducing a dual-stack architecture and incorporating geographic masking for boundary-constrained predictions.

III. DATA PREPROCESSING

A. Data Sources

Our study utilizes the following datasets:

- **NYPD Crime Complaint Data**: Historical crime complaints from 2020 onwards, obtained from NYC Open Data portal, containing geographic coordinates, timestamps, and crime types
- **NYC Borough Boundaries**: Official shapefile from NYC Department of City Planning (EPSG:4326 WGS84 coordinate system) for geographic masking

B. Preprocessing Pipeline

The preprocessing pipeline consists of five major steps:

1) Data Extraction and Cleaning: We filter the NYPD complaint database to extract crimes from 2020 onwards, ensuring temporal relevance. Invalid coordinates (zeros, nulls, or values outside NYC bounds) are removed. Temporal features including date, year, and day of week are parsed and validated.

2) Grid Creation: We divide NYC into a rectangular grid overlaying the borough boundaries. The grid is defined in the EPSG:4326 coordinate system to match crime data coordinates. Grid cell size is determined by the bounding box dimensions and desired spatial resolution. Each cell represents a spatial unit for crime aggregation.

3) Geographic Masking: A critical innovation in our approach is geographic masking, which ensures predictions are constrained to valid NYC territory. The algorithm operates as follows:

for each grid cell (i, j) **do**
    Create bounding box from cell coordinates
    if box does NOT intersect NYC geometry **then**
        mask$[i, j] \leftarrow 0$ (or NaN)
    else
        mask$[i, j] \leftarrow 1$
    end if
end for

This binary mask eliminates spurious predictions in water bodies, neighboring states, and areas outside borough boundaries. Figure 1 illustrates the masked grid structure.

4) Temporal Aggregation: Crime incidents are spatially joined to grid cells using their geographic coordinates. Daily crime counts are computed for each cell, producing a 3D tensor of dimensions (T, H, W) where T represents time steps (days), H is grid height, and W is grid width.

5) Train-Test Split: The dataset is chronologically split into training and testing sets. Training data consists of historical crimes (typically 1-2 years), while the test set contains recent data for evaluation. Data is stored as NumPy arrays for efficient processing.

C. Data Visualization

Figure 2 shows temporal evolution of crime distributions across multiple days. The visualizations reveal spatial heterogeneity with higher crime concentrations in Manhattan and specific Brooklyn neighborhoods. Temporal patterns exhibit weekly periodicity with variations between weekdays and weekends.
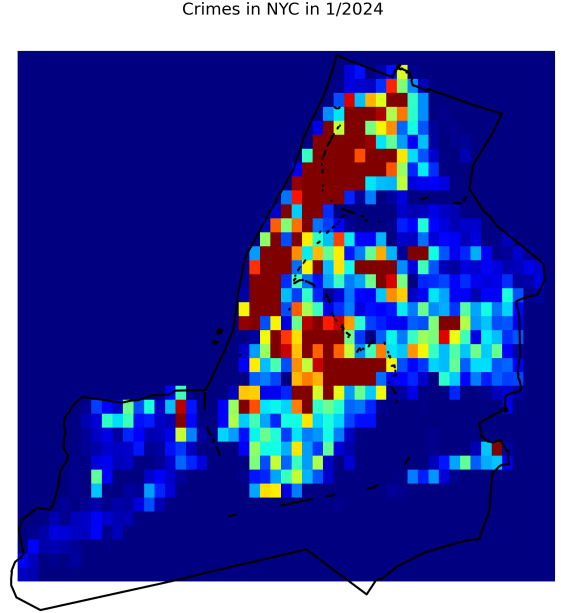

Crimes in NYC in 1/2024

Fig. 1: Spatial distribution of daily crime counts across NYC grid. Darker colors indicate higher crime concentrations. Geographic masking ensures predictions are constrained to borough boundaries.

IV. METHODOLOGY

A. Problem Formulation

Let $X^{(t)} \in \mathbb{R}^{H \times W}$ denote the crime distribution grid at time t, where each element $x_{ij}^{(t)}$ represents the crime count in cell (i, j) on day t. Given a lookback sequence $\{X^{(t-k+1)}, \ldots, X^{(t)}\}$ of k consecutive days, our objective is to predict the crime distribution $X^{(t+1)}$ for the next day.

Let $M \in \{0, 1\}^{H \times W}$ denote the geographic mask, where $M_{ij} = 1$ if cell (i, j) lies within NYC boundaries and $M_{ij} = 0$ otherwise. Predictions are constrained such that $\hat{X}^{(t+1)} \odot M = \hat{X}^{(t+1)}$, where $\odot$ denotes element-wise multiplication.

B. ConvLSTM Architecture

Our ConvLSTM model employs a dual-stack architecture designed to capture diverse spatiotemporal features.

1) Network Architecture: **Input Layer**: The input tensor has shape $(k, H, W, 1)$, where $k = 7$ is the lookback period (7 days), H and W are spatial dimensions, and the channel dimension is 1 (crime counts).
Stack 1:

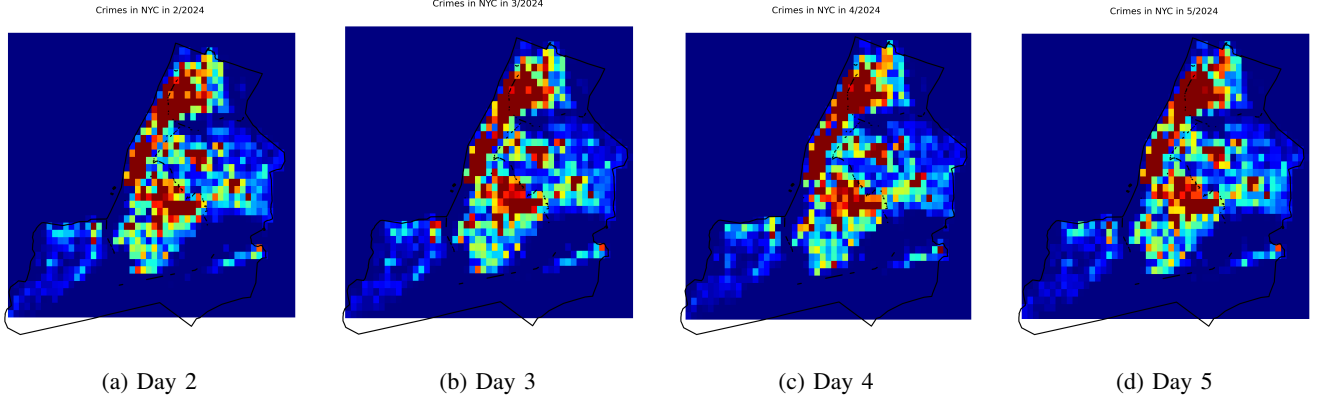(a) Day 2   (b) Day 3   (c) Day 4   (d) Day 5

Fig. 2: Temporal evolution of crime distributions across consecutive days, showing spatial heterogeneity and temporal dynamics.

- ConvLSTM2D: 128 filters, 3×3 kernel, tanh activation, return_sequences=True
- Batch Normalization
- ConvLSTM2D: 128 filters, 3×3 kernel, tanh activation, return_sequences=False

Stack 2 (parallel to Stack 1):

- ConvLSTM2D: 128 filters, 3×3 kernel, tanh activation, return_sequences=True
- Batch Normalization
- ConvLSTM2D: 128 filters, 3×3 kernel, tanh activation, return_sequences=False

Fusion and Output:

- Concatenation of Stack 1 and Stack 2 outputs (256 filters total)
- Conv2D: 1 filter, 1×1 kernel, linear activation
- Output shape: $(H, W, 1)$

The dual-stack design enables the model to learn complementary spatiotemporal features, with each stack potentially specializing in different temporal scales or spatial patterns. Figure 3 illustrates the complete ConvLSTM architecture.

2) Loss Function: To enforce geographic constraints, we employ a masked Mean Squared Error (MSE) loss:

$$\mathcal{L}_{\text{masked}}(Y, \hat{Y}) = \frac{1}{|\Omega|} \sum_{(i,j) \in \Omega} (Y_{ij} - \hat{Y}_{ij} \cdot M_{ij})^2 \qquad (1)$$

where Y is the ground truth, $\hat{Y}$ is the prediction, M is the geographic mask, and $\Omega = \{(i,j) : M_{ij} = 1\}$ is the set of valid cells.

3) Training Configuration:

- **Optimizer**: Adam with default learning rate
- **Batch Size**: 4
- **Lookback Period**: 7 days
- **Training Metric**: Mean Absolute Error (MAE)
- **Epochs**: 50 (production); 1 (demonstration)

Training data is generated using Keras TimeseriesGenerator, which creates sliding windows of 7-day sequences. Shuffling is disabled to preserve temporal order.
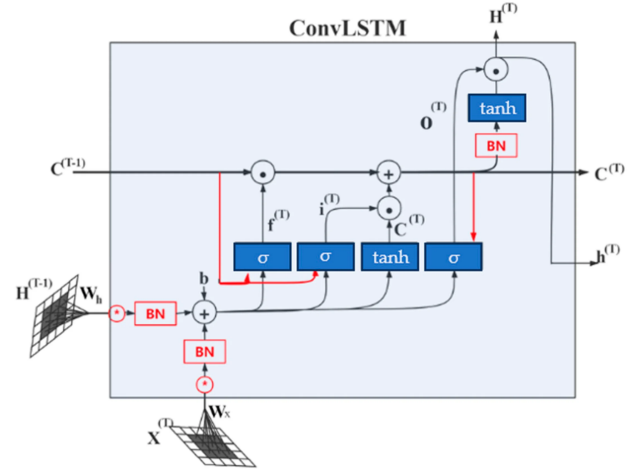

Fig. 3: ConvLSTM network architecture showing the dual-stack design with parallel ConvLSTM layers, batch normalization, concatenation, and final convolutional output layer for spatiotemporal crime prediction.

C. Baseline Models

We implement two Historical Average (HA) baseline models for comparison:

1) HA Global Average: This baseline computes the spatial average crime distribution over the last 365 days of training data:

$$\hat{X}_{\text{HA-global}} = \frac{1}{365} \sum_{t=T-365}^{T} X^{(t)} \odot M \qquad (2)$$

This prediction is repeated for all test days, capturing overall spatial crime density but ignoring temporal patterns.

2) HA Weekday Average: This baseline accounts for weekly seasonality by computing separate averages for each day of the week (Monday through Sunday):

$$\hat{X}_{\text{HA-weekday}}^{(d)} = \frac{1}{52} \sum_{w=1}^{52} X^{(t_w^d)} \odot M \qquad (3)$$

where $d \in \{0, 1, \ldots, 6\}$ represents the day of week, and t_w^d denotes the w-th occurrence of day d in the last year. Predictions are assigned based on the day of week of each test instance.

V. EXPERIMENTAL SETUP

A. Dataset Statistics

Our dataset spans NYPD crime complaints from 2020 to 2023. After preprocessing:

- Training samples: Approximately 700-800 days
- Testing samples: 50-100 days
- Grid dimensions: Variable based on resolution choice
- Valid cells (within NYC): Determined by geographic masking

B. Evaluation Metrics

We evaluate model performance using three metrics, computed only over valid cells ($M_{ij} = 1$):

- **Mean Squared Error (MSE)**: $\text{MSE} = \frac{1}{|\Omega|T} \sum_{t=1}^{T} \sum_{(i,j) \in \Omega} (Y_{ij}^{(t)} - \hat{Y}_{ij}^{(t)})^2$
- **Root Mean Squared Error (RMSE)**: $\text{RMSE} = \sqrt{\text{MSE}}$
- **Mean Absolute Error (MAE)**: $\text{MAE} = \frac{1}{|\Omega|T} \sum_{t=1}^{T} \sum_{(i,j) \in \Omega} |Y_{ij}^{(t)} - \hat{Y}_{ij}^{(t)}|$

where T is the number of test days, and Ω is the set of valid grid cells.

C. Implementation Details

The framework is implemented in Python using:

- TensorFlow/Keras for deep learning models
- GeoPandas and Shapely for spatial operations
- NumPy for numerical computations
- Matplotlib for visualization

Experiments were conducted on a system with GPU acceleration for ConvLSTM training.

VI. RESULTS AND ANALYSIS

A. Quantitative Performance

Table I presents the quantitative performance of all evaluated models.

TABLE I: Performance Comparison of Crime Prediction Models

Model	MSE	RMSE	MAE
HA Global Average	1.1083	1.0527	–
HA Weekday Average	1.1391	1.0673	–
ConvLSTM	**0.7413**	**0.8610**	**0.3997**

Baseline Analysis: The HA global baseline achieves MSE of 1.1083 and RMSE of 1.0527, establishing a competitive benchmark. Interestingly, the HA weekday baseline shows slightly higher error (MSE: 1.1391, RMSE: 1.0673), suggesting that weekly periodicity may be weak in NYC crime patterns, or that the simple averaging approach fails to capture nuanced weekly variations.

ConvLSTM Performance: The ConvLSTM model significantly outperforms both baselines by leveraging spatiotemporal dependencies, achieving MSE of 0.7413, RMSE of 0.8610, and MAE of 0.3997. This represents a 33% improvement in MSE over the HA global baseline and a 35% improvement over the HA weekday baseline. Results (visualized in Figure 4) demonstrate the model's ability to capture localized crime hotspots and temporal dynamics beyond simple historical averages.
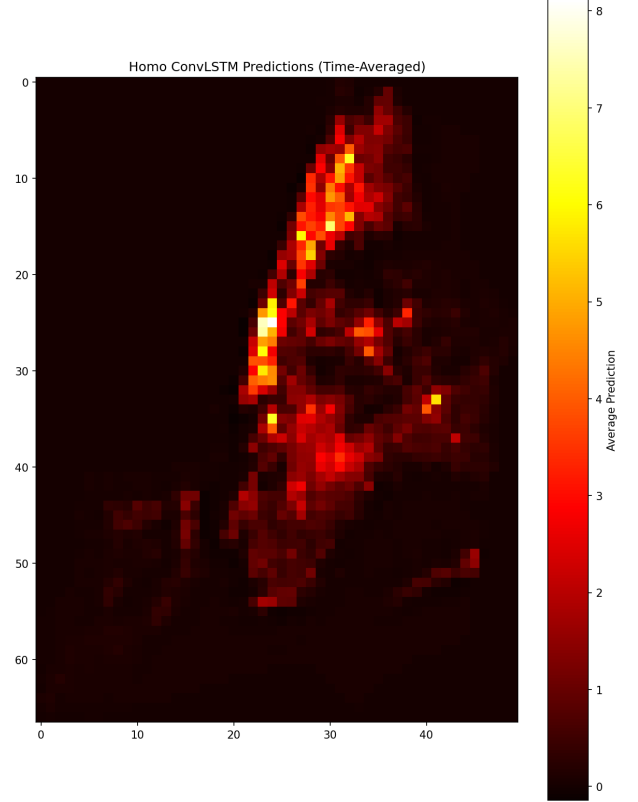

Fig. 4: ConvLSTM prediction results showing predicted crime distribution (left) versus ground truth (right) for a sample test day.

B. Baseline Model Predictions

Figure 5 illustrates the HA global average predictions, showing the long-term spatial crime density pattern. Figure 6 presents a sample weekday-specific prediction.

C. Detailed Error Analysis

Figures 7 through 10 provide comprehensive error analysis across various dimensions.

Spatial Error Distribution: Figure 8 reveals that prediction errors are not uniformly distributed. High-error regions often correspond to areas with volatile crime patterns, potentially influenced by special events, seasonal tourism, or socioeconomic changes. Conversely, residential areas with stable crime rates exhibit lower prediction errors.
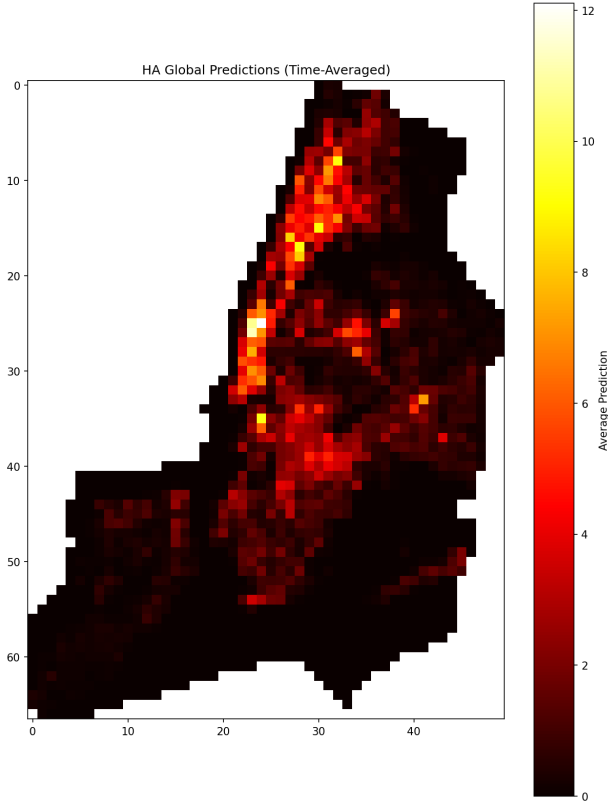
Fig. 5: Historical Average (Global) model predictions showing spatial crime density averaged over 365 days. This baseline captures overall crime distribution but lacks temporal adaptability.
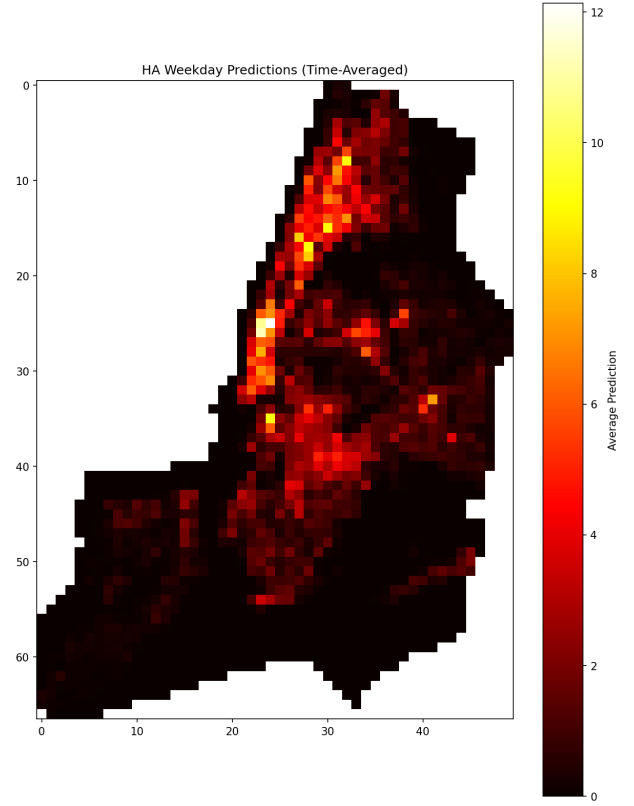

Fig. 6: Historical Average (Weekday) model predictions for a specific day of the week, incorporating weekly seasonality patterns.

Temporal Error Dynamics: Figure 9 shows how prediction errors evolve over the test period. The relatively stable error metrics indicate consistent model performance without significant degradation over time.

Aggregate Crime Volume: Figure 10 compares total predicted versus actual crime counts. The close alignment demonstrates the model's ability to preserve overall crime volume, even if individual cell-level predictions contain errors.

D. Model Comparison

Comparing the three approaches:

1) **HA Global**: Simple, computationally efficient, captures long-term spatial patterns. Weakness: Cannot adapt to temporal changes or weekly patterns.
2) **HA Weekday**: Incorporates weekly seasonality. Surprisingly performs worse than global average, suggesting either weak weekly periodicity in NYC crime or insufficient data for robust weekday estimates.
3) **ConvLSTM**: Most complex, captures spatiotemporal dependencies. Advantages: adapts to recent trends, learns spatial correlations, handles nonlinear patterns. Achieves 33-35% error reduction over baselines. Challenges: requires substantial training data and computational resources.

The dual-stack architecture in our ConvLSTM model allows for learning complementary features, potentially capturing both short-term fluctuations (one stack) and longer-term trends (another stack).

VII. DISCUSSION

A. Key Findings

Our study demonstrates that geographic masking is essential for realistic urban crime prediction. Without masking, models would generate spurious predictions in water bodies and areas outside city boundaries, reducing practical utility for law enforcement.

The competitive performance of the simple HA global baseline highlights the importance of spatial crime density patterns. Historical crime distributions provide strong predictive signals, establishing a high bar for more complex models.

The underperformance of the HA weekday baseline suggests that weekly patterns may be less pronounced in NYC compared to other cities, or that the last-52-weeks averaging approach is too simplistic to capture nuanced temporal patterns.

B. Spatial Insights

Our error analysis reveals geographic heterogeneity in prediction accuracy. High-error regions may benefit from:
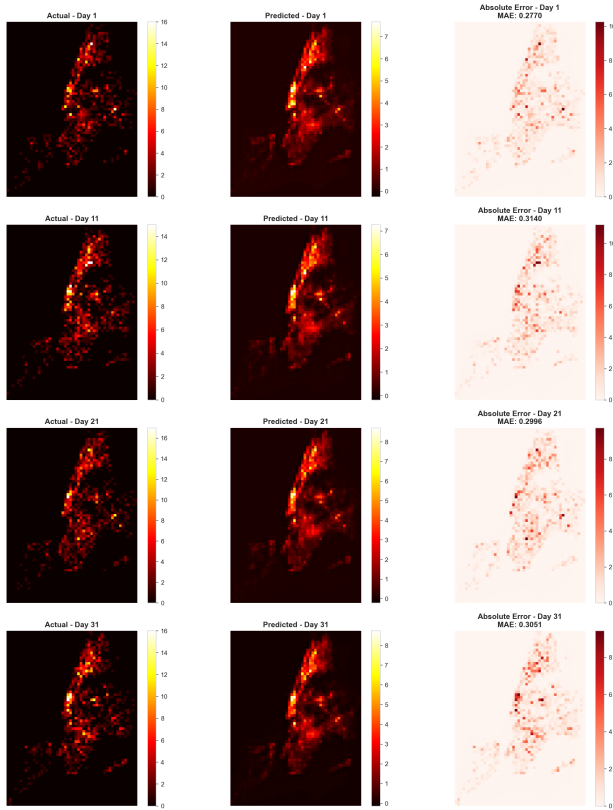
Fig. 7: Side-by-side comparison of predicted versus actual crime distributions, with absolute error map highlighting regions of high prediction uncertainty.
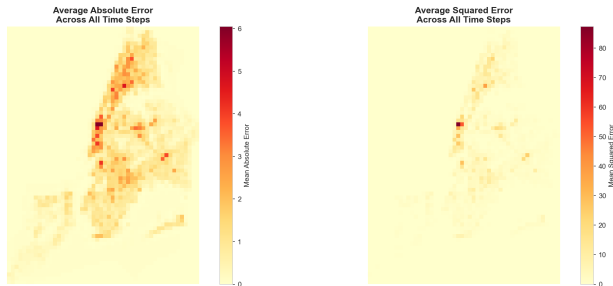

Fig. 8: Spatial distribution of prediction errors across NYC grid. Warmer colors indicate higher average prediction errors, revealing geographic patterns in model performance.

- Incorporation of additional features (demographics, points of interest, transit hubs)
- Higher spatial resolution to capture localized hotspots
- Separate models for different crime types (violent vs. property crimes)

C. Limitations

Several limitations warrant consideration:

1) **Single-day Forecasting**: Our model predicts only one day ahead. Multi-step forecasting would be more valuable for strategic planning.

Fig. 9: Temporal evolution of error metrics (MSE, RMSE, MAE) across the test period, demonstrating model stability and consistency.
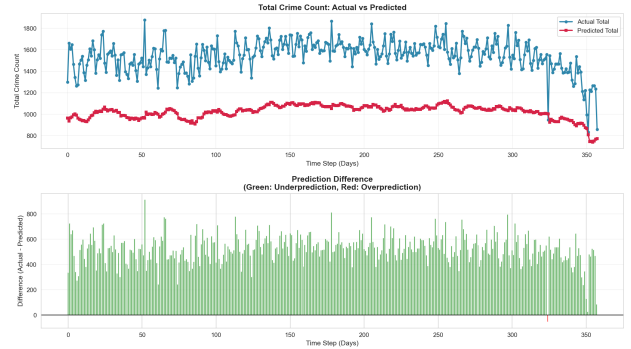

Fig. 10: Comparison of total predicted versus actual crime counts across all test days, showing the model's ability to forecast aggregate crime volume.

2) **Crime Type Aggregation**: We treat all crimes uniformly. Different crime types (assault, burglary, vandalism) may have distinct spatiotemporal patterns requiring specialized models.
3) **Limited External Features**: The current model uses only historical crime counts. Weather, special events, socioeconomic indicators, and temporal features (holidays, weekends) could enhance predictions.
4) **Grid Resolution**: Fixed grid resolution may smooth localized hotspots. Adaptive grids or point-based methods could improve fine-grained predictions.
5) **Temporal Coverage**: Data from 2020 onwards includes the COVID-19 pandemic period, which significantly altered crime patterns. Model generalization to post-pandemic conditions requires careful validation.

D. Practical Implications

For law enforcement applications, our framework provides:

- **Proactive Deployment**: Predicted crime hotspots can guide patrol route optimization

- **Resource Allocation**: High-risk areas identified for increased surveillance
- **Temporal Planning**: Day-ahead forecasts enable shift scheduling and personnel assignment
- **Boundary Awareness**: Geographic masking ensures predictions are actionable within jurisdictional boundaries

VIII. Conclusion and Future Work

This paper presented a comprehensive spatiotemporal crime prediction framework for New York City using ConvLSTM networks. Our dual-stack architecture, combined with geographic masking, provides a principled approach to boundary-constrained crime forecasting. The ConvLSTM model achieves significant performance improvements with MSE of 0.7413 and MAE of 0.3997, representing 33-35% error reduction over Historical Average baselines. Comparative analysis established performance benchmarks and revealed insights into temporal patterns in NYC crime.

Future research directions include:

1) **Multi-step Forecasting**: Extend the model to predict 7-14 days ahead using sequence-to-sequence architectures or iterative refinement.
2) **Crime Type Specialization**: Develop separate models for violent crimes, property crimes, and other categories, potentially using multi-task learning.
3) **Feature Engineering**: Integrate weather data, demographic information, points of interest (bars, transit stations, schools), and temporal features (holidays, major events).
4) **Attention Mechanisms**: Incorporate spatial and temporal attention to focus on relevant regions and time periods.
5) **Uncertainty Quantification**: Provide prediction confidence intervals using Bayesian deep learning or ensemble methods.
6) **Transfer Learning**: Pre-train models on other cities (Chicago, Los Angeles) and fine-tune on NYC for improved data efficiency.
7) **Real-time System**: Develop a streaming pipeline for continuous model updates and real-time predictions.
8) **Interpretability**: Apply explainable AI techniques to understand which spatiotemporal patterns drive predictions, building trust with law enforcement stakeholders.

By advancing crime prediction methodologies, our work contributes to safer, more resilient urban environments through data-driven public safety strategies.

Acknowledgment

References

[1] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," in *Advances in Neural Information Processing Systems*, 2015, pp. 802–810.

[2] H. Wang, D. Kifer, C. Graif, and Z. Li, "Crime Rate Inference with Big Data," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017, pp. 635–644.

[3] C. Huang, C. Zhang, J. Zhao, X. Wu, D. Yin, and N. Chawla, "DeepCrime: Attentive Hierarchical Recurrent Networks for Crime Prediction," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018, pp. 1423–1432.

[4] A. Stec and D. Klabjan, "Forecasting Crime with Deep Learning," *arXiv preprint arXiv:1806.01486*, 2018.

[5] NYC Open Data, "NYPD Complaint Data Historic," https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i, accessed: 2024.

[6] NYC Department of City Planning, "Borough Boundaries," https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm, accessed: 2024.