Crime Forecasting with Satellite Imagery

He Zhou
zhou1354@umn.edu
University of Minnesota

1 INTRODUCTION

The goal of this project is to forecast likely places of future crime scenes, that is, to predict when and where a certain crime type will occur, using past data. Stalidis et al. [13] presents a detailed study on crime classification and prediction using deep learning architectures. Najjar et al. [9] attempts to predict crime rate directly from raw satellite imagery and presents a proof-of-concept study on predicting crime rate from raw satellite imagery. Motivated by these two papers, this project hopes to study crime classification and prediction with the additional support of satellite imagery.

1.1 Related work

The prediction of hotspots, which are areas of varying geographical size with high crime probability, has been studied widely in the literature. There are several mainstream methods for crime hotspot forecasting, including Spatial and Temporal Analysis of Crime [7], Thematic Mapping [15], Kernel Density Estimation [1] and machine learning [2, 5, 10, 13, 14].

Stalidis et al. [13] examine the effectiveness of deep learning algorithms in hotspot forecasting using open data from police reports. They fed their deep learning (DL) methods with the minimum amount of data containing only spatial, temporal, and crime type information. Their spatial and then the temporal (SFTT) model configuration is selected as the winning one, compared with 10 different algorithms in 5 crime incidence datasets. For the models to get a better hotspot classifier, they simultaneously predicted the number of crimes that occurred within the same area in the same future window.

Najjar et al. [9] uses Convolutional Neural Networks (CNNs) to predict crime rate from raw satellite imagery. Their best model could predict the crime rate with an accuracy of 79%. Their results confirmed that visual features contained in satellite imagery can be effectively used as a proxy indicator of crime rate.

As mentioned in the *Conclusions and Future Work* Section of Stalidis et al. [13]'s paper, it remains to investigate if incorporating additional information, like temporal semantics, demographics, street maps and so on can help the model to learn better future. Therefore, this project tries to investigate the additional support of satellite imagery for better hotspot forecasting.

2 APPROACH

2.1 Problem Formulation

We have two sources of data, the crime incident report, and the satellite imagery.

Let D be the dataset of n vectors x_i with $i \in \{1, \cdots, n\}$ where each x_i contains information on the location (given as longitude and latitude), the time, and the type of a reported crime. Following the approach by Lin et al. [8], the city map is split into a two-dimensional grid with cell edge size ℓ, and therefore a $H \times W$ cells grid is produced. Then given a time window of duration t, the data points are aggregated by counting the number of occurrences of each crime type within each cell in this time window. Different crime types are put in separate channels of the incident map I. Considering the whole period T of the time series, multiple aggregated incident maps are generated and will be used as the dynamic input for the DL model.

For each cell, a satellite image centered around the centroid of the cell is obtained from Google Static Maps API (https://developers.google.com/maps/documentation/maps-static). The spatial resolution is set to 256×256 pixels each with zoom level 17. The raw satellite imagery will be used as the static input for the DL model.

Our goal is to classify each cell as a hotspot or not for a certain crime type in a future time period using the multiple incident maps from the past time period as well as the static satellite imagery. To enhance the hotspot probability, we follow the idea by Stalidis et al. [13] and use secondary parallel prediction of the number of occurrences y for the corresponding crime type.

2.2 Proposed Methodology

CNNs is a deep feedforward neural network that is capable of capturing patterns and statistics at multiple levels of abstraction [9]. Long-Short Term Memory (LSTM) networks [4] are a variant of recurrent neural networks (RNNs) that are better suited to long sequences since they could mitigate the vanishing gradient problem. We will combine the CNNs and RNNs for a spatio-temporal forecasting model.

We first use a spatio-temporal model to extract dynamic features from the multiple incident maps. The model configuration follows the winning one in [13], named SFTT (Figure 1). For every timespan t, the corresponding incident map is passed into a CNN submodel to produce a feature vector that encodes the spatial distribution of incidents with a much lower dimension than the origin. Then for the whole time-series period T, the sequence of T/t feature vectors is fed into the LSTM network to extract the temporal features.

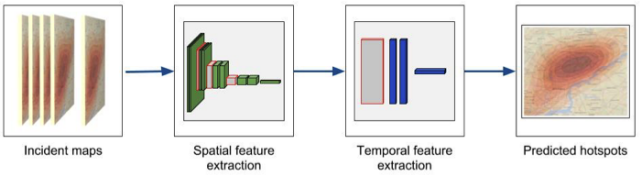


Figure 1: Overview of the sequence of dynamic feature extraction: the spatial features first then the temporal (SFTT) [13].

The CNN submodel is constructed based on the VGGNet [12] architecture (Figure 2), as suggested in [13]. In the VGGNet body, there are a series of 5 pairs of convolution layers. The first 3 are

followed each by a *Max pooling* layer as well as a *Batch Normalization* layer and a *Dropout* layer to avoid the possibility of overfitting. The first pair of convolution layer applies 32 filters of size 3×3, the second pair applies 64 filters of size 3×3, the third pair applies 128 filters of size 3×3, the fourth pair applies 256 filters of size 3×3, and the last pair applies 256 filters of size 1×1. In all layers, the activation function is the *Rectified Linear Units (ReLU)*, which could partly mitigate the vanishing gradient problem [12].

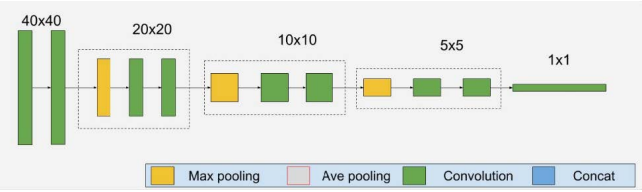


Figure 2: VGGNet body architecture for spatial information extraction [13].

The sequence of vectors containing the spatial features is then fed into an LSTM submodel (Figure 3) to further extract the temporal correlations. The stacked LSTMs model consists of 3 consecutive LSTM layers. The first two layers produce hidden outputs of dimension 500, and the last layer produces outputs of dimension $H * W$, i.e., the number of cells that the model predicts. Thus, for each cell, we generate a dynamic vector containing both the spatial and the temporal features from the incident maps.

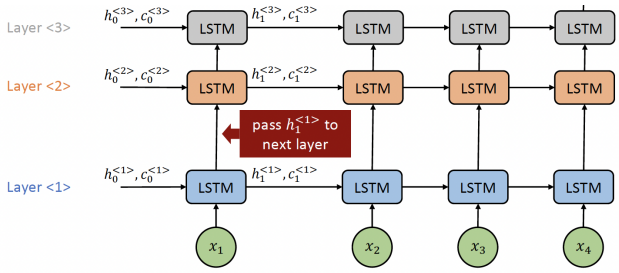


Figure 3: Stacked LSTMs [4] (Figure credit: https://yaoyichi.github.io/spatial-ai.html).

The next parallel step is to extract the static spatial features from the raw satellite imagery. Najjar et al. [9] built their classification model by finetuning a pre-trained AlexNet model [6]. So we will also use the pre-trained AlexNet model to extract the spatial feature from the static satellite imagery. By removing the last source classification layer, the pre-trained AlexNet extracts a 1000-dimensional vector containing the static spatial features of each cell from the corresponding satellite image.

Finally, the dynamic spatio-temporal feature vector from the VGGNet-LSTM model and the static spatial feature vector from the pre-trained AlexNet model are concatenated together to form a whole feature vector representing each cell. The extracted features are finally fed into two parallel fully connected output layers for classification and prediction. In one output layer, the binary cross-entropy (*BCE*) loss function is used for the hotspot classification

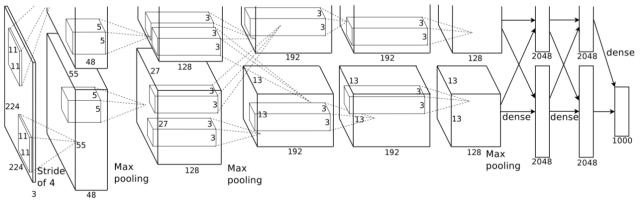


Figure 4: AlexNet [6] as static feature extractor.

task:

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}[\hat{o}_i \log o_i + (1-\hat{o}_i)\log(1-o_i)] \tag{1}$$

In the other output layer, the mean squared error (*MSE*) loss function is used for predicting the number of crimes:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{2}$$

where $N = H * W$ is the total number of cells, y_i and $\hat{y}_i$ are the actual and the predicted number of crimes in the cell respectively. The total loss combines the *BCE* loss and *MSE* loss:

$$Loss = BCE + \lambda \cdot MSE \tag{3}$$

where λ is the regularization parameter for *MSE*. When $\lambda = 0$, we only care about the classification loss. When $\lambda > 0$, the prediction loss takes into effect and drives the classifier to give bigger probabilities to cells where multiple crimes are predicted to occur.

3 EXPERIMENTS

3.1 Dataset

In this project, we use the San Francisco Department Incident Reports (https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783) from the year 2018 to the present. The dataset contains 47 crime categories and are mapped to 10 common crime types, shown in Table 1. The cell edge size ℓ is approximately 450 m and results in grids of 40×40 cells. The timespan t is set to be 1 day and the time period T is set to be 30 days. Thus, 30 daily incident maps are generated from the incident report dataset to forecast hotspots for the next month. There are in total 52 one-month periods from the year 2018 to the present. The first 39 periods are used for training and the last 13 periods are used for testing. Testing is performed by making predictions for every month and the averaged scores are reported.

Some basic statistics from the incident report dataset are presented in Table 2. The second column contains the number of incidents for each crime type from the year 2018 to the present. The third column contains the average number of hotspots per day for every crime type. From these statistics, we could observe a class imbalance problem.

Figure 5 shows the satellite images of four different cells. Figure 5a and 5b correspond to two cells with the top two number of incidents reported, while Figure 5c and 5d correspond to two cells with only one incident reported. One can observe visual similarity between high-risk areas and between low-risk areas, and visual dissimilarity between high-risk areas and low-risk areas.

Table 1: Mapping of 47 Crime Categories to the 10 Crime Types [13].

Crime Type	Crime Category
Assault	Assault Disorderly Conduct Offences Against The Family And Children Weapons Carrying Etc Weapons Offence Weapons Offense
Theft	Larceny Theft Malicious Mischief Stolen Property Vandalism
Robbery	Robbery
Burglary	Burglary
Motor Vehicle	Motor Vehicle Theft Motor Vehicle Theft? Recovered Vehicle Vehicle Impounded Vehicle Misplaced
Arson	Arson Fire Report
Homicide	Homicide
Vice	Human Trafficking (A), Commercial Sex Acts Human Trafficking (B), Involuntary Servitude Human Trafficking, Commercial Sex Acts Prostitution Sex Offense Rape
Narcotics	Drug Offense Drug Violation
Other	...

Table 2: Basic statistics from incident report dataset: *Column 2:* Number of Incidents for each crime type; *Column 3:* Mean number of hotspots per day for each crime type.

Crime Type	Num. of Incidents	Mean Num. of hotspots
Assault	59,773	27.78
Theft	202,506	93.77
Robbery	13,106	7.33
Burglary	32,816	17.99
Motor Vehicle	45,034	25.67
Arson	2,464	1.46
Homicide	64	0.04
Vice	1,791	0.79
Narcotics	13,564	4.48
Other	143,169	62.57
Total	514,287	# CELLS 951

3.2 Metrics

the metric of accuracy is not very trustful when dealing with data that have a class imbalance. So we draw the *ROC* (Receiver Operating Characteristic) curve and the *PR* (Precision-Recall) curve, and report the *AUROC* (Area under *ROC* curve) score and the *AP* (average precision) score that summarize the performance of a classifier for all possible probability thresholds.

(a) Cell 985 with 9, 194 incidents **(b) Cell 1028 with 7, 779 incidents**

(c) Cell 1170 with 1 incidents **(d) Cell 1135 with 1 incidents**

Figure 5: Satellite images of four cells: (a) (b) with top two number of incidents reported and (c) (d) with only one incident reported.

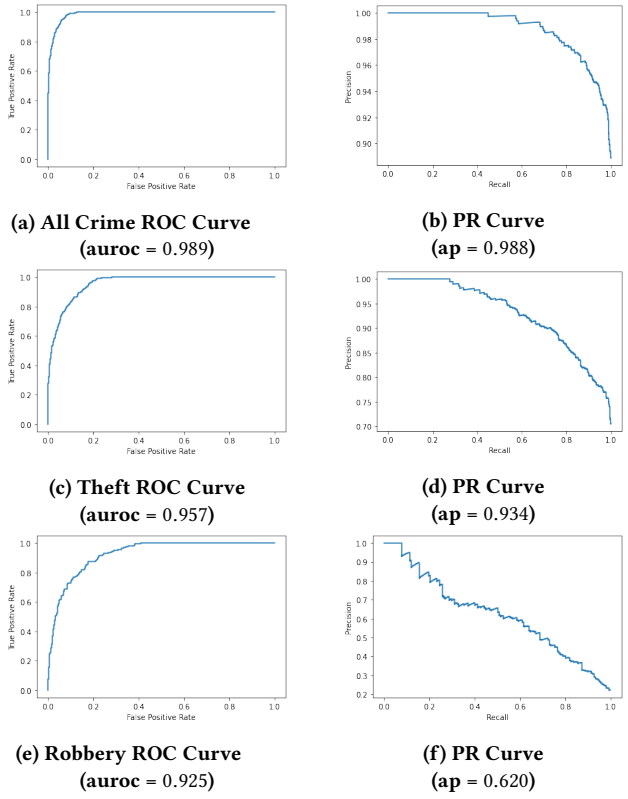
3.3 Results

In Figure 6, we draw the *ROC* curves and *PR* curves, and report the corresponding *AUROC* and *AP* scores for forcasting *All Crimes*, *Theft*, and *Robbery* saperately. That is, we are interested in classifying each cell as the hotspot for a certain crime type or not. From the results, we can see that the classifier from our proposed model performs the best when forecasting *All Crimes* as a whole, with *AUROC* score equal to 0.989 and *AP* score equal to 0.988. Around 40% of the incidents are thefts (shown from Table 2), so the classifier also performs well when forecasting *Theft*, with *AUROC* score equal to 0.957 and *AP* score equal to 0.934. The classifier is capable of achieve high precision and high recall at the same time. But when forecasting crime types that do not happen frequently, for exampe, *Robbery*, the classifier does not perform well with *AUROC* score equal to 0.925 and *AP* score equal to 0.620. We can observe from the *PR* curve (Figure 6f) that the classifier can not achieve high precision and recall at the same time.

In Figure 7, we draw the true hotspots (Figure 7a, 7b), predicted hotspots (Figure 7c, 7d) with probability threshold 0.5, and the predicted hotspot probability (Figure 7e, 7f) for forecasting *Theft* and *Robbery* crime types. Comparing the true hotspots and predicted hotspots, one observes that the classifier makes some false positive errors when forecasting *Theft* hotspots, and misses a lot of hotspots when forecasting *Robbery* hotspots. This shows the limitation of our classifier when encountering class imbalance problem.

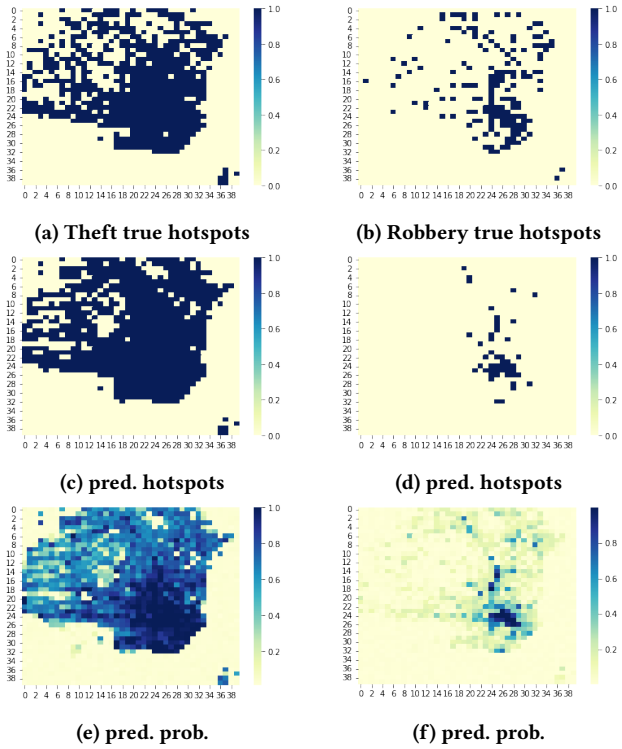
4 DISCUSSION & FUTURE WORK

In this project, we investigated the capability of DL method to forecast hotspot areas with the additional help of satellite imagery. We



(a) All Crime ROC Curve (auroc = 0.989)
(b) PR Curve (ap = 0.988)
(c) Theft ROC Curve (auroc = 0.957)
(d) PR Curve (ap = 0.934)
(e) Robbery ROC Curve (auroc = 0.925)
(f) PR Curve (ap = 0.620)

Figure 6: *ROC* curves and *PR* curves with *AUROC* scores and *AP* scores for forcasting *All Crimes*, *Theft* and *Robbery* crime types.



(a) Theft true hotspots
(b) Robbery true hotspots
(c) pred. hotspots
(d) pred. hotspots
(e) pred. prob.
(f) pred. prob.

Figure 7: True hotspots (a) (b), predicted hotspots (c) (d) with probability threshold 0.5, and predicted hotspot probability (e) (f) for forecasting *Theft* and *Robbery* crime types.

use the VGGNet-LSTM model to extract dynamic spatial and temporal features from the incident maps, and a pre-trained AlexNet to extract static spatial features from the satellite imagery. Forecasting hotspots for *All Crimes* performs better than specific crime types because of the class imbalance problem.

In the future, we will design more experiments to quantitatively investigate the use of satellite imagery in helping the prediction performance. What's more, Instead of using the VGGNet-LSTM model suggested by Stalidis et al. [13], the ConvLSTM model [11] may be more reasonable for extracting the dynamic spatio-temporal features simultaneously from the incident maps. There are also ethical issues with crime forecasting, which are discussed in detail by Gstrein et al. [3].

REFERENCES

[1] Richard A Davis, Keh-Shin Lii, and Dimitris N Politis. 2011. Remarks on some nonparametric estimates of a density function. In *Selected Works of Murray Rosenblatt*. Springer, 95–100.

[2] Mingchen Feng, Jiangbin Zheng, Jinchang Ren, Amir Hussain, Xiuxiu Li, Yue Xi, and Qiaoyuan Liu. 2019. Big data analytics and mining for effective visualization and trends forecasting of crime data. *IEEE Access* 7 (2019), 106111–106123.

[3] Oskar Josef Gstrein, Anno Bunnik, and Andrej Zwitter. 2019. Ethical, legal and social challenges of Predictive Policing. *Católica Law Review, Direito Penal* 3, 3 (2019), 77–98.

[4] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[5] Keivan Kianmehr and Reda Alhajj. 2008. Effectiveness of support vector machine for crime hot-spots prediction. *Applied Artificial Intelligence* 22, 5 (2008), 433–458.

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25 (2012).

[7] Ned Levine and I CrimeStat. 2002. A spatial statistics program for the analysis of crime incident locations. *Ned Levine and Associates, Houston, TX, and the National Institute of Justice, Washington, DC* (2002).

[8] Ying-Lung Lin, Meng-Feng Yen, and Liang-Chih Yu. 2018. Grid-based crime prediction using geographical features. *ISPRS International Journal of Geo-Information* 7, 8 (2018), 298.

[9] Alameen Najjar, Shun'ichi Kaneko, and Yoshikazu Miyanaga. 2018. Crime mapping from satellite imagery via deep learning. *arXiv preprint arXiv:1812.06764* (2018).

[10] Andreas M Olligschlaeger. 1997. Artificial neural networks and crime mapping. *Crime mapping and crime prevention* 1 (1997), 313.

[11] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems* 28 (2015).

[12] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[13] Panagiotis Stalidis, Theodoros Semertzidis, and Petros Daras. 2021. Examining deep learning architectures for crime classification and prediction. *Forecasting* 3, 4 (2021), 741–762.

[14] Bao Wang, Duo Zhang, Duanhao Zhang, P Jeffery Brantingham, and Andrea L Bertozzi. 2017. Deep learning for real time crime forecasting. *arXiv preprint arXiv:1707.03340* (2017).

[15] D Williamson, S McLafferty, P McGuire, T Ross, J Mollenkopf, V Goldsmith, and S Quinn. 2001. Tools in the spatial analysis of crime. Mapping and analysing crime data. *A. Hirschfield and K. Bowers. London and New York, Taylor & Francis* 1 (2001), 187.