This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Unsupervised Domain Adaptation for Crime Risk Prediction Across Cities

Binbin Zhou, Longbiao Chen, Sha Zhao, *Member, IEEE*, Shijian Li, Zengwei Zheng, and Gang Pan, *Member, IEEE*

***Abstract*—Crime risk prediction is crucial for city safety and residents' life quality. However, without labeled data, it is challenging to predict crime risk in cities. Due to municipal regulations and maintenance costs, it is not trivial for many cities to collect high-quality labeled crime data. In particular, some cities have lots of labeled data while others may have few. It has been possible to develop a crime prediction model for a city without labeled crime data by learning knowledge from a city with abundant data. Nevertheless, the inconsistency of relevant context data between cities exacerbates the difficulty of this prediction task. To this end, this article proposes an effective unsupervised domain adaptation model (UDAC) for crime risk prediction across cities while addressing the contexts' inconsistency issue. More specifically, we first identify several similar source city grids for each target city grid. Based on these source city grids, we then construct auxiliary contexts for the target city, to make contexts consistent between the two cities. A dense convolutional network with unsupervised domain adaptation is designed to learn high-level representations for accurate crime risk prediction and simultaneously learn domain-invariant features for domain adaptation. The effectiveness of our model is verified through extensive experiments using three real-world datasets.**

***Index Terms*—Crime prediction, crime risk, unsupervised domain adaptation.**

I. Introduction

CRIME continuously threatens urban safety and undermines citizens' life quality. According to [1], there have been 435 mass shooting events happened in the United States during the year 2019, resulting in 517 dead, 1648 wounded, severe property loss, and inestimable grief. Thus, sensing crime risk is important for individuals and society, to prevent and reduce potential crime events. Fortunately, the availability of various urban data in some cities (e.g., Chicago) fosters unprecedented opportunities for researchers to explore crime-related problems, such as crime hotspot detection [2], [3], crime classification [4], [5], [6], crime rate inference [7], [8], and crime count prediction [4], [5], [9], [10], [11]. An amount of urban data has been investigated to be helpful for performance improvement of crime-related studies. For example, the occurrence of crime events may be affected by human mobility, that more crowd of people may bring an increasing possibility of larceny.

Nevertheless, due to the uneven development level of cities, a number of cities do not disclose data to the public for some possible reasons, i.e., the high cost of data collection and maintenance, the absence of clear-cut regulations, and increasing privacy concerns. Thus, residents need sufficient experience to sense whether there will be risk. But not all residents have such local experience, and this brings more challenges for newcomers, e.g., tourists. Recently, transfer learning [12], [13] provides a new paradigm that enables us to use learned knowledge from a data-rich city (*source city*) to solve similar tasks in a data-scarce city (*target city*), e.g., chain store site recommendation and crowd flow prediction [14], [15]. Therefore, we attempt to resort to unsupervised transfer learning to explore crime risk prediction in cities without labeled crime data.

However, even though adequate labeled data can be collected from a source city, a prediction model trained using these data may fail to predict crime risk in the target city without labeled crime data. Different data collection capabilities may result in inconsistencies in the available relevant context data in different cities. Suppose that the source city is New York City (NYC), and the target city is Los Angeles (LA). Due to its widespread deployment of detection equipment and long-standing open data project, NYC has collected multisource urban data over the past many years and continues to disclose information to the public, e.g., point of interest (POI) distribution and taxi trip records. Some cities also collect many urban data. But due to some concerns on privacy issues or high data collection costs, they do not make some useful and relevant data available to the public, such as taxi mobility data in cities like LA. Thus, the contexts' inconsistency issue hinders crime risk prediction performance in cities with unlabeled data.

An intuitive approach to solve contexts' inconsistency is to only use common context data to train a model from the source city and then fine-tune this model to solve tasks in the target city, leaving inconsistent city-specific context data alone. But this may lose some useful information for crime risk prediction, and even worse when context data are sparse.

Manuscript received 7 February 2022; revised 21 July 2022; accepted 15 September 2022. This work was supported by the National Natural Science Foundation of China under Grant 62102349 and Grant 61925603. *(Corresponding author: Gang Pan.)*

Binbin Zhou is with the Department of Computer Science and Computing, Zhejiang University City College, Hangzhou 310015, China, and also with the College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China (e-mail: bbzhou@zucc.edu.cn).

Longbiao Chen is with the Fujian Key Laboratory of Sensing and Computing for Smart City, Xiamen University, Xiamen 361005, China (e-mail: longbiaochen@xmu.edu.cn).

Sha Zhao, Shijian Li, and Gang Pan are with the College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China (e-mail: szhao@zju.edu.cn; shijianli@zju.edu.cn; gpan@zju.edu.cn).

Zengwei Zheng is with the Department of Computer Science and Computing, Zhejiang University City College, Hangzhou 310015, China (e-mail: zhengzw@zucc.edu.cn).

Digital Object Identifier 10.1109/TCSS.2022.3207987

Due to the prevalence of deep learning technologies, more data may bring better performance since they can extract more useful information in the networks. Therefore, we try to design an effective model that can construct possible source-city-specific context data for the target city to address the contexts' inconsistency problem, and then predict crime in the target city leveraging learned knowledge from the source city.

In this article, we propose an unsupervised domain adaptation model (UDAC) for crime risk prediction across cities while addressing the contexts' inconsistency issue, called UDAC. The objective of UDAC is to solve crime risk prediction in a target city without labeled crime data, by transferring knowledge learned from a source city with abundant labeled data to the target city. Inspired by [15], we design a method to construct possible city-specific context data for the target city grids, based on context data in similar source city grids, to address the issue of inconsistent context. We then present a network to learn effective features for crime risk prediction in the source city and learn domain-invariant features simultaneously for the success of unsupervised domain adaptation. The optimization process would consider three elements simultaneously, i.e., crime risk prediction error, domain classification error, and distribution discrepancy distance. Extensive experiments are conducted to verify the effectiveness of UDAC using three real-world datasets from NYC, Chicago, and LA.

The main contributions of our work are as follows.

1) Our work is a promising step toward unsupervised domain adaptation in crime prediction across cities, while simultaneously addressing the contexts' inconsistency issue between cities.
2) We propose an effective model for crime risk prediction in cities without labeled data, which can facilitate deep unsupervised domain adaptation method leveraging knowledge learned from a source city with abundant labeled data. To address the inconsistent contexts between two cities, we first construct city-specific contexts for the target city, and then present a dense convolutional network to learn effective features for accurate crime prediction and domain-invariant features for unsupervised domain adaptation. The optimized network can be feasible for crime prediction in the target city.
3) We conduct extensive experiments to illustrate the effectiveness of our proposed UDAC model using real-world datasets from three cities. The experimental results show that our strategy outperforms the state-of-the-art comparison methods.

The rest of this article is organized as follows. We begin by reviewing previous studies in Section II. Section III presents the problem formulation and an overview of our proposed framework. In Section IV, we describe the three phases of this framework in detail. Extensive evaluation results are demonstrated in Section V. Finally, we conclude this article and chart future plans in Section VI.

II. Related Work

Our work is related to previous studies on crime prediction and unsupervised domain adaptation. In this section, we briefly introduce some related work from these two categories.

A. Crime Prediction

There have been a few studies on urban crime prediction in the past decades. Identifying relevant external features for crime prediction study is significant. Ranson [16] analyzed meteorological data and found that these data may be relevant to crime, e.g., weather and temperature information. Zhou *et al.* [17] conducted a fine-grained study to understand crime leveraging various urban data, including meteorological data, POI distribution, and taxi trips' data. They found valuable correlation between these data and crime.

With features analyzed, designing effective models to achieve accurate prediction has been popular. A lot of spatio-temporal prediction models have been proposed over the past few decades capturing spatial and temporal dependencies to solve various tasks, such as traffic prediction and inference [18], [19], [20], social event prediction [21], air quality prediction [22], and logistics management optimization [23]. However, these spatio-temporal prediction models paid little effort on the prediction tasks without labeled data. For crime prediction, a large amount of data, e.g., taxi trip, Twitter, demographic, and Foursquare data, have been used in various methods (e.g., linear models, count models, and machine learning models) to improve prediction performance [9], [24], [25]. Huang *et al.* [4] proposed a hierarchical recurrent neural network with an attention layer to capture dynamic patterns and learn temporal relevance for future crime occurrences' prediction, using crime data, POI, and 311 public service complaint data. Yang *et al.* [3] leveraged Twitter and POI data into multiple machine learning models (e.g., random forest and decision tree) to predict crime hotspots in NYC. Yi *et al.* [26] proposed an integrated model using a clustered continuous conditional random field (CCRF) method to extract spatio-temporal features and improve future crime prediction performance. They further incorporated long short-term memory (LSTM) units into the aforementioned CCRF method to learn nonlinear relationship between the input and the output, and stacked denoising autoencoder to learn pairwise interactions between spatial regions [10]. Zhou *et al.* [27] proposed a hierarchical framework for road-level crime prediction, which first established a pattern using spatio-temporal features to estimate crime prior knowledge and then update crime prediction results incorporating recurrence crime features. They further investigated crime dynamics from the perspective of influence propagation and proposed a zero-inflated negative binomial regression model to predict future road-level crime risk [28].

All these relevant studies paid efforts to solve prediction tasks with adequate labeled data and external urban data, with little effort on prediction with unlabeled data. For crime prediction, predicting crime risk in cities without labeled crime data is also significant for citizens and society.

B. Unsupervised Domain Adaptation

Domain adaptation can help improve the crime prediction performance when predicting crime in cities with few crime data since it can learn knowledge from the source domain and apply it to solve new tasks in the target city. For cities without releasing crime data, unsupervised domain adaptation

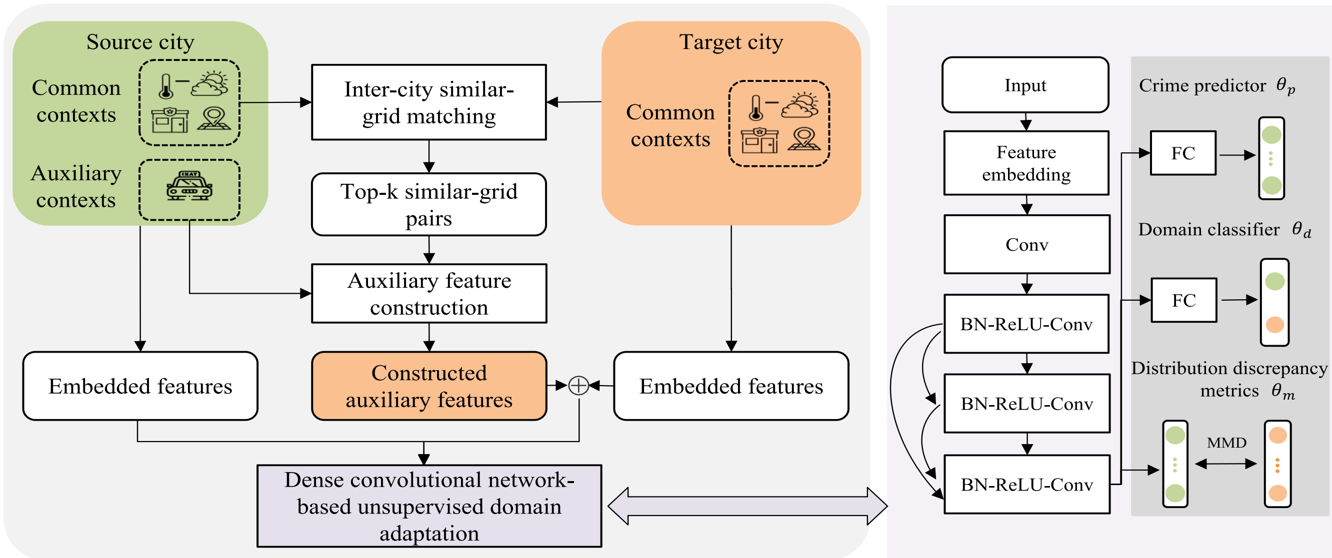


Fig. 1. Overview of the framework.

can play significant roles. It has been widely applied in many areas, such as image recognition, fault diagnosis cellular traffic prediction, transaction fraud detection, vehicle positioning, and POI recommendation [14], [15], [29], [30], [31], [32], [33], [34], [35]. Yao *et al.* [36] investigated spatial–temporal prediction problems in cities with only a few data. They proposed an effective meta-learning based spatial–temporal network using long period data from multiple cities. The experiments on two typical tasks, i.e., traffic and water quality prediction, verified the effectiveness of their proposed model. Zhao and Tang [37] exploited transferring crime knowledge learned from one borough to other boroughs in the same city to improve prediction performance. A novel transfer learning framework had been proposed.

Inspired by previous studies, we choose the unsupervised domain adaptation framework to predict future crime risk in cities without ground-truth values. However, there still exist challenges to apply the existing methods directly in this problem, due to the contexts' inconsistency as aforementioned. So we attempt to design a novel method to construct domain-specific contexts to make contexts consistent from both the domains for further crime prediction in the target domain.

III. Problem Formulation and Framework Overview

A. Problem Formulation

In this article, we define a source city $\mathcal{S}$ with adequate labeled crime risk data while a target city $\mathcal{T}$ without any crime data. There are more types of relevant external urban data collected from city $\mathcal{S}$ than city $\mathcal{T}$. Thus, the context data from city $\mathcal{S}$ can be separated into two categories: common context data $C^{\mathcal{S}}$ and city $\mathcal{S}$-specific auxiliary context data $A^{\mathcal{S}}$. The common context data from city $\mathcal{T}$ can be represented as $C^{\mathcal{T}}$. For example, here, some urban data are available in both the cities $\mathcal{S}$ and $\mathcal{T}$, such as weather conditions, air temperature, POI distribution, and police station distribution. They can be treated as the common context data. Some urban data are available in city $\mathcal{S}$ and unavailable in city $\mathcal{T}$, such as taxi trip records, so that they can be treated as the city $\mathcal{S}$-specific context data. Both the cities are partitioned into equal-sized grids, which is denoted as $r^{\mathcal{S}}$ and $r^{\mathcal{T}}$, respectively. The crime risk data in city $\mathcal{S}$ and city $\mathcal{T}$ are defined as $D^{\mathcal{S}}$ and $D^{\mathcal{T}}$, respectively. Note that these values are numerical values representing crime counts.

We aim to design a prediction model leveraging historical Δt time durations' data in both the cities $\mathcal{S}$ and $\mathcal{T}$ to predict unobserved crime data $\tilde{D}^{\mathcal{T}}_{t+1}$ in the target city $\mathcal{T}$ at time duration $t+1$. Here, the historical data are defined as $\{D^{\mathcal{S}}, C^{\mathcal{S}}, A^{\mathcal{S}}, C^{\mathcal{T}}\}_{[t-\Delta t+1,t]}$, and $[t-\Delta t+1, t]$ is defined as a series of continuous time durations

$$\left\{D^{\mathcal{S}}, C^{\mathcal{S}}, A^{\mathcal{S}}, C^{\mathcal{T}}\right\}_{[t-\Delta t+1,t]} \longrightarrow \tilde{D}^{\mathcal{T}}_{t+1}. \tag{1}$$

B. Framework Overview

Fig. 1 presents an overview of our proposed framework, which can learn knowledge from the source city while addressing the contexts' inconsistency issue, and then apply to the target city to predict crime risk. The framework consists of three phases, i.e., intercity similar-grid matching, auxiliary features construction, and crime risk prediction using a dense convolutional-network-based unsupervised domain adaptation. In the first phase, inspired by the work [15], we design a novel intercity similar-grid matching method leveraging common context data from a source city and a target city, while taking into account the data sparsity caused by the low frequency of crime incidents. In the second phase, we present a novel method to construct auxiliary features for the target city, based on the source-city-specific contexts of several similar source city grids. This method can address the contexts' inconsistency problem between the source city and the target city. In the third phase, as shown at the right part of Fig. 1, to achieve unsupervised crime risk prediction, we propose a dense convolutional network to learn features for accurate crime risk prediction and domain-invariant features for unsupervised domain adaptation. The network parameters can be learned with consideration of three optimization elements simultaneously during the optimization process, i.e., crime risk prediction error, domain classification error, and distribution discrepancy distance.

IV. UDAC for Crime Risk Prediction

In this section, we present our UDAC for crime risk prediction. We first match several similar source city grids for each target city grid, and then we construct auxiliary contexts for the target city to make contexts consistent between two cities. After that, we present a dense convolutional network with unsupervised domain adaptation for crime risk prediction.

A. Intercity Similar-Grid Matching

The objective of this phase is to discover some source city grids $\{r_{i,1}^{\mathcal{S}}, r_{i,2}^{\mathcal{S}}, \ldots, r_{i,k}^{\mathcal{S}}\}$ having similar spatio-temporal patterns with each target city grid $r_i^{\mathcal{T}}$. Due to data sparsity caused by rare crime events, we match top-k similar grids in the source city for each grid in the target city based on their similarity coefficients, bringing k intercity similar-grid pairs. We adopt the same similarity metrics following the work [15], as *Pearson* coefficient. As there are fewer types of relevant data in the target city than the source city, we choose the common context data in both the cities, $C^{\mathcal{S}}$ and $C^{\mathcal{T}}$, to compute similarity coefficients. We denote $c_{r^{\mathcal{S}},t}$ as the common context data of the source city grid $r^{\mathcal{S}}$ in time duration t and $c_{r^{\mathcal{T}},t}$ as the common context data of the target city grid $r^{\mathcal{T}}$ in time duration t. For example, in our crime risk prediction problem, for each target city grid $r_i^{\mathcal{T}}$, we use common contexts (e.g., weather conditions, POI distribution, and police station distribution) from each source city grid and $r_i^{\mathcal{T}}$ and compute similarity coefficients between these two grids. Based on the similarity coefficients' results, we can identify k source city grids as the partners of $r_i^{\mathcal{T}}$

$$\rho_{r^{\mathcal{T}},r^{\mathcal{S}}} = \mathrm{corr}\big(\{c_{r^{\mathcal{T}},t}\}, \{c_{r^{\mathcal{S}},t}\}\big). \tag{2}$$

Thus, each target grid $r^{\mathcal{T}}$ has identified their intercity similar-grid partners in the source city with top-k similarity coefficient values.

B. Auxiliary Features' Construction

This phase aims to construct auxiliary features for each target city grid, using the source-city-specific contexts of several similar source city grids. In particular, we separate the context data in the source city into two groups, i.e., common contexts shared by both the cities and source-city-specific auxiliary contexts which are only collected in the source city and the target city did not release this type data. Note that there is no requirement for city-specific context data selection. Then, we construct auxiliary features for the target city with the same feature dimensionality as the auxiliary contexts in the source city. More specifically, for each target city grid $r^{\mathcal{T}}$, we can find k matched similar-grid partners $\{r_1^{\mathcal{S}}, r_2^{\mathcal{S}}, \ldots, r_k^{\mathcal{S}}\}$ in the source city after similarity coefficients' computation. A simple and effective method is to calculate the average value of the auxiliary contexts of these source city grids $\{r_1^{\mathcal{S}}, r_2^{\mathcal{S}}, \ldots, r_k^{\mathcal{S}}\}$, as presented in (3). $a_{r_i^{\mathcal{S}}}$ is defined as the auxiliary data of the grid $r_i^{\mathcal{S}}$. $\hat{A}^{\mathcal{T}}$ is defined as the set of constructed auxiliary data in the target city grids, $\hat{a}_{r^{\mathcal{T}}}$ is defined as the constructed auxiliary data in the grid $r^{\mathcal{T}}$. $\hat{a}_{r^{\mathcal{T}}} \in \hat{A}^{\mathcal{T}}$

$$\hat{a}_{r^{\mathcal{T}}} = \sum_{i=1}^{k} \rho_{r^{\mathcal{T}},r_i^{\mathcal{S}}} \times a_{r_i^{\mathcal{S}}}. \tag{3}$$

C. Dense Convolutional Network With Unsupervised Domain Adaptation

After auxiliary features' construction for the target city and features' embedding from both the source city and the target city, we propose a dense-convolutional-network-based [38] UDAC to learn knowledge and apply to the target city for unsupervised crime risk prediction. The input data include common and auxiliary features in $\mathcal{S}$, common context features in $\mathcal{T}$, and constructed auxiliary features in $\mathcal{T}$. The first two features are embedded as one-hot representation. The common context features in $\mathcal{T}$ would be concatenated with constructed auxiliary features in $\mathcal{T}$ together, and then fed into the network. Then we use a layer convolution operation (Conv) to capture the latent features. After that, three layers of composite operation are presented, which are defined as a sequential combination of batch normalization (BN), rectified linear units (ReLUs), and Conv. The three layers are following a dense connectivity pattern, which can connect each layer of the network to each other layer in a feed-forward manner. The implementation of the dense connectivity pattern here facilitates feature learning and improves feature propagation, while alleviating the vanishing gradient issue. High level and features are learned effectively after the final composition operation (BN-ReLU-Conv). With the proposed network, we need to optimize the parameters following three objectives simultaneously, to accomplish the unsupervised domain adaptation task as predicting crime risk in cities with unlabeled data.

The first objective is to minimize the crime risk prediction error on $\mathcal{S}$ in the Frobenius norm. We define the loss L_p as in (4), for further optimization of the parameters learned from the source city

$$L_p = \sum_t \big|\big|\tilde{D}_t^{\mathcal{S}} - D_t^{\mathcal{S}}\big|\big|_F^2. \tag{4}$$

The second objective is to maximize the domain classification error with the purpose of learning domain-invariant features. If a domain classifier is unable to identify which specific domain features should belong to, these features can be treated as domain-invariant features [39]. Here, we use binary cross-entropy as the domain classification loss L_d as in (5) and denote $\tilde{d}_i$ as the predicted domain label and d_i as the ground-truth value. $M_{\mathcal{S}}$ and $M_{\mathcal{T}}$ are defined as the number of training samples from $\mathcal{S}$ and $\mathcal{T}$, respectively,

$$L_d = \sum_{i}^{M_{\mathcal{S}}+M_{\mathcal{T}}} \tilde{d}_i \log(d_i) + \big(1 - \tilde{d}_i\big) \log(1 - d_i). \tag{5}$$

The third objective is to minimize the distribution discrepancy distance between the source and target city datasets. Following [14], we use the way to compute the distance as the maximum mean discrepancy (MMD) distance. $\widetilde{\text{MMD}}(f^{\mathcal{S}}, f^{\mathcal{T}})$ is defined as the distance estimated between high-level learned features from two cities, i.e., $f^{\mathcal{S}}$ and $f^{\mathcal{T}}$. $\Phi(\cdot)$ is a kernel function, and we adopt the Gaussian radial basis function to calculate the distance. The computation is shown as follows:

$$\begin{aligned}\widetilde{\text{MMD}}\big(f^{\mathcal{S}}, f^{\mathcal{T}}\big) &= \frac{1}{M_{\mathcal{S}}^2}\sum_{i}^{M_{\mathcal{S}}}\sum_{j}^{M_{\mathcal{S}}}\Phi\big(f_i^{\mathcal{S}}, f_j^{\mathcal{S}}\big)\\ &\quad+\frac{1}{M_{\mathcal{T}}^2}\sum_{i}^{M_{\mathcal{T}}}\sum_{j}^{M_{\mathcal{T}}}\Phi\big(f_i^{\mathcal{T}}, f_j^{\mathcal{T}}\big)\\ &\quad-\frac{2}{M_{\mathcal{S}}M_{\mathcal{T}}}\sum_{i}^{M_{\mathcal{S}}}\sum_{j}^{M_{\mathcal{T}}}\Phi\big(f_i^{\mathcal{S}}, f_j^{\mathcal{T}}\big). \end{aligned}\tag{6}$$

The total optimization function is the combination of aforementioned functions, with optimization on θ_p, θ_d, and θ_m, as shown in (7). μ and λ control the strength caused by domain adaptation from city $\mathcal{S}$ to city $\mathcal{T}$

$$\min L = \min_{\theta_p,\theta_d,\theta_m} L_p\big(\theta_p,\theta_m\big) - \mu L_d(\theta_d,\theta_m) + \lambda\widetilde{\text{MMD}}(\theta_m). \tag{7}$$

To solve this task, we use a popular parameter optimization method, the stochastic gradient descent (SGD) [40], to learn these parameters. Thus, θ_p, θ_d, and θ_m can be updated through the following optimization process:

$$\widetilde{\theta_m} = \arg\min_{\theta_m} L\big(\theta_p,\theta_d,\theta_m\big) \tag{8}$$

$$\widetilde{\theta_p} = \arg\min_{\theta_p} L\big(\theta_p,\theta_d,\theta_m\big) \tag{9}$$

$$\widetilde{\theta_d} = \arg\max_{\theta_d} L\big(\theta_p,\theta_d,\theta_m\big) \tag{10}$$

$$L = L_p\big(\theta_p,\theta_m\big) - \mu L_d(\theta_d,\theta_m) + \lambda\widetilde{\text{MMD}}(\theta_m). \tag{11}$$

V. Experiments

In this section, we first introduce datasets used in this study, implementation and evaluation metrics, as well as comparison methods. After that, we evaluate the performance on both prediction and ranking, respectively. We also conduct a parameter study to investigate the effect of the parameter on the prediction tasks.

A. Dataset

We collect crime data and external relevant data during the year 2015 from three cities, i.e., NYC, Chicago, and LA. The crime datasets contain all the reported crime occurrences with detailed information, e.g., crime types and locations. The external datasets include weather conditions, POI distribution, police station distribution, and taxi mobility. Specifically, we collect meteorological data of the three cities from a public website [41], including weather conditions and temperature information. For POI distribution, we collect POI distribution data from [42], and each belongs to the first-tier categories of Foursquare. We collect the taxi mobility dataset from NYC Taxi and Limousine Commission (NTLC) [43]. For each city, we split the city into multiple grids. The number of grids among the three cities is different, that is, NYC has 899 grids, Chicago has 683 grids, and LA has 1377 grids. Thus, we select 25 × 25 grids in this study. The time duration is set as one day. The summary of these datasets is presented in Table I.

TABLE I

Summary of Datasets

City	**NYC**	**Chicago**	**LA**
# Grid	899	683	1,377
# Crime	467,665	200,575	212,762
# Weather	365 days		
# POI	53,374	18,601	15,102
# Police Station	77	23	21
# Taxi trips	93,975,444	/	/

B. Implementation and Evaluation Metrics

With these cities, we conduct six experiments to show the effectiveness of our model, i.e., NYC → Chicago, Chicago → NYC, NYC → LA, LA → NYC, Chicago → LA, and LA → Chicago. Since we collect more data from NYC than Chicago and LA, the extracted feature dimensionality from NYC and Chicago is different, while the extracted feature dimensionality from LA and Chicago is the same. For each experiment, we define the city before the arrow as the source city and the city after the arrow as the target city. More specifically, for the experiments of NYC → Chicago and NYC → LA, we collect more data from NYC than Chicago and LA, and it is easy to separate the common contexts and auxiliary contexts. For the experiments of Chicago → NYC and LA → NYC, we do not take taxi trip data and police station distribution data in NYC into consideration. For the experiments of Chicago ⇌ LA, we discard the police station distribution data in the target city. In these experiments, police station distribution data are treated as source-city-specific contexts in the source cities.

We have presented a flowchart of our proposed method, as shown in Fig. 2. In particular, after we have collected all the accessible data, we split them into two datasets, i.e., training dataset and testing dataset. The training dataset consists of all the labeled data from the source city and half of the unlabeled data from the target city. The testing dataset consists of the remaining half of the data from the target city. We divide the training datasets into common context data and auxiliary context data, to create auxiliary features for the target city, and then train a dense-convolutional-network-based UDAC. We obtain a trained model after completing training the model, then use the testing data with the trained model to test the performance, and finally achieve the final prediction results in the target city. The detailed implementation parameters are set as follows. The model is trained for 5000 epochs. The batch size is set at 32. Most of the convolution layers

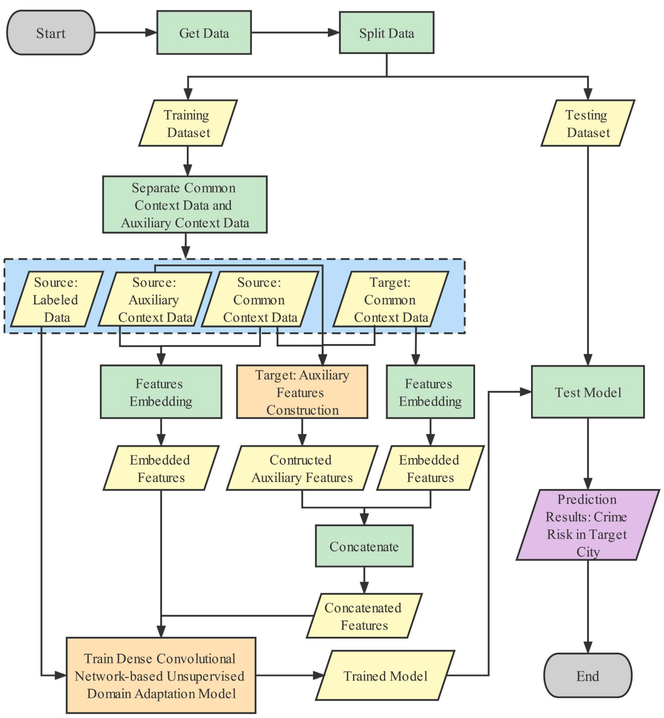


Fig. 2. Flowchart of our UDAC.

have 16 filters with a kernel size 3×3, and the final layer has one filter with 1×1 kernel size. The penalty parameters λ and μ change gradually every ten epochs. The learning rate during optimization is set as 0.001. These parameters' values are chosen based on the needs of the experiments. Alternative values are also possible

$$\text{RMSE} = \frac{1}{N}\sqrt{\sum\nolimits_r \sum\nolimits_t \left(\tilde{d}_{r,t} - d_{r,t}\right)^2} \tag{12}$$

$$\text{MAPE} = \frac{1}{N}\sum_r \sum_t \frac{\left|\tilde{d}_{r,t} - d_{r,t}\right|}{d_{r,t}} \tag{13}$$

$$\text{MAE} = \frac{1}{N}\sum_r \sum_t \left|\tilde{d}_{r,t} - d_{r,t}\right| \tag{14}$$

$$\text{HR@}K = \frac{1}{K * T}\sum_t I\left(\text{list}_K(t), \widehat{\text{list}}_K(t)\right). \tag{15}$$

For the evaluation metrics, we use the root mean square error (RMSE), the mean absolute percentage error (MAPE), the mean absolute error (MAE), and the hitting rate @K (HR@K) as the evaluation metrics. Thus, we can evaluate both the prediction performance and the ranking performance. Since crime is a low-frequent and sparse event, there have been lots of grids with value zero. Here, we only consider the scenarios where there is a nonzero value between the ground truth and the predicted value, the number of which is denoted as N. The computation of these metrics is defined in (12)–(15), respectively. For HR@K, we define $\text{list}_K(t)$ as a ranking list of most threatening grids, and $\widehat{\text{list}}_K(t)$ as the predicted ranking list. We also use $I()$ [10] as the binary function to calculate the number of identical pairs. T is defined as the total time duration in the experiments. The computation of HR@K is shown in (15). A smaller RMSE/MAPE/MAE value indicates a more accurate prediction performance, while a higher HR value brings a better ranking performance.

C. Comparison Methods

To demonstrate the superiority of our proposed model, we compare our model with several methods. Due to the contexts' inconsistency between the source city and target city, these methods train their prediction models after discarding the auxiliary context data in the source city.

1) *CNN*: This method is a typical three-layer convolutional neural network (CNN) model trained only using context data and crime data from the source city.
2) *DDC* [44]: This method incorporates a CNN model with an adaptation layer and MMD to identify domain-invariant features.
3) *Deep Adaptation Network (DAN)* [45]: This method adds three adaptation layers on the basis of the deep domain confusion (DDC) method and adopts multikernel MMD for better representation abilities.
4) *Domain-Adversarial Neural Network (DANN)* [46]: This method proposes a deep neural network with a domain discriminative component so that the classifier cannot identify the specific domain.
5) *RegionTrans* [15]: This method is able to achieve cross-city transfer learning to learn a predictive model while addressing the representation divergence with matched region pairs.
6) *STCNet* [32]: This method develops a transfer learning mechanism to improve knowledge reuse in various domains and leverages a convolutional LSTM network to represent the spatial–temporal dependencies.
7) *DATN* [47]: This method designates task-specific feature learning networks and domain adversarial training techniques to cope with the domain distribution discrepancy issue.
8) *ARG-STNet* [48]: This method proposes a generation strategy to learn long-term spatial–temporal dependencies and subsequently transfer them to the target domain with few-shot learning.
9) *UDAC-aux*: This method is a variant of our proposed UDAC. It discards the auxiliary feature construction module and then trains a prediction model based on the dense convolutional network with unsupervised domain adaptation for further crime prediction tasks in the target city.

D. Prediction Performance

We present the experimental results of the prediction performance comparison by our UDAC model and comparison methods in terms of RMSE, MAPE, and MAE, through six experiments between different city pairs, i.e., NYC $\rightarrow$ Chicago, Chicago $\rightarrow$ NYC, NYC $\rightarrow$ LA, LA $\rightarrow$ NYC, Chicago $\rightarrow$ LA, and LA $\rightarrow$ Chicago. These results are demonstrated in Table II.

Horizontally, we observe that our proposed model UDAC outperforms all the comparison methods, with an improvement of 62.20%, 15%, 13.6%, 9.97%, 5.81%, 5.11%, 6.11%,

TABLE II
PREDICTION PERFORMANCE COMPARISON

Source City	Target City	Metric	CNN	DDC	DAN	DANN	Region -Trans	STC-Net	DATN	ARG-STNet	UDAC -aux	UDAC
NYC	Chicago	RMSE	1.619	0.724	0.704	0.676	0.660	0.651	0.643	0.639	0.645	**0.618**
		MAPE	0.448	0.298	0.273	0.308	0.280	0.268	0.294	0.286	0.217	**0.121**
		MAE	0.736	0.276	0.279	0.257	0.259	0.268	0.264	0.301	0.293	**0.214**
Chicago	NYC	RMSE	1.717	0.752	0.757	0.713	0.697	0.700	0.701	0.678	0.668	**0.652**
		MAPE	0.466	0.311	0.288	0.303	0.287	0.309	0.293	0.277	0.195	**0.140**
		MAE	0.610	0.354	0.261	0.249	0.258	0.256	0.332	0.254	0.291	**0.243**
NYC	LA	RMSE	1.646	0.737	0.721	0.697	0.652	0.640	0.662	0.640	0.652	**0.626**
		MAPE	0.453	0.308	0.277	0.294	0.278	0.265	0.293	0.264	0.179	**0.130**
		MAE	0.657	0.332	0.331	0.272	0.318	0.261	0.322	0.302	0.286	**0.218**
LA	NYC	RMSE	1.728	0.773	0.757	0.737	0.696	0.695	0.697	0.688	0.685	**0.635**
		MAPE	0.473	0.303	0.302	0.315	0.279	0.288	0.290	0.290	0.170	**0.140**
		MAE	0.747	0.314	0.250	0.283	0.253	0.246	0.337	0.277	0.286	**0.213**
Chicago	LA	RMSE	2.142	0.955	0.942	0.890	0.852	0.851	0.862	0.859	0.846	**0.813**
		MAPE	0.526	0.359	0.357	0.349	0.312	0.330	0.311	0.327	0.196	**0.143**
		MAE	0.942	0.414	0.450	0.299	0.325	0.345	0.418	0.311	0.352	**0.270**
LA	Chicago	RMSE	2.134	0.944	0.925	0.899	0.849	0.837	0.859	0.855	0.831	**0.809**
		MAPE	0.486	0.340	0.316	0.327	0.315	0.342	0.327	0.318	0.246	**0.156**
		MAE	0.723	0.435	0.439	0.451	0.324	0.294	0.385	0.386	0.336	**0.303**

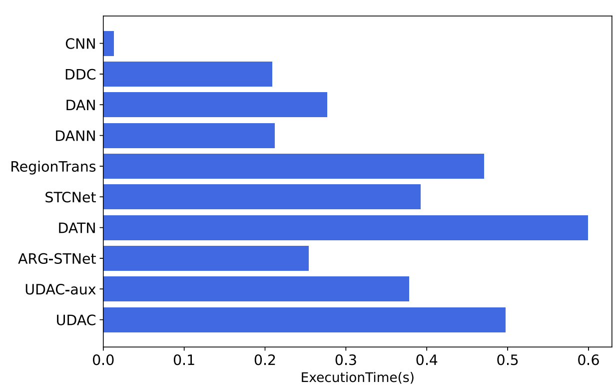


Fig. 3. Execution time comparison.

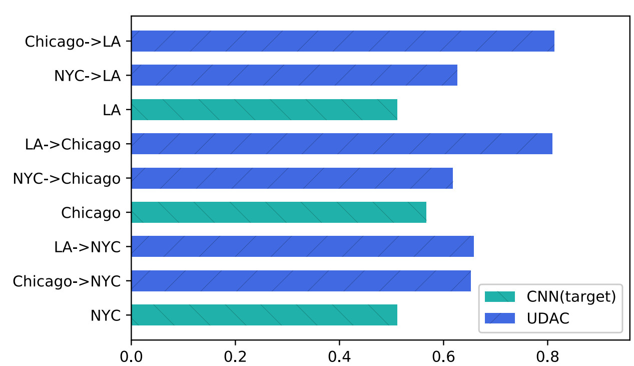


Fig. 4. Prediction performance comparison between UDAC and CNN (target).

4.63%, and 4.07% in RMSE, respectively, among all the six experiments. In terms of MAPE, our model obtains a decrease of 70.88%, 56.69%, 54.05%, 56.19%, 52.62%, 53.76%, 54.12%, 52.81%, and 30.11%, respectively, compared with the baseline methods. Besides, with consideration of MAE, the UDAC model achieves an improvement of 66.47%, 30.87%, 25%, 17.68%, 15.64%, 12.28%, 28.59%, 19.8%, and 21.01%, respectively. These experimental results demonstrate the superiority of our model for predicting crime risk in an unsupervised manner, which significantly decreases the distribution discrepancy between two cities' data. Moreover, we also observe that the comparison method CNN has the worst performance among all the baseline methods. Here, this CNN method only learns the knowledge from the source city and applies it directly to the target city for crime risk prediction. All the transfer learning-based methods outperform the CNN method, indicating that transfer learning technologies would be a helpful and valuable tool to solve the crime risk prediction tasks with unlabeled data. Besides, we observe that the methods RegionTrans and UDAC-aux obtain better prediction performance over the other transfer-learning-based methods. This implies that the processing of intercity similar-grid matching is beneficial for subsequent knowledge transfer. Furthermore, UDAC-aux achieves superior prediction

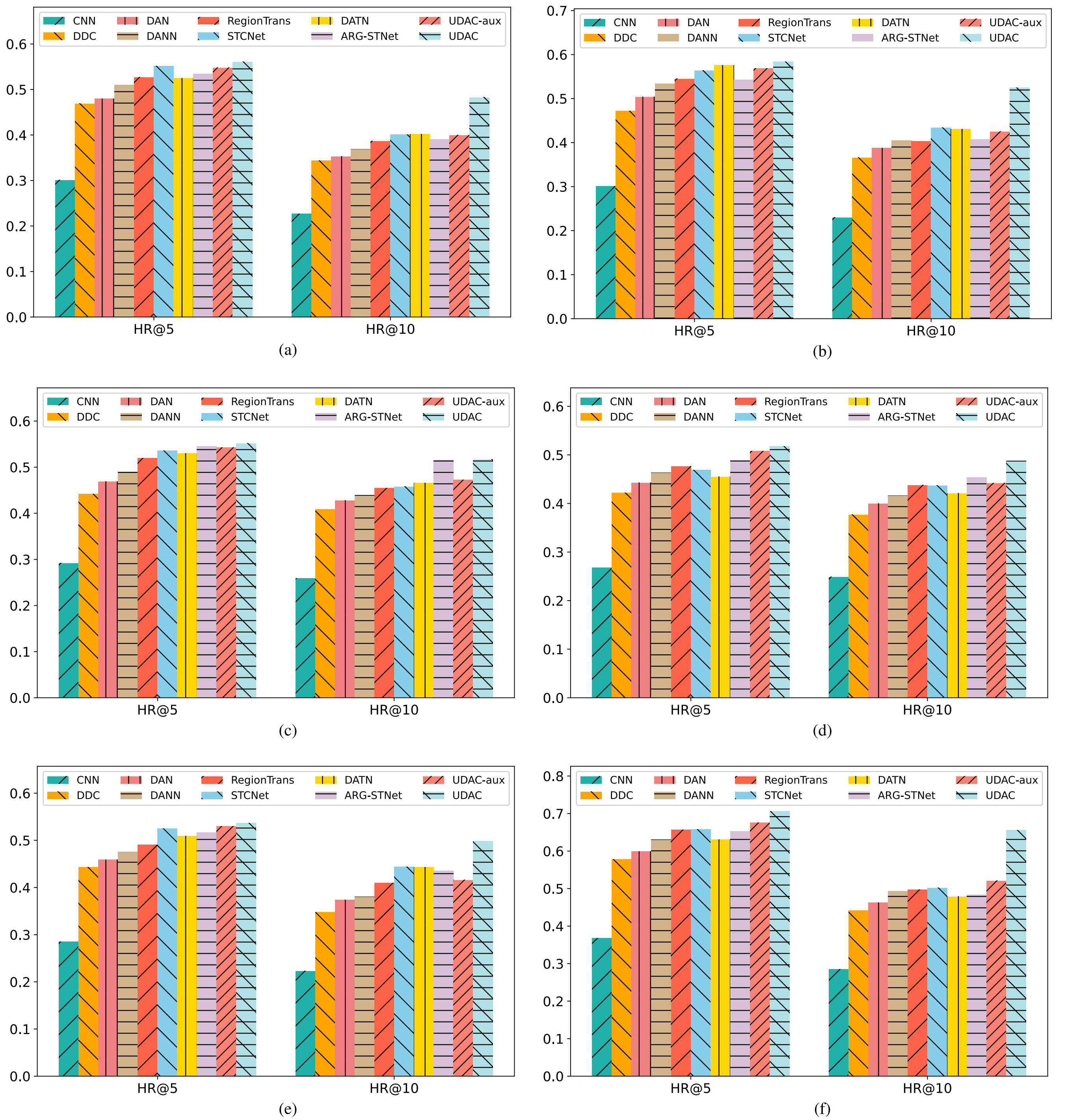


Fig. 5. Ranking performance comparison. (a) NYC → Chicago. (b) Chicago → NYC. (c) NYC → LA. (d) LA → NYC. (e) Chicago → LA. (f) LA → Chicago.

performance than RegionTrans. The explanation may be that crime incidents are infrequent and is more sparse than the data used in the original RegionTrans [15], i.e., crowd flow data. Thus, to obtain a robust intercity similar-grid matching, the design of top-k similar-grid matching is more applicable here. We also observe that STCNet and attention-reptile-generation spatial-temporal network (ARG-STNet) achieve considerable prediction performance, which indicates their power of learning domain-invariant features. In addition, compared with the variant method UDAC-aux, our UDAC can obtain an average decrease of 4.07%, 30.11%, and 21.01% in terms of RMSE, MAPE, and MAE, respectively, among conducted six experiments. This highlights the significance and value of the auxiliary feature construction module.

Vertically, we note that UDAC is more robust under different configurations of experiments. Among the three cities, Chicago and LA are more similar than Chicago and NYC, or LA and NYC, in terms of population, geographic size, POI distribution, and so on. More specifically, the experiment of NYC → Chicago achieves superior performance to the experiment of Chicago → NYC, with lower RMSE, MAPE, and MAE values. The possible explanation is that the model trained in the former experiment may be better able to accurately predict crime risk than the model trained in the latter experiment, due to more context data in NYC, such as weather conditions, POI distribution, police station distribution, and taxi mobility. The experiments of NYC ⇌ LA also have similar performance that the experiment of NYC → LA obtains better prediction performance than the experiment of LA → NYC, with lower RMSE and MAPE values and considerable MAE values, due to the same possible reason. Besides, we note that experiments of Chicago ⇌ LA obtain comparable performance on crime risk prediction. This may be because of the similarity between Chicago and LA in many aforementioned city characteristics.

We further compare our proposed method with the baseline methods in execution time. Here, the execution time refers to the average running time of each test sample. It was computed by dividing the overall running time by the total number of test samples, and then averaging the results across all the runs with the same experiment settings. The average execution time of our proposed UDAC is below 0.5 s on these experiments. domain adversarial transfer network (DATN) requires approximately 1.2 times the execution time of UDAC for crime risk prediction, whereas RegionTrans takes a comparable amount of execution time, as shown in Fig. 3. The possible reason behind may be the implementation of the auxiliary feature construction module.

We train a CNN in the target city with all the labeled data as well, to achieve an accurate crime risk prediction model. To distinguish with the aforementioned comparison method CNN, we name this CNN as CNN (target). We then compare our UDAC model with this CNN (target) on the three cities, to evaluate the crime risk prediction performance. The experimental results are presented in Fig. 4. From the figure, we can observe that the supervised method CNN (target) performs better than our unsupervised domain adaptation method UDAC among these experiments. Thus, the experimental results generated by CNN (target) can be treated as the upper bound of our unsupervised crime risk prediction study. In a word, our UDAC performs better than all the aforementioned comparison methods, while being some distance from the single-domain CNN (target).

E. Ranking Performance

We also conduct experiments to examine the accurate prediction capability of our proposed UDAC model and other comparison methods, using the HR metric. Specifically, we predict m grids with the highest m level of high crime risk in the target city and compare them with the actual high crime risk grids in the source city. The experimental results are presented in Fig. 5. Each figure demonstrates the results by these methods on one pair of a source and a target city, with two HR values as HR@5 and HR@10, respectively.

We would like to elaborate on the explanation using the experiment of NYC → Chicago as an illustration. From Fig. 5(a), it can be shown that our method achieves the highest HR with the predicted crime values, successfully predicting the 5 and 10 most threatening grids in the target city. Besides, another interesting observation is that for each method, the HR value obtained by comparing the most ten threatening grids is less than obtained by comparing the most five threatening grids. This is not difficult to understand. Crime incidents are low-frequent events, and the majority of grids may witness few crimes for a long time. Only a small percentage of grids would experience threatening events frequently, such as places with few security forces and valuable property. Thus, predicting five most threatening grids would yield more accurate results than predicting ten grids.

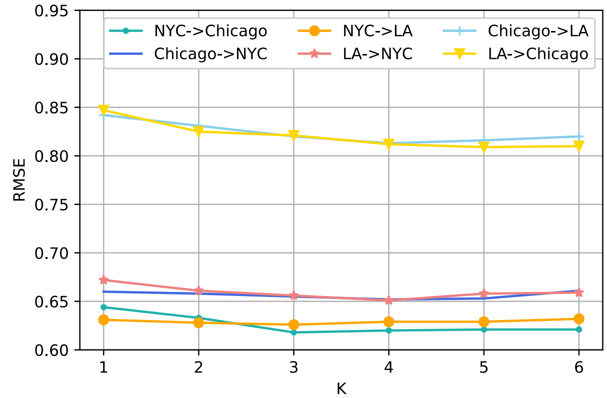


Fig. 6. Parameter impact analysis.

F. Parameter Study

We also study how the parameter k affects the crime prediction performance. The parameter impact analysis results are presented in Fig. 6. Here, k refers to the number of source city grids used to select intercity similar-grid pairs. From the figure, we discover that optimal selection of k is different among these experiments to predict crime risk in cities with unlabeled data. It would be more accurate to make a prediction using k equal to 3 for the experiments of NYC → Chicago and NYC → LA. Number 4 is the optimal choice when considering the results of the experiments conducted as Chicago → NYC and LA → NYC. A larger value for 5 would bring about improved performance for the experiments carried out for LA ⇌ Chicago.

We take the experiment of NYC → Chicago as an example to investigate the parameter impact, with number 3 representing optimal selection for k. This indicates that when we match three source city grids that are most similar to each target city grid, and then average the auxiliary context data among these three source city grids, the auxiliary features constructed for the target grid are effective and useful for future crime prediction in the target city. This is accomplished by averaging the data from the three source city grids. As a result, we select a different optimal value for k on different experiments.

VI. Conclusion

In this article, to solve the prediction tasks with unlabeled data, we propose an unsupervised domain adaptation method. This method can learn knowledge from a source city with labeled data and transfer it to the target city while addressing the context inconsistency issue between cities, to predict crime risk. We partition cities into multiple equal-sized grids and identify several similar source city grids for each target grid. Based on these pairs, we construct auxiliary features for the target city, to address the contexts' inconsistency problem across cities. We then propose a dense-convolutional-network-based unsupervised domain adaptation module to learn knowledge from the source city and apply it to the target city for future crime risk prediction. Domain-invariant features are learned to facilitate knowledge transfer. Extensive experiments are conducted to verify the effectiveness of our method using real-world data from NYC, Chicago, and LA. The experimental results reveal the superiority of our proposed method over various state-of-the-art comparison methods.

In the future, we intend to improve our work from several perspectives. First, we plan to explore prediction performance when more serious contexts' inconsistency exists between the source and target cities. Second, we plan to investigate a fine-grained unsupervised crime risk prediction, such as predicting crime risk in roads. This would be more challenge due to severe data sparsity problem.

Acknowledgment

The authors would like to acknowledge the Supercomputing Center of Zhejiang University City College for the support of advanced computing resources.

References

[1] (2020). GunViolenceArchive. *List of Mass Shootings in the United States in 2019*. [Online]. Available: https://www.gunviolencearchive.org/reports/mass-shooting

[2] M. S. Gerber, "Predicting crime using Twitter and kernel density estimation," *Decision Support Syst.*, vol. 61, pp. 115–125, May 2014.

[3] D. Yang, T. Heaney, A. Tonon, L. Wang, and P. Cudré-Mauroux, "CrimeTelescope: Crime hotspot prediction based on urban and social media data fusion," *World Wide Web*, vol. 21, pp. 1323–1347, Sep. 2018.

[4] C. Huang, J. Zhang, Y. Zheng, and N. V. Chawla, "DeepCrime: Attentive hierarchical recurrent networks for crime prediction," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, 2018, pp. 1423–1432.

[5] L. Xia *et al.*, "Spatial–temporal sequential hypergraph network for crime prediction with dynamic multiplex relation learning," in *Proc. 30th Int. Joint Conf. Artif. Intell.*, Aug. 2021, pp. 1631–1637.

[6] P. Das, A. K. Das, J. Nayak, D. Pelusi, and W. Ding, "Incremental classifier in crime prediction using bi-objective particle swarm optimization," *Inf. Sci.*, vol. 562, pp. 279–303, Jul. 2021.

[7] H. Wang, D. Kifer, C. Graif, and Z. Li, "Crime rate inference with big data," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 635–644.

[8] H. Wang, H. Yao, D. Kifer, C. Graif, and Z. Li, "Non-stationary model for crime rate inference using modern urban data," *IEEE Trans. Big Data*, vol. 5, no. 2, pp. 180–194, Jun. 2019.

[9] L. Vomfell, W. K. Härdle, and S. Lessmann, "Improving crime count forecasts using Twitter and taxi data," *Decis. Support Syst.*, vol. 113, pp. 73–85, Sep. 2018.

[10] F. Yi, Z. Yu, F. Zhuang, and B. Guo, "Neural network based continuous conditional random field for fine-grained crime prediction," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 10–16.

[11] X. Zhao and J. Tang, "Modeling temporal-spatial correlations for crime prediction," in *Proc. ACM Conf. Inf. Knowl. Manage.*, 2017, pp. 497–506.

[12] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," 2014, *arXiv:1409.7495*.

[13] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2009.

[14] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316–7325, Sep. 2018.

[15] L. Wang, X. Geng, X. Ma, F. Liu, and Q. Yang, "Cross-city transfer learning for deep spatio-temporal prediction," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 1893–1899.

[16] M. Ranson, "Crime, weather, and climate change," *J. Environ. Econ. Manage.*, vol. 67, no. 3, pp. 274–302, May 2014.

[17] B. Zhou, L. Chen, S. Zhao, F. Zhou, S. Li, and G. Pan, "Spatio–temporal analysis of urban crime leveraging multisource crowdsensed data," *Pers. Ubiquitous Comput.*, pp. 1–14, Jan. 2021, doi: 10.1007/s00779-020-01456-6.

[18] Y. Liang *et al.*, "Fine-grained urban flow prediction," in *Proc. Web Conf.*, 2021, pp. 1833–1845.

[19] Y. Liang *et al.*, "UrbanFM: Inferring fine-grained urban flows," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 3132–3142.

[20] M. Fang, L. Tang, X. Yang, Y. Chen, C. Li, and Q. Li, "FTPG: A fine-grained traffic prediction method with graph attention network using big trace data," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5163–5175, Jun. 2021.

[21] X. Wu, C. Huang, C. Zhang, and N. V. Chawla, "Hierarchically structured transformer networks for fine-grained spatial event forecasting," in *Proc. Web Conf.*, Apr. 2020, pp. 2320–2330.

[22] B. Liu, S. Yan, J. Li, Y. Li, J. Lang, and G. Qu, "A spatiotemporal recurrent neural network for prediction of atmospheric PM2.5: A case study of Beijing," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 3, pp. 578–588, Jun. 2021.

[23] B. Huang, H. Zhu, D. Liu, N. Wu, Y. Qiao, and Q. Jiang, "Solving last-mile logistics problem in spatiotemporal crowdsourcing via role awareness with adaptive clustering," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 3, pp. 668–681, Jun. 2021.

[24] A. Bogomolov, B. Lepri, J. Staiano, N. Oliver, F. Pianesi, and A. Pentland, "Once upon a crime: Towards crime prediction from demographics and mobile data," in *Proc. 16th Int. Conf. Multimodal Interact.*, 2014, pp. 427–434.

[25] S. K. Rumi, K. Deng, and F. D. Salim, "Crime event prediction with dynamic features," *EPJ Data Sci.*, vol. 7, no. 1, p. 43, Dec. 2018.

[26] F. Yi, Z. Yu, F. Zhuang, X. Zhang, and H. Xiong, "An integrated model for crime prediction using temporal and spatial factors," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2018, pp. 1386–1391.

[27] B. Zhou *et al.*, "ESCORT: Fine-grained urban crime risk inference leveraging heterogeneous open data," *IEEE Syst. J.*, vol. 15, no. 3, pp. 4656–4667, Sep. 2021.

[28] B. Zhou, L. Chen, F. Zhou, S. Li, S. Zhao, and G. Pan, "Dynamic road crime risk prediction with urban open data," *Frontiers Comput. Sci.*, vol. 16, no. 1, pp. 1–13, Feb. 2022.

[29] L. Wang, X. Geng, X. Ma, D. Zhang, and Q. Yang, "Ridesharing car detection by transfer learning," *Artif. Intell.*, vol. 273, pp. 1–18, Aug. 2019.

[30] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 1, pp. 136–144, Oct. 2017.

[31] F. Yuan, L. Yao, and B. Benatallah, "DARec: Deep domain adaptation for cross-domain recommendation via transferring rating patterns," 2019, *arXiv:1905.10760*.

[32] C. Zhang, H. Zhang, J. Qiao, D. Yuan, and M. Zhang, "Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1389–1401, Jun. 2019.

[33] T. Ni, C. Zhang, and X. Gu, "Transfer model collaborating metric learning and dictionary learning for cross-domain facial expression recognition," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 5, pp. 1213–1222, Oct. 2020.

[34] L. Zheng, G. Liu, C. Yan, C. Jiang, and M. Li, "Improved TrAdaBoost and its application to transaction fraud detection," *IEEE Trans. Computat. Social Syst.*, vol. 7, no. 5, pp. 1304–1316, Oct. 2020.

[35] X. Kong, H. Gao, G. Shen, G. Duan, and S. K. Das, "FedVCP: A federated-learning-based cooperative positioning scheme for social Internet of Vehicles," *IEEE Trans. Computat. Social Syst.*, vol. 9, no. 1, pp. 197–206, Feb. 2021.

[36] H. Yao, Y. Liu, Y. Wei, X. Tang, and Z. Li, "Learning from multiple cities: A meta-learning approach for spatial–temporal prediction," in *Proc. World Wide Web Conf. (WWW)*, 2019, pp. 2181–2191.

[37] X. Zhao and J. Tang, "Exploring transfer learning for crime prediction," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 1158–1159.

[38] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[39] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Mach. Learn.*, vol. 79, nos. 1–2, pp. 151–175, May 2010.

[40] T. Zhang, "Solving large scale linear prediction problems using stochastic gradient descent algorithms," in *Proc. 21st Int. Conf. Mach. Learn. (ICML)*, 2004, p. 116.

[41] *WeatherUnderground*. (2015). [Online]. Available: https://www.wunderground.com/weather/api

[42] *Foursquare*. (2015). [Online]. Available: https://foursquare.com

[43] *NTLC*. (2015). [Online]. Available: https://www.nyc.gov/taxi

[44] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," 2014, *arXiv:1412.3474*.

[45] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," 2015, *arXiv:1502.02791*.

[46] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2030–2096, Apr. 2016.

[47] Z. Chen, G. He, J. Li, Y. Liao, K. Gryllias, and W. Li, "Domain adversarial transfer network for cross-domain fault diagnosis of rotary machinery," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8702–8712, Nov. 2020.

[48] C. Tian, X. Zhu, Z. Hu, and J. Ma, "A transfer approach with attention reptile method and long-term generation mechanism for few-shot traffic prediction," *Neurocomputing*, vol. 452, pp. 15–27, Sep. 2021.

Binbin Zhou received the B.Math. degree from Hangzhou Normal University, Hangzhou, China, in 2008, the M.Phil. degree in Computer Science (CS) from The Hong Kong Polytechnic University, Hong Kong, in 2011, and the Ph.D. degree in CS from Zhejiang University, Hangzhou, in 2021.

She is currently an Associate Research Fellow with the Department of CS and Computing, Zhejiang University City College, Hangzhou. Her research interests are mainly in urban computing, big data applications, and ubiquitous computing.

Longbiao Chen received the B.Sc. and Ph.D. degrees in Computer Science (CS) from Zhejiang University, Hangzhou, China, in 2010 and 2016, respectively.

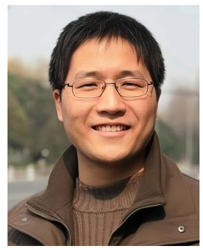
He is currently an Associate Professor with the Fujian Key Laboratory of Sensing and Computing for Smart City, Xiamen University, Xiamen, China. His research interests include urban big data mining and ubiquitous computing, mobility behavior analytics, medicines, and children.

Dr. Chen has served as a reviewer or a member of several conferences.

Sha Zhao (Member, IEEE) received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2017.

She was a Post-Doctoral Research Fellow with Zhejiang University from July 2017 to January 2021. She visited the Human–Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, USA, as a visiting Ph.D. student, from 2015 to 2016. She is currently a Research Professor with the College of Computer Science (CS) and Technology, Zhejiang University. Her research interests include pervasive computing, data mining, and machine learning.

Dr. Zhao received the Best Paper Award from Association for Computing Machinery (ACM) UbiComp'16.



Shijian Li received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2006.

He is currently a Professor with the College of Computer Science (CS) and Technology, Zhejiang University. His research interests include sensor networks, ubiquitous computing, and social computing.

Prof. Li serves as an Editor of the *International Journal of Distributed Sensor Networks* and as a reviewer or a PC member of more than ten conferences.

Zengwei Zheng received the B.S. and master's degrees in Computer Science (CS) and western economics from Hangzhou University, Hangzhou, China, in 1991 and 1998, respectively, and the Ph.D. degree in CS and technology from Zhejiang University, Hangzhou, in 2005.

He is currently a Professor with the School of Computer and CS and the Director of the Intelligent Plant Factory of Zhejiang Province Engineering Laboratory and the Hangzhou Key Laboratory for IoT Technology and Application, Zhejiang University City College, Hangzhou. His research interests include wireless sensor networks, location-based service, the Internet of Things, digital agriculture, and pervasive computing.

Dr. Zheng is a Senior Member of China Computer Federation (CCF).

Gang Pan (Member, IEEE) received the B.Eng. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 1998 and 2004, respectively.

He is a Professor with the College of Computer Science (CS) and Technology and the Vice-Director of the State Key Laboratory of CAD&CG, Zhejiang University. He has coauthored more than 100 refereed articles and has 39 patents granted. His interests include artificial intelligence, brain-inspired computing, brain–machine interfaces, pervasive computing, and computer vision.

Dr. Pan was a recipient of Natural Science Foundation (NSF) for Distinguished Young Scholars in 2019, the IEEE Technical Committee on Scalable Computing (TCSC) Award for Excellence (Middle Career Researcher) in 2018, and the China Computer Federation (CCF)-IEEE CS Young Computer Scientist Award in 2016. He has received many technical awards, including the TOP-10 Achievements in Science and Technology in Chinese Universities in 2016, the National Science and Technology Progress Award in 2015, the Best Paper Award of Association for Computing Machinery (ACM) UbiComp'16, and the IEEE Ubiquitous Intelligence and Computing (UIC) Test-of-Time Paper Award in 2019. He serves as an Associate Editor for IEEE SYSTEMS JOURNAL, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, and *Pervasive and Mobile Computing*.