

# DeepDPM: Dynamic Population Mapping via Deep Neural Network

Zefang Zong\*, Jie Feng\*, Kechun Liu, Hongzhi Shi, Yong Li†

Beijing National Research Center for Information Science and Technology  
Department of Electronic Engineering, Tsinghua University, Beijing, China  
{zzf15,feng-j16,lkc15,shz17}@mails.tsinghua.edu.cn, liyong07@tsinghua.edu.cn

## Abstract

Dynamic high resolution data on human population distribution is of great importance for a wide spectrum of activities and real-life applications, but is too difficult and expensive to obtain directly. Therefore, generating fine-scaled population distributions from coarse population data is of great significance. However, there are three major challenges: 1) the complexity in spatial relations between high and low resolution population; 2) the dependence of population distributions on other external information; 3) the difficulty in retrieving temporal distribution patterns. In this paper, we first propose the idea to generate dynamic population distributions in full-time series, then we design dynamic population mapping via deep neural network(DeepDPM), a model that describes both spatial and temporal patterns using coarse data and point of interest information. In DeepDPM, we utilize super-resolution convolutional neural network(SRCNN) based model to directly map coarse data into higher resolution data, and a time-embedded long short-term memory model to effectively capture the periodicity nature to smooth the finer-scaled results from the previous static SRCNN model. We perform extensive experiments on a real-life mobile dataset collected from Shanghai. Our results demonstrate that DeepDPM outperforms previous state-of-the-art methods and a suite of frequent data-mining approaches. Moreover, DeepDPM breaks through the limitation from previous works in time dimension so that dynamic predictions in all-day time slots can be obtained.

## Introduction

Obtaining high-resolution population distribution(HRPD) is of great importance for urban applications of business locating, transportation planning, city service managing, etc. However, the HRPD is only available to some specific companies like the internet service providers. The general businesses have to pay a lot for it or rely on some coarse and static population data to make decisions. Meanwhile, the real-time collection of HRPD data is also computing-consuming and generates unnecessary burdens to computing systems. Thus, more efficient methods are highly required in obtaining HRPD data.

\*Equal contribution.

†Corresponding author.

Some previous studies utilized external information like remote-sensing data to generate HRPD (Wu, Qiu, and Wang 2005; Gaughan et al. 2013; Stevens et al. 2015). As the state-of-the-art for these works, R. Stevens. et.al. (Stevens et al. 2015) used random forest algorithm via lots of data sources. However, these works heavily rely on many ancillary data sets that are expensive to obtain, difficult to process, and limited to certain time slots.

Other studies developed interpolation algorithms for super-resolution images, which provided powerful tools to convert low-resolution data into high-resolution data, independent of other ancillary data (Nasrollahi and Moeslund 2014; Yang, Ma, and Yang 2014; Vandal et al. 2017). One of the state-of-the-art method is SRCNN (Dong et al. 2016), which first utilized convolutional neural network as the interpolate function to model complex spatial relations in image. However, the amplification of this model is limited to  $2 \sim 4$ , which is too small for our population application. Additionally, the HRPD is not only related to the coarse population distribution but also the structure of the urban, which can not be considered directly by these existing methods.

Several challenges exist in dynamic population mapping. First, it is not easy to define a simple math function to describe the complex spatial relations. Second, the HRPD is influenced by other external knowledge like urban structures. Third, generating temporal trends from scratch is difficult.

In this paper, we propose DeepDPM, a deep learning based model that consists of augmented stacked super-resolution convolution neural network(SRCNN) as the static part, and time-embedded LSTM as the dynamic part, to evaluate population mapping in both spatial and temporal dimensions. Our static part predicts HRPD at different time slots, while the static output is further used by the dynamic part to generate temporal trends. PoI(Point of Interest) is also considered as an ancillary data which is easy to obtain. Our experiments showed that DeepDPM outperforms other traditional methods, and generates high-resolution results for urban population structures.

Our contributions can be summarized as follows:

- We present the idea to generate, in a scalable manner, dynamic population distributions in full time series from static disaggregate data sets into finer scales. To the best of our knowledge, this is the first time to analyze both spatial and temporal patterns in an urban population mapping

research work.

- We propose DeepDPM, an augmented structure that consists of static prediction and dynamic generation parts, using super-resolution convolutional neural network and time-embedded LSTM separately, based on observational and augmented PoI data.
- We perform extensive experiments based on real-life mobility dataset in Shanghai. Our results demonstrate that DeepDPM outperforms the previous state-of-the-art models and a suite of frequent data-mining and machine learning methods in terms of several metrics in predictive performance.

## Preliminaries

In this section, we review the population mapping problem and introduce the special scenarios to be investigated in this paper. Then we briefly overview our solutions.

### Definitions and Problem Formulation

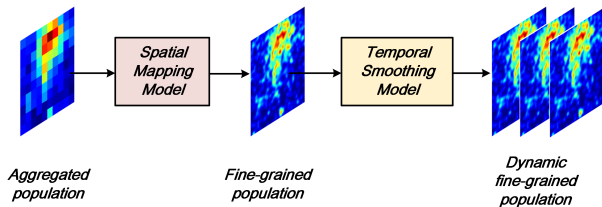
**Definition 1 (Grid Region)** *In this study, we partition a city into an  $M \times N$  grid map based on the longitude and latitude, where each grid denotes a region called grid region. As the basic space unit, we investigate the population distributions among these grid regions and a typical grid region is a  $1km \times 1km$  grid in the map.*

**Definition 2 (Population Distribution)** *In our paper, the population distribution  $X_i$  represented in terms of grid map depends on different aggregation level, where  $i \in \{1, 2, 3\}$ .  $i$  can be equal to 1, 2 or 3 each in district level, street-block level and fine-grained level, where  $X_3$  can also be denoted as  $X_{fg}$ . For any pair of  $X_p$  and  $X_q$ ,  $p < q$ , it obeys the following relationship:*

$$\sum_{i=x_{start}}^{x_{end}} \sum_{j=y_{start}}^{y_{end}} X_p(i, j) = \sum_{i=x_{start}}^{x_{end}} \sum_{j=y_{start}}^{y_{end}} X_q(i, j) \quad (1)$$

Further more, taking the regularity and mobility of population into consideration, we also utilize typical population trend along time to generate dynamic population distribution  $X_i^t$  of time  $t$ .

**Problem 1** *Given the static aggregated population data  $X_i$  and the regularity pattern of population, generate the dynamic finer-scaled grid region population distribution  $X_j = X_j^1, X_j^2, \dots, X_j^{24}$  by a certain mapping function  $F$  in a day.*



**Figure 1:** Basic framework of our solution for aggregated population mapping problem.

## Solution Overview

As Figure 1 shows, we divide the total task into two sub-tasks as follows.

**Problem 1: Spatial modelling.** Previous studies are limited in the reality. Existing studies (Gaughan et al. 2013; Stevens et al. 2015) rely on various ancillary data like high resolution imagery to obtain the spatially weighted density. However, these large-scale data are expensive to obtain and challenging to process. Furthermore, because of the strong dependency on the spatial information from the input data, previous studies used semi-automated classification algorithms combined with simple dasymetric mapping approach like random forest (Stevens et al. 2015), which is limited in directly modelling the spatial relations.

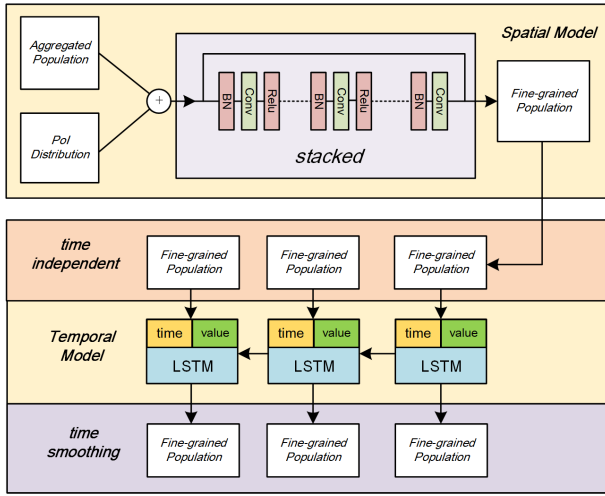
**Solution Overview:** As a powerful spatial modelling tool, convolution neural network (CNN) (LeCun et al. 1989; He et al. 2016) is widely used in many tasks. Particularly, CNN has been applied into image super-resolution task, which is aimed to generate high-resolution image based on the low-resolution image. Based on the formulation in the previous section, the aggregated population mapping task follows the similar goal and working manner with image super-resolution task. Inspired by this, we introduce CNN-based model into our task to enhance the modelling of spatial relations and constrains that exist.

**Problem 2: Temporal modelling.** Although the temporal trend is important, little previous studies of aggregated population mapping task has considered this because of the lack of dynamic data and proper methods for generating the temporal trend of population distribution. They regarded the population mapping problem as a static process only considering the night scenarios. However, the fact is that population distributions in the day in city are totally different from those in the night. Thus, solutions proposed by previous studies failed in generating the fine-grained population distribution during a full day long period.

**Solution Overview:** According to Xu. et.al. (Xu, Zhang, and Li 2016), the temporal pattern of population for a certain region can be divided into several typical classes based on its function. For example, the population of residence region first decreases to a low level in the morning, keeps during the day and finally comes back to the original level. Thus, the temporal pattern of population can be regarded as a function of time. Based on the observation, we consider to utilize recurrent neural network (RNN) (Lipton, Berkowitz, and Elkan 2015; Hochreiter and Schmidhuber 1997) as the basic sequential model to generate temporal trends for population of each region. Particularly, we introduce the time factor into the model by embedding to control the generation process.

## The Spatial-temporal Mapping Model

As presented in the Figure 2, our model consists of two main components: 1) spatial mapping model, which is designed to map the aggregated low-resolution population into high-resolution population; 2) temporal generation model, which is designed to capture the typical temporal trend and generate smoothed dynamic population results.



**Figure 2:** The Network structure of our spatial-temporal mapping model.

### Spatial Mapping Model

Mapping the aggregated population into higher-resolution population is similar to image super resolution problem, in which SRCNN is one of the state-of-the-art methods.

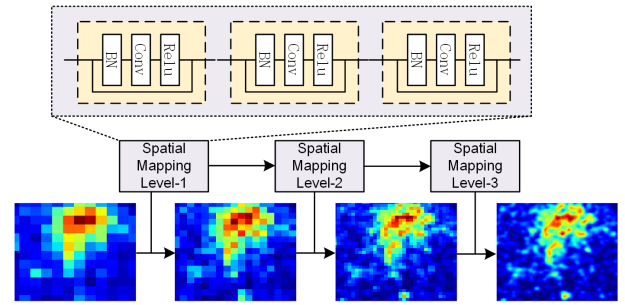
SRCNN consists of three operations (Dong et al. 2016): patch extraction, non-linear mapping, and reconstruction. In SRCNN, each operation is implemented as a three-layer convolution network including a batch-norm layer, a convolution layer and a non-linear activation layer (e.g., Relu). By stacking these three operation units, we formulate basic spatial mapping unit in our model. With the low-resolution image  $X$  as the input and the high-resolution image  $Y$  as the target, the mapping function  $F$  is optimized with the following objective function:

$$\arg \min_{\Theta} \sum_1^n \|F(X_i; \Theta) - Y_i\|_2^2, \quad (2)$$

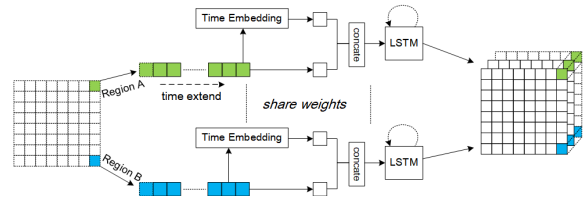
where  $\Theta$  denotes the parameters of the network and  $n$  denotes the instances of image pairs.

The basic mapping unit SRCNN can only handle the resolution enhancement ratio between 2 to 4, while in population mapping this ratio can be up to 15. One possible solution is to stack more convolution networks to directly meet with the higher enhancement requirement, which however, fails in learning complex mapping functions because of the lengthy propagation path and weak supervised signals. Thus, we decompose the high-enhancement ratio mapping task into several low-enhancement ratio tasks. In practice, we train several independent mapping network units for each of these sub-tasks by providing their related mapping ground-truth. Finally, we stack these trained independent mapping units to form a comprehensive spatial mapping model to complete the whole mapping task. In detail, two SRCNNs are stacked in generating  $X_2$  from  $X_1$ , and another two are stacked in generating  $X_3$  from  $X_2$ . The basic mapping units and the whole procedure are presented in Figure 3.

It's worthy to mention that cascading super-resolution networks altogether directly is also a considerable method, which is to train the model in one end-to-end way. However, during experiments we found that stacking outperforms cascading. Generating distributions in a lower resolution by upscaling fine-grained ground truth allows us to train independent input/output pairs and stack them together at test time as well as keeping accuracy. While in a cascading one, the fact that each output is exactly the input of the following level may lead to some error in propagation (Vandal et al. 2017). Besides, according to Xu. et.al. (Xu, Zhang, and Li 2016), the function of a region can play an import role in forming its population pattern. Meanwhile, the function of a region can be represented by the Point of Interests (POIs) distribution to some extent. Hence, we introduce POI distribution matrix to describe the function of regions. Different types of POI matrix are regarded as specific channels to form the multi-channel input matrix in aggregated level.



**Figure 3:** Stacked network structure of the spatial mapping model.



**Figure 4:** Structure of temporal smoothing model.

### Temporal Generation Model

Except the spatial distribution, the temporal trend is also an import characteristic of population distribution. By smoothing population along time, we can reduce the spatial estimation error and better understand the distribution. According to Xu. et.al. (Xu, Zhang, and Li 2016), while the concrete temporal variations of population in different regions are different, their temporal trends can be classified into several typical types based on the function of each region. And these temporal trends are representative and shared by different regions and cities. In this paper, we use long-short term memory network (LSTM), a population variation of RNN as the basic recurrent unit to model these typical temporal trend series.

During the training, we first extract the temporal population series of every region from the output of spatial mapping model. Then, we use one LSTM model to train all the population series. In this way, the parameters of the neural network are shared by all regions, which makes the model small, robust and can be generalized in modeling the temporal trends of population. To model the influence of time factor (e.g., hour of day), we introduce time embedding into our temporal generation model. Particularly, we first discrete the time of day into 24 hours and encode them into 24-dimension one-hot vectors. Then, we build a linear network layer as embedding table to project the one-hot vector into dense vector. Then, the original population value (unknown value can be set as 0) and the time vector are grouped together as a whole to feed into LSTM network in every time step. The details of the temporal structure is presented in Figure 4.

## Performance Evaluation

In this section, we conduct extensive experiments on mobility dataset in Shanghai to answer the following research questions:

- **RQ1: Spatial modelling to obtain the population distribution in higher resolutions at a fixed time slot**
- **RQ2: Temporal modelling to obtain the dynamic population distribution in a fixed resolution**

## Experimental Settings

**Datasets** We collect our representative real-life mobility dataset from ISP, which contains cellular network access records in 9685 different base stations in Shanghai for 4464 different time slots, from 1<sup>st</sup> July, 2017 to 31<sup>st</sup> July, 2017 (the data usage is recorded every 10 minutes for 31 days). Considering the periodicity difference between weekdays and weekends, we manually drop the data on weekends and focus on weekdays’ data in experiments.

We use the address of base station to estimate cellular data users’ location distribution. Since mobile devices keep accessing cellular data as long as their data connection are kept on, our dataset well represents the population distribution in Shanghai. Similar data types are used in urban research as well, such as call detail records (Isaacman et al. 2012; Ficek and Kencl 2012) or GPS data (Zheng et al. 2008). However, these data above are event driven, which update only when a user acquires service. While our dataset passively captures users’ newest location information, which guarantees the credibility of our analysis.

Also, we collect our PoI dataset from Tencent, which contains 618296 PoI records in 17 categories. We manually classify them into 4 categories, entertainment, business, transportation junctions and residence, based on their functions.

**Preprocessing** One obstacle in using mobility records to represent population distribution is that base stations lose the track when devices are turned off, or are disconnected due to other various factors, like weak signal strength. Therefore, an augmented algorithm is in need to recover the missing fingerprints.

Aggregated Categories	Original Categories
Entertainment	Hotel, Entertainment, Shopping, Catering, Culture, Sports, Tourist Spots
Business and Education	Departments, Industries, Education, Medical
Transportation Junctions	Transportation Junctions
Residence	Housing, Residential Services

**Table 1:** PoI Classification from original categories into aggregated categories.

We define the record user number of base station  $i$  in time slot  $t$  as  $X_i[t]$ , the actual number as  $Y_i[t]$ , the total time slot number as  $T$ , and the total station number as  $N$ . We first compute the sum of all activated devices at each time slot  $t$ , and find out the maximum of it ever recorded as the representation of population amount in the city. Then we estimate the percentage of activated devices denoted by  $R_i[t]$ . Finally we obtain the estimated user number  $Y_i[t]$ . The formulation in math is described below:

$$S[t] = \sum_{i=1}^N X_i[t] \text{ for } t = 1 \text{ to } T, \quad (3)$$

$$R[t] = \frac{S[t]}{\max_{j=1}^T S[j]} \text{ for } t = 1 \text{ to } T, \quad (4)$$

$$Y[t] = \frac{X_i[t]}{R[t]}, \text{ for } t = 1 \text{ to } T, i = 1 \text{ to } N \quad (5)$$

Since the base stations are located irregularly in geometry, we need to further generate the grid regions. We generate Voronoi diagrams based on the distribution of base stations. The contribution of population from each polygon to each grid is determined by the ratio, which is the intersection area divided by the polygon area. Finally, the population distribution at each time slot can be successfully mapped into grids, which is an 83\*114 grid map based on longitude and latitude. PoIs mapped into the same grid are counted together, and the sum is assigned as the grid value according to different PoI categories.

After obtaining the fine-grained grid region distribution denoted as  $X_{fg}$  or  $X_3$ , we further generate the aggregated distribution at another two levels,  $X_1$  and  $X_2$ . The grid maps remain the same size, while grids in the same district or the same street area are equalized using their average. Grid values outside the boundary are set to 0. Those grids where more than half of the area are out of boundary are dropped from the patch set.

**Baselines** Automated Statistical Downscaling (ASD) (Hessami et al. 2008) is a traditional method for statistical downscaling. ASD requires regression methods to predict population density pixel by pixel. We compared three ASD methods, which are logistic and lasso regression, support vector machine (SVM) regression and artificial neural network (ANN) regression. Each method uses the density of lower resolution and PoIs to predict the higher

resolution population map. Due to the time complexity of SVM, we randomly chose 80000 pixels to train SVM model. A second set of methods, random forest-based dasymetric mapping approach (Stevens et al. 2015) and decision tree algorithm are applied to compare to our spatial mapping model. According to the approach described by Stevens et al. (Stevens et al. 2015), all the population data are transformed into log density. Higher resolution map is predicted by applying random forest regression on log population density of lower resolution and PoIs. Decision tree algorithm uses the same data processing approach.

**Metrics and Parameter Settings** We use 5-fold cross-validation in the experiment. For static model that consists of SRCNNs, the input data is obtained by concatenating population-grid maps with PoI-grid maps. The depth of the final matrix relies on how many PoIs categories we use, with all 4 categories as default. Except that  $38 \times 38$  patches are used in  $X_2-X_{fg}$  level and  $58 \times 58$  in  $X_1-X_2$  level, all SRCNNs are trained with the same set of parameters. Layer 1 consists of 64 filters of  $9 \times 9$  kernels, layer 2 consists of 32 filters of  $1 \times 1$  filters, and the output layer uses a  $5 \times 5$  kernel. Higher resolution models which have a greater number of sub-images may gain from larger kernel sizes and an increased number of filters. Each network is trained using Adam optimization with a learning rate of  $10^{-4}$  for the first two layers and  $10^{-5}$  for the last layers, and MSE loss as the loss function for every training step.

Each model is trained for  $10^5$  iterations with a batch size of 512. Tensorflow is utilized to build and train DeepDPM. Each SRCNN is trained independently on a Titan X GPU, and the inference is then executed sequentially on a single Titan X GPU.

In order to measure the performance of our structure and other traditional methods in comparison, we use a few key metrics to show static model’s applicability. Root mean square error(RMSE) and Pearson’s correlation(CORR) are used to measure the prediction quality. We also use normalized root mean square error(NRMSE) for inner comparison later.

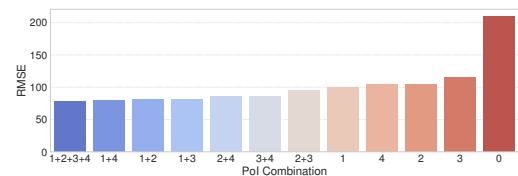
### RQ1: static population mapping performance

**Overall Model** Without considering dynamic changes in distribution during the day, we first train an overall model with all data available in weekdays to compare DeepDPM with other baseline methods.

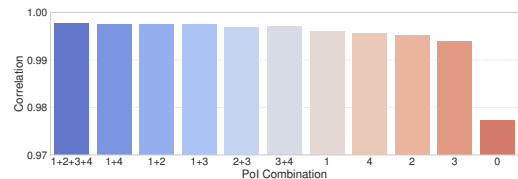
Our experiment compares performance with another six approaches, static model, random forest, decision tree, svm, ann and lasso, presented on Table 2 The three metrics discussed above are computed at all time slots in the test set and the averages are collected. We find that our static model outperforms all other methods in both  $X_1-X_2$  level and  $X_2-X_3$  level in terms of all three metrics. In detail, random forest gives the best prediction among all traditional methods, and is slightly outperformed by our static model by a difference in RMSE for no more than 3.0 in  $X_1-X_2$  level. While the difference enlarges as the number of stacked SRCNNs increases, which is about 14.6 in  $X_1-X_3$  level. In terms of correlation, decision tree, random forest and our model all per-

form well. Ann costs the longest run time, while it performs poorly compared to others. Stacked convolution neural network shows its strong ability in describing spatial structure.

**Poi Influence on Model Performance** It is important to choose correct type and amount of PoIs as augmentation before training. We run our experiment based on different combinations of PoI and with completely no PoI as presented in Figure 5 Generally, performance gets promoted rapidly as PoI usage increases. The result verifies the hypothesis that the more information augmented in PoIs we add into our model, the more precise our predictions will be. Using all four categories gives the best result. In detail, we find out that the entertainment PoIs play the most important role, with residential PoIs following when different categories are considered alone. The combination usage of such two PoIs also prove to outperform other bi-combinations. The model without using any PoI performs terribly. Local functions of different regions prove to be an important factor in describing population distribution pattern.



(a) RMSE performance on different poi combinations.



(b) Results on different poi distribution.

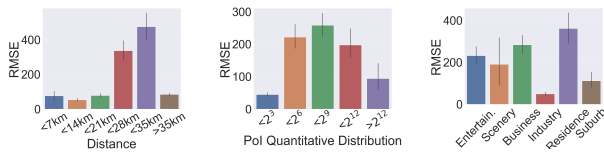
**Figure 5:** Comparison of predictive ability using DeepDPM with different PoI combinations. #1, #2, #3, #4 stand for entertainment, business, transportation, residence PoIs separately. #0 stands for prediction with completely no PoI usage.

Besides measuring global predictive ability based on different PoI usage, we also test local performance in different regions in our default model considering all 4 PoIs, shown in Figure 6. Figure 6(a) shows the relationship between the RMSE with the distance to the center of downtown (the grid where there is the highest population density). It turns out that we can reach high mapping accuracy in both suburbs and downtown areas, however the performance descends rapidly in the joint places. This is for population distribution change frequently in these places, which makes it hard to predict population distribution in finer scales. Figure 6(b) shows the relationship between the performance and the amount of PoIs locally. Figure 6(c) shows the relationship between local performances with the functions of local region. Industrial parks and suburbs turns out to have a much

Method	District to Street-Block( $X_1-X_2$ )			District to Fine-Grained( $X_1-X_3$ )		
	RMSE	NRMSE	Corr	RMSE	NRMSE	Corr
Lasso	513.2714	2.5975	0.7966	697.3197	3.5237	0.7144
ANN	465.7865	2.3573	0.8362	679.8032	3.4352	0.7334
SVM	850.5768	4.3045	0.2658	1002.3468	5.0650	0.2215
DecisionTree	51.5804	0.2610	0.9982	117.1829	0.5921	0.9931
Random Forest	47.0408	0.2381	0.9985	93.5023	0.4725	0.9956
Static Model	<b>44.5574</b>	<b>0.2255</b>	<b>0.9987</b>	<b>78.9081</b>	<b>0.3987</b>	<b>0.9978</b>

**Table 2:** Comparison of predictive ability between all six methods for all time slots in the dataset. All four PoIs are used in the experiments.

better performance than other regions, for the population distributions in these areas are much steadier than those in other regions as time changes, while residence regions suffer from frequent population movement.



(a) Results on down-town and suburb. (b) Results on different PoI distribution. (c) Results on different functional zones.

**Figure 6:** Comparison of local prediction performance using DeepDPM in different regions.

**Segmented Model** Considering different distribution patterns at different time slots in a day, temporal changes might have a strong impact on the prediction. We manually separate the entire dataset into three parts, which represent a specific period each, to further investigate the influence of different time in a day on our model precision. Period intervals include 0:00-7:00, 7:00-17:00 and 17:00-24:00. Considering the length of this paper, we only show the comparison results in Period2(7:00-17:00) in Table 3.

We find that DeepDPM still outperforms other baseline algorithms in three segmented models. While compared to the overall model using all time slots, the segmented one reached better performance. It is because population is more steady within a limited period. It proves that a flatter temporal trend in a fixed period in a day helps to improve predictive ability in the static model, for the static model itself doesn't take temporal changes into account.

However, the improvement from time slot segmentation is not enough to evaluate temporal changes in population distribution. We further put forward our dynamic population mapping model and conduct experiments to solve the problem addressed.

## RQ2: dynamic population mapping performance

**Quantitative Results** We use time-embedded LSTM to generate our temporal model. Fine-grained results from static model in  $X_2 - X_3$  level are the input sent to the model. Table 4 shows the prediction performance in terms

of RMSE, NRMSE and MAE. The initial results from the static model and prediction using flat LSTM are shown as baselines.

Compared with our static model, both two LSTM models showed their advantage of their powerful ability in sequence modeling. We find that the the LSTM+Time Embedding model reduces NRMSE by 14.5%. This suggests that the strong time-sequence regularity in population that our static model doesn't captured can be modelled well in our temporal model. As shown in (a) in Figure 7, NRMSE changes in all three methods in all time slots of a day are illustrated. Our temporal model outperforms other two models at almost every time slot, except that LSTM performs as well as it at several moments at around 4:30 and 17:00. The general LSTM model has more accurate prediction than static mapping except from around 7:00 to 11:00. The temporal model trained based on the input from our static one can be able to model the complex sequential transition as well as holding attention on spatial patterns.

**Illustrated Cases** After a generalized analysis, we focus on detailed prediction performance in local areas. We choose a typical grid to study the predicted population series in a day time. (b) in Figure 7 shows the population change in a day time. The red curve stands for the result from our temporal prediction, from which we can tell the similarity more visually. The blue curve represents prediction result from our static model, which almost remains at the same quantity, even though it shares the same undulating trend as the ground truth. The changing range in a day time of ground truth exceeds 1000 in population, while the static model only ranges no more than 100. It shows that the spatial modelling using super-resolution structure does lose sequential structures when capturing spatial regularity. While the population sequence pattern of our temporal prediction is much more similar to the ground truth.

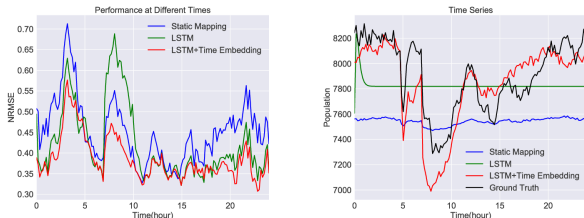
The case study explains the reason static model fails to predict temporal trend that the temporal one is able to. Since the distribution in all time slots are regarded as the same into SRCNNs, the system tends to average all population signals from different time slots, which results in a great temporal pattern loss. While our time-embedded LSTM temporal model overcomes the shortcoming, and retrieves it by time-based training. The whole DeepDPM system thus retains both spatial and temporal patterns in urban population distribution.

Method	District to Street-Block( $X_1 - X_2$ )			District to Fine-Grained( $X_1 - X_3$ )		
	RMSE	NRMSE	Corr	RMSE	NRMSE	Corr
Lasso	491.4463	2.4870	0.8025	667.8841	3.3749	0.7214
ANN	441.9888	2.2368	0.8440	641.2524	3.2404	0.7762
SVM	824.2706	4.1713	0.2733	969.5835	4.8995	0.2285
DecisionTree	44.0956	0.2231	0.9986	97.1856	0.4911	0.9953
Random Forest	42.8038	0.2166	0.9986	84.4569	0.4268	0.9961
Static Model	<b>40.7466</b>	<b>0.2062</b>	<b>0.9989</b>	<b>76.9615</b>	<b>0.3788</b>	<b>0.9980</b>

**Table 3:** Comparison of predictive ability between all six methods for time slots in period 2, from 7:00 to 17:00 every day.

Method	RMSE	NRMSE	MAE
Static Mapping	86.87	0.4517	32.28
LSTM	81.46	0.4236	31.60
LSTM+Time Embedding	74.27	0.3862	32.01

**Table 4:** Performance comparison of static and dynamic population mapping in Shanghai.



(a) Performance comparison in terms of global NRMSE at different time slots. (b) Case study: population series at one typical grid for all three model predictions and ground truth.

**Figure 7:** Performance of Dynamic population Mapping in Shanghai.

## RELATED WORK

Two major fields are related to our study.

*Fine-Grained Population Mapping:* Early studies (Anderson and Anderson 1973; Hessami et al. 2008; Sutton et al. 2001) used remotely sensed information, such as satellite imagery, as the main data source. While Azar et al.; Chen chose to refine census population distribution using ancillary data. As the state-of-the-art method in the field, R.Stevens (Stevens et al. 2015) used random forest (Liaw, Wiener, and others 2002; Breiman 2001) algorithm as a dasymetric redistribution approach based on both census and remotely sensed data. Compared to their studies, our DeepDPM uses coarse population data like census data, and PoI data as augmented data, which are much easier to obtain. Besides, our study breaks through the limitation in the time dimension.

*Image Super-Resolution:* Early studies used filtering approaches, e.g. linear, bicubic or Lanczos (Duchon 1979) filtering. Freeman, Jones, and Pasztor and Freeman, Pasztor, and Carmichael firstly sought to construct mapping algorithm between training patches and corresponding known

high-resolution counterparts. In recent years, convolutional neural networks(CNN) based SR algorithms have shown excellent performance (Wang et al. 2015; Dong et al. 2016; Wang et al. 2016; Kim, Kwon Lee, and Mu Lee 2016), where SRCNN (Dong et al. 2016) is one of the state-of-the-art for the problem. Vandal et al. successfully used the SRCNN based DeepSD structure in climate prediction. Our study also learns from the advantage of SRCNN to construct our static part of DeepDPM structure, and furthermore implements the dynamic part to learn the temporal pattern.

## LIMITATION AND FUTURE WORK

Currently, we are using mobile dataset to represent HRPD. However, the premise of our preprocess method is the hypothesis that the urban area has no explicit inside or outside population flows. This may be a major source of systematic error. We will consider more practical approaches to quantify the gain and loss of our current method, and explore more to reduce the error.

## Conclusion

In this paper, we investigate population mapping in both static and dynamic view using PoI as augmented information. We propose a deep learning based model to generate a complete population mapping structure, DeepDPM, which has two novel characteristics compared to previous studies and methods: 1) a stacked SRCNN based static model that evaluates static population prediction; and 2) a time-embedded LSTM based dynamic model that smooths the temporal change. Extensive experiments on the dataset of mobility data collected from Shanghai showed that DeepDPM significantly improves the performance compared to all other baselines. Meanwhile, our structure also breaks through the limitation in time dimension that previous studies had. As a result, population distribution in full time series is generated.

## Acknowledgments

This work was supported in part by The National Key Research and Development Program of China under grant 2017YFE0112300, the National Nature Science Foundation of China under 61861136003, 61621091 and 61673237, Beijing National Research Center for Information Science and Technology under 20031887521, and research fund of Tsinghua University - Tencent Joint Laboratory for Internet Innovation Technology.

## References

- Anderson, D. E., and Anderson, P. N. 1973. Population estimates by humans and machines. *Photogrammetric Engineering* 39(2).
- Azar, D.; Graesser, J.; Engstrom, R.; Comenetz, J.; Leddy Jr, R. M.; Schechtman, N. G.; and Andrews, T. 2010. Spatial refinement of census population distribution using remotely sensed estimates of impervious surfaces in haiti. *International Journal of Remote Sensing* 31(21):5635–5655.
- Breiman, L. 2001. Random forests. *Machine learning* 45(1):5–32.
- Chen, K. 2002. An approach to linking remotely sensed data and areal census data. *International Journal of Remote Sensing* 23(1):37–48.
- Dong, C.; Loy, C. C.; He, K.; and Tang, X. 2016. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38:295–307.
- Duchon, C. E. 1979. Lanczos filtering in one and two dimensions. *Journal of applied meteorology* 18(8):1016–1022.
- Ficek, M., and Kencl, L. 2012. Inter-call mobility model: A spatio-temporal refinement of call data records using a gaussian mixture model. In *INFOCOM, 2012 Proceedings IEEE*, 469–477. IEEE.
- Freeman, W. T.; Jones, T. R.; and Pasztor, E. C. 2002. Example-based super-resolution. *IEEE Computer graphics and Applications* 22(2):56–65.
- Freeman, W. T.; Pasztor, E. C.; and Carmichael, O. T. 2000. Learning low-level vision. *International journal of computer vision* 40(1):25–47.
- Gaughan, A.; Stevens, F. R.; Linard, C.; Jia, P.; and Tatem, A. J. 2013. High resolution population distribution maps for southeast asia in 2010 and 2015. In *PloS one*.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778.
- Hessami, M.; Gachon, P.; Ouarda, T. B.; and St-Hilaire, A. 2008. Automated regression-based statistical downscaling tool. *Environmental Modelling & Software* 23(6):813–834.
- Hochreiter, S., and Schmidhuber, J. 1997. Long short-term memory. *Neural Comput.* 9(8):1735–1780.
- Isaacman, S.; Becker, R.; Cáceres, R.; Martonosi, M.; Rowland, J.; Varshavsky, A.; and Willinger, W. 2012. Human mobility modeling at metropolitan scales. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, 239–252. Acm.
- Kim, J.; Kwon Lee, J.; and Mu Lee, K. 2016. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1637–1645.
- LeCun, Y.; Boser, B.; Denker, J. S.; Henderson, D.; Howard, R. E.; Hubbard, W.; and Jackel, L. D. 1989. Backpropagation applied to handwritten zip code recognition. *Neural Computation* 1(4):541–551.
- Liaw, A.; Wiener, M.; et al. 2002. Classification and regression by randomforest. *R news* 2(3):18–22.
- Lipton, Z. C.; Berkowitz, J.; and Elkan, C. 2015. A critical review of recurrent neural networks for sequence learning. *Computer Science*.
- Nasrollahi, K., and Moeslund, T. B. 2014. Super-resolution: a comprehensive survey. *Machine Vision and Applications* 25(6):1423–1468.
- Stevens, F. R.; Gaughan, A. E.; Linard, C.; and Tatem, A. J. 2015. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PloS one* 10(2):e0107042.
- Sutton, P.; Roberts, D.; Elvidge, C.; and Baugh, K. 2001. Census from heaven: An estimate of the global human population using night-time satellite imagery. *International Journal of Remote Sensing* 22(16):3061–3076.
- Vandal, T.; Kodra, E.; Ganguly, S.; Michaelis, A.; Nemani, R.; and Ganguly, A. R. 2017. DeepSD: Generating high resolution climate change projections through single image super-resolution. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '17*, 1663–1672. New York, NY, USA: ACM.
- Wang, Z.; Liu, D.; Yang, J.; Han, W.; and Huang, T. 2015. Deep networks for image super-resolution with sparse prior. In *Proceedings of the IEEE International Conference on Computer Vision*, 370–378.
- Wang, Y.; Wang, L.; Wang, H.; and Li, P. 2016. End-to-end image super-resolution via deep and shallow convolutional networks. *arXiv preprint arXiv:1607.07680*.
- Wu, S.-S.; Qiu, X.; and Wang, L. 2005. Population estimation methods in gis and remote sensing: A review. 42:80–96.
- Xu, F.; Zhang, P.; and Li, Y. 2016. Context-aware real-time population estimation for metropolis. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '16*, 1064–1075. New York, NY, USA: ACM.
- Yang, C.-Y.; Ma, C.; and Yang, M.-H. 2014. Single-image super-resolution: A benchmark. In Fleet, D.; Pajdla, T.; Schiele, B.; and Tuytelaars, T., eds., *Computer Vision – ECCV 2014*, 372–386. Cham: Springer International Publishing.
- Zheng, Y.; Li, Q.; Chen, Y.; Xie, X.; and Ma, W.-Y. 2008. Understanding mobility based on gps data. In *Proceedings of the 10th international conference on Ubiquitous computing*, 312–321. ACM.