



## Article

# Mapping Multi-Temporal Population Distribution in China from 1985 to 2010 Using Landsat Images via Deep Learning

Haoming Zhuang <sup>1</sup> , Xiaoping Liu <sup>1,2,\*</sup>, Yuchao Yan <sup>3</sup>, Jinpei Ou <sup>1</sup>, Jialyu He <sup>1</sup> and Changjiang Wu <sup>1</sup>

<sup>1</sup> Guangdong Key Laboratory for Urbanization and Geo-Simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China; zhuanghm3@mail2.sysu.edu.cn (H.Z.); oujinpei3@mail.sysu.edu.cn (J.O.); hesysugis@foxmail.com (J.H.); wuchj9@mail2.sysu.edu.cn (C.W.)

<sup>2</sup> Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China

<sup>3</sup> Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing 100091, China; yuchaoyan@pku.edu.cn

\* Correspondence: liuxp3@mail.sysu.edu.cn

**Abstract:** Fine knowledge of the spatiotemporal distribution of the population is fundamental in a wide range of fields, including resource management, disaster response, public health, and urban planning. The United Nations' Sustainable Development Goals also require the accurate and timely assessment of where people live to formulate, implement, and monitor sustainable development policies. However, due to the lack of appropriate auxiliary datasets and effective methodological frameworks, there are rarely continuous multi-temporal gridded population data over a long historical period to aid in our understanding of the spatiotemporal evolution of the population. In this study, we developed a framework integrating a ResNet-N deep learning architecture, considering neighborhood effects with a vast number of Landsat-5 images from Google Earth Engine for population mapping, to overcome both the data and methodology obstacles associated with rapid multi-temporal population mapping over a long historical period at a large scale. Using this proposed framework in China, we mapped fine-scale multi-temporal gridded population data (1 km × 1 km) of China for the 1985–2010 period with a 5-year interval. The produced multi-temporal population data were validated with available census data and achieved comparable performance. By analyzing the multi-temporal population grids, we revealed the spatiotemporal evolution of population distribution from 1985 to 2010 in China with the characteristic of concentration of the population in big cities and the contraction of small- and medium-sized cities. The framework proposed in this study demonstrates the feasibility of mapping multi-temporal gridded population distribution at a large scale over a long period in a timely and low-cost manner, which is particularly useful in low-income and data-poor areas.

**Keywords:** population mapping; Landsat; deep learning; multi-temporal; ResNet-N; Google Earth Engine; China; SDGs



**Citation:** Zhuang, H.; Liu, X.; Yan, Y.; Ou, J.; He, J.; Wu, C. Mapping Multi-Temporal Population Distribution in China from 1985 to 2010 Using Landsat Images via Deep Learning. *Remote Sens.* **2021**, *13*, 3533. <https://doi.org/10.3390/rs13173533>

Academic Editor: Ronald C. Estoque

Received: 3 August 2021

Accepted: 3 September 2021

Published: 6 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Understanding the spatiotemporal distribution of the population is fundamental in a wide range of fields, including resource management [1,2], disaster response [3–6], public health [7–9], and urban planning [10,11]. The United Nations' Sustainable Development Goals (SDGs) also require the accurate and timely assessment of where people live to formulate, implement, and monitor sustainable development policies [12,13].

Census data released by an official body are authoritative and vital data about population distribution [14]. However, census data are based on administrative units; thus, they have several inherent limitations and are ill-suited to many spatial studies. Firstly, there is significant spatial heterogeneity in population distribution, which cannot be reflected by census data, which assumes a completely uniform distribution of the population within census units [15]. Secondly, the size of administrative units varies significantly in

urban and rural areas, which results in the Modifiable Areal Unit Problem [16]. Thirdly, administrative boundaries may change over time and are seldom compatible with practical applications, making census data challenging to integrate with other spatial data sets, preventing interdisciplinary research and temporal dynamic analysis [17]. In order to overcome the limitations of census data, fine-grained gridded population data, which are spatially continuous, are produced to supplement census data [18,19].

Several approaches have been developed to produce fine-scale gridded population data in the past few decades, such as areal weighting [20], spatial interpolation [21–23], and dasymetric mapping [24–32]. Among them, dasymetric mapping technology [33], which uses fine-scale auxiliary variables and specific weighting schemes to re-allocate census counts to grid cells, is the most widely used and effective one [19]. Commonly adopted ancillary data include land use/cover data [34–36], nighttime light data [26,37], terrain data [38], and social sensing data (e.g., points-of-interest [39], mobile phone records [40], and social media data [41]). Multiple methods, including empirical rules [15], statistical models (e.g., linear regression [34] and geographically weighted regression [17]), and machine learning models (e.g., random forest [42], expectation-maximization [43], and neural network [44]), have been proposed to estimate the distribution weight of grid cells. Various gridded population data at regional and global scales have been produced and published using the methods mentioned above, including the Gridded Population of the World (GPW) [45], Global Human Settlement Population layer (GHS-POP) [46], Global Rural-Urban Mapping Project (GRUMP) [47], WorldPop [48], and LandScan [49].

The accuracy of gridded population data is determined by the quality of auxiliary data [19]. Numerous previous studies focused on integrating novel auxiliary variables related to population distribution to improve the quality of the produced population grids [15,50,51]. The thematic, spatial, and temporal accuracy of auxiliary data themselves is also crucial to the quality of the final gridded data [19]. For example, classification error in land use/land cover data will be propagated to the produced gridded population data. Furthermore, involving more auxiliary variables contributes more uncertainty to the final result [34]. In order to produce multi-temporal gridded population data, temporally explicit and consistent auxiliary data are essential, whose availability and sustainability are questionable, especially at a large scale, precluding the production of continuous multi-temporal data over a long historical period [34].

Remote sensing (RS) data (e.g., satellite imagery) that can capture the physical characteristics of the ground at low cost, broad coverages, and high spatiotemporal resolution are becoming increasingly available with improvements in imaging technology over time [36,52]. The physical characteristics of the ground and human activities interact with each other. Human activities can lead to distinct spatial landscapes, which inversely constrain how people live, produce, and travel, providing the possibility of consistent and sustainable population estimation with RS imagery as auxiliary data without the problems mentioned above [53]. However, the raw RS imagery is highly unstructured, and its association with population count is complex and nonlinear, making it challenging to construct a mapping from raw RS imagery to population count [54]. An emerging supervised deep learning approach, convolutional neural networks (CNN), which are capable of extracting the hidden hierarchical structures of RS images [55], have shown outstanding performance in obtaining knowledge from RS images in the domain of geography (e.g., land use classification [56], spatial interpolation [57], and poverty mapping [58]). Therefore, it is possible that CNN can form a mapping from RS imagery to population count.

A few studies have tried to estimate population counts from RS imagery directly. Doupe proposed the use of a VGG-like network to estimate population density in Tanzanian and Kenya from Landsat images and achieved remarkable performance and generalizability [54]. Robinson regarded the population estimation task as a classification problem and used a similar VGG-like network to classify RS image patches into 14 population density levels. They produced gridded population data for the United States in 2010, achieved high performance, and qualitatively explained the predictions in terms of the input RS

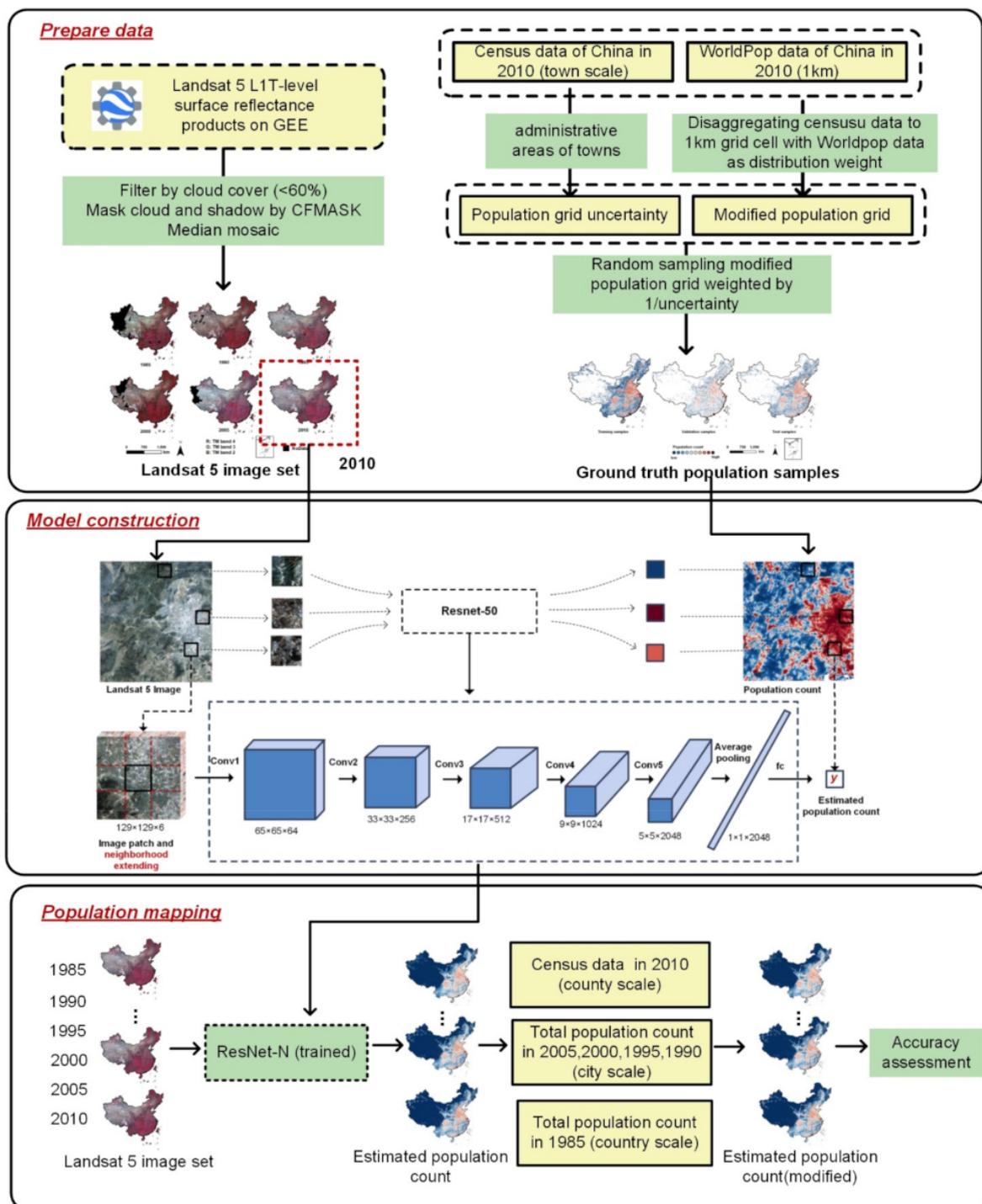
imagery [59]. Xing proposed a Neighbor-ResNet architecture by embedding the neighbor knowledge into ResNet in order to estimate the volumes of human activity from Google imagery in 18 cities in China [53]. The attempts mentioned above verify the feasibility and superiority of integrating CNN and RS imagery for population mapping. However, the established models have not been used to map historical population distributions and understand their spatiotemporal evolution.

China is the world's most populous developing country. Fine-scale population distribution data are crucial for China's sustainable development [13]. Numerous gridded population data of China, with various spatial resolutions, have been developed [25,39]. A few studies have also used time-invariant and time-explicit auxiliary variables to produce multi-temporal gridded population distribution data [17,30,60]. However, due to the lack of appropriate auxiliary datasets and effective methodological frameworks, there are rarely continuous multi-temporal gridded population data for China over a long historical period to aid in our understanding of the spatiotemporal evolution of the population.

In this study, we developed a framework integrating a ResNet-N deep learning architecture with the consideration of neighborhood effects with a vast number of Landsat-5 images from Google Earth Engine (GEE) [61] for population mapping to overcome both the data and methodology obstacles of rapid multi-temporal population mapping over a long historical period at a large scale. Once the framework was constructed, we developed multi-temporal gridded population data with a 1 km resolution for China (excluding Taiwan, Hong Kong, and Macao) for the 1985–2010 period with a 5-year interval and analyzed its spatiotemporal evolution.

## 2. Materials and Methods

This study aimed to develop a framework integrating a deep learning model with Landsat-5 RS images from GEE to estimate population count. Once the framework is established, large-scale population mapping can be achieved only with easily accessible and regularly updatable RS imagery. Furthermore, we produce multi-temporal gridded population data (1 km × 1 km) of China for the 1985–2010 period with a 5-year interval and analyze the spatiotemporal evolution of the population distribution of China in this period. The flowchart of this study is illustrated in Figure 1, containing three main parts: (1) we collected ground-truth population count grid cells and corresponding Landsat-5 RS image patches as reference datasets for training, validating, and testing the developed deep learning model; (2) a ResNet-N architecture considering neighborhood effects was developed to establish the end-to-end mapping between population count and RS image patches; (3) based on the trained model, we estimated the gridded population count of China with corresponding Landsat-5 image patches from GEE as input from 1985 to 2010. Furthermore, the produced raw estimations were adjusted by available census data to acquire the final gridded population data. Finally, we validated the produced datasets and analyzed the spatiotemporal evolution of China's population distribution.

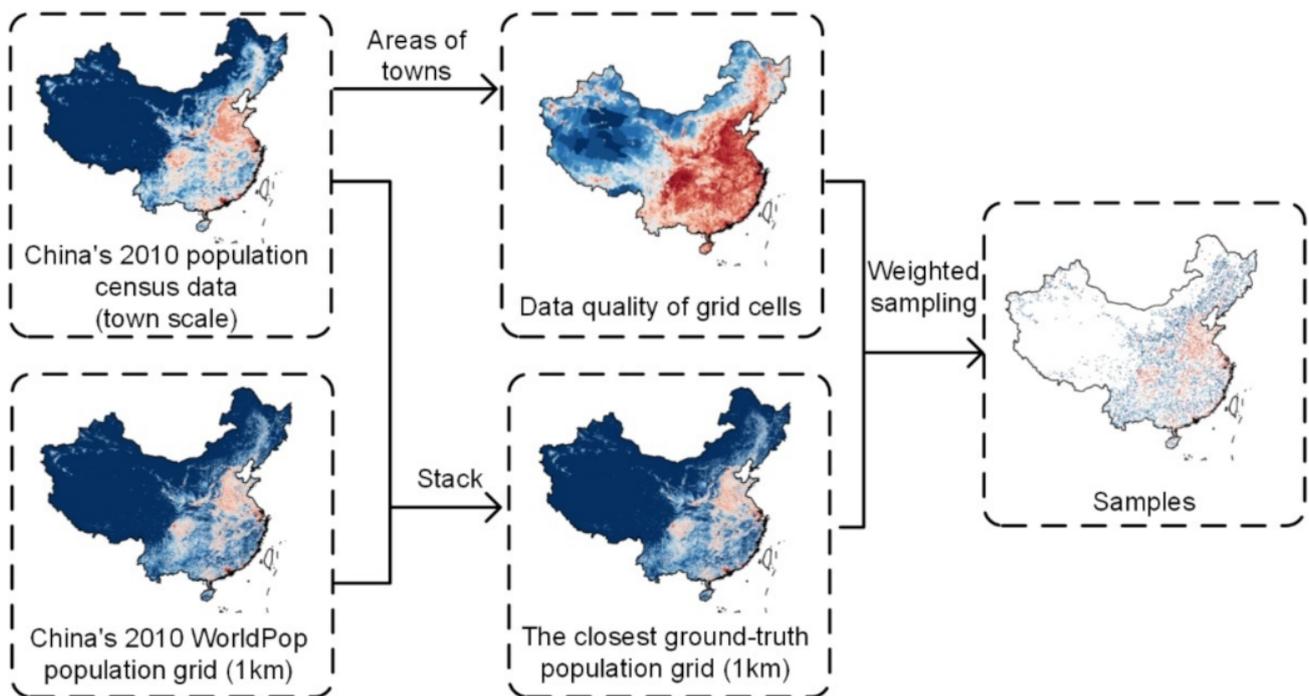


**Figure 1.** The flowchart of the proposed framework for mapping population distribution of China by integrating the ResNet-N model and Landsat-5 images from GEE.

## 2.1. Data Sources and Preprocessing

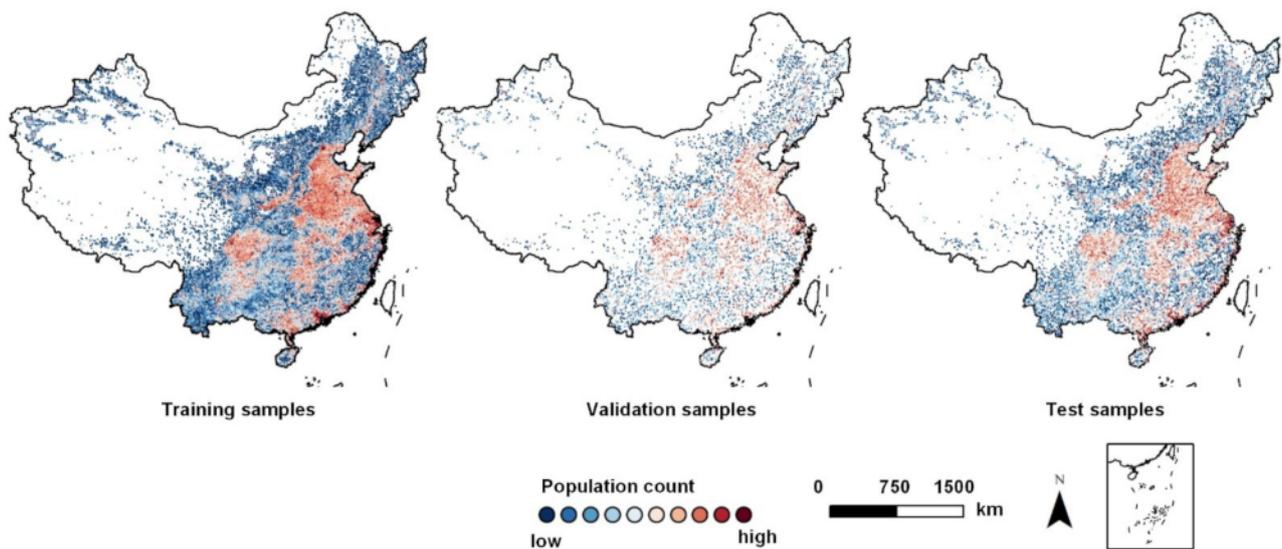
### 2.1.1. Ground-Truth Population Grid

In order to establish an end-to-end mapping between Landsat-5 RS image patches and population count by deep learning architecture, it is necessary to collect ground-truth population grid cells as training samples. However, the ground-truth population grid does not exist [19]. In this study, an alternative method (Figure 2) was utilized to collect the closest ground-truth population grid samples with a resolution of 1 km.



**Figure 2.** The flowchart of collecting the closest ground-truth population grid cell samples with a resolution of 1 km.

We obtained China's 2010 population census data at the town level (level 4 administrative unit), the finest-scale census data publicly available, from China's Sixth National Population Census. Towns are the fundamental administrative units in China, with relatively small jurisdiction areas, 58% of which are less than 100 km<sup>2</sup>, so that the spatial heterogeneity of population distribution is tiny within towns. However, it is not adequate to use the average population density of towns as references due to heterogeneity within towns [62]. We obtained the WorldPop gridded population data [48] with a resolution of 1 km for China in 2010 to remedy this problem. The WorldPop data are produced by coupling a random forest algorithm with various auxiliary data to disaggregate county-level (level 3 administrative units) census data, recognized as some of the finest gridded population data to date [5]. In this study, we used WorldPop data in 2010 as a weighting layer to redistribute the total population count of each town to grid cells to account for the spatial heterogeneity within towns in part. Numerous towns are small in area. Therefore, this modified population map represents the closest ground-truth population grid that is available to use as training data. Finally, we sampled 100,000 grid cells from the ground-truth population grid weighted by the quality of grid cells to tradeoff the reliability and representativeness of the samples. The administrative areas of towns act as a data quality metric of grid cells [19]. Let  $area^k$  represent the area of the  $k^{th}$  town; then, the weight of selecting a grid cell inside the  $k^{th}$  town is given as  $\frac{1}{area^k}$ . We discarded the grid cells with a population count of less than 10. It is unnecessary and intractable to distinguish the subtle change in population count via RS images in a 1 km<sup>2</sup> area [59]. The distribution of ground-truth population samples is heavily tailed, with kurtosis of 1312.41 and skewness of 26.31. To balance the dataset and ease the training of the deep learning model, the population count was logarithmized [53]. The collected samples were randomly divided into three groups: training (70%), validation (10%), and testing (20%). Figure 3 presents the spatial distribution of the collected population samples.



**Figure 3.** The spatial distribution of the ground-truth population samples.

### 2.1.2. Landsat-5 RS Imagery

This study used RS images from Landsat-5 collected by the Thematic Mapper (TM) sensor, covering the 1985–2010 period [52]. The full Landsat-5 L1T-level surface reflectance archive [63] covering China with a cloud score of less than 60 was preprocessed and downloaded effortlessly from GEE, a cloud-based platform for processing petabyte-scale geospatial datasets [61]. The L1T-level products have undergone geometric, radiation, and atmospheric corrections and are ready for use [64,65]. After masking clouds and shadows using Landsat quality flag information [66], a composite for a given year was produced in the form of a median mosaic of all available Landsat scenes. To address the shortage of cloud-free images, we included the Landsat scenes of the year before and the year after the target year in the composite. By referring to previous research, six bands were retrieved, i.e., Band 1 (blue), Band 2 (green), Band 3 (red), Band 4 (near-infrared), Band 5 (shortwave infrared 1), and Band 7 (shortwave infrared 2), all with a spatial resolution of 30 m [65]. Figure 4 presents the cloud-free Landsat composites with standard false-color band combination from 1985 to 2010. Due to the shortage of cloud-less images, there are missing data in western areas of China in some target years. As these areas are usually sparsely populated with slight variation, we used the valid data in the nearest adjacent year to supplement these areas.

Previous studies have revealed that the detailed characteristics of various landscapes can be well reflected by these 6 visible and invisible bands [54]. Figure 5 presents the probability density distribution of population count in the ground-truth samples and the example RS image patches that correspond to various population counts. Obviously, different magnitudes of population count correspond to distinct landscape characteristics in the RS image patches. The interplay between population count and RS images indicates the potential of estimating population count based only on RS images from Landsat-5 via a deep learning architecture.

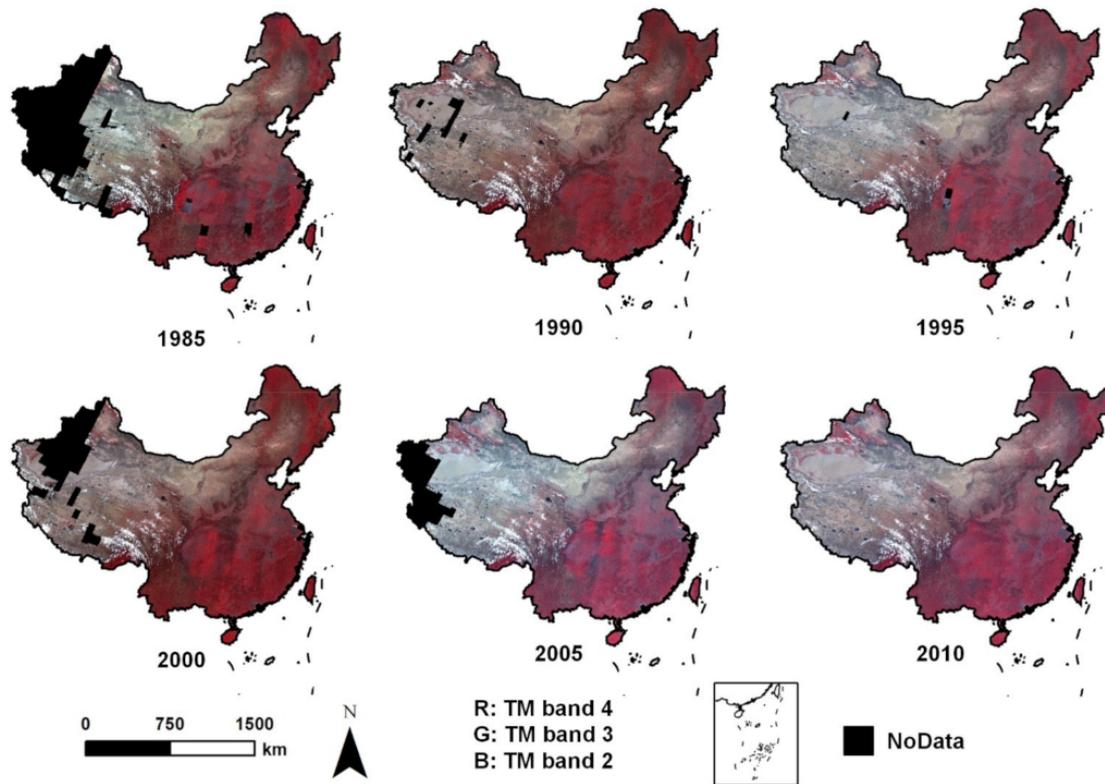


Figure 4. Cloud-free Landsat-5 composites of China from 1985 to 2010.

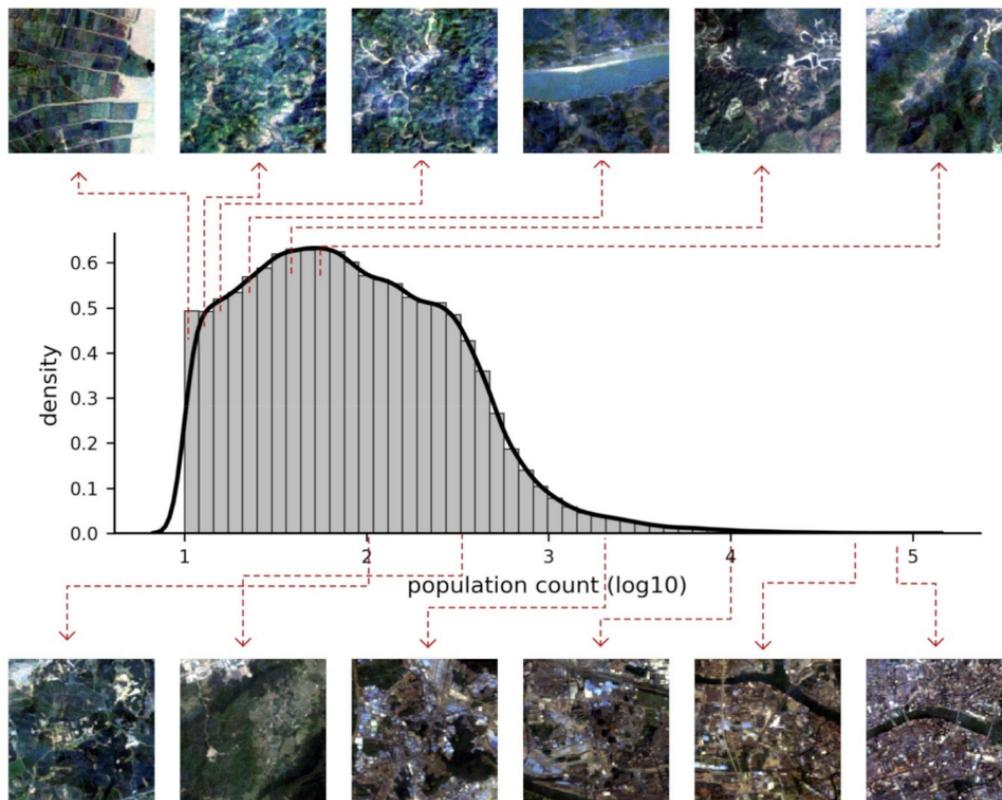


Figure 5. Probability density distribution of population count in the ground-truth samples and example RS image patches that correspond to various population counts.

## 2.2. Methods

### 2.2.1. Building a Mapping from RS Image Patches to Population Counts via ResNet-N Model

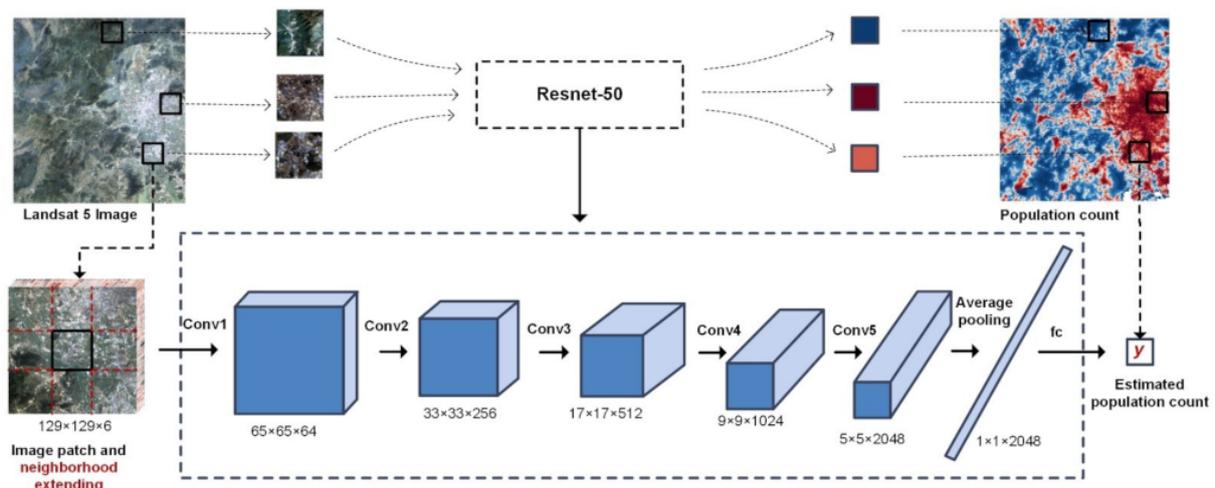
In this study, we view the gridded population estimation task as a regression problem. The method framework is shown in Figure 6. Given an image patch  $\theta^i$  of the grid cell  $i$  and the corresponding logarithmized population count  $p_i$ , we express our learning task as building a mapping function:

$$p_i = f(\theta_i) \quad (1)$$

where  $f(\cdot)$  is the mapping function to be learned by deep learning models. Acknowledging the highly nonlinear and complex relationship between RS images and population count, a ResNet (specifically, ResNet-50 was adopted) model considering neighborhood effects (ResNet-N) was utilized to approximate such a complex mapping relationship [53]. The ResNet model is one of the state-of-the-art CNN architectures and has been widely adopted to mine geographical knowledge from RS images [55,56]. The fundamental building block of Resnet-50 is the bottleneck, a convolution layer with an identity shortcut connection, which solves the problem of gradient vanishing [55]. As shown in Figure 6, ResNet-50 contains 7 layers (groups). Conv1 is a plain convolution layer with 64 convolution kernels of size  $3 \times 3$ , which slide on the RS image to extract hidden features and output 64 feature maps. Conv2 contains 3 bottleneck blocks, each with 128 convolution kernels of size  $3 \times 3$ , which slide on the feature maps generated by Conv1 to extract higher-level features. Likewise, Conv3 contains 4 bottleneck blocks, each with 512 convolution kernels; Conv4 contains 6 bottleneck blocks, each with 1024 convolution kernels, and Conv5 contains 3 bottleneck blocks, each with 2048 convolution kernels. In the network, deeper layers excavate more abstract and informative features related to the task from previous feature maps. Between each convolution layer (or bottleneck block), the feature map is reduced by half to aggregate information. Finally, the average pooling layer squeezes the feature map to 1 dimension, which is inputted into the fully connected layer (fc) to regress the population count. The ReLU activation function and batch normalization are used in all convolution layers to facilitate the training of networks [67]. Figure A1 illustrates how the input RS image evolves to the output population count in the network. Because of the autocorrelation of population distribution, the center grid cell population count may be affected by landscapes in the neighborhood. Hence, we constructed extended image patches by extending the center image patch to include its  $3 \times 3$  neighboring patches to embed neighborhood knowledge [53,68]. Hence, the layer-wise convolutional operations of the ResNet model can extract interior and neighborhood and integrate latent features for population estimation when sliding on the extended image patches. In order to regress the population count directly, the softmax activation function in the final fully connected layer was removed. We used the log-cosh function for back-propagation training:

$$Loss(\hat{p}, p) = \sum_{i=1}^s \log_{10}(\cosh(\hat{p}_i - p_i)) \quad (2)$$

where  $Loss(\hat{p}, p)$  is the log-cosh loss function,  $p_i$  is the ground-truth population count of grid cell  $i$ ,  $\hat{p}_i$  is the estimated population count of grid cell  $i$ , and  $\cosh(\cdot)$  is the hyperbolic cosine function [68]. The log-cosh loss is similar to the L1 loss, commonly used in regression problems, but is more tolerant of anomalous estimations and achieves better performance [68]. Hyperparameters were tuned empirically based on 1/10 of the available samples. A stochastic gradient descent (SGD) optimizer with a momentum of 0.9 and a learning rate of  $10^{-4}$  was used for weight updating. The batch size and the maximum number of epochs were set to 32 and 1000, respectively. The framework was implemented using the Tensorflow 2.0 library on a Linux server with a 2.50 GHz Intel Xeon E5-2680 CPU, an NVIDIA GTX 2080Ti GPU, and 128GB RAM.



**Figure 6.** An end-to-end ResNet-N model to estimate population count from RS images by embedding the neighbor knowledge into ResNet.

### 2.2.2. Mapping Multi-Temporal Population Distributions in China via ResNet-N Model and Landsat-5 RS Images

This study aims to produce multi-temporal gridded population maps with 1 km spatial resolution for China via establishing a framework integrating a deep learning model with Landsat-5 RS images from GEE. Our research area, mainland China, is covered by a grid of  $7346 \times 4507$  consisting of  $1 \text{ km} \times 1 \text{ km}$  cells. We excluded grid cells with a population count of  $<10$  in the ground-truth population grid in 2010, which can be regarded as uninhabited areas, to reduce the computational burden, resulting in 5,508,904 grid cells being retained. A  $1 \text{ km} \times 1 \text{ km}$  cell in the population grid approximately covers a  $34 \times 34$  image patch with a spatial resolution of 30 m on Landsat-5 composites. To consider the contribution of neighborhood effects on population count, we constructed extended image patches with a width and height of 102, including the center patch and its  $3 \times 3$  neighboring patches. The extended image patch was then resized to a fixed size of  $129 \times 129$  for inputting into the deep learning model. We obtained a centroid for each cell in the population grid and extracted a  $102 \times 102$  image patch center around the obtained centroid from the Landsat-5 composites for each target year from 1985 to 2010. A total of 33,053,424 RS image patches were extracted and normalized to 0–1. Among them, the image patches that corresponded to the 2010 ground-truth population samples were utilized for training, evaluating, and testing the deep learning model. Once the model was trained, all RS image patches were inputted into the model to measure the population count of each position of each target year.

### 2.2.3. Modifying Raw Population Estimation via Census Data

Ensuring that the aggregated grid population counts at census units match the known official total population count is necessary. The dasymetric mapping method is used to achieve this goal. For a census unit  $s$  with a known official total population count  $p_s$ , the following equations are used to modify the raw population estimations:

$$w_i = \frac{p_i^r}{\sum_{i \in S} p_i^r} \quad (3)$$

$$p_i^m = p_s \times w_i \quad (4)$$

where  $p_i^r$  is the raw population count of cell  $i$  estimated by the deep learning model,  $w_i$  is the distribution weight of cell  $i$ , and  $p_i^m$  represents the modified population count of cell  $i$ . In 2010, we used county-scale census data from the National Bureau of Statistics of China to modify the estimation. Due to data limitations, we used the city-scale (level 2

administrative unit) total population count from WorldPop generated using the dasymetric method based on county-scale census data to modify estimations for 2005 and 2000 [48]. For 1995 and 1990, we used the city-scale total population count from GPWv3 produced by the areal weighting method based on official census data at the county scale to modify the estimations [45]. For 1985, due to the unavailability of census data, a single country-wide population count from the World Bank Database was used for modification. The data source and administrative unit level of the census data or total population count are summarized in Table A1.

#### 2.2.4. Accuracy Assessment

We used six quantitative metrics to assess the performance of the proposed population mapping framework and the produced multi-temporal gridded population data, including Pearson's correlation coefficient (R), the coefficient of determination ( $R^2$ ), mean absolute error (MAE), percentage mean absolute error (%MAE), root mean squared error (RMSE), and percentage root mean squared error (%RMSE):

$$R = \sum_{i=1}^n \frac{(p_{i,o} - \bar{p}_o)(p_{i,s} - \bar{p}_s)}{\sqrt{\sum_{i=1}^n (p_{i,o} - \bar{p}_o)^2} \sqrt{\sum_{i=1}^n (p_{i,s} - \bar{p}_s)^2}} \quad (5)$$

$$R^2 = \frac{n \sum_{i=1}^n p_{i,o} p_{i,s} - \sum_{i=1}^n p_{i,o} \sum_{i=1}^n p_{i,s}}{\sqrt{n \sum_{i=1}^n p_{i,o}^2 - (\sum_{i=1}^n p_{i,o})^2} \sqrt{n \sum_{i=1}^n p_{i,s}^2 - (\sum_{i=1}^n p_{i,s})^2}} \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_{i,o} - p_{i,s}| \quad (7)$$

$$\%MAE = \frac{1}{n} \sum_{i=1}^n \frac{|p_{i,o} - p_{i,s}|}{p_{i,o}} \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_{i,o} - p_{i,s})^2} \quad (9)$$

$$\%RMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (p_{i,o} - p_{i,s})^2}}{\bar{p}_o} \quad (10)$$

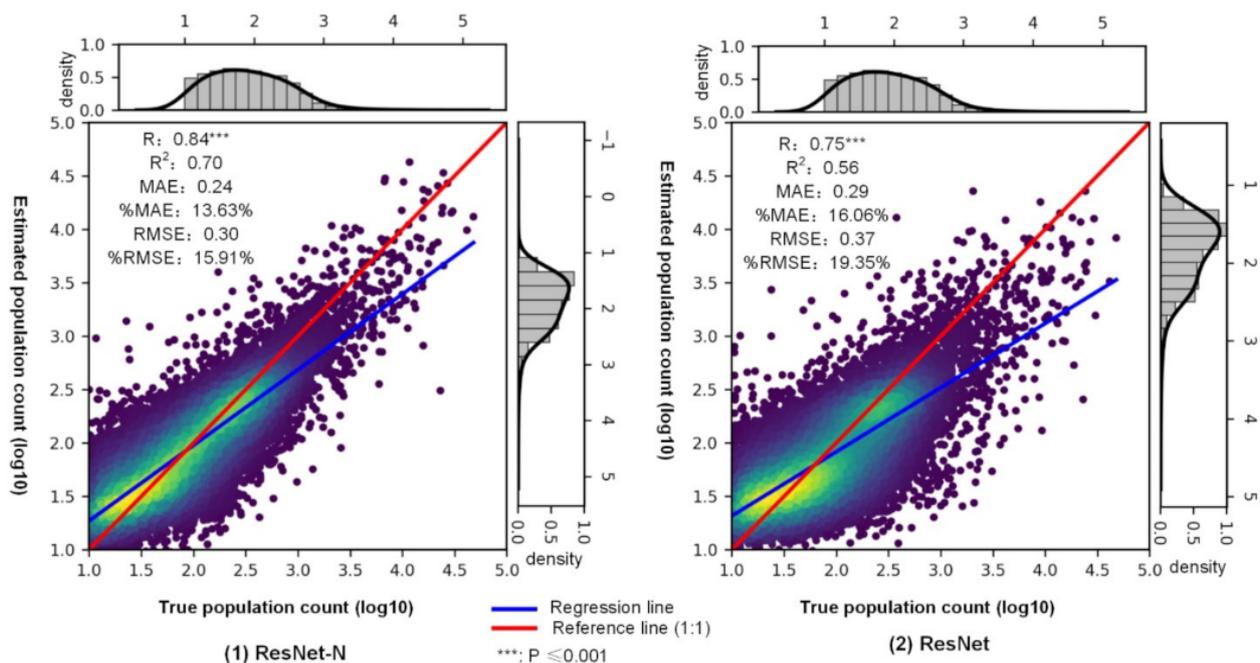
where  $p_{i,o}$  is the ground-truth population count of the  $i^{th}$  sample,  $p_{i,s}$  denotes the estimated population count of the  $i^{th}$  sample,  $n$  represents the total number of samples,  $\bar{p}_o$  is the average of the ground-truth population count, and  $\bar{p}_s$  is the average of the estimated population count. The indicator  $R$ , ranging from  $-1$  to  $1$ , measures the linear correlation between actual values and estimated values to evaluate the relative magnitude fitting performance [62]. The indicator  $R^2$ , with a value from  $-\infty$  to  $1$ , measures how much variance in actual values is captured by the predicted values, assessing the absolute magnitude fitting performance [53].  $R$  and  $R^2$  evaluate the explainability of estimated values to actual values. MAE designates the average absolute error between actual values and estimated values. In order to highlight large errors, absolute errors are squared in RMSE. Since MAE and RMSE are not as understandable, the percentage errors (%MAE and %RMSE) assessing the proportion of the error to the actual value are also presented [54]. These 4 error metrics evaluate the absolute and percentage estimation error together. The mentioned 6 metrics complement each other and provide a comprehensive assessment of the proposed framework and the produced data [69].

### 3. Results

#### 3.1. Accuracy Assessment of ResNet-N Model for Population Estimation

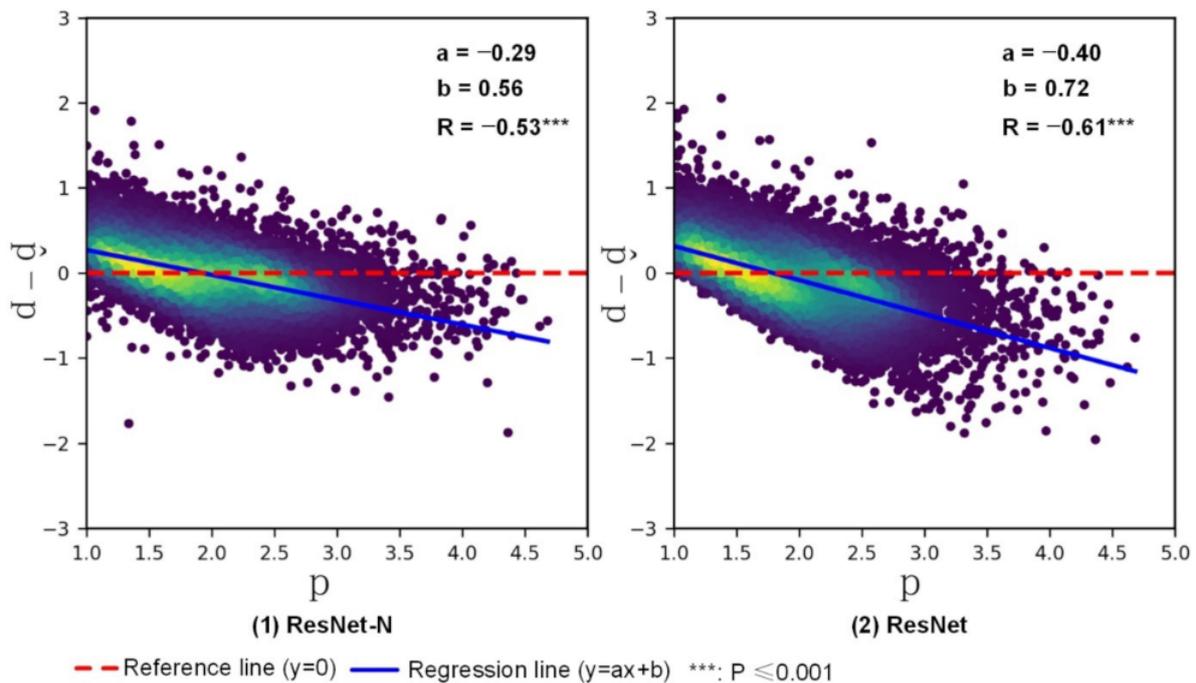
In this study, a ResNet-N model with neighbor augmentation was utilized to establish the end-to-end mapping between Landsat-5 RS image patches and population count. The model's performance of directly estimating the population count from RS images

was evaluated by the collected 20,000 testing samples. Figure 7 shows the scatterplots of ground-truth population count ( $p$ ) and estimated population count ( $\hat{p}$ ) with their probability density distributions. As shown in Figure 7, the scatterplots of  $p$  and  $\hat{p}$  present a clustered distribution pattern along the 1:1 reference line, validating that the deep learning architectures can effectively establish the mapping from RS image patches to population count. The probability density distributions of  $p$  and  $\hat{p}$  exhibit similar shapes and also confirm this conclusion. Compared to the ResNet model without neighbor augmentation, the ResNet-N model with neighbor augmentation used in this study displays superior performance in terms of the six evaluation metrics. ResNet-N ( $R = 0.84$ ,  $R^2 = 0.70$ ) exhibits higher explainability of landscape characteristics extracted from the RS images on population count compared to ResNet ( $R = 0.70$ ,  $R^2 = 0.56$ ). The  $R^2$  indicates that 70% of the variance population count can be explained by the ResNet-N, compared to 56% by the ResNet. ResNet-N (%MAE = 13.63%, %RMSE = 15.91%) also has higher absolute accuracy than ResNet (%MAE = 16.06%, %RMSE = 19.35%). The %RMSE of ResNet-N is lower than that of ResNet by 21.62% and %MAE by 17.93%. The comparatively low %RMSE and %MSE of both models reveal the capacity of the deep learning model to capture the heterogeneity in population distribution from RS images, and improved estimation performance can be achieved considering neighbor effects.



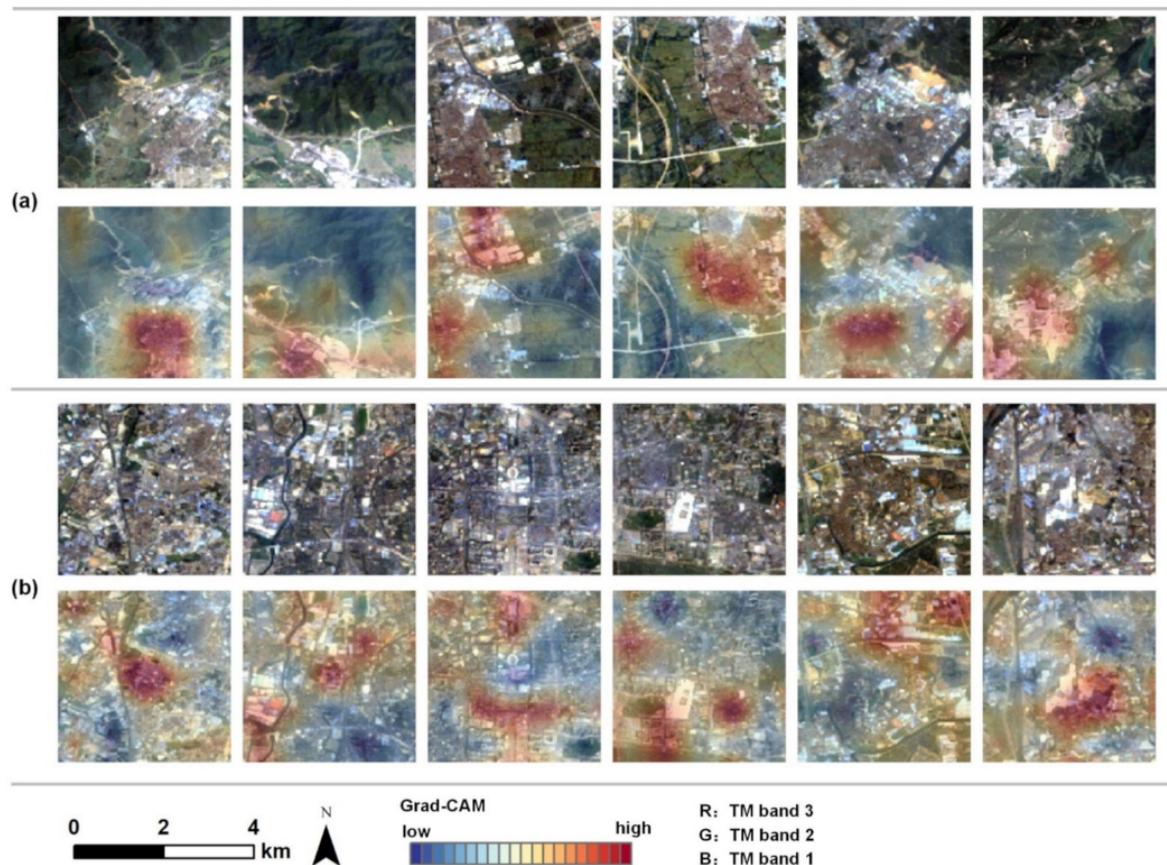
**Figure 7.** Scatterplots and probability density distributions of ground-truth population count and estimated population count from ResNet-N and ResNet.

For true population count ( $p$ ) and estimated population count ( $\hat{p}$ ), an investigation of the relationship between  $p$  and  $\hat{p} - p$  was conducted to explore the systematic bias of estimating population count from RS images via deep learning technologies. Figure 8 shows the scatterplots of  $p$  and  $\hat{p} - p$  from ResNet-N and ResNet. The results reveal that both models tend to underestimate densely populated samples and overestimate sparsely populated samples, evidenced by the significant negative correlation coefficient and the negative slope. The observed bias can be ascribed to the inherent limitations of multispectral RS images, which cannot identify the social-economic factors that affect population distribution (i.e., the high utilization efficiency of space in densely populated areas). However, consideration of neighbor effects leads to reduced biases and better estimation performance [53].



**Figure 8.** Scatterplots of test samples between  $p$  and  $\hat{p} - p$  from ResNet-N and ResNet. ( $p$ : true population count;  $\hat{p}$ : estimated population count).

Interpretability is a critical aspect of a model [53,59]. A model with good interpretability usually has good performance. In this study, the ResNet-N model considers only RS images as input to estimate population count. Therefore, all estimations can be explained in terms of the landscape details from RS images. We used gradient-weighted class activation mapping (Grad-CAM), a visual explanation technology for deep learning models, to figure out what features our model learns to estimate population count [70]. Grad-CAM can output a heatmap for an RS image patch. The heat value quantifies the relative contribution of input pixels in the original patch to the estimated population count [70]. For analysis, we selected 12 typical grid cells with different magnitudes of population count. Figure 9 presents the RS image patches in the top rows and corresponding heatmaps in the bottom rows. As shown in Figure 9a, built-up areas are highlighted in heatmaps when they border natural areas. The explanation for this is that built-up areas are usually more densely populated than natural areas. Figure 9b proves the ability of our model to recognize different buildings by capturing hidden hierarchic features of RS images in the interior of the built-up area to estimate population count. As densely populated buildings (i.e., residential) and sparsely populated buildings (i.e., factories) are staggered in the built-up area, distinguish different buildings contributes to accurate population estimation. The heatmaps offer insights into how human activities interact with the underpinning physical environment and prove that our model can learn valuable features for population estimation.



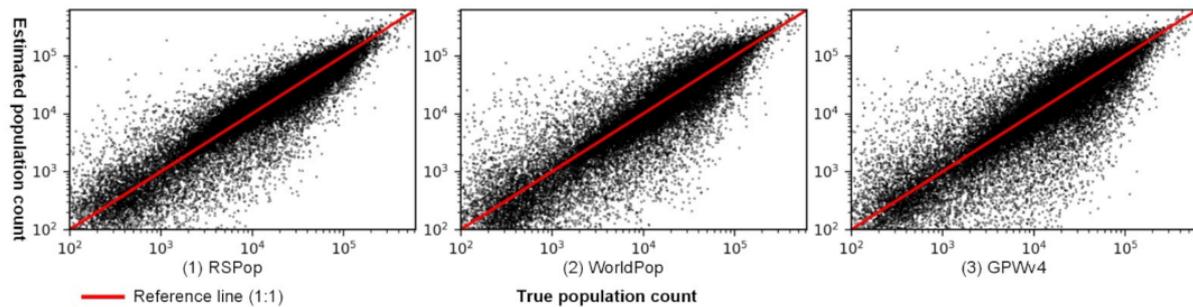
**Figure 9.** RS image patches (**top row**) and corresponding heatmaps (**bottom row**) produced by Grad-CAM in 12 typical grid cells. (a) Built-up areas border natural areas; (b) Interiors of built-up areas.

### 3.2. Validating Multi-Temporal Gridded Population Data via Census Data

A stable end-to-end mapping from RS image patches to population count was established by the ResNet-N model. It is promising that population distribution mapping can be achieved with only the formed mapping and RS images. However, it is necessary to ensure that the aggregated grid population counts at census units match the known official total population count. Furthermore, the grid cell estimation will be more accurate when scaled to match the true population value [59]. We used county-level census data to modify the raw population count estimated from RS images by the model in 2010. Validation of the modified population map was conducted using town-level census data. It is a common practice in dasymetric mapping to use census data of a finer scale to evaluate the accuracy of the produced gridded population data [39]. Two well-known gridded population datasets, WorldPop [48] and GPWv4 [45], were selected as baselines to highlight the performance of the produced data. We collected towns with a population of >100 to assess the comparative performance of the produced gridded data. As shown in Table 1, our new population map produced by coupling RS images and deep learning technologies (referred to as RSPop) achieved the best performance, with the lowest absolute and relative errors and the highest explainability and correlation with the true population count. Figure 10 presents scatterplots of the true population count and estimated population count of each town from RSPop, WorldPop, and GPWv4. Compared to other gridded population data, the scatterplot of RSPop presents a more concentrated distribution pattern along the 1:1 reference line, with the highest accuracy. In contrast, points are scattered and distributed away from the 1:1 reference line in GPWv4, which has the lowest accuracy.

**Table 1.** Accuracy assessment of RSPop at town scale comparing WorldPop and GPWv4.

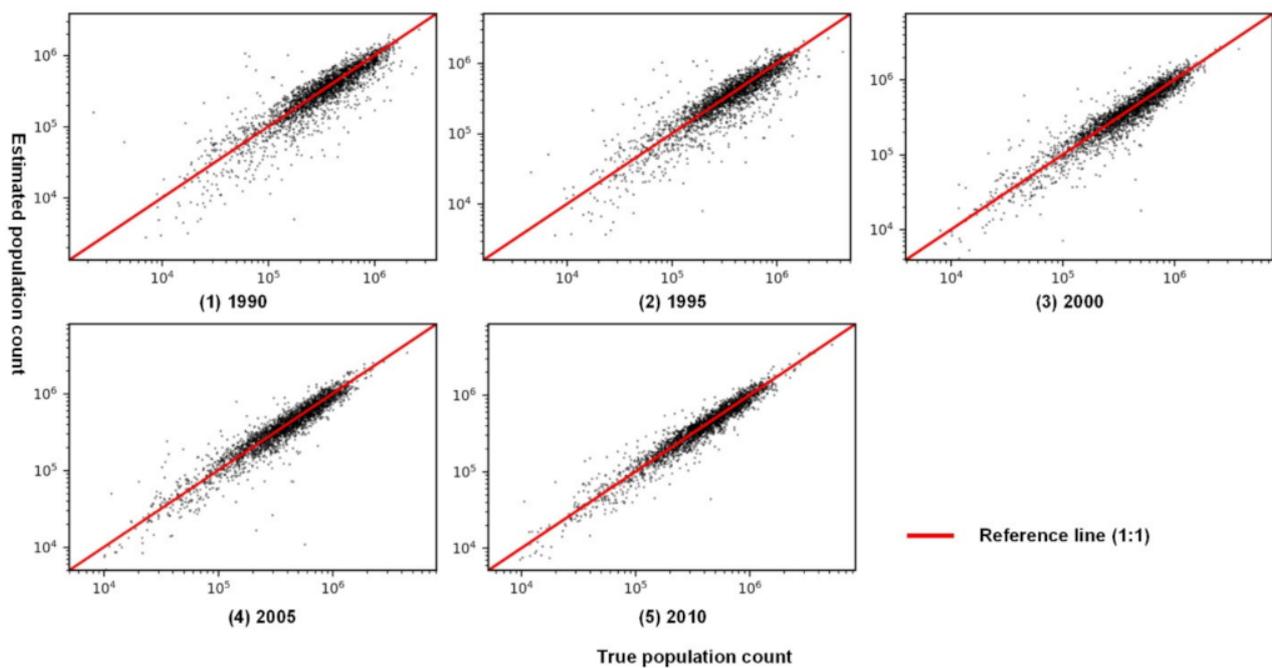
	RSPop	WorldPop	GPWv4
R	0.89	0.87	0.82
R <sup>2</sup>	0.77	0.69	0.61
MAE	7846.62	8138.20	9463.33
%MAE	46.21	51.19	62.48
RMSE	15,686.74	18,277.52	20,448.11
%RMSE	56.03	65.28	73.03

**Figure 10.** Scatterplots of the true population count and estimated population count from RSPop, WorldPop, and GPWv4 at town scale.

The gridded population data in 2010 produced by the proposed framework were validated and achieved the highest performance compared to other datasets. Due to the consistency of Landsat-5 images and the relative stability of human activity patterns, it can be expected that accurate gridded population data from 1985 to 2005 can be produced by the same framework, using corresponding RS images at target years as input. For the period of 1990–2005, because town-scale census data are challenging to collect, we used the city-scale total population count to modify the estimated population count and applied the county-scale total population count to verify the accuracy of the data. Total population counts at city scale and county scale in 2000 and 2005 were obtained from WorldPop, while total population counts at city scale and county scale for 1990 and 1995 were obtained from GPWv3. Both WorldPop and GPWv3 were produced based on county-scale census data [45,48]. Therefore, it would be impractical to use them for comparative analysis. Instead, as the accuracy of the gridded population data in 2010 has been verified, the population estimation in 2010 modified by city-scale census data was used for comparison at the county scale. As shown in Table 2, overall performance reductions exist for each target year in 1990–2005 compared to 2010. For example, the R<sup>2</sup> is reduced from 0.93 in 2010 to 0.91 in 2005, 0.88 in 2000, 0.73 in 1995, and 0.74 in 1990, with an average reduction of 12.37%. Figure 11 shows scatterplots of the true population count and estimated population count at the county level, which present clustered patterns along the 1:1 reference line. These results imply that the model trained in 2010 is generalizable to other years.

**Table 2.** Accuracy assessment of RSPop at county scale from 1990 to 2010.

	1990	1995	2000	2005	2010
R	0.86	0.86	0.94	0.95	0.97
R <sup>2</sup>	0.74	0.73	0.88	0.91	0.93
MAE	93,260.92	103,362.10	83,413.22	77,668.16	69,052.32
%MAE	30.68	28.48	22.11	19.67	16.64
RMSE	163,431.57	182,624.46	127,106.24	116,733.57	105,319.00
%RMSE	38.45	40.74	27.43	24.47	21.49

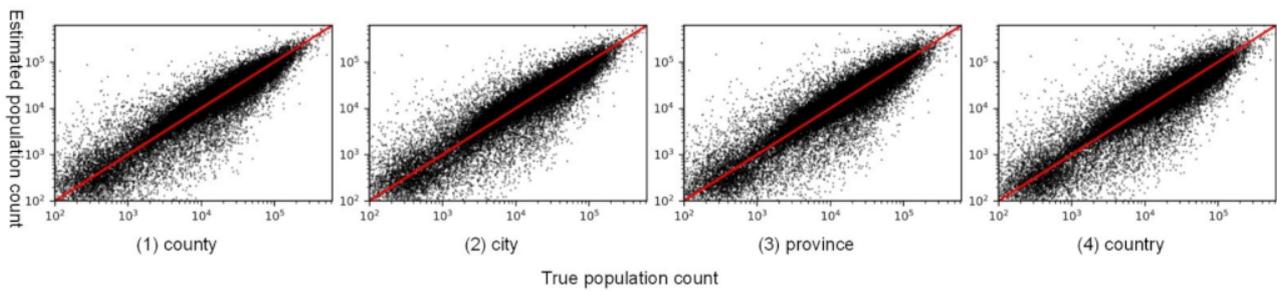


**Figure 11.** Scatterplots of the true population count and estimated population count at county scale from 1990 to 2010.

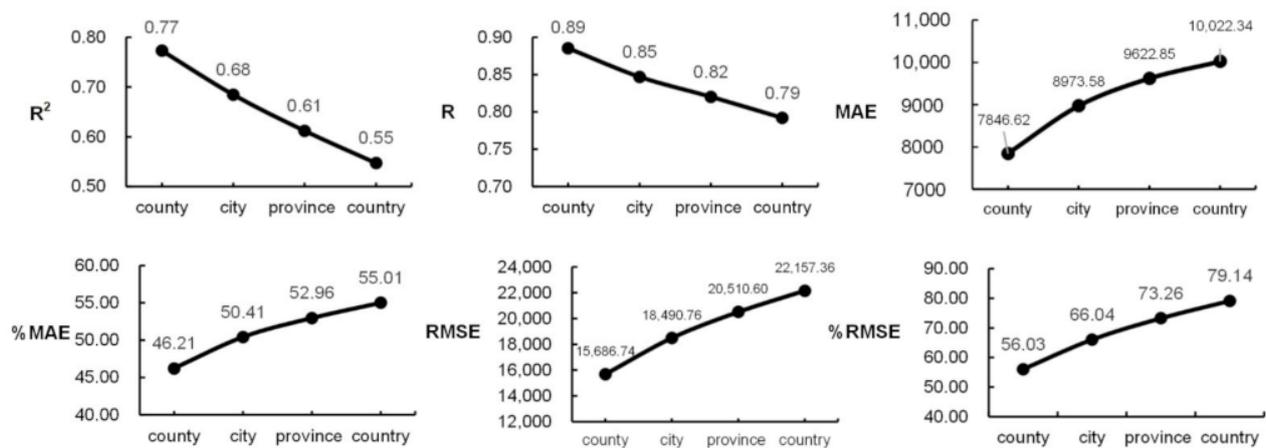
For 1985, as the corresponding census data were unavailable, we used a single country-wide population count from the World Bank Database to modify gridded population data and have not verified it. Due to the consistency of the proposed population mapping framework, we argue that data accuracy in 1985 is comparable to that in other years.

### 3.3. Accuracy Analysis of Gridded Population Data to Scales of Census Data

The availability of census data restrains the production of gridded population data, and fine-scale census data benefit accurate population mapping [48]. However, census surveys are time-consuming and labor-intensive, and, in many cases, only coarse-grained census data can be obtained [71]. Here, we utilized the population distribution in 2010 to investigate the difference in the accuracy of gridded population data based on census data of different scales. The true population and the estimated population at the town scale were compared to evaluate the accuracy of the modified data. Figure 12 shows scatterplots of the true population count and estimated population count at the town scale from gridded population data based on county-scale, city-scale, province-scale, and country-scale census data. The points of true and estimated values present clustered distribution along the 1:1 reference line at all scales, suggesting that the produced gridded population data based on all census scales can capture the heterogeneity in population distribution. Figure 13 shows the variation in the accuracy of gridded population distribution based on census data of four different scales in terms of six accuracy metrics. It is shown that with the increase in the scale of census units, data accuracy decreases. Therefore, when census data are available, it is necessary to use them to modify the raw estimations and obtain better accuracy. Comparable to GPWv4 ( $R^2 = 0.61$ , %RMSE = 73.03), based on county-census data, the  $R^2$  and %RMSE of gridded population data based on a single country-wide population count is 0.55 and 79.14, with a difference of 9.84% and 8.37%, respectively. As the difference is relatively low and a single country-wide population count is easily accessible, it is promising that the constructed framework can generate reliable gridded population data from RS images without census data efficiently.



**Figure 12.** Scatterplots of the true population count and estimated population count at town scale based on county-scale, city-scale, province-scale, and country-scale census data.



**Figure 13.** Variation in the accuracy of gridded population data based on county-scale, city-scale, province-scale, and country-scale census data in terms of 6 accuracy metrics.

### 3.4. Evolution of China's Population Distribution from 1985 to 2010

Figure 14 shows the produced gridded population maps of China with the resolution of 1 km for the years 1985, 1990, 1995, 2000, 2005, and 2010. Although the total population of China grew from 105,104,000 in 1985 to 133,770,500 in 2010, the pattern of population distribution has not changed significantly. The famous Hu-Line pattern [72], characterized by a dense population in the southeast part and a sparse population in the northwest areas of China, remains. From 1985 to 2010, the population gravity center [73] of China lay roughly at the point (113.89° E, 32.97° N), which showed a slight movement to the southeast, with a moving distance of fewer than 33 km (Figure A2), suggesting that China's population and economic center was moving towards the southeast area. In line with previous studies, China's population density is classified into eight levels in this study [34]. Among them, grid cells with a population density greater than 1500 persons/km<sup>2</sup> are regarded as high-density regions, cells with a population density between 200 and 1500 are regarded as medium-density regions, and cells with a population density less than 200 are regarded as low-density regions. Table 3 lists the percentage values of the area and population for different levels, reflecting the evolution of China's population distribution from 1985 to 2010.

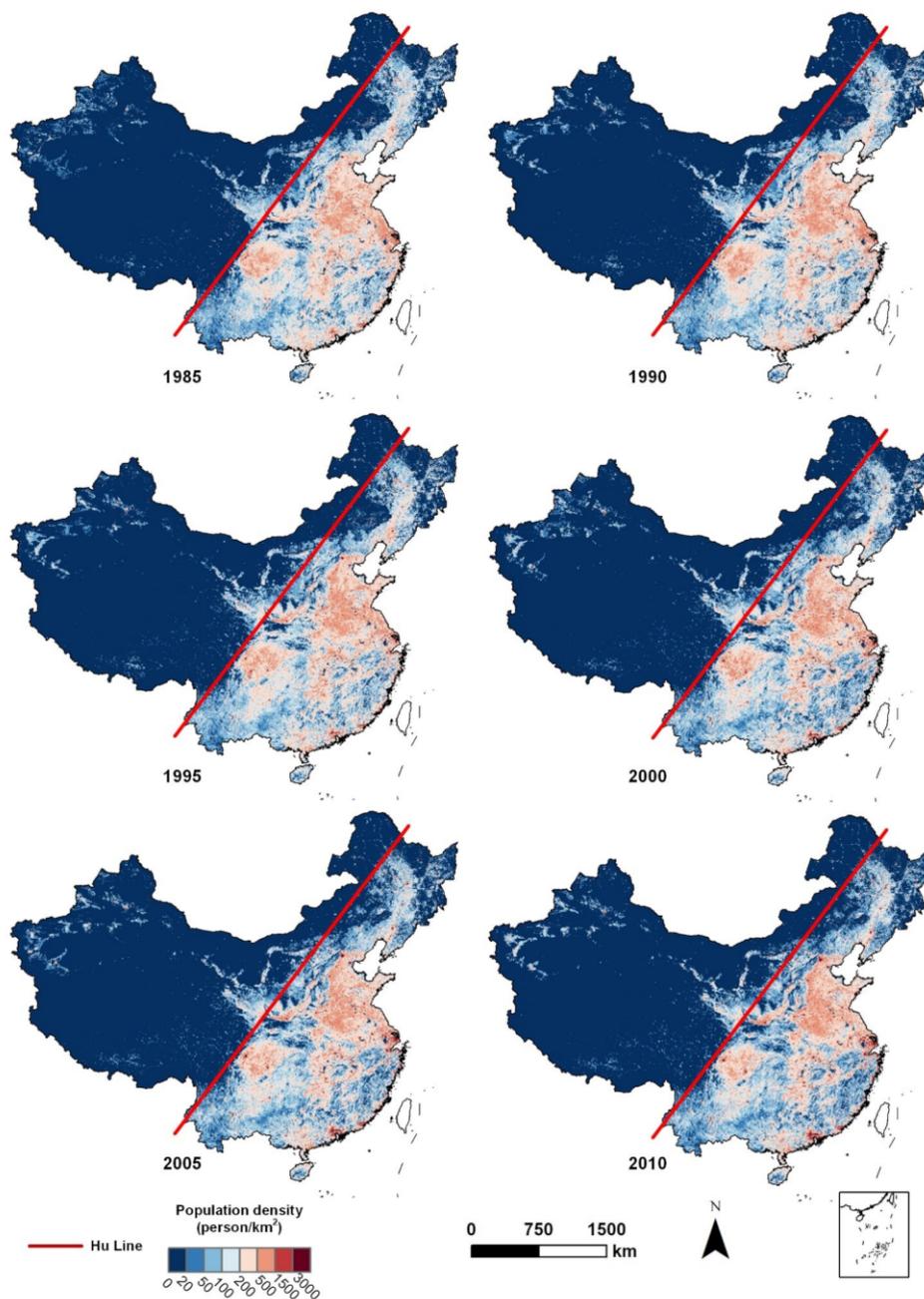


Figure 14. Gridded population data (1 km × 1 km) of China from 1985 to 2010.

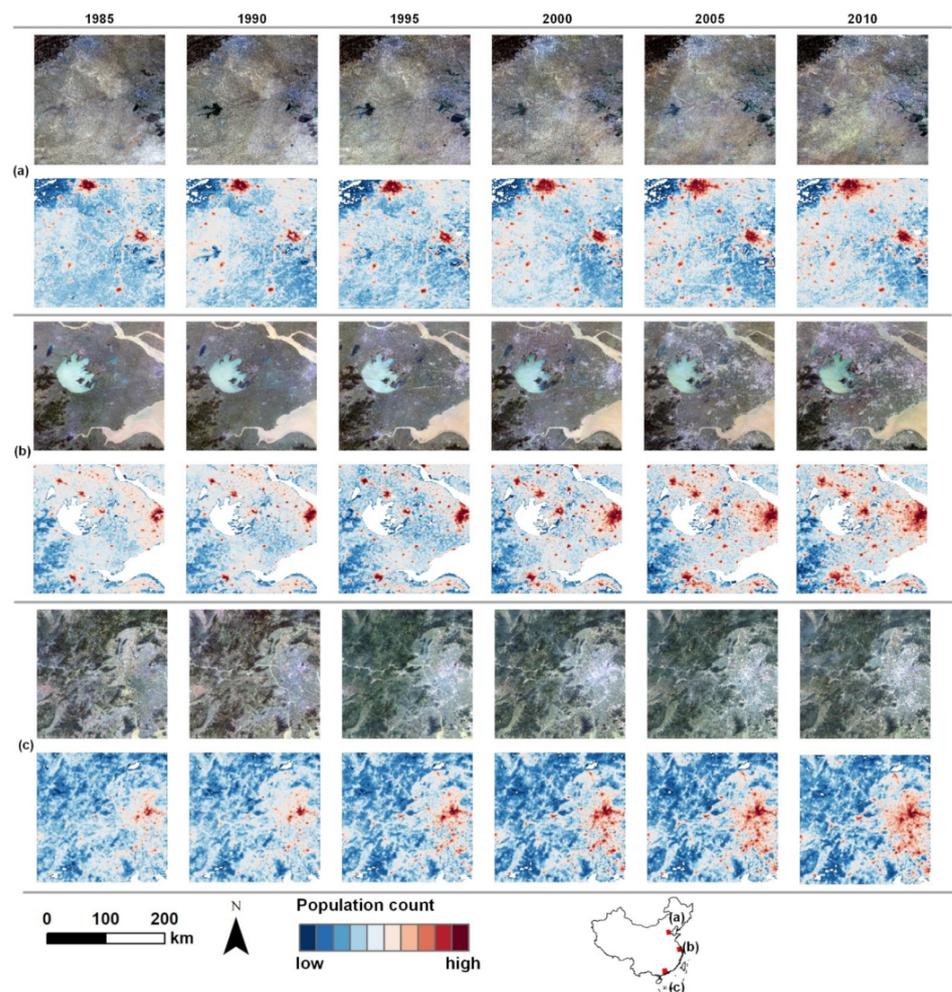
Table 3. Percentage values of area and population for different density levels.

Density	1985		1990		1995		2000		2005		2010	
	Area	Population										
Low	83.19	22.49	82.33	20.48	82.76	19.46	83.79	18.26	83.97	17.62	83.91	16.44
Medium	16.40	61.18	17.2	60.54	16.64	56.21	15.42	50.96	15.12	48.22	15.07	47.66
High	0.42	16.33	0.48	18.98	0.61	24.33	0.79	30.78	0.92	34.17	1.02	35.90

From 1985 to 2010, the area proportion of high-density regions increased from 0.42% to 1.02%, increasing by 145%, and the population proportion increased from 16.33% to 35.90%, increasing by 119.84%. Previous researches have suggested that high-density regions with a population density of >1500 persons/km<sup>2</sup> can be regarded as urbanized regions [34]. The expansion of regions with high population density can be ascribed to rapid urbanization

and the emergence of megacities due to China's reform and opening-up policy. The area proportion of medium-density regions decreased from 16.40% in 1985 to 15.07% in 2010, a decrease of 4.46%, and the population proportion decreased from 61.18% to 47.66%, decreasing by 22.09%. The expansion of megacities can explain the reduction in regions with medium population density as the concentration of the population in megacities leads to the contraction of small- and medium-sized urban regions. The area proportion of low-density regions increased from 83.19% in 1985 to 83.91% in 2010, while the population proportion decreased from 22.49% to 16.44%. The expansion of low-density regions may be attributed to immigration measures in some mountainous areas to protect the ecological environment and alleviate poverty [24]. However, with urbanization, the population becomes gradually concentrated in urban regions, leading to a reduced population in low-density regions.

Figure 15 shows the population distributions and landscape variations of three regions in large urban agglomerations in China from 1985 to 2010: (a) Beijing-Tianjin-Hebei, (b) the Yangtze River Delta, and (c) the Pearl River Delta. During this period, these areas experienced rapid urban expansion and consequent population growth, which further led to the transformation of the urban landscape. The produced continuous multi-temporal gridded population data with high spatial resolution provide support to track the co-evolution of the human population and physical landscape.



**Figure 15.** Population distributions (**bottom row**) and landscape variations (**top row**) of three regions in large urban agglomerations in China from 1985 to 2010. (a) Beijing-Tianjin-Hebei; (b) The Yangtze River Delta; (c) The Pearl River Delta.

#### 4. Conclusions and Discussion

China, as the most populous developing country in the world, has experienced rapid economic development, population growth, and urbanization in recent decades. Fine-scale population distribution data and their dynamics are a crucial component in many fields, including resource management, disaster response, public health, urban planning, and climate change; they are also fundamental in monitoring and achieving sustainable development goals (e.g., SDG 11.6.2—annual mean levels of fine particulate matter (e.g., PM<sub>2.5</sub> and PM<sub>10</sub>) in cities (population-weighted)) [74]. However, due to the lack of adequate methodology and appropriate data, there are rarely continuous multi-temporal gridded population data available for China over a long historical period to aid in our understanding of the evolution of population distribution.

The continuously improving remote sensing technology provides low-cost, broad-coverage, and high spatiotemporal resolution ground information, which, in conjunction with deep learning technology that can mine hidden geographical knowledge, enables continuous population distribution mapping. We introduced a framework integrating a ResNet-N deep learning architecture with the consideration of neighborhood effects with a vast number of Landsat-5 images from GEE for rapid multi-temporal population mapping over a long historical period in this study. The ResNet-N model was developed to establish the end-to-end mapping between population count and RS image patches. Based on the trained model, we estimated the gridded population count (1 km × 1 km) of China with corresponding Landsat-5 image patches from GEE as input from 1985 to 2010. The produced raw estimations were adjusted by available census data to acquire the final gridded population data.

The ResNet-N model with neighbor augmentation achieved  $R^2$  0.70 and %RMSE 15.91%, with a better explainability and higher absolute accuracy than ResNet, which can model the interaction between the physical environment and population and capture the heterogeneity in population distribution from RS images. An interpretation analysis revealed that the constructed deep learning model could provide valuable features for population estimation since it can distinguish the differences between natural and built-up areas and between densely populated and sparsely populated buildings. The produced gridded population data in 2010 was validated via town-scale census data and showed higher accuracy than WorldPop and GPWv4. The produced gridded population data from 1990 to 2005 were validated via county-scale total population count and achieved comparable performance to data in 2010, suggesting that the produced gridded population map can analyze spatiotemporal characteristics of China's population distribution over a long period with acceptable accuracy.

The spatiotemporal analysis of multi-temporal gridded population data showed that China's population distribution pattern did not change significantly from 1985 to 2010, and the famous Hu-Line pattern remains. With China's urbanization process and the emergence of megalopolises, the high-density population regions have dramatically expanded, with the area expanding by approximately 145% and the population expanding by approximately 120%. The concentration of the population in big cities has led to the contraction of cities with medium and small sizes. China's medium-density regions have shrunk by around 4.46%, and their population has decreased by approximately 22.09%. China's low-density regions have expanded slightly with China's poverty alleviation and mountain migration strategy [24], but the population has decreased.

The coupling of deep learning technologies and easily accessible, regularly updated, and analysis-ready remote sensing data from GEE unquestionably establishes a novel avenue that promotes multi-temporal population mapping over a long period at a large scale. However, there are several limitations of this framework. First, although informative knowledge of the population distribution can be extracted from RS images directly, socioeconomic information cannot be identified. For example, the vacancy rate of buildings is difficult to capture, making it impossible to distinguish between vacant buildings and occupied buildings [54]. Especially in China, unreasonable urban expansion has led to

the appearance of ghost cities characterized by high vacancy rates of buildings, which cause overestimation of the population [75,76]. Social sensing data and nighttime light data can depict multiple facets of human society, capturing related socioeconomic information [69,75]. In the future, integrating multi-source RS data and time-series social sensing data can further improve the framework [23]. Second, we produced the gridded population data for each target year independently. However, as population distribution is continuous in the time dimension, specific time-series analysis techniques are needed to stabilize temporal variation in population distribution [17]. Third, the deep learning model ResNet-N was trained based on samples collected from the entirety of China in 2010. Although the generalization performance to other years of the model trained in 2010 has been validated, further efforts are needed in considering generalization errors. As China has a large territory and exhibits significant internal variations, in the future, we will investigate whether using regionally parameterized models will improve the performance of population mapping [59].

The framework proposed in this paper demonstrates the feasibility of mapping multi-temporal gridded population distribution at a large scale over a long period in a timely and low-cost manner, which is particularly useful in low-income and data-poor regions. The framework can also be easily extended to a global scale or to map other gridded socioeconomic variables (e.g., GDP) for monitoring and assessing progress toward fulfillment of the SDGs [12].

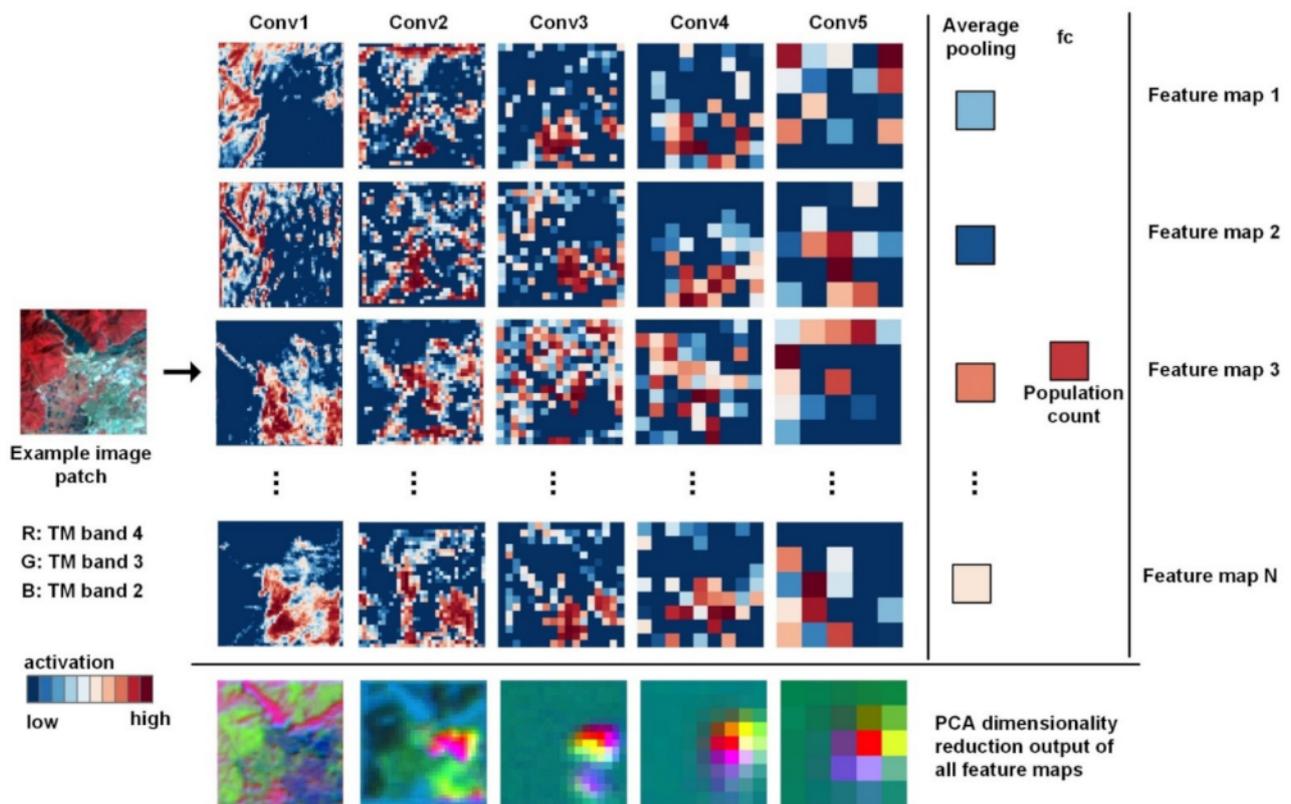
**Author Contributions:** Conceptualization, H.Z. and X.L.; methodology, H.Z. and X.L.; validation, H.Z., J.H. and C.W.; formal analysis, H.Z., X.L., Y.Y. and J.O.; resources, X.L. and J.O.; writing—original draft preparation, H.Z.; writing—review and editing, H.Z., X.L., Y.Y., J.O., J.H. and C.W.; supervision, X.L.; project administration, X.L.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key R&D Program of China (Grant No. 2019YFB2103102), the National Natural Science Foundation of China (Grant No. 41801304), and the Natural Science Foundation of Guangdong Province of China (Grant NO. 2018A030310313, 2021A1515011192).

**Data Availability Statement:** The data presented in this study are openly available in FigShare at <https://doi.org/10.6084/m9.figshare.15095748>.

**Conflicts of Interest:** The authors declare no conflict of interest.

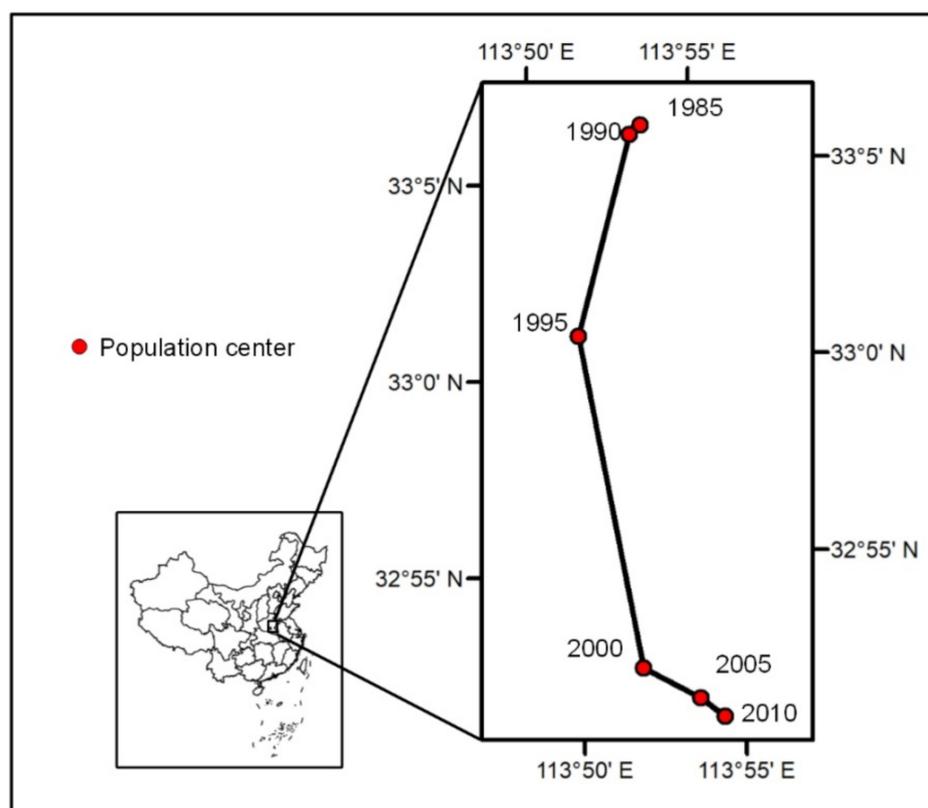
## Appendix A



**Figure A1.** Illustration of how the input RS image evolves to the output population count in the ResNet-N by an example image patch. The activations of the first three and the last feature map of each network layer were visualized. The principal component analysis (PCA) dimension-reduction technique [77] was used to compress all feature maps of each layer to 3 RGB channels for visualization. It is shown that the shallow neural layers (Conv1 and Conv2) excavate concrete features such as texture, shape, and edge from natural landscapes. Then, the deep layers (Conv3, Conv4, and Conv5) extract informative abstract features based on the shallow features for population estimation.

**Table A1.** Source and administrative unit level of census data or total population count for modifying raw population estimation of each year.

Year	Administrative Unit Level	Source
1985	Country	World Bank Database
1990	City	GPWv3
1995	City	GPWv3
2000	City	WorldPop
2005	City	WorldPop
2010	County	National Bureau of Statistics of China



**Figure A2.** The movement path of population center in China from 1985 to 2010.

## References

1. Parish, E.S.; Kodra, E.; Steinhäuser, K.; Ganguly, A.R. Estimating future global per capita water availability based on changes in climate and population. *Comput. Geosci.* **2012**, *42*, 79–86. [[CrossRef](#)]
2. Deichmann, U.; Meisner, C.; Murray, S.; Wheeler, D. The economics of renewable energy expansion in rural Sub-Saharan Africa. *Energy Policy* **2011**, *39*, 215–227. [[CrossRef](#)]
3. Ehrlich, D.; Melchiorri, M.; Florczyk, A.J.; Pesaresi, M.; Kemper, T.; Corbane, C.; Freire, S.; Schiavina, M.; Siragusa, A. Remote sensing derived built-up area and population density to quantify global exposure to five natural hazards over time. *Remote Sens.* **2018**, *10*, 1378. [[CrossRef](#)]
4. Chen, Y.; Xie, W.; Xu, X. Changes of Population, Built-up Land, and Cropland Exposure to Natural Hazards in China from 1995 to 2015. *Int. J. Disaster Risk Sci.* **2019**, *10*, 557–572. [[CrossRef](#)]
5. Chen, Y.; Li, X.; Huang, K.; Luo, M.; Gao, M. High-Resolution Gridded Population Projections for China Under the Shared Socioeconomic Pathways. *Earth's Future* **2020**, *8*. [[CrossRef](#)]
6. Mohanty, M.P.; Simonovic, S.P. Understanding dynamics of population flood exposure in Canada with multiple high-resolution population datasets. *Sci. Total Environ.* **2021**, *759*, 143559. [[CrossRef](#)] [[PubMed](#)]
7. Song, Y.; Huang, B.; Cai, J.; Chen, B. Dynamic assessments of population exposure to urban greenspace using multi-source big data. *Sci. Total Environ.* **2018**, *634*, 1315–1325. [[CrossRef](#)] [[PubMed](#)]
8. Wang, H.; Li, J.; Gao, Z.; Yim, S.H.L.; Shen, H.; Ho, H.C.; Li, Z.; Zeng, Z.; Liu, C.; Li, Y.; et al. High-spatial-resolution population exposure to PM<sub>2.5</sub> pollution based on multi-satellite retrievals: A case study of seasonal variation in the Yangtze River Delta, China in 2013. *Remote Sens.* **2019**, *11*, 2724. [[CrossRef](#)]
9. Hay, S.I.; Noor, A.M.; Nelson, A.; Tatem, A.J. The accuracy of human population maps for public health application. *Trop. Med. Int. Health* **2005**, *10*, 1073–1086. [[CrossRef](#)]
10. Song, J.; Tong, X.; Wang, L.; Zhao, C.; Prishchepov, A.V. Monitoring finer-scale population density in urban functional zones: A remote sensing data fusion approach. *Landsc. Urban Plan.* **2019**, *190*, 103580. [[CrossRef](#)]
11. Dong, N.; Yang, X.; Cai, H.; Xu, F. Research on Grid Size Suitability of Gridded Population Distribution in Urban Area: A Case Study in Urban Area of Xuanzhou District, China. *PLoS ONE* **2017**, *12*, e0170830. [[CrossRef](#)] [[PubMed](#)]
12. Estoque, R.C. A Review of the Sustainability Concept and the State of SDG Monitoring Using Remote Sensing. *Remote Sens.* **2020**, *12*, 1770. [[CrossRef](#)]
13. Wang, Y.; Huang, C.; Feng, Y.; Zhao, M.; Gu, J. Using earth observation for monitoring SDG 11.3.1-ratio of land consumption rate to population growth rate in Mainland China. *Remote Sens.* **2020**, *12*, 357. [[CrossRef](#)]

14. Zeng, C.; Zhou, Y.; Wang, S.; Yan, F.; Zhao, Q. Population spatialization in china based on night-time imagery and land use data. *Int. J. Remote Sens.* **2011**, *32*, 9599–9620. [[CrossRef](#)]
15. Huang, X.; Wang, C.; Li, Z.; Ning, H. A 100 m population grid in the CONUS by disaggregating census data with open-source Microsoft building footprints. *Big Earth Data* **2021**, *5*, 112–133. [[CrossRef](#)]
16. Fotheringham, A.S.; Wong, D.W.S. The modifiable areal unit problem in multivariate statistical analysis. *Environ. Plan. A* **1991**, *23*, 1025–1044. [[CrossRef](#)]
17. Wang, L.; Wang, S.; Zhou, Y.; Liu, W.; Hou, Y.; Zhu, J.; Wang, F. Mapping population density in China between 1990 and 2010 using remote sensing. *Remote Sens. Environ.* **2018**, *210*, 269–281. [[CrossRef](#)]
18. Mesev, V. *Remotely-Sensed Cities*; CRC Press: Boca Raton, FL, USA, 2003; ISBN 9780415260459.
19. Leyk, S.; Gaughan, A.E.; Adamo, S.B.; De Sherbinin, A.; Balk, D.; Freire, S.; Rose, A.; Stevens, F.R.; Blankespoor, B.; Frye, C.; et al. The spatial allocation of population: A review of large-scale gridded population data products and their fitness for use. *Earth Syst. Sci. Data* **2019**, *11*, 1385–1409. [[CrossRef](#)]
20. Tobler, W.; Deichmann, U.; Gottsegen, J.; Maloy, K. World population in a grid of spherical quadrilaterals. *Int. J. Popul. Geogr.* **1997**, *3*, 203–225. [[CrossRef](#)]
21. Bracken, I.; Martin, D. The generation of spatial population distributions from census centroid data. *Environ. Plan. A* **1989**, *21*, 537–543. [[CrossRef](#)]
22. Chen, Y.; Zhang, R.; Ge, Y.; Jin, Y.; Xia, Z. Downscaling Census Data for Gridded Population Mapping with Geographically Weighted Area-To-Point Regression Kriging. *IEEE Access* **2019**, *7*, 149132–149141. [[CrossRef](#)]
23. Cheng, Z.; Wang, J.; Ge, Y. Mapping monthly population distribution and variation at 1-km resolution across China. *Int. J. Geogr. Inf. Sci.* **2020**, 1–19. [[CrossRef](#)]
24. Lu, D.; Wang, Y.; Yang, Q.; Su, K.; Zhang, H.; Li, Y. Modeling spatiotemporal population changes by integrating dmsp-ols and npp-viirs nighttime light data in chongqing, china. *Remote Sens.* **2021**, *13*, 284. [[CrossRef](#)]
25. Wang, Y.; Huang, C.; Zhao, M.; Hou, J.; Zhang, Y.; Gu, J. Mapping the population density in mainland china using npp/viirs and points-of-interest data based on a random forests model. *Remote Sens.* **2020**, *12*, 3645. [[CrossRef](#)]
26. Wang, L.; Fan, H.; Wang, Y. Fine-resolution population mapping from international space station nighttime photography and multisource social sensing data based on similarity matching. *Remote Sens.* **2019**, *11*, 1900. [[CrossRef](#)]
27. He, M.; Xu, Y.; Li, N. Population spatialization in Beijing city based on machine learning and multisource remote sensing data. *Remote Sens.* **2020**, *12*, 1910. [[CrossRef](#)]
28. Luo, P.; Zhang, X.; Cheng, J.; Sun, Q. Modeling population density using a new index derived from multi-sensor image data. *Remote Sens.* **2019**, *11*, 2620. [[CrossRef](#)]
29. Zhao, Y.; Li, Q.; Zhang, Y.; Du, X. Improving the accuracy of fine-grained population mapping using population-sensitive POIs. *Remote Sens.* **2019**, *11*, 2502. [[CrossRef](#)]
30. Yu, S.; Zhang, Z.; Liu, F. Monitoring population evolution in China using time-series DMSP/OLS nightlight imagery. *Remote Sens.* **2018**, *10*, 194. [[CrossRef](#)]
31. Li, L.; Zhang, Y.; Liu, L.; Wang, Z.; Zhang, H.; Li, S.; Ding, M. Mapping changing population distribution on the qinghai-tibet plateau since 2000 with multi-temporal remote sensing and point-of-interest data. *Remote Sens.* **2020**, *12*, 4059. [[CrossRef](#)]
32. Yang, X.; Ye, T.; Zhao, N.; Chen, Q.; Yue, W.; Qi, J.; Zeng, B.; Jia, P. Population mapping with multisensor remote sensing images and point-of-interest data. *Remote Sens.* **2019**, *11*, 574. [[CrossRef](#)]
33. Eicher, C.L.; Brewer, C.A. Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartogr. Geogr. Inf. Sci.* **2001**, *28*, 125–138. [[CrossRef](#)]
34. Tan, M.; Li, X.; Li, S.; Xin, L.; Wang, X.; Li, Q.; Li, W.; Li, Y.; Xiang, W. Modeling population density based on nighttime light images and land use data in China. *Appl. Geogr.* **2018**, *90*, 239–247. [[CrossRef](#)]
35. Lo, C.P. Automated population and dwelling unit estimation from high-resolution satellite images: A GIS approach. *Int. J. Remote Sens.* **1995**, *16*, 17–34. [[CrossRef](#)]
36. Patela, N.N.; Angiuli, E.; Gamba, P.; Gaughan, A.; Lisini, G.; Stevens, F.R.; Tatem, A.J.; Trianni, G. Multitemporal settlement and population mapping from landsat using google earth engine. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *35*, 199–208. [[CrossRef](#)]
37. Elvidge, C.D.; Baugh, K.E.; Kihn, E.A.; Kroehl, H.W.; Davis, E.R.; Davis, C.W. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *Int. J. Remote Sens.* **1997**, *18*, 1373–1379. [[CrossRef](#)]
38. Wang, F.; Lu, W.; Zheng, J.; Li, S.; Zhang, X. Spatially explicit mapping of historical population density with random forest regression: A case study of Gansu province, China, in 1820 and 2000. *Sustainability* **2020**, *12*, 1231. [[CrossRef](#)]
39. Ye, T.; Zhao, N.; Yang, X.; Ouyang, Z.; Liu, X.; Chen, Q.; Hu, K.; Yue, W.; Qi, J.; Li, Z.; et al. Improved population mapping for China using remotely sensed and points-of-interest data within a random forests model. *Sci. Total Environ.* **2019**, *658*, 936–946. [[CrossRef](#)]
40. Deville, P.; Linard, C.; Martin, S.; Gilbert, M.; Stevens, F.R.; Gaughan, A.E.; Blondel, V.D.; Tatem, A.J. Dynamic population mapping using mobile phone data. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15888–15893. [[CrossRef](#)]
41. Zhao, S.; Liu, Y.; Zhang, R.; Fu, B. China's population spatialization based on three machine learning models. *J. Clean. Prod.* **2020**, *256*, 120644. [[CrossRef](#)]

42. Stevens, F.R.; Gaughan, A.E.; Linard, C.; Tatem, A.J. Disaggregating census data for population mapping using Random forests with remotely-sensed and ancillary data. *PLoS ONE* **2015**, *10*, e0107042. [[CrossRef](#)]
43. Harvey, J.T. Population estimation models based on individual TM pixels. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 1181–1192.
44. Cheng, L.; Wang, L.; Feng, R.; Yan, J. Remote sensing and social sensing data fusion for fine-resolution population mapping with a multi-model neural network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **2021**, *14*, 5973–5987. [[CrossRef](#)]
45. Doxsey-Whitfield, E.; MacManus, K.; Adamo, S.B.; Pistolesi, L.; Squires, J.; Borkovska, O.; Baptista, S.R. Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4. *Pap. Appl. Geogr.* **2015**, *1*, 226–234. [[CrossRef](#)]
46. Melchiorri, M.; Florczyk, A.J.; Freire, S.; Schiavina, M.; Pesaresi, M.; Kemper, T. Unveiling 25 years of planetary urbanization with remote sensing: Perspectives from the global human settlement layer. *Remote Sens.* **2018**, *10*, 768. [[CrossRef](#)]
47. Balk, D.L.; Deichmann, U.; Yetman, G.; Pozzi, F.; Hay, S.I.; Nelson, A. Determining Global Population Distribution: Methods, Applications and Data. *Adv. Parasitol.* **2006**, *62*, 119–156.
48. Tatem, A.J. WorldPop, open data for spatial demography. *Sci. Data* **2017**, *4*, 2–5. [[CrossRef](#)]
49. Dobson, J.E.; Bright, E.A.; Coleman, P.R.; Durfee, R.C.; Worley, B.A. LandScan: A global population database for estimating populations at risk. *Photogramm. Eng. Remote Sens.* **2000**, *66*, 849–857.
50. Yu, B.; Lian, T.; Huang, Y.; Yao, S.; Ye, X.; Chen, X.; Yang, C.; Wu, J. Integration of nighttime light remote sensing images and taxi GPS tracking data for population surface enhancement. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 687–706. [[CrossRef](#)]
51. Yao, Y.; Liu, X.; Li, X.; Zhang, J.; Liang, Z.; Mai, K.; Zhang, Y. Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 1220–1244. [[CrossRef](#)]
52. Liu, X.; Hu, G.; Chen, Y.; Li, X.; Xu, X.; Li, S.; Pei, F.; Wang, S. High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform. *Remote Sens. Environ.* **2018**, *209*, 227–239. [[CrossRef](#)]
53. Xing, X.; Huang, Z.; Cheng, X.; Zhu, D.; Kang, C.; Zhang, F.; Liu, Y. Mapping Human Activity Volumes Through Remote Sensing Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5652–5668. [[CrossRef](#)]
54. Doupe, P.; Bruzelius, E.; Faghmous, J.; Ruchman, S.G. Equitable development through deep learning: The case of sub-national population density estimation. In Proceedings of the 7th Annual Symposium on Computing for Development ACM DEV-7 2016, Nairobi, Kenya, 18–20 November 2016. [[CrossRef](#)]
55. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26–30 June 2016; pp. 770–778.
56. Helber, P.; Bischke, B.; Dengel, A.; Borth, D. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2217–2226. [[CrossRef](#)]
57. Zhu, D.; Cheng, X.; Zhang, F.; Yao, X.; Gao, Y.; Liu, Y. Spatial interpolation using conditional generative adversarial neural networks. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 735–758. [[CrossRef](#)]
58. Jean, N.; Burke, M.; Xie, M.; Davis, W.M.; Lobell, D.B.; Ermon, S. Combining satellite imagery and machine learning to predict poverty. *Science* **2016**, *353*, 790–794. [[CrossRef](#)]
59. Robinson, C.; Hohman, F.; Dilkina, B. A deep learning approach for population estimation from satellite imagery. In Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities, Redondo Beach, CA, USA, 7–10 November 2017; pp. 47–54.
60. Gaughan, A.E.; Stevens, F.R.; Huang, Z.; Nieves, J.J.; Sorichetta, A.; Lai, S.; Ye, X.; Linard, C.; Hornby, G.M.; Hay, S.I.; et al. Spatiotemporal patterns of population in mainland China, 1990 to 2010. *Sci. Data* **2016**, *3*, 1–11. [[CrossRef](#)]
61. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [[CrossRef](#)]
62. Hu, W.; Patel, J.H.; Robert, Z.A.; Novosad, P.; Asher, S.; Tang, Z.; Burke, M.; Lobell, D.; Ermon, S. Mapping Missing Population in Rural India: A Deep Learning Approach with Satellite Imagery. *arXiv* **2019**, arXiv:1905.02196.
63. Masek, J.G.; Vermote, E.F.; Saleous, N.E.; Wolfe, R.; Hall, F.G.; Huemmrich, K.F.; Gao, F.; Kutler, J.; Lim, T.K. A landsat surface reflectance dataset for North America, 1990–2000. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 68–72. [[CrossRef](#)]
64. Dwyer, J.L.; Roy, D.P.; Sauer, B.; Jenkerson, C.B.; Zhang, H.K.; Lyburner, L. Analysis ready data: Enabling analysis of the landsat archive. *Remote Sens.* **2018**, *10*, 1363. [[CrossRef](#)]
65. Liu, H.; Gong, P.; Wang, J.; Wang, X.; Ning, G.; Xu, B. Production of global daily seamless data cubes and quantification of global land cover change from 1985 to 2020—iMap World 1.0. *Remote Sens. Environ.* **2021**, *258*, 112364. [[CrossRef](#)]
66. Qiu, S.; Lin, Y.; Shang, R.; Zhang, J.; Ma, L.; Zhu, Z. Making Landsat time series consistent: Evaluating and improving Landsat analysis ready data. *Remote Sens.* **2019**, *11*, 51. [[CrossRef](#)]
67. Nguyen, G.; Dlugolinsky, S.; Bobák, M.; Tran, V.; López García, Á.; Heredia, I.; Malík, P.; Hluchý, L. Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: A survey. *Artif. Intell. Rev.* **2019**, *52*, 77–124. [[CrossRef](#)]
68. Huang, X.; Zhu, D.; Zhang, F.; Liu, T.; Li, X.; Zou, L. Sensing Population Distribution from Satellite Imagery via Deep Learning: Model Selection, Neighboring Effect, and Systematic Biases. Available online: <http://arxiv.org/abs/2103.02155> (accessed on 1 September 2021).
69. Yao, Y.; Zhang, J.; Hong, Y.; Liang, H.; He, J. Mapping fine-scale urban housing prices by fusing remotely sensed imagery and social media data. *Trans. GIS* **2018**, *22*, 561–581. [[CrossRef](#)]
70. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2020**, *128*, 336–359. [[CrossRef](#)]

71. Wardrop, N.A.; Jochem, W.C.; Bird, T.J.; Chamberlain, H.R.; Clarke, D.; Kerr, D.; Bengtsson, L.; Juran, S.; Seaman, V.; Tatem, A.J. Spatially disaggregated population estimates in the absence of national population and housing census data. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 3529–3537. [[CrossRef](#)] [[PubMed](#)]
72. Chen, D.; Zhang, Y.; Yao, Y.; Hong, Y.; Guan, Q.; Tu, W. Exploring the spatial differentiation of urbanization on two sides of the Hu Huanyong Line—Based on nighttime light data and cellular automata. *Appl. Geogr.* **2019**, *112*, 102081. [[CrossRef](#)]
73. Liang, L.; Chen, M.; Luo, X.; Xian, Y. Changes pattern in the population and economic gravity centers since the Reform and Opening up in China: The widening gaps between the South and North. *J. Clean. Prod.* **2021**, *310*, 127379. [[CrossRef](#)]
74. UN IAEG-SDGs Global Indicator Framework for the Sustainable Development Goals and Targets of the 2030 Agenda for Sustainable Development. Available online: [https://unstats.un.org/sdgs/indicators/Global%2520Indicator%2520Framework%2520after%25202020%2520review\\_Eng.pdf](https://unstats.un.org/sdgs/indicators/Global%2520Indicator%2520Framework%2520after%25202020%2520review_Eng.pdf) (accessed on 1 September 2021).
75. Zeng, Q.; Zhang, W. Research on the Development of “Ghost City” Based on Night Light Data: Taking Sichuan Province as an Example. *Open J. Soc. Sci.* **2019**, *7*, 176–188. [[CrossRef](#)]
76. Mingye, L. Evolution of Chinese Ghost Cities. *China Perspect.* **2017**, *2017*, 69–78. [[CrossRef](#)]
77. Jolliffe, I.T.; Cadima, J.; Cadima, J. Principal component analysis: A review and recent developments Subject Areas. *Philos. Trans. R. Soc. A* **2016**, *374*, 20150202. [[CrossRef](#)] [[PubMed](#)]