Remote Sensing and Social Sensing Data Fusion for Fine-Resolution Population Mapping With a Multimodel Neural Network

Luxiao Cheng, Lizhe Wang, *Fellow, IEEE*, Ruyi Feng, *Member, IEEE*, and Jining Yan

***Abstract*—Mapping population distribution at fine spatial scales is significant and fundamental for resource utilization, assessment of city disaster, environmental regulation, and urbanization. Multisource data produced by remote and social sensing have been widely used to disaggregate census information to map population distributions at fine resolution. However, it is challenging to achieve accurate high-spatial-resolution population mapping by combining multisource data and considering geographic spatial heterogeneity. The existing approaches do not consider global and local spatial information simultaneously, resulting in low accuracy. This article proposes a multimodel fusion neural network for estimating fine-resolution population estimates from multisource data. Our approach takes into account the local spatial information and global information of each geographic unit. Specifically, a first-order space matrix of a geographic unit is used to characterize its local spatial information. We propose a multimodel neural network, which combines a convolutional neural network and a multilayer perceptron (MLP) model to estimate a fine-resolution population mapping. Using Shenzhen, China, as the experimental setting, a population distribution map was generated at a 100-m spatial resolution. The model was quantitatively validated by showing that it captured the relationship between the estimated population and the census population at the township level ($R^2 = 0.77$) more accurately than the WorldPop dataset ($R^2 = 0.51$) and the MLP-based model ($R^2 = 0.63$). Qualitatively, the proposed model can identify differences in population density in densely populated areas and some remote population clusters more accurately than the WorldPop population dataset.**

***Index Terms*—Convolutional neural network (CNN), multimodel neural network, population mapping, population spatialization, remote sensing.**

I. Introduction

THE spatial distribution of population refers to the distribution of population in a certain period. For the purposes of resource utilization, city disaster assessment, regulatory environment, and urbanization, it is of fundamental importance to know precisely where people live and how many people live there. While a survey-based census can provide a comprehensive overview of a country's population, the spatial resolution of population distribution divided by administrative divisions is low, not suitable for fine-grained urban management, and expensive to implement [1], [2]. Finer spatial population mapping beyond the limits of geography enables the spatial transformation of census data based on administrative units to a regular grid [3], which is useful for resolving problems involving both natural resources and population distribution [3], [4]. High-resolution population mapping is, therefore, a key tool for sustainable urban development.

To properly address the questions "Where do people live?" and "How many people live there?" the essential requirement is fine-resolution population spatialization and the creation of a population distribution grid to supplement demographic data. The traditional solution involves population decomposition and population projection. Based on different data resources and research goals, population spatialization methods fall into two general types: areal interpolation and statistical modeling. Areal interpolation for population spatialization assigns the population of the statistical unit area into grid cells at a specific scale according to a calculated weight based on a specific data model [5], [6]. This approach has been extensively studied over the past few decades. Urban geography theory has revealed that population distribution in urban areas is closely related to multiple socioeconomic factors such as business districts, road density, and building density [7]. With advances in image processing and geospatial data enrichment, researchers can mine information from remote sensing imagery and socioeconomic data to produce accurate gridded population datasets at high spatial resolution [8]–[11]. The statistical model method utilizes remote sensing and geographic information system technology for population spatialization through multiple underlying factors. These two methods consider local spatial information and global information, respectively. According to Tobler's first law of geography, everything is related to everything else, but near things are more related to each other [12]. As well as the influence of global factors, population distribution is also affected by neighboring regions. Therefore, population distribution is affected by global and local spatial factors. In addition, there is a complex nonlinear relationship between population density distribution and multiple geographical factors. Making fuller use of local and global spatial information to improve the accuracy of population spatialization is a matter of urgency.

Manuscript received October 10, 2020; revised January 30, 2021 and April 17, 2021; accepted May 31, 2021. Date of publication June 3, 2021; date of current version June 23, 2021. This research was supported by National Natural Science Foundation of China under Grants U1711266 and 41925007. *(Corresponding author: Lizhe Wang.)*

The authors are with the Hubei Key Laboratory of Intelligent Geo-Information Processing and the School of Computer Science, China University of Geosciences, Wuhan 430074, China (e-mail: Chenglx@cug.edu.cn; Lizhe.Wang@gmail.com; fengry@cug.edu.cn; yanjn@cug.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3086139

The rapid development of deep learning has led to developments in many research fields, including image processing [13], image recognition [14], and semantic analysis [15], [16]. The main advantage of deep learning methods is their powerful capacity to automatically learn high-level features from large volumes of data, which is an essential step in bridging the gap among different data patterns at feature level. As convolutional neural networks (CNNs) exhibit excellent spatial feature extraction performance, they have great potential for fusion of multisource data to extract important spatial information. The main advantage of deep learning methods is their powerful capacity to automatically learn high-level features from large volumes of data, which is an essential step in linking different data patterns at feature level [17]. As CNNs exhibit excellent spatial feature extraction performance, they have great potential for fusion of multisource data to extract important spatial information.

To address the problem of fine-resolution population distribution based on multisource data, we designed a schema for population spatialization that includes multisource data processing, spatial data representation, a multimodel neural network for population estimation, and model verification methods. Specifically, the proposed approach uses a first-order space matrix of a geographic unit to represent local spatial information and to construct local spatial features of the units in question. The proposed multimodel neural network, which combines a CNN and a multilayer perceptron (MLP) model, takes account of both local and global spatial information by integrating multisource data. The CNN-based model extracts spatial dependence features from the first-order adjacency matrix, and the MLP-based model estimates a fine-resolution population mapping at a 100-m spatial resolution. After multisource data processing, the data of the two representations are inputted to the multimodel neural network, using the CNN-based and the MLP-based models to extract significant features. Those features are then fused, and the outputs are assigned to fully connected (FC) layers and the regression layer for predictive purposes. Available datasets (WorldPop) were used to evaluate the model both qualitatively and quantitatively, and the results confirm the model's effectiveness and efficiency. Finally, a thorough analysis of the method and results was conducted to gain insights into the fusion of local and global features of multisource spatial data. The contributions of this article can be summarized as follows.

1) We propose a schema for population spatialization including data processing, spatial data representation, population estimation modeling, and verification methods.
2) We propose a multimodel fusion neural network for effective automatic extraction of spatial signature features: a CNN-based approach that extracts spatial dependence features from the first-order adjacency matrix and an MLP-based approach that extracts global nonlinear features from factors.
3) To address the spatial representation of multiple heterogeneous data, we propose the use of a first-order adjacency matrix of each spatial unit to represent the local spatial information, making the proposed multimodel neural network more robust for population estimation without prior knowledge.

The remainder of this article is organized as follows. We review related works on fine-resolution population distribution with remote and social sensing data in Section II. In Section III, the methodology of population spatialization proposed in this article is introduced. Furthermore, it introduces the representation method of spatial data and multimodel neural network in detail. Section IV introduces the case study area and the multisource data, and the method to convert remote sensing data, points of interest (POIs), and road network of different categories into the raster layer. Our method was evaluated on an open dataset and the results are analyzed in Section V. Section VI concludes this article.

II. Related Work

The population distribution is a hot topic of multiple disciplines, including demography and geography. As accurate population spatialization is essential for effective use of social statistics, many methods of population spatialization have been proposed and applied.

The most basic approach is the areal interpolation method, which treats the population within an administrative area as evenly distributed [18]. This method imposes a discrete grid on the management area, and each cell in the grid is assigned a filling value equal to the total number of cells in the management area. The areal interpolation method can transform demographic data into finer spatial units [19], such as areal weighting [6] and dasymetric mapping [20]. Wright [21] proposed the zoning density method, combining auxiliary location information and iteratively dividing the area into multiple areas to estimate population density. These methods further assume an absence of spatial heterogeneity in population distribution, but spatial heterogeneity is a common issue in population spatialization.

The weight distribution scheme distributes the known population by creating a weight rather than distributing it evenly. This method extends the area interpolation method. Weighting schemes incorporate land surface characteristics (e.g., impervious surface coverage, slope, average rainfall, and land use/land cover) according to specific rules. Although the existing method has made some progress, it has the disadvantage of low generalization. However, as the traditional multiple regression model is global modeling, local regression models are used to study population distribution. For example, the geographically weighted regression approach to population spatialization takes account of local spatial features and has achieved better results [22].

Machine learning has recently been used to improve this approach and has become an effective tool for learning patterns from data [23]. Machine learning is not limited by statistical assumptions and can model complex and hidden relationships between multiple factor and target variables for classification, regression, and so on [24], [25]. As geospatial data are gradually enriched, machine-learning-based methods determine weights based on specific rules related to spatial characteristics at different levels (e.g., nighttime light (NTL) intensity [26], and POIs [27], [28]). Social sensing data can be used to improve the accuracy of population spatialization [29]. Remote and social sensing data can complement each other by offering

different urban surface information. These methods have produced many well-known high-resolution population datasets covering large geographic areas, including the Gridded Population of the World [30], the LandScan [31], [32], WorldPop [33], and the Global Human Settlement Population Gridded datasets [34].

Deep learning is increasingly used to solve problems in the field of computational urban planning [35]. As the development branch of machine learning, deep learning involves neural networks with multiple hidden layers extended by artificial neural networks. A recent local feature information extraction algorithm based on CNNs has shown excellent performance [36]–[38]. By fusing natural image models as a rich data target for transfer learning tasks, CNNs can be used for land use mapping [39]. Similarly, CNNs have been used in the United States to estimate PM2.5, to gather predictors of nearby locations, and to use the spatial correlation between predictors to estimate PM2.5 at unknown locations [40]. In similar fashion, we used the spatial correlation between predictors to realize the estimation of a high-resolution grid population map for planning applications. The excellent spatial feature extraction performance of CNNs can be exploited to extract local spatial features. For present purposes, we selected CNNs to construct a population spatialization model for extracting local spatial features.

This article proposes a multimodel fusion neural network to address the problem of population spatialization using multisource data. This model integrates first-order adjacent matrix information about geographic units, mapping them into latent representation vectors and low-dimensional representations of complex nonlinearities and fusing these to estimate the population at a 100-m spatial resolution.

III. Methodology

The workflow of schema is shown in the flowchart in Fig. 1, which includes four parts: multisource data preprocessing, models fit for training, and verification.

In this article, we consider remote sensing data, road network data, and POI data. After preprocessing the ancillary data (including POIs, NTL data, terrain data, and land cover data), the 100-m$\times$100-m grid value of the corresponding population distribution's influencing factors was extracted, including land use/land cover, slope, brightness of NTL, normalized difference vegetation index (NDVI) value, and distance to the nearest road and POIs. And data preprocessing and POI data details are described in detail in Section IV. The geographic attributes of each grid geographic unit comprise eight factors. Those factors were then inputted to the population spatialization models as an independent variable for training and the WorldPop dataset as the training label. The spatial representation of each geographic unit is used as input to the training model. Then, the output is the corresponding population value of the grid. Finally, we validated our method by fitting the results to township-level census data. The model results were compared with the published gridded population dataset (WorldPop) for accuracy verification and analysis.

A. Spatial Data Representation

In geography, the different dimensional information describing the geographic unit is called the first-order information [41]. The first-order information of a geographic unit refers more specifically to a given geographic unit with attributes or attribute vectors corresponding to it. General spatial data can be abstracted as a mapping $f : V \rightarrow S$. V represents a collection of geographic units, and S represents attribute vector collection. The corresponding data structure can be expressed as $< x, Attri >$. x represents the location of the geographic unit, and $Atrri$ represents the attribute value. From the first-order spatial matrix, the geographical pattern manifests as the spatial differentiation of different attributes. For present purposes, social media spatial data or remote sensing images were used in flexible spatial tensor data form for the set spatial unit as the input for the neural network model. To ensure the consistency of data units, the relative position of each input data unit was specified in the space-time dimension. The spatial tensor is a way of organizing various types of spatiotemporal data. Fig. 2(a) represents the multispatial data. The representation of spatial units is shown in Fig. 2(b): $v_1, v_2, \ldots, v_8, \ldots, v_i$ represent the irregular geographic units in the study area, and the dashed line represents the relationship (e.g., distance) between geographic units. It shows the first-order adjacency matrix geographic units of v_i. The first-order adjacency matrix of each geographic unit contains eight geographic units. Each geographic unit comprises eight attributes Atrri_{ijn}; i and j are spatial location attributes, i.e., longitude and latitude, respectively, and n are information attributes (e.g., slope, NDVI, and POIs). Thus, each geographic unit V_i can be represented as $V_i = (Attri_{ij1}, \ldots, Attri_{ijn})$. The first-order matrix used here is calculated on the basis of Euclidean distance. The first-order space matrix of each geographic unit is then represented as a high-dimensional tensor [see Fig. 2(c)]. The unit identified in red is any geographic unit [V_i in Fig. 2(b)], and the surrounding blue units [$V_1 \ldots V_8$ in Fig. 2(b)] represent the first-order adjacency geographic units of the red unit. The high-dimensional tensor representation of spatial data is used as input into the model.

B. Multimodel Neural Network

To take account of the effects of influencing factors on population distribution, the experiment used the WorldPop grid dataset and various influencing factors with high-dimensional tensor as model inputs. Representation, nonlinear mapping and reconstruction, and model output were obtained for all predictors. In short, to produce the fine-resolution population spatialization map, we designed a multimodel fusion neural network that includes a CNN with a limited number of layers, an MLP model, and three FC layers fusing the front two parts. The proposed multimodal fusion neural network includes an efficient and trainable structure, with several convolutional, pooling, and FC layers.

The overall architecture of the proposed multimodal fusion network model is shown in Fig. 3. The multimodel neural network F comprises three parts: an encoder layer to extract

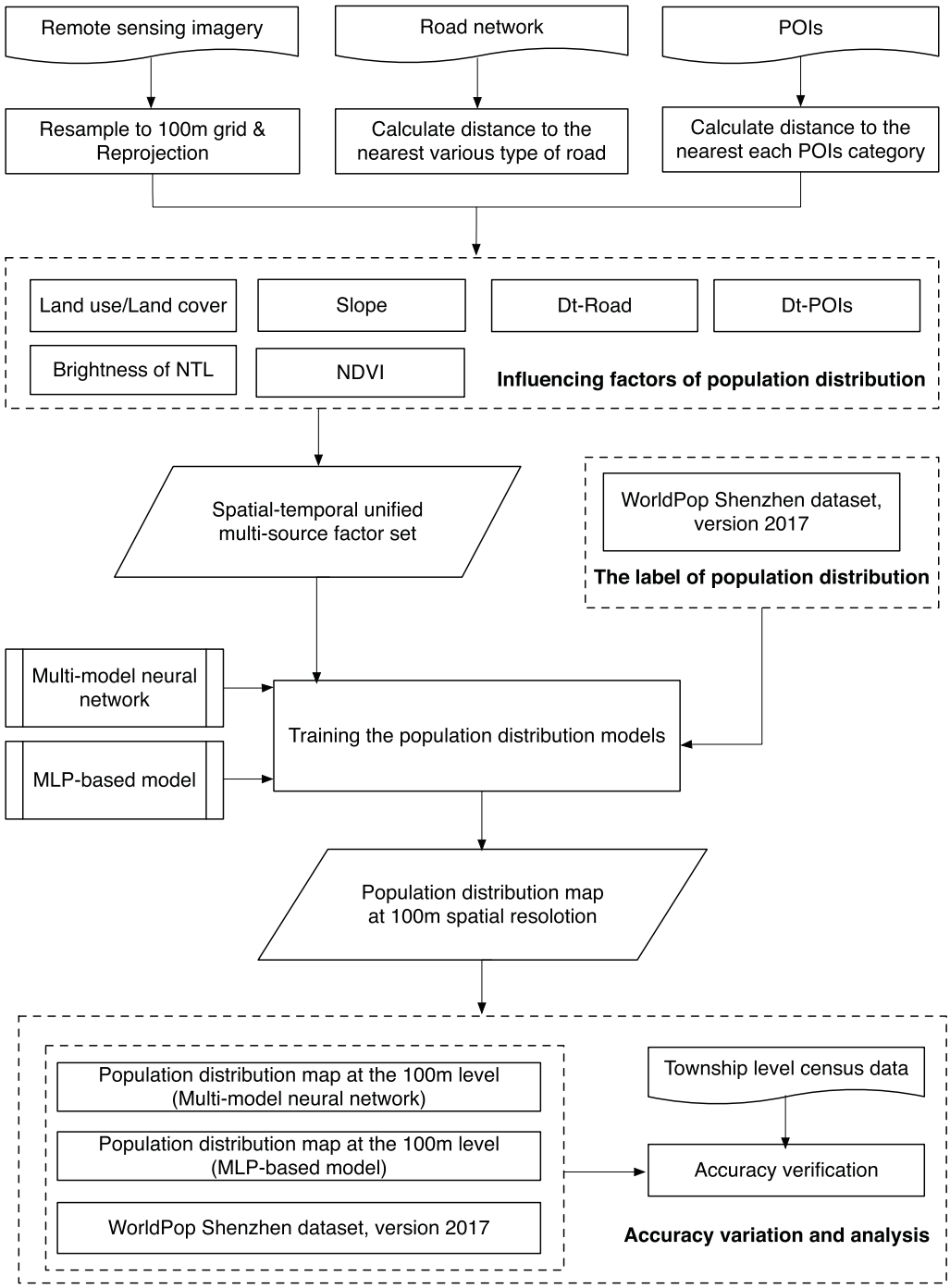


Fig. 1. Flowchart of the population spatialization process employed in the study.

features from the first-order adjacency matrix F_s, the central geographic unit encoder F_d, and the data fusion model F_f. The network takes the first-order spatial matrix S and central units D as inputs and outputs the predicted probability distribution P for all categories, which can be defined as $P = F(S, D) = F_f(F_s(S), F_d(D))$. Significantly, the six-layer CNN [see Fig. 3(a)] is designed as a spatial matrix feature encoder F_s, and the MLP model [see Fig. 3(b)] is the proposed global feature encoder F_d. The two extracted features are further fused and inputted to the FC layers and for regression prediction. The key role of the model is to learn a joint space, in which adjacent spatial features and global attribute features can be effectively combined to estimate population distribution. Auxiliary loss was used to enable the proposed multimodal fusion model to learn more coordinated multimodel features, making the model more robust to missing modalities. The composition and related configuration of the multimodel neural network model are described in detail in the following.

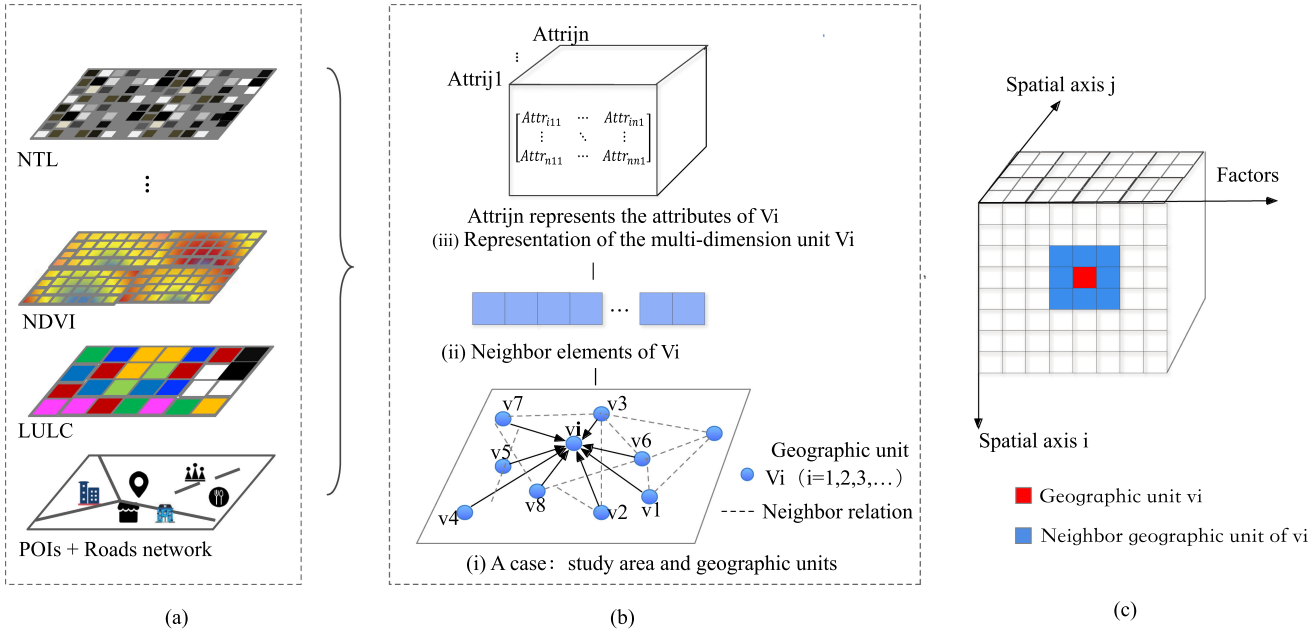


Fig. 2. Spatial data representation. (a) Multisource data. (b) Presentation of spatial units. (c) Input data representation of the model.

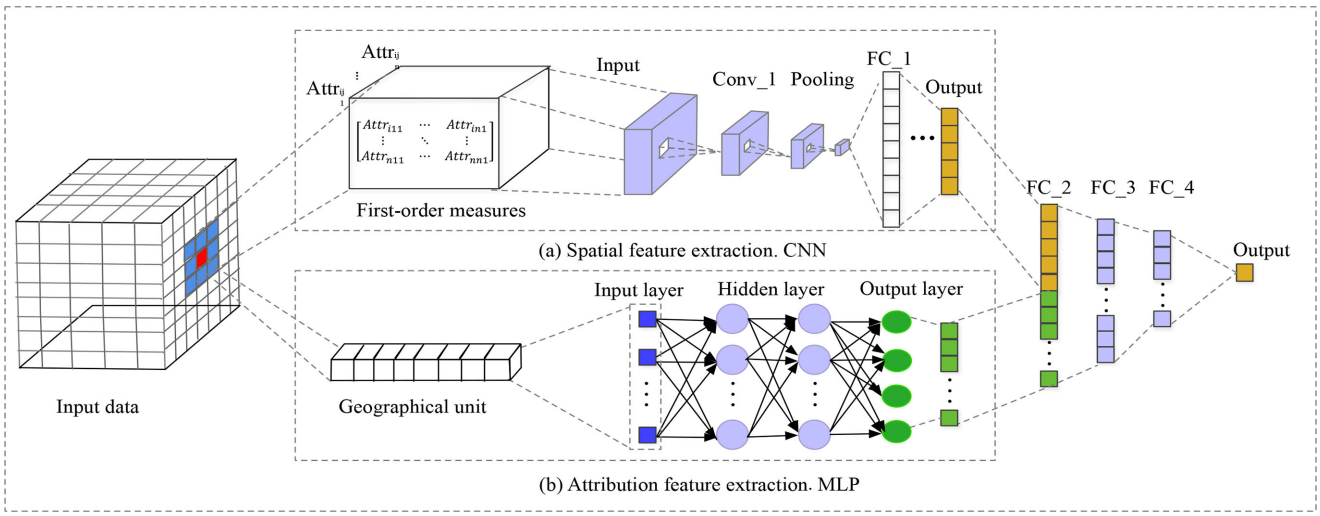


Fig. 3. Overview of the proposed multimodel fusion neural network, which includes (a) the CNN as the spatial feature model and (b) the MLP as the predictor attribute feature exaction of spatial units.

Unlike traditional linear regression models, the CNN is a nonparametric method that can model complex nonlinear relationships between predictions and heterogeneous predictor variables [14]. CNNs employ a layer-by-layer pretraining method to obtain initial parameters, adjusting and updating parameters continuously through backpropagation. A typical CNN consists of four layers: convolutional, activation, pooling, and FC layers. Convolutional layers have the characteristics of sparse local connections and weight sharing, using moving convolution filters that significantly reduce the number of network parameters and improve efficiency [42]. A weight-shared convolution layer is usually followed by a nonlinear activation layer to capture more complex characteristics of the input data. Based on relative invariance properties, pooling layers can aggregate information from adjacent regions with moving pooling filters, which reduce sensitivity to input offset and improve the robustness of characteristic data [43]. Finally, the CNN structure connects one or more FC layers to integrate, calculate, and output many of the features from the above processes.

1) *Convolutional layer:* Local connection and weight sharing in the convolutional layer play an essential role in extracting spatial features. The convolutional layer input consists of several convolution units, whose parameters are optimized by a backpropagation algorithm. As the input to the convolutional layer is an $m \times n \times l$ feature map, where $m \times n$ is the size of the feature map, and l is

the number of feature map channels. A convolutional layer has k filters of size $k \times k \times l$. The output is an $m \times n \times k$ feature map with k channels.

2) *Pooling layer:* The pooling layer takes the activations within a small spatial region of the feature map. It includes the maximum pooling and average pooling layers, which use the maximum or average operator, respectively, to extract the value of the spatial region. The pooling layer is also called the subsampling layer, which can extract high-level spatial features to represent the spatial context.
3) *FC layer:* The FC layer can be understood as a special convolutional layer that weights and sums the features learned in the previous convolutional and pooling layers.

The architecture of the proposed multimodel fused neural network is presented in Fig. 3. To overcome the limitations of multiple linear regression, we used the CNN, as shown in Fig. 3(a). The CNN takes input data as spatially corrected after predictors are spatially corrected. The model incorporates 18-D predictor vector for the grid cell that the model is trying to estimate and predictors from the neighboring grid cells. As shown in Fig. 2(b), when estimating the population value for the red grid cell of the input, we take all the predictors and the first-order neighbor grid cells of the red grid cell, about N blocks of the cells closest to the red grid cell. Note that these N block grid cells attribute input data representing the local spatial information of the estimated cell (red cell), not including the population value at the center. Now, give such $N \times M$ as input data. As shown in Fig. 3, the CNN model operates by convolution, pooling, and FC layers. All FC layers use dropout technology to improve the generalization ability of the model [44]. The convolution layer consists of multiple weight-shared kernels and a nonlinear function activation element. These layers are essentially nonlinear representations of the input data. More relative information presentation of the input data predictors can be obtained for grid cell estimation.

The number of convolution kernels in each convolution layer is 16, 32, 64. Moreover, each convolution layer consists of 3×3 kernel and zero padding to keep the channel size fixed. The output of each convolutional layer is sent to the FC layer to obtain a fixed-length representation and then processed by the FC layer. Finally, the final output 10-D features are used as the extracted local spatial features. Convolutional and FC layers can further extract higher level semantic features from convolutional features.

The purpose of population spatialization is to ensure that the model output resembles the actual population data as closely as possible. To that end, a standard backpropagation gradient descent method is used to minimize the root-mean-square error (RMSE) between the model output and the ground truth dataset to obtain high-resolution population distribution results. The township-level census data are constrained as ground truth data.

The CNN structure is initially configured as input layer–convolutional layer–pooling layer–FC layer–output layer. The size of the input matrix is 3×3, and the size of each convolution kernel and pooling kernel is 2×2. Both the CNN and the MLP adopt the Tanh and ReLU activation functions; an Adam optimizer, a learning efficiency of 0.01, and a dropout of 0.5, and the models are trained 500 times.

The population distribution estimation model in this study is based on grid-level variables (such as impact factor values and grid-level population distribution map). However, the total population of each county controls the actual population distribution. To reduce the error of the predicted population at the county level, we use the ratio of the county-level statistical population to the predicted population to correct the pixel-level predicted population. For each grid, population at county level can be corrected as follows:

$$P_i = D_i \times \frac{P_j}{D_j} \tag{1}$$

where i represents a geographic grid unit, j stands for one county-level administrative region, P_i is the corrected value of the geographic grid unit i, and D_i is the corresponding estimated value of the grid i before correction. P_j stands for a county's demographic value, and D_j is the total estimated gridded population of the corresponding county.

IV. Experiments

This article considers four datasets, i.e., census tract data, remote sensing data, road network, and POIs, which are available open datasets. This section introduces the study area and data information on the experiment in detail.

A. Study Area

China is a country with a large population, and rapid urbanization has substantially increased the population flow [45]. Shenzhen city is representative of the rapidly developing cities in China. A detailed description of the population distribution is indispensable in regional planning toward sustainability. In this article, we defined the study domain as Shenzhen city, which is a coastal city situated at the southern edge of Guangdong Province, South of China, is adjacent to Hong Kong, and covers an area of 1997.47 km^2. The region consists of nine districts with a total resident population of 13.0266 million. Shenzhen is located at a lower latitude, south of the Tropic of Cancer (113°43' E to 114°38' E, and 22°24' N to 22°52' N). The landforms in Shenzhen are mainly low mountains, flat platform, and terraced hills, and it has a subtropical marine climate with a mean annual temperature of 22.4°. The location of the study area is shown in Fig. 4. In addition, Shenzhen is the first special economic zone of China and has been developed into an international city with a certain degree of influence, creating a world-renowned "Shenzhen Speed." As China's first fully urbanized city and one of the most economically efficient cities in mainland China, Shenzhen city has witnessed rapid land expansion and urbanization in the past 30 years and has attracted a large number of people from all over the country in recent years.

B. Data and Data Processing

The effect of population distribution on the city is a complicated process involving multiple factors [46]. We extracted

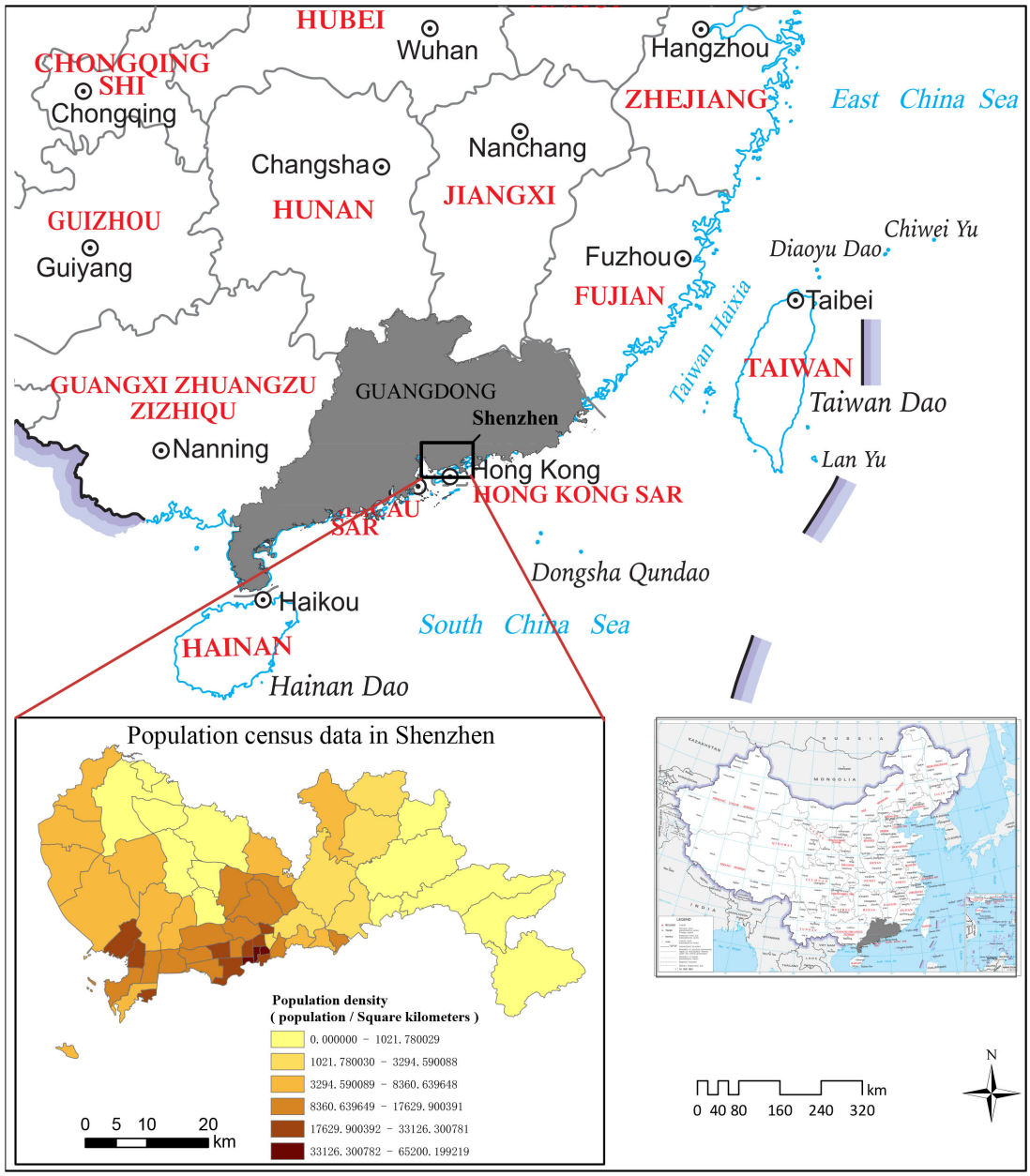


Fig. 4. Location of the Shenzhen in China and the population census data in Shenzhen.

complex relationships related to population distribution from remote sensing data, using POI data and road networks, to estimate fine-resolution population distribution. For the satellite-derived dataset, we collected NTL products, which to some extent characterize a region's economy [47], [48]. Additionally, land cover, land use, and terrain data partly capture human activities and livability, respectively. The NDVI dataset partly captures vegetation coverage, reflecting an area's greenness; humans tend to live in places with good air and green characteristics [49]. Social sensing data, POIs, and road networks can indicate an area's prosperity. All of the above factors are related, in complicated ways, to population distribution [50]. Given the imbalance and spatial heterogeneity of POIs and road networks in the data collection process, we followed previous work [46] and separately calculated the distance to the nearest POIs and road network from the center of each geographic unit (abbreviated as Dt-POIs and Dt-Road). Predictors from multisource data include land-use variables, digital elevation model (DEM) variables, NDVI variables, Dt-POIs, and Dt-Road network, as detailed in Table I.

1) Population Census Dataset and Administrative Boundaries: Census data are the best available baseline data on populations [51]. This study collected the 2017 resident demographic data at the township level, which is available from the yearbook. Moreover, the yearbook can be available from the Shenzhen Statistics official website. The population of Shenzhen is 13.0266 million, covering 73 township-level regions in 2017. The population census data include the township-level resident population of Shenzhen, which is shown in Fig. 4 (Population

TABLE I
MULTISOURCE DATA FOR EVALUATING POPULATION DISTRIBUTION

Indicator	Spatial	Data source	Temporal
Night time light (NTL) intensity	500-m	https://ngdc.noaa.gov/eog/viirs/	2017
NDVI	1-km	http://www.resdc.cn/	2017
Land use and land cover (LULC)	30-m	http://data.ess.tsinghua.edu.cn/fromglc2017v1.html	2017
Slope slope from DEM	90-m	http://www.resdc.cn/	2000
Distance to nearest POIs (Dt-POIs)	/	http://ditu.amap.com/	2017
Distance to nearest road (Dt-Road)	/	https://www.openstreetmap.org/	2017

census data in Shenzhen). The census data are used to verify the experimental results.

The boundary map at county and township levels (scale of 1:4 000 000) were obtained from the Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences. As the census dataset exhibited some inconsistencies with county-level administrative boundary maps due to administrative changes such as the change of county name and county range adjustment, we align the census population with the areas on the administrative boundary map. Then, we spatially combine the census data with the corresponding administrative boundary vector elements to realize the spatialization of the census data, which is shown in Fig. 4. The administrative boundary vector data are used to extract the research area in the data preprocessing part.

2) Remote Sensing Data: The NTL products were obtained from the United States National Oceanic and Atmospheric Administration's Earth Observation Group (https://ngdc.noaa.gov/eog/viirs/). The images represent the average intensity of NTL with digital number value ranging from 0 to 63. Compared with other widely used Defense Meteorological Satellite Program (Operational Line-Scan System) products, the Visible infrared Imaging Radiometer NTL product comprises different fixed-gain images to avoid several saturation problems from emerging in urban areas while maintaining dim lights in suburban areas. This product has a spatial resolution of 500 m. Thus, the NTL product was calibrated to allow for cross-year analysis and resampled to 100 m using the neighbor approach to avoid changing any pixel value during the resampling process.

The NDVI products can accurately reflect the land surface's vegetation coverage. The annual NDVI products of China were obtained from the Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (available from http://www.resdc.cn/DOI /). The annual NDVI products are based on the continuous-time series of SPOT/VEGETATION NDVI satellite remote sensing data generated by the maximum value composite method [7], which has a spatial resolution of 1 km. Both the NTL and NDVI products were resampled to 100 m using the nearest neighbor approach.

Land cover data are essential data of global resource surveys, land cover study, land surface monitoring, and land-use patterns. The land cover products of Guangdong province in 2017 were download from the Finer Resolution Observation and Monitoring of Global Land Cover (FROM-GLC) dataset (available from http://data.ess.tsinghua.edu.cn /), which was produced by Gong *et al.* [52]. The FROM-GLC is a 30-m spatial resolution global remote sensing land cover product. It includes ten categories: cropland, forest, grassland, shrubland, wetland, water, tundra, impervious surface, bare land, and snow/ice. In this study, cropland, grassland, and impervious surface land types were extracted through attribute values. The area proportion of each type in each county level was calculated as the land cover factor. The extracted data were resampled to 100-m grids. The proportion of different land types in each grid was calculated as the land cover factor for population estimation, prepared for the following experimental section.

The terrain data (DEM) of spatial distribution in China come from Shuttle Radar Topography Mission (SRTM). Consider that the SRTM dataset has the advantages of vital reality and free access. Many researchers around the world use it to carry out the environmental analysis. The radar image data obtained by the SRTM system cover more than 80% of the Earth's land surface. After more than two years of data processing, a global digital terrain elevation model has been created. This dataset is based on the latest SRTM V-4.1 to organize and splice Chinese provincial data with a spatial resolution of 90 m. The terrain data (DEM) is obtained from the Resource and Environment Science and Datacenter, Chinese Academy of Sciences (available from http://www.resdc.cn/DOI /). Although these datasets were collected in 2000, it is considered that the terrain generally has not undergone much deformation in the past two decades. Those remote sensing images were calibrated to allow for cross-year analysis and resampled to 100 m using the nearest neighbor approach [46], [53] to avoid changing any pixel value during the resampling process.

3) Infrastructure Data: POIs capture geographic location attributes that indicate population concentration, including information such as the name, category, and location of geographic objects. POIs represent people's understanding of the functions and attributes of a specific place, and it is an important source of social sensing data [54]. Therefore, the social sensing data used in this article mainly refer to POI data. POI data contain urban infrastructure information. We collected POI data from the Gaode map service (http://ditu.amap.com/), which is China's largest navigation map service [55]. Those POI data used in our article are obtained through the open application programming interface (API) of Gaode Map (https://lbs.amap.com). This process requires applying for key authentication that binds the Gaode Map web service. According to the name of the city and POIs, the data can be downloaded from the certified

TABLE II
STATISTICS OF POIS CATEGORIES IN SHENZHEN

Category	Count
Education service	26271
Health and social service	46450
Commercial building	101298
Commercial facilities and services	147144
Park	1380
Government agency	15106
Railway station	23003
Hotel service	33287
Restaurant service	105603
Residential community	100269

open API. We obtained 827 768 records for 2017 using the Gaode Map (Amap) API. POI category statistics are detailed in Table II. The POI information here mainly includes education service (including school and educational institutions), health and social service (including hospital and clinics), commercial building (including companies and office buildings), commercial facilities and services (including shopping malls, shops, and banking institutions), parks, government agency, railway station (includes bus stations and subway stations), hotel service (including hotel buildings), restaurant service, and residential community.

In general, the range of POIs reflects the type of region as determined by urban development and historical factors, which, in turn, indirectly determine population distribution. Spatial heterogeneity becomes increasingly significant as spatial scale decreases. Urban road networks also indicate economic situation, which is another important influence on population distribution. We collected road network data from Open Street Map, including state highways, provincial highways, country roads, and township-level roads.

4) Data Processing: The remote sensing data are cropped by using the administrative boundary data collected above. Thus, the corresponding basic data of Shenzhen can be obtained. After performing area clipping and unifying the coordinate projection, we calculated all Dt-POIs and Dt-Road values, as referred to above. All distance layers for ten categories of POIs were output as raster layers at the 100-m spatial resolution. Road network data were used to generate the Euclidean distance to the nearest road. Finally, kernel density estimation with an initial bandwidth of 500 m was used to convert individual Dt-POIs and Dt-Road values to continuous and smooth density surfaces for each category. Density surface and distance were output as raster layers at the 100-m spatial resolution.

Given the spatial heterogeneity of units and the evaluation criteria for different factors, each factor was normalized in the 100-m grid resolution. These heterogeneous spatial data were projected consistently using a Universal Transverse Mercator (UTM) map projection with a unified spatial coordinate system: GCS_WGS_1984 and UTM zone 49N. In this way, we established a unified spatial dataset at the 100-m resolution. There are eight predictors considered in this article, including slope, brightness of NTL, NDVI, Dt-Road, Dt-POI density, and proportion of grassland, cropland, and impervious surface. Overall, the eight predictors for each grid cell were used to estimate the gridded population. The number of predictors was subsequently optimized by selecting the most important predictors from the list.

The open-access gridded population dataset was used: WorldPop Population data. The WorldPop population is produced combined AfriPop, AsiaPop, and AmeriPop mapping and started in 2013 with a spatial resolution of 100 m (http://www.worldpop.org.uk/). This open archive of spatial demographic datasets for Central and South America, Africa, and Asia is intended to support development, disaster response, and health applications.

C. Experimental Setup

All the experiments employed a TensorFlow framework. For the multimodel neural network comparison, we conducted two groups of experiments with the same data settings. In the first group, we used the proposed multimodel fusion network, taking account of spatial adjacency information. In the other group, we used the MLP-based model to train the dataset. For both groups, the processed data were used as training, and township-level population density was used as the validation set. The training and testing sets were randomly divided in a 3:7 ratio, while the generalization test set included the whole datasets. The trained model was then applied to population simulation, and the population of the corresponding location was outputted based on the global and local spatial features of each geographic unit.

The 100-m $\times$ 100-m raster layers, including slope, brightness of NTL, NDVI, Dt-Road, Dt-POI density, and proportion of grassland, cropland, and impervious surface, were aggregated and then linked with the natural logarithm of the WorldPop dataset to train the CNN model. The nonlinear relation was then used to disaggregate township-level census population (see Fig. 4) to pixels. To ensure the reliability of results, we conducted both sets of experiments three times, taking the average of the three as the final result for spatialization of the demographic data.

V. RESULTS AND ANALYSIS

A. Accuracy Assessment

This study simulated population distribution using a 100-m grid for population spatialization based on township-level census data. To assess the predictive ability and generalizability of the population spatialization models, Shenzhen was selected as the sampling area. Census data for 73 township-level areas in Shenzhen were selected for accuracy assessment.

The model was evaluated both quantitatively and qualitatively. For quantitative assessment, average absolute error (MAE) [see (2)], RMSE [see (3)], and relative root mean square error (%RMSE) [see (4)] were used to measure the accuracy of the models and public dataset. Among them, MAE is the relative absolute error and is averaged to avoid positive and negative

TABLE III
ACCURACY EVALUATION INDEX CALCULATION RESULT OF EACH MODEL

Model	MAE	RMSE	%RMSE
WorldPop	5409.53	16034.47	28.96%
MLP-based	4344.37	15849.63	27.87%
Multi-model neural network	3439.05	12269.30	22.36%

offset. %RMSE is obtained by dividing the RMSE by the average of the census numbers, which can reflect the accuracy of the model simulation

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y} - y| \tag{2}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y} - y)^2} \tag{3}$$

$$\%\text{RMSE} = \frac{\text{RMSE}}{\frac{1}{N}\sum r_i} \tag{4}$$

where $\hat{y}$ is the estimated value at the township level, y is the census density actual value at the township level that is the estimated population density obtained after population spatialization in this article, N represents the number of townships ($N = 74$), and r_i is the reference value of the ith group of data, i.e., the value obtained from census data.

The population density spatialization results for 73 township-level regions in Shenzhen were compared and analyzed using the WorldPop and MLP-based models. Error calculation results are shown in Table III. It can be concluded that the simulation achieved a higher overall accuracy for the chosen location than that of the WorldPop dataset and the MLP-based method.

The accuracies of the results of the three population models are shown in Table III. The errors of the three models are all within an acceptable range. In general, the WorldPop model has the lowest accuracy, and the average RMSE for the result is 28.96%. The MLP-based model is slightly better than the WorldPop on the whole. The multimodel neural network has the highest accuracy, with an %RMSE of 22.36%. Moreover, the MAE and RMSE are consistent with the performance of the RMSE.

To evaluate multimodel neural networks in terms of fine spatial resolution, we utilized county-level census data from the Guangdong provincial Bureau of Statistics for validation purposes. Similarly, for quantitative evaluation, we integrate the result mapping data at township level and fit the township data with the census data (see Fig. 5). This illustrates the relationship between predicted and census population data, where each data point corresponds to township-level census data. As compared to the WorldPop dataset ($R^2 = 0.51$) and the MLP-based model ($R^2 = 0.63$), the accuracy of the multimodel neural network is acceptable ($R^2 = 0.77$). Errors were derived mainly from underestimation in townships with large populations, while overestimation was common in smaller population areas. The estimated value reported here is also close to the census figure, indicating that these results are scientific and reasonable.

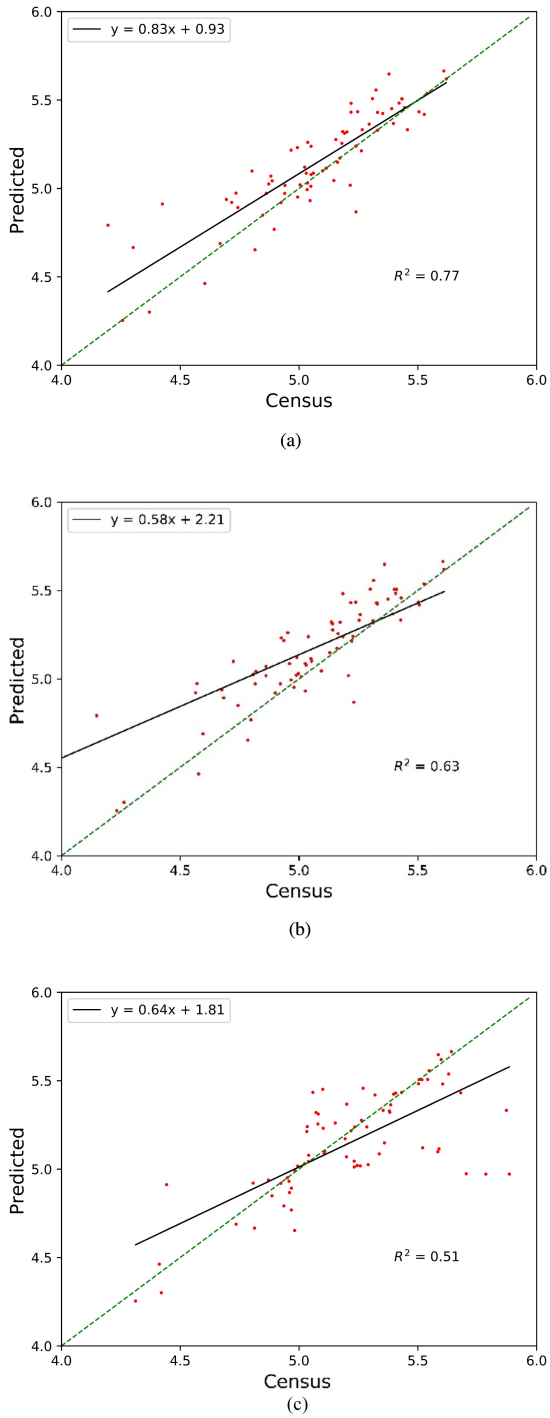


Fig. 5. Linear regression of the predicted and the census population density at the township level. The black line in each figure is the result of linear fitting. (a) Multimodel neural network estimation result. (b) MLP-based estimation result. (c) WorldPop estimation result.

This confirms that population distribution data with a spatial resolution of 100 m can facilitate comprehensive and effective management of population, resources, environment, and social economics, with important practical and theoretical implications for refined urban management.

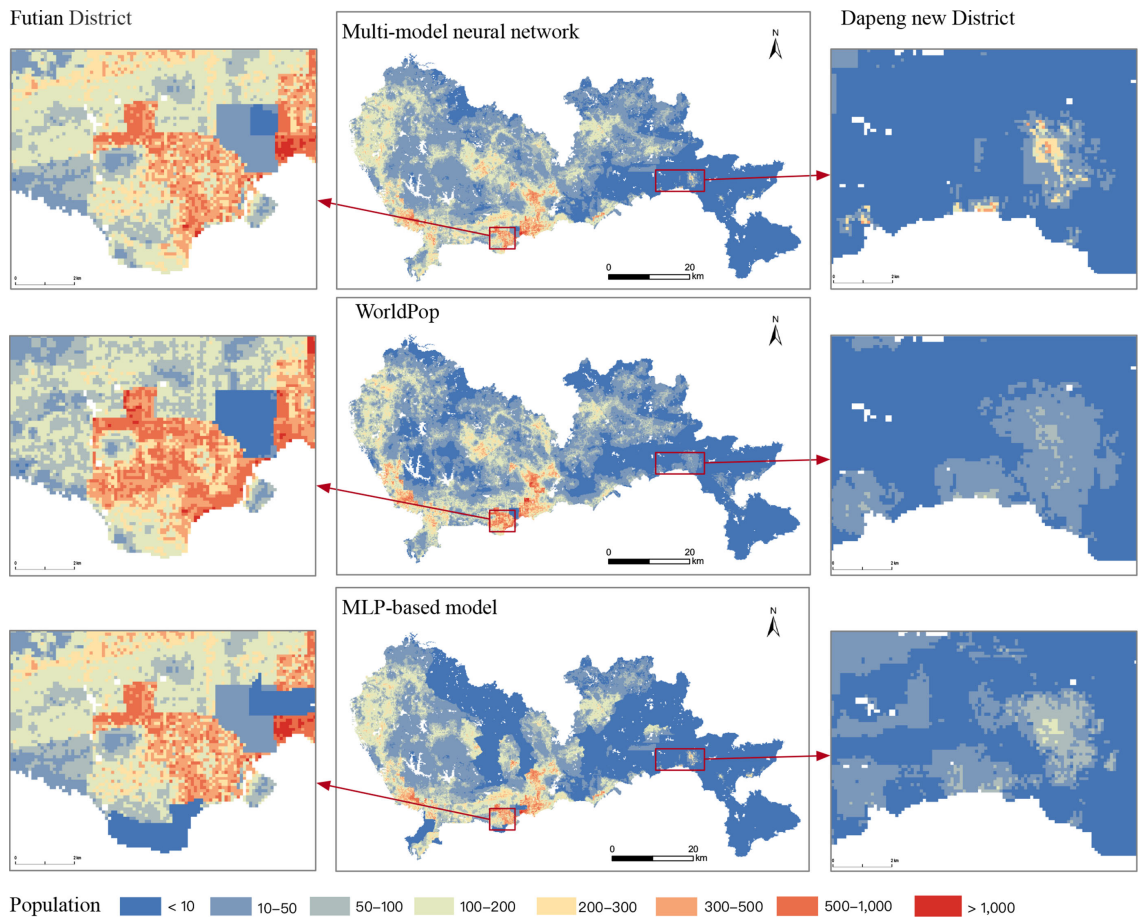


Fig. 6. Estimated population result at the 100-m spatial resolution displayed in Shenzhen (multimodel neural network, WorldPop, and MLP-based model), and the Futian District is highlighted on the left, and Dapeng New District is on the right.

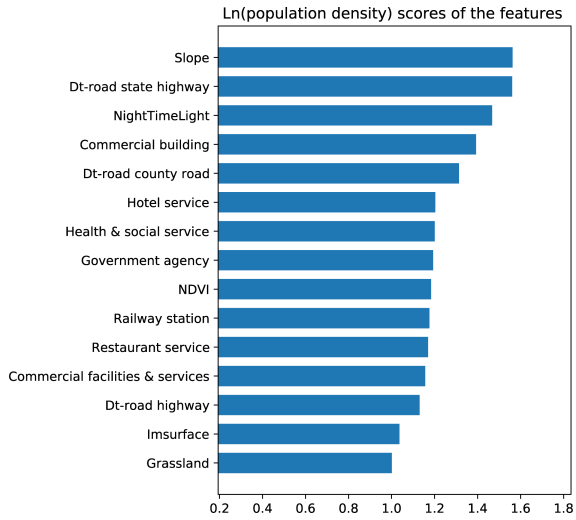


Fig. 7. Important predictor lists for the multimodel neural network model.

B. Qualitative Population Spatialization Map

Qualitatively, the residual error between the label and the estimation can be used to measure spatial error. Fig. 6 shows the Shenzhen population distribution as captured by different modeling methods, including the multimodel neural network, the random-forest-based WorldPop dataset, and the MLP-based model. The existing literature reports that the WorldPop gridded population dataset with fine spatial resolution is accurate for China. The general trend of these results is similar and can be said to describe the population distribution well. We compared and analyzed high-density and low-density areas of the study region. The high-density population distribution regions are mainly found in Futian District, Luohu District, and Longhua District, while the low-density regions are in Yantian District, Pingshan District, and Dapeng New District.

To highlight the competitive accuracy of our method at the 100-m spatial resolution, we selected results from Futian District (high-density) and Dapeng New District (low-density) (see Fig. 6). The three mapping results are roughly the same. However, when compared to the WorldPop and MLP-based mappings, the jagged edges of our model reveal a transition that is more natural and realistic. Additionally, our results provide more abundant information at the same scale about more heterogeneous population distribution in high-, medium-, and low-density regions. In the typical regions we selected (in Futian District and Dapeng New District), the results of the multimodel neural network show more spatial details and pattern differences for comparison. When compared to the functional zoning of the

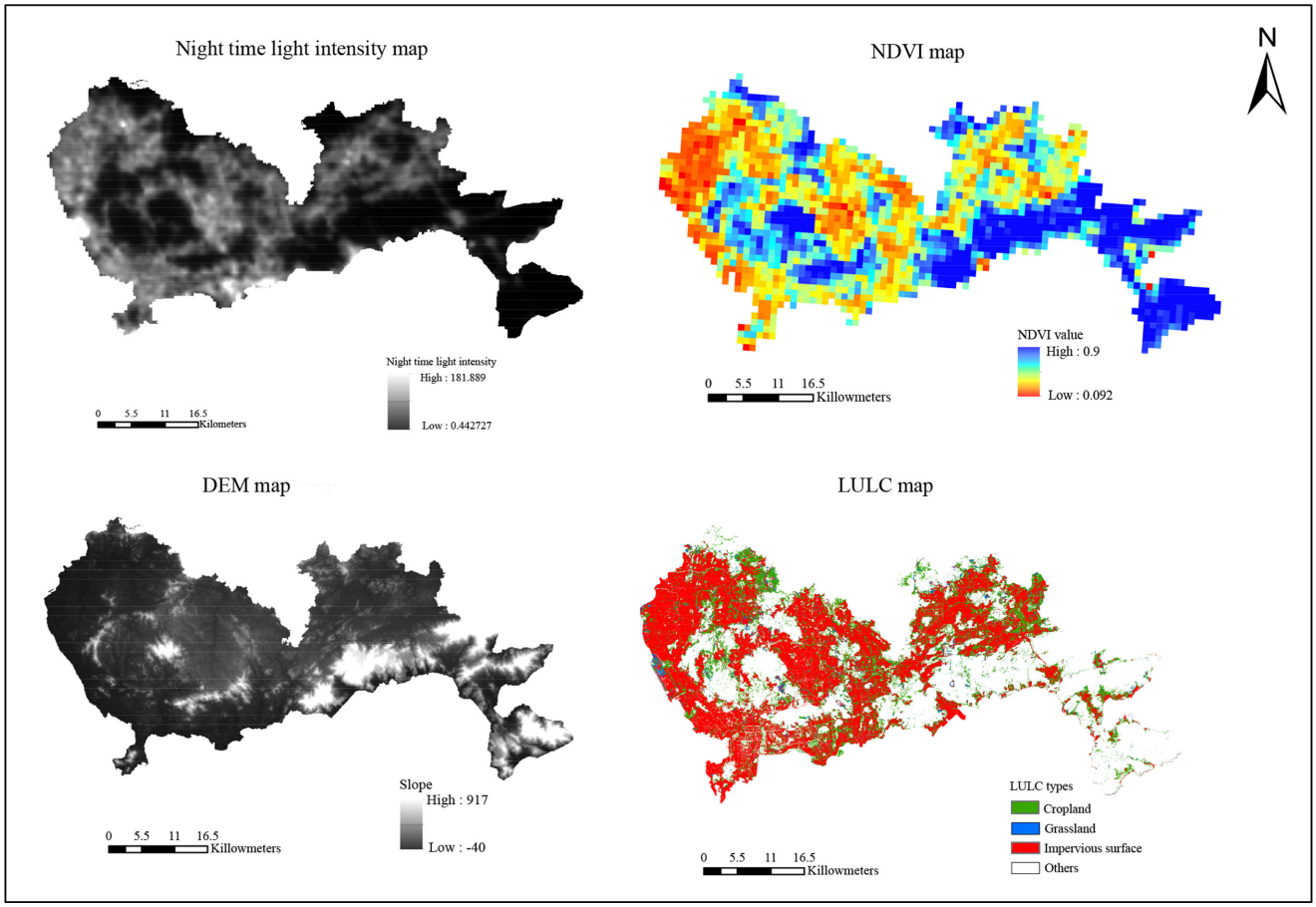


Fig. 8. Remote sensing data product map in Shenzhen.

Futian district, our model results are closer in spatial granularity. Compared with Google Maps urban functional divisions, from the spatialization results of the Futian District, our model results are closer in spatial granularity, while the distribution of the WorldPop dataset is too uniform and does not show the changes in population distribution well. From the spatialization results of the Dapeng New District, our model also clearly captures changing population density trends from communities and school areas to suburbs. However, neither the WorldPop dataset nor the MLP-based model captures high-density areas, resulting in a more even distribution from the high-density area (communities and schools areas) to suburban. The low-value population distribution of the MLP-based model is significantly higher than the WorldPop and multimodel neural network results. The distribution of high values and low values has district boundaries, while the multimodel neural network and WorldPop mapping results exhibit very similar morphological distributions.

C. *Responses of Population Distribution to the Variables*

Based on these multisource data, the multimodel neural network captures the importance of variable factors, as shown in Fig. 7, and confirms the strong positive correlation between population density and impacts. It shows that various influencing factors have different importance to population density distribution.

The above analysis aligns with earlier findings that terrain, road network, night lights, and commercial buildings all play an essential role in population distribution [46].

The brightness of night lights has a positive effect on population distribution. Comparing the results of the experiment with Fig. 8 (NTL intensity map), areas with high light brightness are often densely populated, but it does not rule out the light abnormal values in a special region. The NDVI indicates vegetation coverage on the Earth's surface. Human beings also live in areas with plenty of water resources and plants [56]. Fig. 8 (NDVI map of Shenzhen) shows that a region's NDVI value changes in population density area, but these changes are not significant, indicating that the NDVI has a certain impact on population distribution. However, when the NDVI exceeds a certain value, population density decreases with any further increase in the NDVI (which can be defined as dense forest) [57].

As Shenzhen's terrain is high in the southeast and low in the northwest, with the highest altitude of 943.7 m, we can see the topography trend from Fig. 8 (DEM map of Shenzhen). Most of Shenzhen's topography is low hilly land, with gentle terraces in the middle and a coastal plain to the west. Clearly, this terrain is an important factor in population distribution. As the high-slope areas are not suitable for building, living

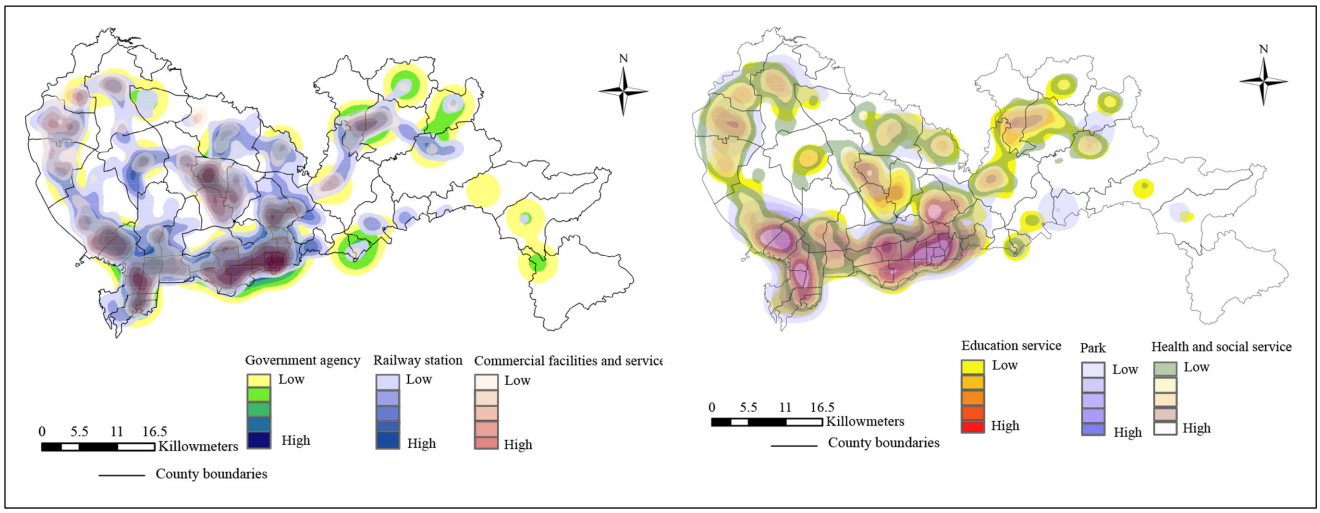


Fig. 9. POI data distribution map in Shenzhen.

infrastructure in these areas is relatively scarce. Instead, most human settlements are found in the plains. Compared with the above influencing factors, the area of impervious surface and grassland [see Fig. 8 (LULC map)] does not have great impact on population distribution.

In today's society, humankind's existence inevitably produces a demand for various services, which promotes the construction of various infrastructures. POI data are generated by human activities and are closely related to humans. In this article, POI data are regarded as a type of social sensing data. Due to the limitation of the length of this article, we visualized six types of POI distribution, as shown in Fig. 9. It can be seen that commercial building and hotel service together with social service have an important impact on population distribution.

Comparing the above influencing factors with the population distribution results, it can be concluded that the factors extracted from the remote sensing products and social sensing POI data have different degrees of impact on the population distribution. However, the POI data are biased and are not evenly distributed in the area, so we combine remote sensing data products that can fully cover the study area to avoid the results' bias. The results prove that the combination of multiple data can improve accuracy.

Since the reform and opening-up policy, the secondary and tertiary industries in the Pearl River Delta region with Shenzhen as the center have developed rapidly, and the city has also expanded rapidly. Shenzhen has become a hotspot for economic growth, attracting many immigrants, and implementation of urban infrastructures such as road construction and entertainment facility construction has improved. As transportation of materials and urban residents depends on roads, it follows that road network density is a key factor in population distribution; higher road network density improves accessibility and attracts more residents. Equally, as more residents move in, road network density increases to facilitate increased travel.

City development has an agglomerating effect. The built-up area formed by contiguous construction land attracts municipal facilities and stimulates commerce, leading ultimately to higher population densities. According to the distance attenuation theory, population becomes sparser as one moves farther from the built-up area, and Shenzhen confirms this theory.

VI. Conclusion

Constructing a fine-resolution population distribution map is a complicated process. While existing research makes extensive use of multisource data, global and local spatial information needs to be more fully used for accurate assessment of population distribution data, requiring improved data processing methods and spatialization models. In this article, we have designed a schema for population spatialization that includes data processing, spatial data representation, population estimation models, and model verification methods. Specifically, utilizing a first-order space matrix of a geographic unit to characterize spatial information and a spatial multimodel neural network, combining the CNN and an MLP model, we estimated a fine-resolution (100-m grid) population mapping. When compared to the MLP-based model and the WorldPop dataset, the proposed model performed more accurately than both in both quantitative and qualitative terms. Our results highlighted the importance of terrain slope, road networks, and commercial centers, which are widely accepted as good proxies of population distribution. Meanwhile, the generalization of our proposed model is mainly based on regions with similar economic development levels. Specifically, for urban areas with similar economic development, the trained model can be directly migrated. However, for regions with relatively large differences in economic development, we recommend using local relevant data to retrain the multimodel to estimate the spatial distribution of the population.

The population distribution product in 2017 with a 100-m spatial resolution in Shenzhen developed by this study is more consistent with the real population distribution under a complex surface environment. These data can provide more detailed data support for sustainability research. In the future, we will pay more attention to using historical surface information to construct dynamic spatial changes in the population.

REFERENCES

[1] C. Aubrecht, D. Özceylan, K. Steinnocher, and S. Freire, "Multi-level geospatial modeling of human exposure patterns and vulnerability indicators," *Natural Hazards*, vol. 68, no. 1, pp. 147–163, 2013.

[2] C. Zeng, Y. Zhou, S. Wang, F. Yan, and Q. Zhao, "Population spatialization in China based on night-time imagery and land use data," *Int. J. Remote Sens.*, vol. 32, no. 24, pp. 9599–9620, 2011.

[3] D. L. Balk, U. Deichmann, G. Yetman, F. Pozzi, S. I. Hay, and A. Nelson, "Determining global population distribution: Methods, applications and data," *Adv. Parasitol.*, vol. 62, pp. 119–156, 2006.

[4] A. J. Tatem, A. M. Noor, C. Von Hagen, A. Di Gregorio, and S. I. Hay, "High resolution population maps for low income nations: Combining land cover and census in East Africa," *PloS One*, vol. 2, no. 12, 2007, Art. no. e1298.

[5] W. R. Tobler, "Smooth pycnophylactic interpolation for geographical regions," *J. Amer. Statist. Assoc.*, vol. 74, no. 367, pp. 519–530, 1979.

[6] W. R. Tobler, U. Deichmann, J. Gottsegen, and K. Maloy, "World population in a grid of spherical quadrilaterals," *Int. J. Population Geography*, vol. 3, no. 3, pp. 203–225, 1997.

[7] D. Lu, H. Tian, G. Zhou, and H. Ge, "Regional mapping of human settlements in southeastern China with multisensor remotely sensed data," *Remote Sens. Environ.*, vol. 112, no. 9, pp. 3668–3679, 2008.

[8] P. C. Sutton, D. A. Roberts, C. D. Elvidge, and K. E. Baugh, "Census from Heaven: An estimate of the global human population using night-time satellite imagery," *Int. J. Remote Sens.*, vol. 22, no. 16, pp. 3061–3076, 2001.

[9] F. R. Stevens, A. E. Gaughan, C. Linard, and A. J. Tatem, "Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data," *PloS One*, vol. 10, no. 2, 2015, Art. no. e0107042.

[10] N. N. Patel *et al.*, "Multitemporal settlement and population mapping from landsat using Google earth engine," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 35, pp. 199–208, 2015.

[11] C. Robinson, F. Hohman, and B. Dilkina, "A deep learning approach for population estimation from satellite imagery," in *Proc. 1st ACM SIGSPATIAL Workshop Geospatial Humanities*, 2017, pp. 47–54.

[12] W. R. Tobler, "A computer movie simulating urban growth in the Detroit region," *Econ. Geography*, vol. 46, pp. 234–240, 1970.

[13] C. Tian, Y. Xu, and W. Zuo, "Image denoising using deep CNN with batch renormalization," *Neural Netw.*, vol. 121, pp. 461–473, 2020.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[15] C. Zhang *et al.*, "Joint deep learning for land cover and land use classification," *Remote Sens. Environ.*, vol. 221, pp. 173–187, 2019.

[16] P. Ghamisi *et al.*, "Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art," *IEEE Trans. Geosci. Remote Sens.*, vol. 7, no. 1, pp. 6–39, Mar. 2019.

[17] A. Voulodimos, N. Doulamis, A. D. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, 2018, Art. no. 7068349.

[18] M. F. Goodchild, L. Anselin, and U. Deichmann, "A framework for the areal interpolation of socioeconomic data," *Environ. Plan. A*, vol. 25, no. 3, pp. 383–397, 1993.

[19] C. L. Eicher and C. A. Brewer, "Dasymetric mapping and areal interpolation: Implementation and evaluation," *Cartography Geography Inf. Sci.*, vol. 28, no. 2, pp. 125–138, 2001.

[20] P. Deville *et al.*, "Dynamic population mapping using mobile phone data," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 45, pp. 15888–15893, 2014.

[21] J. K. Wright, "A method of mapping densities of population: With cape cod as an example," *Geograph. Rev.*, vol. 26, no. 1, pp. 103–110, 1936.

[22] C. P. Lo, "Population estimation using geographically weighted regression," *GISci. Remote Sens.*, vol. 45, no. 2, pp. 131–148, 2008.

[23] J. D. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature*, vol. 549, no. 7671, pp. 195–202, 2017.

[24] C. Hung, W. Chen, P. Lai, C. Lin, and C. Lee, "Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jeju Island, South Korea, Jul. 11–15, 2017, pp. 3110–3113.

[25] J. Song, X. Tong, L. Wang, C. Zhao, and A. V. Prishchepov, "Monitoring finer-scale population density in urban functional zones: A remote sensing data fusion approach," *Landscape Urban Planning*, vol. 190, 2019, Art. no. 103580.

[26] S. Amaral, G. Câmara, A. M. V. Monteiro, J. A. Quintanilha, and C. D. Elvidge, "Estimating population and energy consumption in Brazilian Amazonia using DMSP night-time satellite data," *Comput. Environ. Urban Syst.*, vol. 29, no. 2, pp. 179–195, 2005.

[27] M. Bakillah, S. H. L. Liang, A. Mobasheri, J. J. Arsanjani, and A. Zipf, "Fine-resolution population mapping using openstreetmap points-of-interest," *Int. J. Geograph. Inf. Sci.*, vol. 28, no. 9, pp. 1940–1963, 2014.

[28] W. Fan, C. Wu, and J. Wang, "Improving impervious surface estimation by using remote sensed imagery combined with open street map points-of-interest (POI) data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 11, pp. 4265–4274, Nov. 2019.

[29] Y. Yao *et al.*, "Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data," *Int. J. Geograph. Inf. Sci.*, vol. 31, no. 6, pp. 1220–1244, 2017.

[30] E. Doxsey-Whitfield *et al.*, "Taking advantage of the improved availability of census data: A first look at the gridded population of the world, version 4," *Papers Appl. Geography*, vol. 1, no. 3, pp. 226–234, 2015.

[31] J. E. Dobson, E. A. Bright, P. R. Coleman, R. C. Durfee, and B. A. Worley, "LandScan: A global population database for estimating populations at risk," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 7, pp. 849–857, 2000.

[32] B. Bhaduri, E. Bright, P. Coleman, and M. L. Urban, "LandScan USA: A high-resolution geospatial and temporal modeling approach for population distribution and dynamics," *GeoJournal*, vol. 69, nos. 1/2, pp. 103–117, 2007.

[33] Tatem and J. Andrew, "Worldpop, open data for spatial demography," *Sci. Data*, vol. 4, 2017, Art. no. 170004.

[34] M. Pesaresi *et al.*, "Operating procedure for the production of the global human settlement layer from landsat data of the epochs 1975, 1990, 2000, and 2014," Pub. Office Eur. Union, Luxembourg City, Luxembourg, Tech. Rep. JRC97705, 2016, pp. 1–62.

[35] L. Ru, B. Du, and C. Wu, "Multi-temporal scene classification and scene change detection with correlation based fusion," *IEEE Trans Image Process.*, vol. 30, pp. 1382–1394, 2020.

[36] J. Kim, J. Kwon Lee, and K. Mu Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1637–1645.

[37] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images," *IEEE Trans Geosci Remote Sens.*, vol. 57, no. 12, pp. 9976–9992, Dec. 2019.

[38] R. Fan, R. Feng, L. Wang, J. Yan, and X. Zhang, "Semi-MCNN: A semi-supervised multi-CNN ensemble learning method for urban land cover classification using sub-meter HRRS images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4973–4987, 2020.

[39] B. Huang, B. Zhao, and Y. Song, "Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery," *Remote Sens. Environ.*, vol. 214, pp. 73–86, 2018.

[40] Y. Park, B. Kwon, J. Heo, X. Hu, Y. Liu, and T. Moon, "Estimating PM2. 5 concentration of the conterminous united states via interpretable convolutional neural networks," *Environ. Pollut.*, vol. 256, 2020, Art. no. 113395.

[41] X. Zhu and D. Guo, "Mapping large spatial flow data with hierarchical clustering," *Trans. GIS*, vol. 18, no. 3, pp. 421–435, 2014.

[42] X. Feng, X. Su, J. Shen, and H. Jin, "Single space object image denoising and super-resolution reconstructing using deep convolutional networks," *Remote Sens.*, vol. 11, no. 16, 2019, Art. no. 1910.

[43] Q. Wang *et al.*, "What deep CNNs benefit from global covariance pooling: An optimization perspective," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Seattle, WA, USA, Jun. 13–19, 2020, pp. 10768–10777.

[44] N.Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[45] J. Li, C. Xu, M. Chen, and W. Sun, "Balanced development: Nature environment and economic and social power in China," *J. Clean Prod.*, vol. 210, pp. 181–189, 2019.

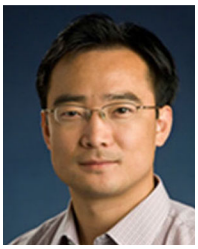
[46] T. Ye *et al.*, "Improved population mapping for China using remotely sensed and points-of-interest data within a random forests model," *Sci. Total Environ.*, vol. 658, pp. 936–946, 2019.

[47] B. Yu, K. Shi, Y. Hu, C. Huang, Z. Chen, and J. Wu, "Poverty evaluation using NPP-VIIRS nighttime light composite data at the county level in

China," *J. Sel. Topics Appl. Earth Observ. Remote. Sens.*, vol. 8, no. 3, pp. 1217–1229, 2015.
[48] C. Mellander, J. Lobo, K. Stolarick, and Z. Matheson, "Night-time light data: A good proxy measure for economic activity?," *PloS One*, vol. 10, no. 10, 2015, Art. no. e0139779.
[49] T. Wellmann, F. Schug, D. Haase, D. Pflugmacher, and S. van der Linden, "Green growth? On the relation between population density, land use and vegetation cover fractions in a city using a 30-years landsat time series," *Landscape Urban Planning*, vol. 202, 2020, Art. no. 103857.
[50] M. Xu, C. Cao, and P. Jia, "Mapping fine-scale urban spatial population distribution based on high-resolution stereo pair images, points of interest, and land cover data," *Remote. Sens.*, vol. 12, no. 4, 2020, Art. no. 608.
[51] L. Wang and L. Chen, "Spatiotemporal dataset on chinese population distribution and its driving factors from 1949 to 2013," *Sci. Data*, vol. 3, no. 3, 2016, Art. no. 160047.
[52] P. Gong *et al.*, "Finer resolution observation and monitoring of global land cover: First mapping results with landsat TM and ETM+ data," *Int. J. Remote Sens.*, vol. 34, no. 7, pp. 2607–2654, 2013.
[53] A. E. Gaughan *et al.*, "Spatiotemporal patterns of population in mainland China, 1990 to 2010," *Sci. Data*, vol. 3, no. 1, pp. 1–11, 2016.
[54] Z. Huang, H. Qi, C. Kang, Y. Su, and Y. Liu, "An ensemble learning approach for urban land use mapping based on remote sensing imagery and social sensing data," *Remote. Sens.*, vol. 12, no. 19, 2020, Art. no. 3254.
[55] Y. Yao *et al.*, "Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model," *Int. J. Geograph. Inf. Sci.*, vol. 31, no. 4, pp. 825–848, 2017.
[56] J. J. Nieves *et al.*, "Examining the correlates and drivers of human population distributions across low- and middle-income countries," *J. Roy. Soc. Interface*, vol. 14, no. 137, 2017, Art. no. 20170401.
[57] L. Zhuo, T. Ichinose, J. Zheng, J. Chen, P. Shi, and X. Li, "Modelling the population density of china at the pixel level based on DMSP/OLS nonradiance-calibrated night-time light images," *Int. J. Remote Sens.*, vol. 30, no. 4, pp. 1003–1018, 2009.

Luxiao Cheng is working toward the Ph.D. degree with the School of Computer Science, China University of Geosciences, Wuhan, China.

Her research interests include remote sensing data processing, intelligent information extraction from remote sensing images, and spatiotemporal data analysis.



Lizhe Wang (Fellow, IEEE) received B.E. and M.E degrees from Tsinghua University, Beijing, China, in 1998 and 2001 respectively and the D.E. degree from the University of Karlsruhe, Karlsruhe, Germany in 2007.

He is currently the Dean of the School of Computer Science, China University of Geosciences, Wuhan, China. His research interests include remote sensing data processing, Digital Earth, and Big Data Computing. He is a Fellow of IEEE, IET and BCS, and associate editor of Remote Sensing, IJDE, IEEE J-MASS.

Ruyi Feng (Member, IEEE) received the B.S. degree in geographic information system from Hunan Normal University, Changsha, China, in 2011, and the M.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2013 and 2016, respectively.

Since 2016, she has been with the School of Computer Science, China University of Geosciences, Wuhan, where she is currently an Associate Professor. Her research interests include sparse representation, deep learning, hyperspectral image analysis, high-resolution remote sensing understanding, and intelligent interpretation of remote sensing imagery.

Jining Yan received the Ph.D. degree in signal and information processing from the University of Chinese Academy of Sciences, Beijing, China, in 2017.

He is currently an Associate Professor with the School of Computer Science, China University of Geosciences, Wuhan, China. His research interests include remote sensing data processing, time-series analysis and change detection, cloud computing in remote sensing, and applied oceanography.