Resources, Environment and Sustainability 14 (2023) 100136

Contents lists available at ScienceDirect

Resources, Environment and Sustainability

journal homepage: www.elsevier.com/locate/resenv

Research article

Local–global dual attention network (LGANet) for population estimation using remote sensing imagery

Yanxiao Jiang [a,b], Zhou Huang [a,b,*], Linna Li [c], Quanhua Dong [a,b]

[a] Institute of Remote Sensing and Geographical Information Systems, School of Earth and Space Sciences, Peking University, Beijing, 100871, China
[b] Beijing Key Lab of Spatial Information Integration & Its Applications, Peking University, Beijing, 100871, China
[c] Department of Geography, California State University, Long Beach, CA, 90840, USA

ARTICLE INFO

Keywords:
Population
Remote sensing
CNN
Attention network

ABSTRACT

Accurate and rapid censuses can provide detailed basic information for a country, which is useful for resource allocation, disease control, disaster prevention, urban planning, and business management. However, traditional censuses often take up much time, manpower, and financial resources. Population maps are created by national statistical institutes at statistical units. Remote sensing imagery combined with end-to-end deep learning models makes it possible to estimate a wide range of populations at a low cost. This study demonstrates the effectiveness of a local–global dual attention network (LGANet) for population estimation using remote sensing images. The LGANet contains a local attention embranchment and a global attention embranchment on the top of the backbone to adaptively learn and integrate two discriminative features simultaneously. To enhance the precision of population estimation, the outputs from the two attention modules are combined. This method utilizes daytime remote sensing images as input, complemented by nighttime light data, to estimate the population on 1 km grids. Our method exhibits superior accuracy compared to other deep learning methods, as evidenced by an experimental comparison between the estimated population and the ground-truth population in 1 km grids.

1. Introduction

Accurate demographic statistics are crucial for achieving these goals efficiently and rapidly, providing essential information for economic development and societal transformations (Ferrie, 1996). Population distribution significantly impacts resource allocation within a country (Kummu et al., 2011), influencing decision-making processes at all government levels. In emergency situations, such as pandemics or natural disasters, a rational distribution of human and material resources is vital. Understanding population distribution in areas with high incidences forms the basis for effective disease control (Balk et al., 2006; Hay et al., 2005; Tatem et al., 2012, 2011).

Obtaining an accurate population distribution map of a country is crucial for measuring economic prosperity and urban vitality quickly and cost-effectively. However, creating a high-precision and high-resolution population map poses numerous challenges. Traditional census methods are time-consuming, resource-intensive, and expensive. Many countries conduct censuses every 5–10 years, while some low-income countries may do so once in decades. For instance, China conducts a census every ten years, with the third census in 1982 costing approximately 400 million Yuan (approximately 60.5 million US dollars). The U.S. Agency for International Development's Demographic and Health Survey (DHS) program typically conducts surveys every five years in developing countries, with costs ranging from $1.1 million to $9.7 million (Doupe et al., 2016). In financially constrained or politically unstable countries, censuses may occur even less frequently, possibly every few decades. Timely updates of population data are essential benchmarks for measuring a country's progress towards sustainable development goals. Regular updates can prevent losses resulting from decision-making errors and accelerate progress towards achieving these objectives.

High-resolution satellite imagery enables diverse applications like population distribution mapping and built environment monitoring. Thomson and Hardin (2000) utilized remote sensing and GIS techniques to map urban residential land use. However, in underdeveloped countries, obtaining accurate demographic data is challenging due to a lack of government documentation, leading to undercounting of populations in unincorporated settlements (Uzun and Cete, 2004). Typically, national statistical institutes collect population data at small enumeration units, such as census tracts or building units. But in many countries, data are only available at broader geographical units like

* Corresponding author at: Institute of Remote Sensing and Geographical Information Systems, School of Earth and Space Sciences, Peking University, Beijing, 100871, China.
E-mail address: huangzhou@pku.edu.cn (Z. Huang).

https://doi.org/10.1016/j.resenv.2023.100136
Received 21 June 2023; Received in revised form 30 August 2023; Accepted 5 September 2023
Available online 9 September 2023

municipalities, which limits analysis in various fields and introduces distortions due to significant heterogeneity in unit size. To address these challenges, Wang et al. (2020) proposed a population density mapping method based on random forest (RF) for mainland China in 2015. They developed a high-precision dataset called "Popi", offering a grid resolution of 100 m × 100 m. This dataset incorporated commonly used data sources such as elevation, slope, NDVI, land use/land cover, roads, and NPP/VIIRS, as well as 16,101,762 Points of Interest (POIs) records and 2867 county-level censuses.

In recent years, an increasing number of scientists have proposed deep learning techniques, specifically convolutional neural network (CNN) methods, to estimate population using remote sensing images (Ball et al., 2017; Liu et al., 2018b; Balk et al., 2005). These methods have been applied to various tasks, such as mapping global urban land use with MODIS data (Schneider et al., 2009), explaining local-level economic outcomes using satellite images (Jean et al., 2016), identifying patterns in urban environments on a large scale (Albert et al., 2017), and detecting offshore platforms with a time-series remote sensing approach (Liu et al., 2018a). Researchers have successfully utilized not only daytime satellite images but also nighttime satellite images and land cover/land use types as supplementary data sources for population estimation. These efforts have significantly improved the techniques for timely and accurate population estimation with diverse data sources (Cho et al., 2015; Guo et al., 2019; Kim et al., 2019; Li et al., 2019b,a; Liu et al., 2018a; Suganuma et al., 2019; Ye et al., 2019a). Unlike censuses with lengthy production cycles, satellite images are easily obtained and enable frequent updates of population data whenever new imagery becomes available. Moreover, due to their global coverage, this method can be universally applied to estimate population in any location on Earth.

In this study, we propose an improved CNN network called local–global dual attention network (LGANet) to enhance population estimation performance, making it faster and more refined. While a fundamental CNN is capable of acquiring features from diverse perspectives, its proficiency in extracting features specifically tied to population characteristics may not be prominent. Networks that incorporate an attention mechanism, however, are better equipped to focus on areas of population concentration, thus capturing more relevant features for the task of population estimation. LGANet introduces a self-attention mechanism to learn features at both local and global scales. It comprises a local attention embranchment and a global attention embranchment on top of the backbone Inception-ResNet2 (Szegedy et al., 2016), allowing simultaneous adaptive learning and integration of discriminative features. The local attention module utilizes the self-attention mechanism, supervised by night-light images, to focus on crucial areas with concentrated population. While night light images may be less effective in estimating population density in African countries (Hu et al., 2019), we include them in our analyses for identifying population distribution in major cities. The global attention module also employs a similar self-attention mechanism, selectively integrating interdependent channels and learning associated global features in all channel maps. The outputs of both attention modules are combined to improve feature representations and achieve more precise population estimation. We applied LGANet to 10-meter resolution remote sensing images and corresponding 1-kilometer grid population datasets to estimate population in Hebei Province, China. The results demonstrate higher accuracy compared to several other methods in the literature and effectively characterize the uneven distribution of residents in the area.

Our contributions are summarized as follows: (1) We propose a local–global dual attention network (LGANet) with the self-attention mechanism for population estimation using remote sensing images. The model's outstanding performance indicates its potential for estimating other socioeconomic indicators with high granularity. (2) We provide a reliable estimator for population mapping using easily accessible remote sensing imagery, this is particularly significant in low-income regions characterized by sparse infrastructures or limited availability of mobile positioning data. Our network utilizes nighttime light data as auxiliary data to provide more accurate estimation in the size and distribution of population. (3) Our method has the best performance in the experiments when compared with other deep learning methods, and in the ablation experiment, the dual attention mechanism model also achieves the best estimation.

2. Related work

Population data are traditionally obtained through censuses, which are expensive and time-consuming. To address this limitation, researchers have sought to estimate population quickly using available datasets. However, existing methods that evenly distribute population data on maps do not accurately describe population distribution (Goodchild et al., 1993). To improve population estimation, various researchers (Briggs et al., 2007; Cohen and Small, 1998; Linard et al., 2011; Long and McMillen, 2015; Mennis, 2003; Smith, 1987) have proposed models that consider additional data, such as birth rates, the number of cars, car brands, and land use data. Despite these efforts, downscaled population maps using resampling and distance weighting methods still lack precision. To address the need for accurate population distribution maps, Gaughan et al. (2015) introduced a quantitative allocation policy that uses regional parametric models for more accuracy. However, this method increases computational costs due to the requirement for additional data. Other studies (Doxsey-Whitfield et al., 2015; Tobler et al., 1997) have explored population decomposition and more reasonable population data distribution on a grid. Yet, the heterogeneity of population distribution necessitates different interpolation methods and weighting schemes for different regions (Flowerdew and Green, 1994), resulting in increased computational complexity and limited generalization capability. Traditional spatial population estimation methods include interpolation and statistical modeling. Interpolation methods assume a uniform distribution of socio-economic data within a given area (areal interpolation) or decay with distance (point interpolation). These assumptions often deviate from reality. On the other hand, statistical modeling relies on various geographic factors and multi-source auxiliary data. The complex relationships between different datasets and the significant disparities among them make it challenging to achieve accurate spatial population estimates.

The advent of the big data era has introduced new possibilities through machine learning methods (Bao et al., 2023; Yin et al., 2023). These approaches no longer rely on low-temporal-resolution traditional statistical data. Instead, they offer the potential for near-real-time population estimation, thereby overcoming the limitations of traditional methods.

With the advent of artificial intelligence, machine learning algorithms have been increasingly used for population estimation (Stevens et al., 2015; Ye et al., 2019b). For instance, Sorichetta et al. (2015) applied the random forest method to obtain a population density-weighted layer, achieving estimation results through aggregated decision tree models. Gebru et al. (2017) utilized Convolutional Neural Network (CNN) on Google Street View images to estimate socio-economic characteristics of urban areas in the United States, enabling fine spatial resolution and near real-time demography monitoring.

CNN-based models have proven effective for population density estimation using satellite images, achieving high accuracy at various geographical levels (Hu et al., 2019; Robinson et al., 2017a). Doupe et al. (2016) used a CNN-based approach for Tanzania's data training and Kenya's satellite imagery to estimate population density and distribute known population at a lower regional level. Remote sensing images, within end-to-end deep learning frameworks, are increasingly applied for accurate estimates of human activities. Incorporating the neighbor effect, capturing spatial autocorrelation, has further enhanced the network's performance (Xing et al., 2020; La et al., 2019).

Deep learning has been developing rapidly, and more scholars have noticed the efficient performance of the dual-attention mechanism.

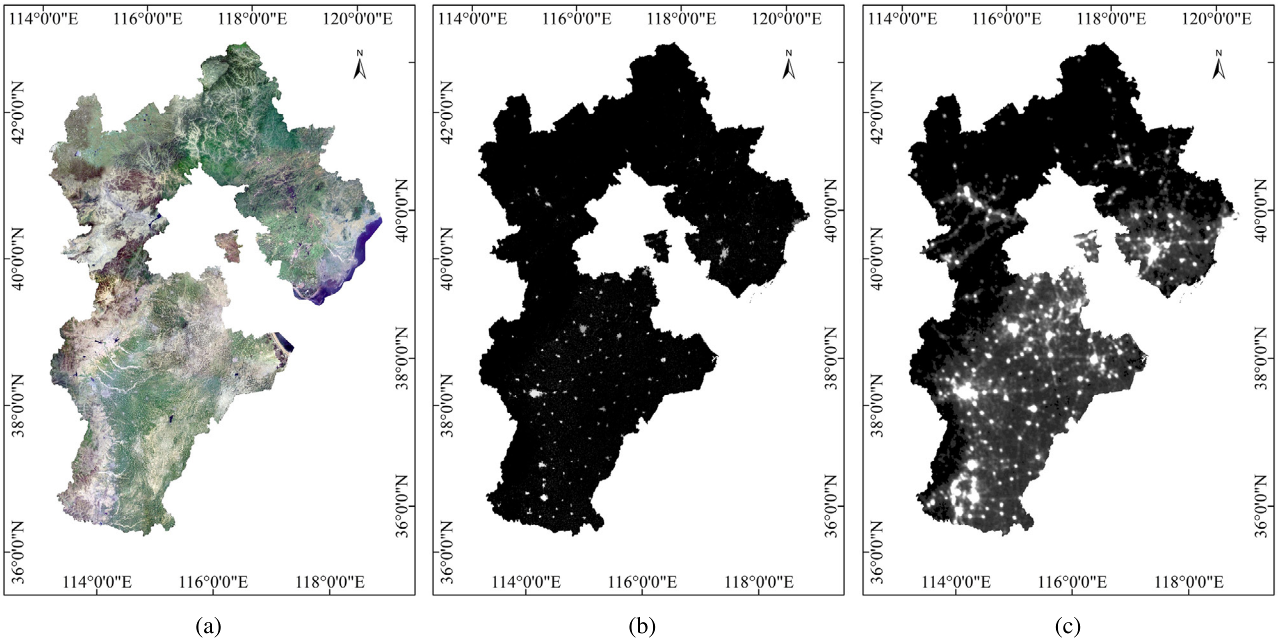


Fig. 1. Study area and data types. (a) Daytime satellite imagery (Sentinel-2). (b) WorldPop dataset. (c) Night light imagery (NPP).

It has been applied to tasks such as scene segmentation and change detection, which further demonstrates the strong capabilities and wide applicability of the dual-attention model (Fu et al., 2019; Zhang et al., 2022).

NightTime Light (NTL) and Land Use Land Cover (LULC) data are also proving valuable for estimating urban population growth (Xing et al., 2020; La et al., 2019). These studies demonstrate the usefulness of combining NTL and LULC data in analyzing urban growth and estimating population changes.

3. Datasets

In this study, we utilize two primary data sources: population data and remote sensing images. The population data from Worldpop serves as the ground truth for model verification. As for the remote sensing images, we use daytime remote sensing images as the main input and supplement them with night light data. Fig. 1 provides an overview of the data types, and Table 1 offers detailed information.

3.1. Population dataset

This study utilizes the WorldPop dataset (https://www.worldpop.org), with a spatial resolution of 1 km. It combines the most recent official census population data and various spatial ancillary datasets using a semi-automated dasymetric modeling approach by Stevens et al. (2015), employing a flexible “Random Forest” estimation technique. The WorldPop dataset offers two products: one showing the population density per hectare and the other displaying the population count per grid. For our study, we use the latter dataset.

3.2. Satellite images

Our proposed model incorporates two types of satellite imagery: daytime satellite images and night light data (NPP). Daytime satellite images, acquired from Google Earth Engine with county boundaries, are selected as both the train and validate datasets, as well as the test dataset. These images, captured in 2020, consist of three bands (B4, B3, and B2 — RGB) with a time resolution of one year and a spatial resolution of 10 m. Notably, they offer advantages over census data in terms of production cycle and timeliness.

Table 1
Data sources.

Type	Source	Spatial resolution	Temporal range
Daylight satellite images	Sentinel-2	10 m	2015-Present
Nightlight satellite images	NPP	500 m	2013–2022
Population	Worldpop	1 km	2017–Present

We utilize night light data (NPP) as supplementary data, obtained from the US Air Force Meteorological Agency and processed by the National Geophysical Data Center of the National Oceanic and Atmospheric Administration. The night lighting data has a time resolution of one year and a spatial resolution of 500 m, presented as a 30 arc-second grid. Unlike the commonly used DMSP-OLS dataset (Hu et al., 2017; Zhang et al., 2017) for stable night-time light analysis, the NPP-VIIRS DNB composite imagery provides the advantage of being calibrated in orbit, ensuring higher spatial accuracy and eliminating brightness saturation issues (Elvidge et al., 2013).

4. local–global dual attention network

We propose a satellite-based method for population estimation. The model is trained with ground-truth population data as labels and can be applied in Hebei province. This section outlines our network framework and introduces two attention mechanism modules: local attention and global attention. Finally, we discuss their integration.

4.1. Overview

LGANet takes satellite images of Hebei province as input and provides population estimations as output. Given that people typically reside in residential areas with buildings, our focus is primarily on estimating population in these local regions. However, remote sensing images vary between densely populated and sparsely populated areas, making it essential to capture global information as well. The convolution operation creates a local receptive field, but it may introduce inconsistencies across different areas, affecting estimation accuracy. To overcome this issue, we incorporate an attention mechanism to enable the network to learn more useful and discriminative information.

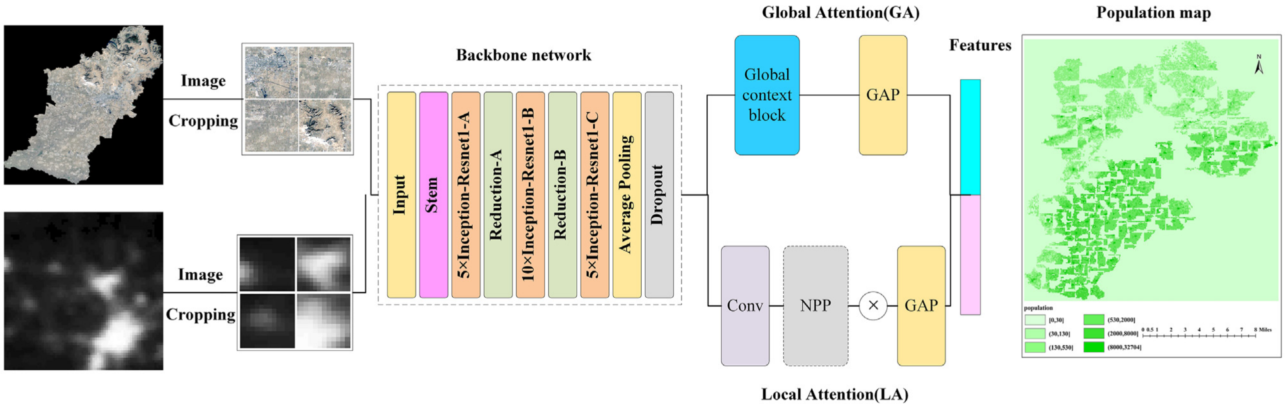


Fig. 2. The overall architecture of the proposed local–global dual attention network (LGANet).

In Fig. 2, we utilize daylight remote sensing images as input and nightlight data as auxiliary information, both cropped by population label to a 1 × 1 km grid. The backbone network chosen is Inception-ResNet2, and image inputs are resized to 224 × 224 × 3 (width × height × channels). These cropped data are fed into the backbone network. To enable the network to learn global and local features and enhance estimation results, we incorporate dual attention mechanism modules.

Specifically, for the global attention module, we remove the global average pooling in the last layer of the backbone and directly connect its penultimate activation layer to the global context block. We then add a global average pooling layer after the global context block.

The bottom part of Fig. 2 represents the local attention module. To begin, we use a convolution layer for dimension reduction to obtain features. These features are then fed into the local attention module, allowing the network to capture local long-range contextual information. This is achieved through a matrix multiplication between the local attention matrix and the original features. Subsequently, we apply a global average pooling layer to the resulting output of the matrix multiplication. These steps enable the network to emphasize areas where a large number of people gather, improving the accuracy of population estimation.

Finally, we concatenate and fuse the features outputted by the global attention module and the local attention module. This process enhances feature representations and significantly improves the accuracy of population estimation.

4.2. Local attention module

A good network structure should not only possess discriminant features for task understanding but also be capable of capturing long-range contextual information. However, traditional Fully Convolutional Networks (FCNs) may lead to incorrect classification of objects and stuff (Peng et al., 2017; Zhao et al., 2017). To tackle this issue and capture rich contextual relationships among local features, we introduce a local attention module, which effectively addresses the problem at hand.

We utilize a convolution layer to transform the channel number of the feature map to 1, which is supervised by the night light data (NPP). As a result, we get the local weight matrix. The following formula is defined to obtain the local attention feature map:

$$\mathbf{F}_L = (1 + \varnothing(\mathbf{X})) \cdot \mathbf{X} \tag{1}$$

where $\varnothing(\mathbf{X})$ is the local weight matrix obtained using night light data supervision and $\mathbf{X}$ is the feature map obtained by the Inception-ResNet2. Then we obtain the local attention feature map $\mathbf{F}_L$ of $\mathbf{X}$.

Once the local attention feature map is obtained, a two-dimensional feature map is extracted using a global average pooling layer.

4.3. Global attention module

Global context is essential for computer vision tasks. Global context (GC) blocks can be easily integrated into various network architectures, making the network perform better across different visual recognition tasks. To capture global contextual information, we introduce the global attention module.

Fig. 3(a) illustrates the three steps of the simplified non-local block: (1) Global attention pooling uses a 1x1 convolution $\mathrm{W_k}$ with the attention weights acquired using the softmax function to capture global context features; (2) Feature transformation is performed by a 1x1 convolution $\mathrm{W_v}$; (3) Feature aggregation combines each position's features with the global context aggregation features using addition.

Compare GC block with the SE block, shown in Fig. 3(b). The SE block comprises three main components: (a) the squeeze operation, which models global context using global average pooling; (b) the excitation operation, involving a bottleneck transform module with a sequence of operations: one 1x1 convolution, one ReLU activation, one 1x1 convolution, and a sigmoid function, to calculate channel importance; and (c) the fusion process, which uses a rescaling function to recalibrate channel-wise features. The SE block is designed to be lightweight, suitable for applying to all layers with minimal computational overhead, unlike the non-local block.

The detailed architecture of the GC block is shown in Fig. 3(c), which is represented as:

$$\mathbf{z}_i = \mathbf{x}_i + W_{v2} ReLU \left[LN \left[W_{v1} \sum_{j=1}^{N_P} \frac{e^{W_k \mathrm{x}_j}}{\sum_{m=1}^{N_P} e^{W_k \mathrm{x}_m}} \mathrm{x}_j \right] \right] \tag{2}$$

where $\alpha_j = \frac{e^{W_k \mathrm{x}_j}}{\sum_{m=1}^{N_P} e^{W_k \mathrm{x}_m}}$ indicates the weight for global attention pooling and $\delta(\cdot) = W_{v2} ReLU \left[LN \left[W_{v1} \right] \right]$ means the bottleneck transform. The GC block comprises three key components: global attention pooling for contextual modeling, bottleneck transform for channel-wise dependencies, and broadcast element-wise addition for feature fusion. These components synergistically enhance the block's performance.

Compared to the non-local block, the GC block significantly reduces computational cost, allowing for its application to multiple layers with minimal computation increase. It effectively captures long-range dependencies and aids network training. The GC block differs from the SE block in two main aspects. First, the fusion module has different goals; the SE block focuses on recalibrating channel importance through rescaling, while the GC block uses addition to capture long-range dependencies, followed by the NL block for further enhancement. Second, the layer normalization in the bottleneck transform is distinct; the GC block employs it to optimize the two-layer architecture for improved performance, while the SE block uses global average pooling, a specific form of global attention pooling, and addition for fusion.

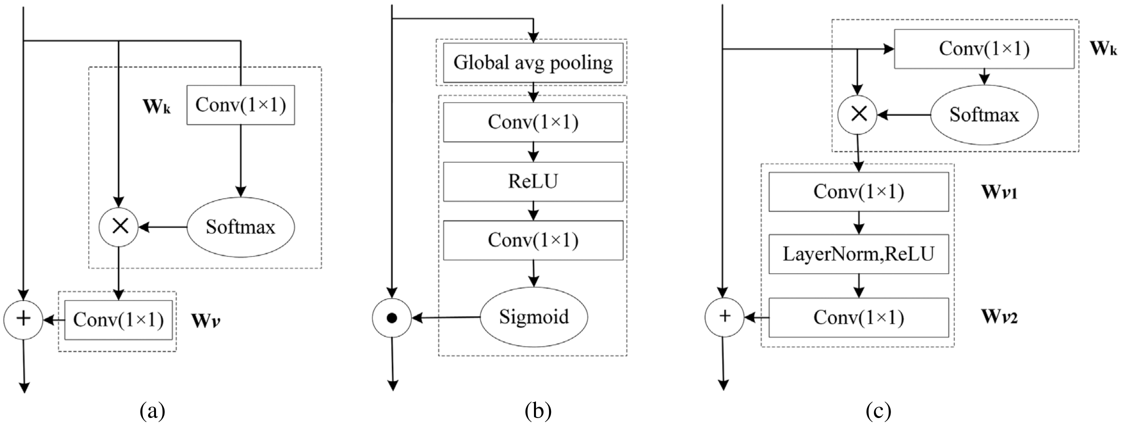


Fig. 3. Comparison of different blocks. (a) Simplified NL block; (b) SE block; (c) GC block.

4.4. Combination of attention modules

To leverage long-term contextual information effectively, we fuse features from both attention modules. The features from the local and global attention modules are concatenated to obtain the final discriminant features. During training, the proposed dual local–global attention network is optimized using two types of loss. The first one is the Mean Absolute Error (MAE) loss, utilized to measure predicted population, formulated as follows:

$$L_{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|b_i - \widehat{b_i}\right| \tag{3}$$

where b_i denotes the normalized ground truth of population, $\widehat{b_i}$ is the output of the network, and n is the batch size.

Another loss is the cross-entropy loss supervised by the night-light images in the local attention model. The cross-entropy loss is formulated as follows:

$$L_{cel} = -[y\log\hat{\mathbf{y}} + (1-y)\log(1-\hat{\mathbf{y}})] \tag{4}$$

where y denotes the normalized ground truth of population, $\hat{\mathbf{y}}$ is the output of the network. Then backpropagation is used to update the parameters of the LGANet.

5. Experimental results

5.1. Implementation details

To improve population size estimation in Hebei Province, we cropped the population grid data to 1x1 km and aligned it with corresponding daytime remote sensing images and nighttime illumination data in the same coverage area (each data has geographic attributes and is cropped accordingly). Approximately 10% of the daytime satellite images are ignored due to irregular county boundaries. The cropped population ground truth indicates that 96.24% of the population sizes fall within the range of [0, 1000], with a small portion exhibiting extremely high population densities, reaching up to 50,249. However, due to the limited range of population data in the training dataset, population estimates in densely populated areas may be less accurate. To mitigate this, we normalize the values and convert them to normal estimates during the validation phase. The distribution of the ground truth population data is presented in Fig. 4.

In this paper, we split the images and labels into a 90% training partition and a 10% validation partition, ensuring consistent distributions in both sets. This results in 233,598 training samples and 25,956 validating samples. During validation, we verify population estimation results for individual images and aggregate total population at the county-level area for validation. For the training process, we set the base learning rate to 0.00001 and use a batch size of 16 for the training dataset and 8 for the validation datasets. A dropout rate of 0.8 is applied. When both attention modules are used, we employ a multi-loss approach at the end of the network. We use the Adam optimization method (Kingma and Ba, 2017) from the Python Keras library. The model architecture consists of two fully connected layers with ReLU as the activation function. We implement the model using Keras, which is built on the TensorFlow machine learning system. Training is performed on an NVIDIA GeForce GTX 1080 Ti GPU with 11 GB of onboard memory.

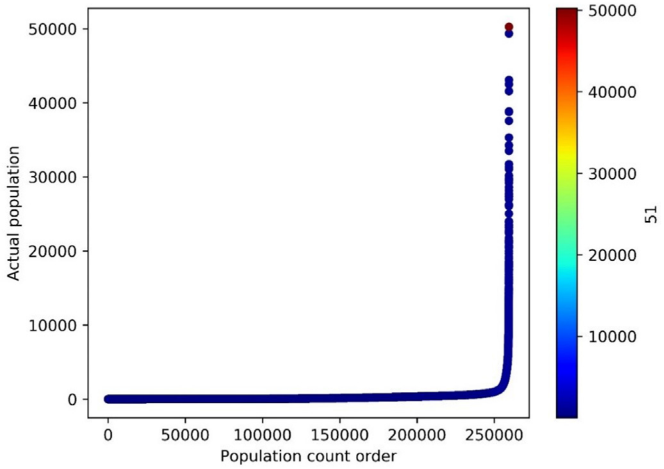


Fig. 4. Distribution of population label. The sample distribution is extremely uneven. The *x*-axis indicates the order of the population from small to large, At least 250,000 population labels are under 10,000 people.

5.2. Feature visualization

We extract high-dimensional feature maps from remote sensing images using the CNN model. To evaluate the discriminative ability of these learned features, we map them to a two-dimensional space. Fig. 5 displays representative county images and their corresponding feature maps. Notably, the extracted features are more pronounced in densely populated areas, indicating the network's focus on regions with larger populations. The population of the illustrated areas increases sequentially, reflecting the network's capability to capture population variations.

Fig. 5(a–d) are daylight remote sensing images and Fig. 5(e–h) are corresponding features extracted using the network. As seen in Fig. 5(a), Zhuozhou is sparsely populated, so the network has captured

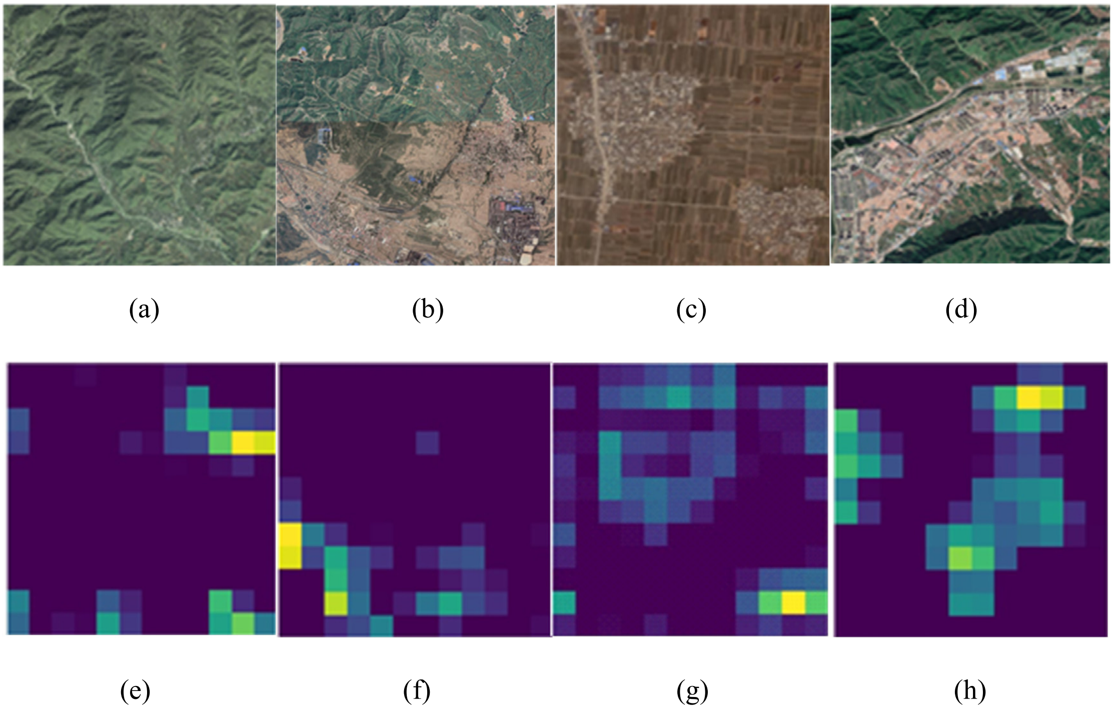


Fig. 5. Feature visualization. Figures (a)–(d) are remote sensing images of Zhuozhou, Wu an, Bai xiang, and Cheng de County in 2020, respectively, where the population is increasing in order. Figures (e)–(h) are features of Zhuozhou, Wu an, Bai xiang, and Cheng de County in 2020, respectively.

areas where population is small (Fig. 5e). This means that the network does not only extract features based on the population size, but also learns other geographical semantic information. When the model is applied to areas with a sparse population, the estimated value may be higher than the true value, which is one of the main causes for estimation errors. It is observed that more built-up areas where population is more in Fig. 5(b) and (c), so the extracted features in Fig. 5(f) and (g) are also reasonable, and the main areas of population distribution are extracted, and in Fig. 5(d), there is a large of population. Fig. 5(h) shows that the network easily extracted the population aggregation areas, but the population obtained by the regression model is not as high as the ground truth. Because the sample size of the area with a large population is very little, it may not achieve high accuracy in estimating the data with a very large population. Although the network will not be particularly accurate in estimating the number of samples with large populations, the network can do a good job of capturing areas with large population distributions. It is obtained that the obvious position in the feature map is located in a place where population gathered, so the network estimates for images with large populations well.

5.3. Evaluation method

We evaluate the model at two levels: grid level and county level. At the grid level, we compare the model's output values with the ground truth from WorldPop. For the county level, we aggregate the output values within each county boundary and compare them with the corresponding label data for that county.

To evaluate the performance of different models, we primarily use the Pearson correlation coefficient as the evaluation metric. This coefficient measures the linear correlation between two variables, quantifying the strength and direction of their relationship. We also used relative error and Mean Absolute Error (MAE) indicators to demonstrate experimental performance. The smaller the two indicators, the better the experimental performance.

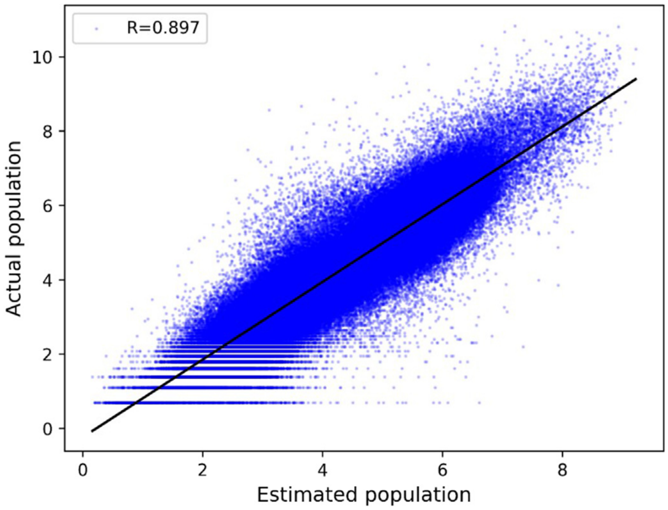
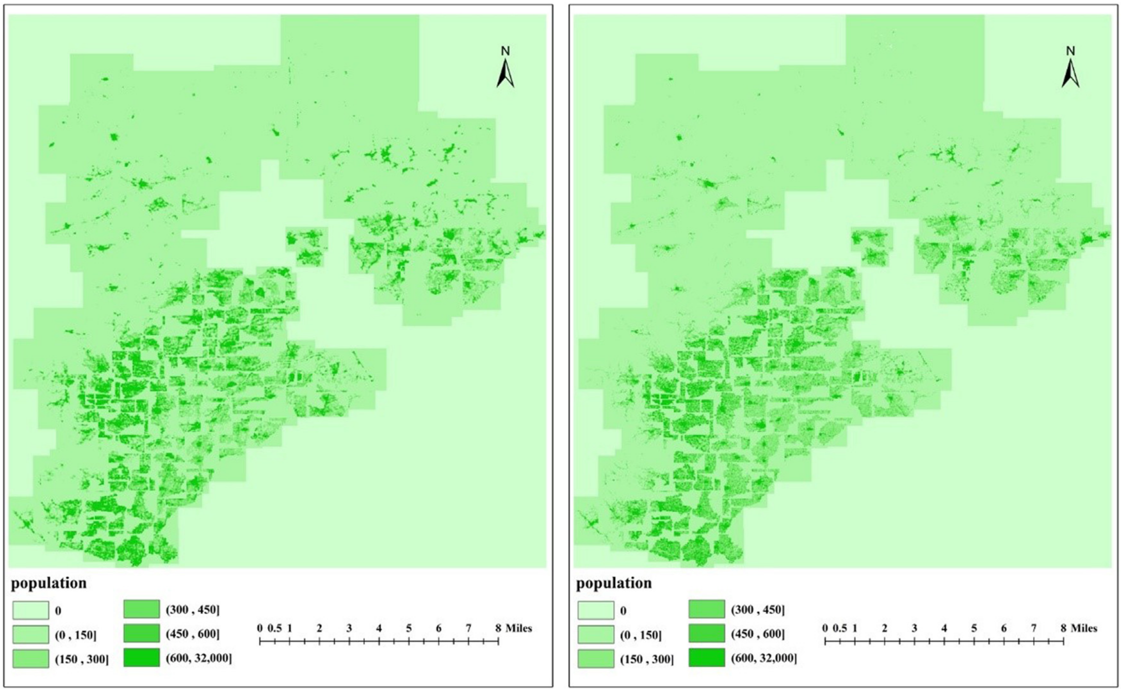


Fig. 6. The relevance of population estimation in Hebei province at grid level.

5.4. Evaluation of grid level

To establish the most intuitive validation of the model, the output value of each image is compared with the true value of WorldPop. This input image level is the most fine-grained comparison in our datasets. The results show that the Pearson correlation coefficient reaches 0.897 in all images (see Fig. 6). The proposed population estimation method demonstrates high accuracy in acquiring population information at a spatial resolution of 1 × 1 km. It provides detailed insights into the evolution of population distribution, capturing fine-grained patterns with precision.

We utilize the image population output to produce an updated grid map. As a result, we present a population grid map for Hebei



(a)

(b)

Fig. 7. The distribution of grid-level population in Hebei province. (a) The population label, (b) The estimated population.

Table 2
Performance of the model in the assessed counties.

Counties	Pearson	Ground truth	Predicted
Zhangbei city	0.921	24.4863	31.7522
Kangbao city	0.909	16.3362	20.4062
Shangyi city	0.894	57.4635	60.2118
Zhuolu city	0.844	23.6236	32.3970
Laishui city	0.863	24.4486	32.7200

province in the year 2020, utilizing a spatial resolution of 1 × 1 km. The accuracy of the population estimates in the grid map is validated by comparing them against reference data. This validation process ensures the reliability and quality of the population distribution depicted in the map. In Fig. 7, 35.1% of the grids have a relative error between [0,0.25]. The population estimation is accurate in these grids because they have less than 10,000 people, whose distribution is consistent with the overall distribution of the whole dataset. 28.1% of relative error of grid-level situate [0.25,0.5], 16.6% of relative error of grid-level situate [0.5,0.75], 20.2% of relative error of grid-level more than 0.75, 96.7% of these grids distribute at somewhere population less than 1000, For these grids with fewer people, slight deviations result in larger relative errors in population estimation.

5.5. Evaluation at the county level

To comprehensively validate the effectiveness of our model, we summarize the results at the county level, including 172 counties and districts in Hebei Province. As shown in Table 2, the Pearson correlation coefficient of Zhangbei City and Kangbao City at the county level is as high as 0.9. In the distribution range of relative errors, 71.3% of the counties are located between [−0.25, 0.25] in Fig. 9, which shows that our model performs well.

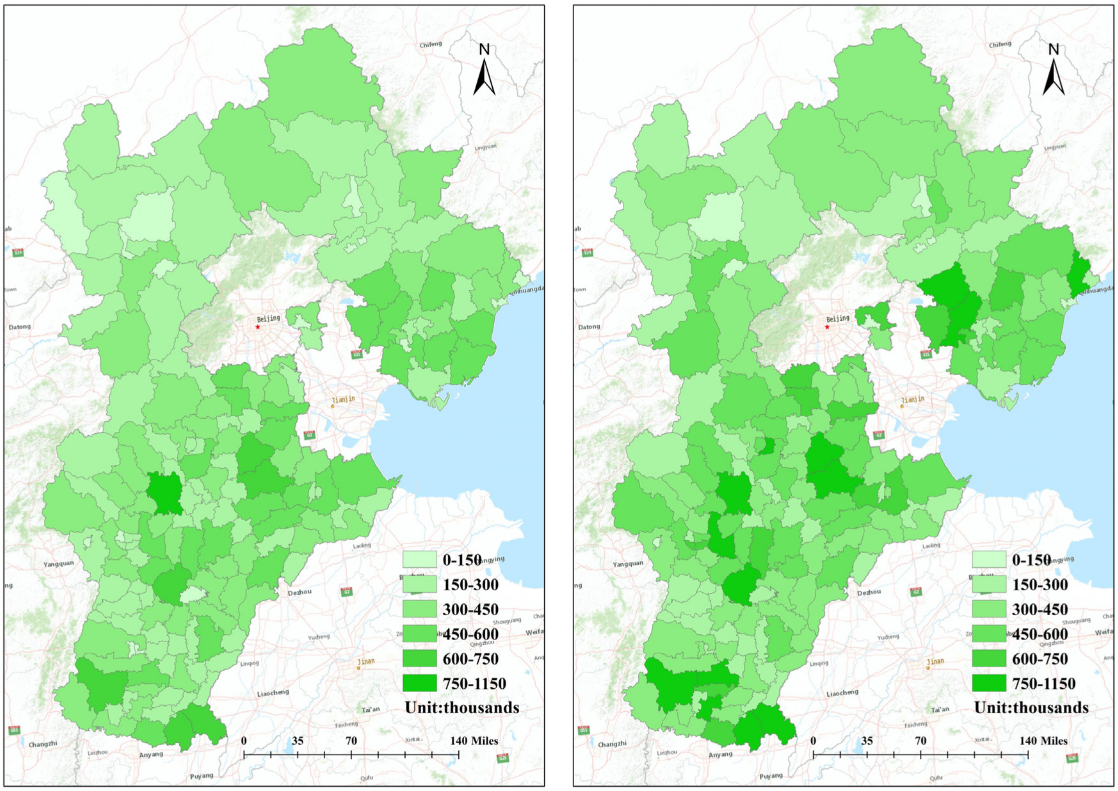
Fig. 8(a) and (b) show the distribution of the actual population and the predicted population in the Hebei province. It is noted that the population in the middle and northeastern parts of Hebei Province is higher. Our model predicts accurate population estimation in most places of Hebei, such as the northwestern, middle, and southern parts of the province. Larger relative errors appear in areas with smaller populations.

In Fig. 9, the confusion matrix illustrates that the accuracy of estimation increases with the increase of the sample size. As shown in Fig. 9, the counties in Hebei province are classified into six categories according to the population size: Category 1 has the smallest population while Category 6 has the largest population. The accuracy of classification of the proportion of 1 and 5 categories is not high and easily be identified to other categories proportion because the proportion of 1 and 5 categories samples are relatively less, the model does not learn the features of recognition well. The proportion of the 2–4 categories is relatively more, the accuracy of classification is higher as the sample size of each category increases. Population estimates for category 6 are low due to the influence of the sample, which is severely unevenly distributed, with very small sample sizes for high-density populations. Obvious estimation differences in spatial characteristics existed in six categories of population, from the aspects of population distribution. The northeastern and southwestern regions are the most populous areas, followed by the central, southeastern, and northern regions. Our population estimation results are generally consistent with the true values.

As shown in Fig. 10, the estimated value of our model is higher than the true value in most regions, only 2.38% counties of the county population estimation is small. The regions with large relative errors occur in the northeast of Hebei Province, where the population is large. In some areas with very sparse population, the relative error is greater because of the small population.

5.6. Ablation studies

To better illustrate the role and effect of local and global modules in our model, we conduct an ablation experiment with different settings. Pearson coefficient and relative error are chosen as evaluation metrics, with an evaluation conducted at the grid level and county level. The performances are shown in Tables 3 and 4.



(a) (b)

Fig. 8. The distribution of county-level population in Hebei province. (a) The population label aggregate to county. (b) The estimated population aggregate to county.

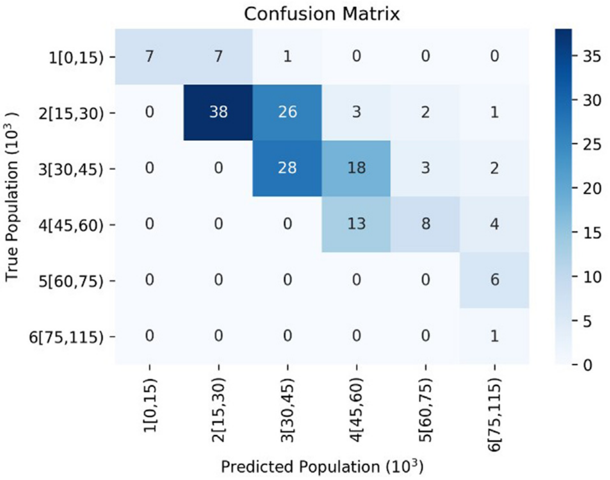


Fig. 9. Confusion matrix for aggregation to county results.

Table 3
Evaluation of models.

Model	Grid level		
	Pearson	Relative error	MAE
Base Model	0.642	0.219	174.172
Global Attention Model	0.787	0.188	159.576
Local Attention Model	0.837	0.157	43819082
Local–global Attention Model	**0.897**	**0.134**	**125.963**

Table 4
Evaluation of models.

Model	County level	
	Pearson	Relative error
Base Model	0.642	0.219
Global Attention Model	0.787	0.188
Local Attention Model	0.837	0.157
Local–global Attention Model	**0.897**	**0.134**

After a comprehensive comparison, we find that Inception-ResNet2 as our basic model is sufficient to complete the task of estimating the population. It demonstrates good performance in population estimation at both levels and the relative error at the county level is inferior only to our local model. Compared with the basic model, the Pearson coefficient of the global attention model is higher at both levels, because it utilizes the global attention pooling for context modeling and fusion feature better. Compared with the basic model and the global attention model, the local model has an even higher Pearson correlation coefficient, because this network identifies regions with clustered population through the local attention model. However, the relative error is large. This is because the network produces some large results when predicting some pictures with small population, which greatly affects the overall relative error results. The MAE indicator performs best in our dual attention model, followed by the global attention model, the basic model, and finally the local attention model. This may be due to the model's excessive focus on population gathering areas, where estimates tend to be relatively high.

At the grid level, the Pearson coefficient of the local–global model is the highest and the base model is the lowest in our four models. In addition, the Pearson coefficients of all four models at the county level are higher than those at the image level. Based on the experiments, the addition of either the local model or the global model to the basic model results in higher accuracy in population estimation. When the local and global models are integrated to enhance the performance of

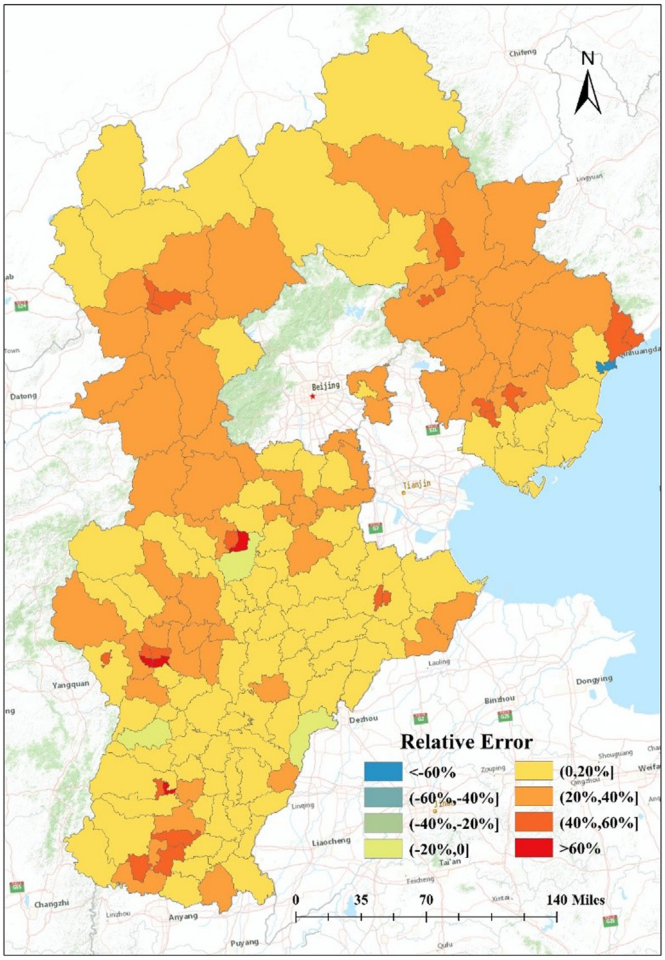


Fig. 10. The estimation errors of the population estimation.

the base network, the resulting combined model achieves the highest Pearson coefficient at both local and global levels.

5.7. Comparison with other methods

To better evaluate the performance of our model, we choose commonly used population estimation methods in recent years for comparison, including a method from a study in the United States in 2017 (Robinson et al., 2017b), a method applied in Kenya in 2016 (Doupe et al., 2016), and a method in applied in Kenya in 2014 (Stevens et al., 2015). To guarantee a fair comparison, the same sample size of the satellite image is used as other methods, without DMSP data. We choose the same assessment level in the experiments. The results are listed in Table 5.

Robinson et al. (2017b) adopt a random forest model (RFM) to generate a gridded population dataset. As shown in Table 4 the performance of this method is not as good as our method in terms of Pearson correlation coefficient, MAE and the relative error. When the n_estimators parameter increases in their model, although the estimation results are improved, the computational time is longer, and it is easier to over-fit (Liaw and Wiener, 2002). Our method is less time-consuming with good performance.

The second network under comparison is the modified VGG-A model, which is relatively simple with ordinary estimation performance (Robinson et al., 2017b). A network model for population estimation called LANDSAT landstats (LL) is also compared, but the results of population estimation at the image level and county level are poor, which may be related to the depth of the network and the difference in research areas (Doupe et al., 2016).

Table 5
Comparison with prior works.

	Grid level			County level	
	Pearson	MAE	Relative error	Pearson	Relative error
RFM	0.774	150.969	0.185	0.831	0.339
VGG-A	0.631	169.865	0.214	0.675	0.415
LL	0.628	174.485	0.220	0.620	0.371
Ours	**0.897**	**125.963**	**0.134**	**0.921**	**0.225**

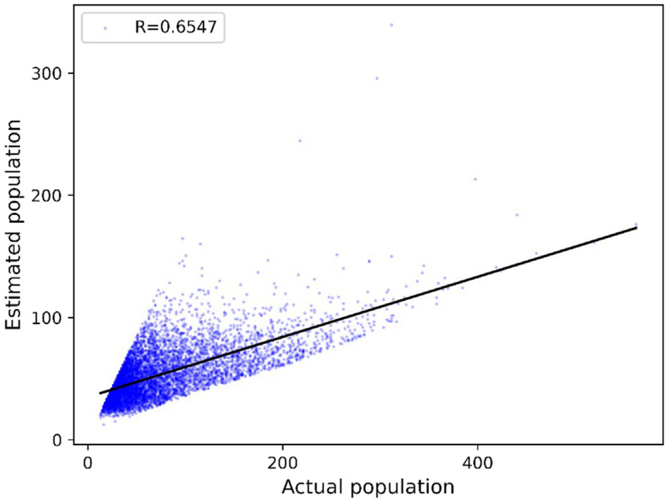


Fig. 11. Generalization experiment correlation.

Here we sum up the best performance of our method. At the grid level, only our method's Pearson coefficient reaches more than 0.9, and the Pearson correlation coefficient at the county level is at least 0.09 higher than other methods, MAE indicators are at least 25.006 less than other models. Our local–global attention network better extracts the features of the population to provide better population estimation.

5.8. Generalization

Furthermore, we conducted a generalization experiment utilizing imagery from the Beijing region. The objective of this experiment was to evaluate the model's capacity to execute tasks beyond the scope of its training data—applying acquired knowledge to diverse and potentially unfamiliar environments. We handpicked a set of 6079 images from Beijing and normalized the input imagery to cover an identical 1 km area. Through this endeavor, we aimed to assess the model's adaptability and performance across geographically distinct regions, each harboring its own distinctive characteristics and challenges.

This undertaking allowed us to delve into the model's robustness and its ability to extrapolate its predictions to disparate geographical contexts. The outcomes, as illustrated in Fig. 11, revealed an estimated Pearson correlation coefficient of 0.6547.

6. Discussion and conclusion

The CNN model, which directly estimates population from satellite images, has a good application prospect. Our model shows a high Pearson correlation coefficient in estimating population at the county level and grid level, with relative errors controlled in a small range. The advantages of our model are summarized as follows:

In the background of remote sensing images, the use of deep learning methods provides better technical support for population estimation and planning because it is faster and cheaper than censuses.

NPP-assisted night-light data are added to remote sensing images, and local attention mechanism is used as a constraint condition to learn more discriminant features to better estimate population.

With gridded population data, we provide more fine-grained population estimation, which may be useful in a wide range of applications.

This paper utilizes daytime remote sensing imagery and nighttime light data as inputs to estimate population counts, achieving near-real-time population estimation tasks. However, there are inherent limitations to this approach. Population estimation lacks a genuine validation dataset, often making direct accuracy validation of each grid unfeasible due to the absence of ground survey data. Future endeavors should focus on enhancing field surveys and data collection within grid cells, along with GIS integration, to provide effective data support for result validation.

Limited exploration has been conducted regarding the choice of the backbone network. Further investigation into whether a more suitable backbone could enhance performance would be beneficial. The inclusion of nighttime light data as auxiliary information and the utilization of attention mechanism models have contributed to accuracy improvement. Furthermore, the prevalence of large-scale models in today's landscape opens up broader avenues for experimentation and exploration.

Our model achieves the goal of estimating population directly from satellite images, and experimental results show population estimates are more accurate and refined compared to several other methods. For the general county and urban areas, the correlation and relative error evaluation indicators are excellent. Our estimation method not only compensates for the high cost of census but also estimates population quickly and accurately. The resolution of the daytime satellite images used in this research is 15 m, and the accuracy of the results may be improved if images with higher resolution are used. In addition, we only utilized daytime satellite images and night light images. If Synthetic Aperture Radar (SAR) images or more remote sensing data are utilized, the results may be further improved. LGANet makes a great contribution to population estimation and may help governments to better serve citizens through effective and equitable material distribution. For example, fast and accurate population information is not only critical to implement effective rescue measures in natural disasters, but also essential for sustainable development in China.

CRediT authorship contribution statement

Yanxiao Jiang: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Zhou Huang:** Methodology, Writing – review & editing, Funding acquisition. **Linna Li:** Writing – review & editing. **Quanhua Dong:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by grants from the We National Natural Science Foundation of China (42271471, 41971331, 41830645), and the National Natural Science Foundation of China (41830645). We also appreciate the detailed comments from the Editor and the anonymous reviewers.

References

Albert, A., Kaur, J., Gonzalez, M., 2017. Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. ArXiv170402965 Cs.

Balk, D.L., Deichmann, U., Yetman, G., Pozzi, F., Hay, S.I., Nelson, A., 2006. Determining global population distribution: Methods, applications and data. In: Advances in Parasitology. Elsevier, pp. 119–156. http://dx.doi.org/10.1016/S0065-308X(05)62004-0.

Balk, D., Pozzi, F., Yetman, G., Deichmann, U., Nelson, A., 2005. The distribution of people and the dimension of place: methodologies to improve the global estimation of urban extents. In: International Society for Photogrammetry and Remote Sensing, Proceedings of the Urban Remote Sensing Conference. pp. 14–16.

Ball, J.E., Anderson, D.T., Chan, C.S., 2017. A comprehensive survey of deep learning in remote sensing: Theories, tools and challenges for the community. J. Appl. Remote Sens. 11, 1. http://dx.doi.org/10.1117/1.JRS.11.042609.

Bao, Y., Huang, Z., Wang, H., Yin, G., Zhou, X., Gao, Y., 2023. High-resolution quantification of building stock using multi-source remote sensing imagery and deep learning. J. Ind. Ecol. 27, 350–361. http://dx.doi.org/10.1111/jiec.13356.

Briggs, D.J., Gulliver, J., Fecht, D., Vienneau, D.M., 2007. Dasymetric modelling of small-area population distribution using land cover and light emissions data. Remote Sens. Environ. 108, 451–466. http://dx.doi.org/10.1016/j.rse.2006.11.020.

Cho, K., Courville, A., Bengio, Y., 2015. Describing multimedia content using attention-based encoder-decoder networks. IEEE Trans. Multimed. 17, 1875–1886. http://dx.doi.org/10.1109/TMM.2015.2477044.

Cohen, J.E., Small, C., 1998. Hypsographic demography: The distribution of human population by altitude. Proc. Natl. Acad. Sci. 95, 14009–14014. http://dx.doi.org/10.1073/pnas.95.24.14009.

Doupe, P., Bruzelius, E., Faghmous, J., Ruchman, S.G., 2016. Equitable development through deep learning: The case of sub-national population density estimation. In: Proceedings of the 7th Annual Symposium on Computing for Development. Presented at the ACM DEV '16: Annual Symposium on Computing for Development. ACM, Nairobi Kenya, pp. 1–10. http://dx.doi.org/10.1145/3001913.3001921.

Doxsey-Whitfield, E., MacManus, K., Adamo, S.B., Pistolesi, L., Squires, J., Borkovska, O., Baptista, S.R., 2015. Taking advantage of the improved availability of census data: A first look at the gridded population of the world, version 4. Pap. Appl. Geogr. 1, 226–234. http://dx.doi.org/10.1080/23754931.2015.1014272.

Elvidge, C.D., Baugh, K.E., Zhizhin, M., Hsu, F.-C., 2013. Why VIIRS data are superior to DMSP for mapping nighttime lights. Proc. Asia-Pac. Adv. Netw. 35, 62. http://dx.doi.org/10.7125/APAN.35.7.

Ferrie, J.P., 1996. A new sample of males linked from the public use microdata sample of the 1850 U.S. federal census of population to the 1860 U.S. federal census manuscript schedules. Hist. Methods J. Quant. Interdiscip. Hist. 29, 141–156. http://dx.doi.org/10.1080/01615440.1996.10112735.

Flowerdew, R., Green, M., 1994. Areal interpolation and types of data.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H., 2019. Dual attention network for scene segmentation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Long Beach, CA, USA, pp. 3141–3149. http://dx.doi.org/10.1109/CVPR.2019.00326.

Gaughan, A.E., Stevens, F.R., Linard, C., Patel, N.N., Tatem, A.J., 2015. Exploring nationally and regionally defined models for large area population mapping. Int. J. Digit. Earth 8, 989–1006. http://dx.doi.org/10.1080/17538947.2014.965761.

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E.L., Fei-Fei, L., 2017. Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the United States. Proc. Natl. Acad. Sci. 114, 13108–13113. http://dx.doi.org/10.1073/pnas.1700035114.

Goodchild, M.F., Anselin, L., Deichmann, U., 1993. A framework for the areal interpolation of socioeconomic data. Environ. Plan. A 25, 383–397.

Guo, Y., Ji, J., Lu, X., Huo, H., Fang, T., Li, D., 2019. Global-local attention network for aerial scene classification. IEEE Access 7, 67200–67212. http://dx.doi.org/10.1109/ACCESS.2019.2918732.

Hay, S.I., Noor, A.M., Nelson, A., Tatem, A.J., 2005. The accuracy of human population maps for public health application. Trop. Med. Int. Health 10, 1073–1086. http://dx.doi.org/10.1111/j.1365-3156.2005.01487.x.

Hu, W., Patel, J.H., Robert, Z.-A., Novosad, P., Asher, S., Tang, Z., Burke, M., Lobell, D., Ermon, S., 2019. Mapping missing population in rural India: A deep learning approach with satellite imagery. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. Presented at the AIES '19: AAAI/ACM Conference on AI, Ethics, and Society. ACM, Honolulu HI USA, pp. 353–359. http://dx.doi.org/10.1145/3306618.3314263.

Hu, Y., Peng, J., Liu, Y., Du, Y., Li, H., Wu, J., 2017. Mapping development pattern in Beijing-tianjin-hebei urban agglomeration using DMSP/OLS nighttime light data. Remote Sens. 9, 760. http://dx.doi.org/10.3390/rs9070760.

Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. Science 353, 790–794. http://dx.doi.org/10.1126/science.aaf7894.

Kim, J., Ma, M., Kim, K., Kim, S., Yoo, C.D., 2019. Progressive attention memory network for movie story question answering. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Long Beach, CA, USA, pp. 8329–8338. http://dx.doi.org/10.1109/CVPR.2019.00853.

Kingma, D.P., Ba, J., 2017. Adam: A method for stochastic optimization.

Kummu, M., de Moel, H., Ward, P.J., Varis, O., 2011. How close do we live to water? A global analysis of population distance to freshwater bodies. PLoS ONE 6, e20578. http://dx.doi.org/10.1371/journal.pone.0020578.

La, Y., Bagan, H., Takeuchi, W., 2019. Explore urban population distribution using nighttime lights, land-use/land-cover, and population census data. In: Presented at the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. pp. 1554–1557.

Li, Y., Chen, X., Zhu, Z., Xie, L., Huang, G., Du, D., Wang, X., 2019a. Attention-guided unified network for panoptic segmentation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Long Beach, CA, USA, pp. 7019–7028. http://dx.doi.org/10.1109/CVPR.2019.00719.

Li, L., Xu, M., Wang, X., Jiang, L., Liu, H., 2019b. Attention based glaucoma detection: A large-scale database and CNN model. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Long Beach, CA, USA, pp. 10563–10572. http://dx.doi.org/10.1109/CVPR.2019.01082.

Liaw, A., Wiener, M., 2002. Classification and regression by random forest. R News 2, 18–22.

Linard, C., Gilbert, M., Tatem, A.J., 2011. Assessing the use of global land cover data for guiding large area population distribution modelling. GeoJournal 76, 525–538. http://dx.doi.org/10.1007/s10708-010-9364-8.

Liu, Y., Hu, C., Sun, C., Zhan, W., Sun, S., Xu, B., Dong, Y., 2018a. Assessment of offshore oil/gas platform status in the northern Gulf of Mexico using multi-source satellite time-series images. Remote Sens. Environ. 208, 63–81. http://dx.doi.org/10.1016/j.rse.2018.02.003.

Liu, Y., Hu, C., Zhan, W., Sun, C., Murch, B., Ma, L., 2018b. Identifying industrial heat sources using time-series of the VIIRS Nightfire product with an object-oriented approach. Remote Sens. Environ. 204, 347–365. http://dx.doi.org/10.1016/j.rse.2017.10.019.

Long, J.F., McMillen, D.B., 2015. A survey of Census Bureau population projection methods. Clim. Change 11, 141–177.

Mennis, J., 2003. Generating surface models of population using dasymetric mapping. Prof. Geogr. 55 (1), 31–42.

Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J., 2017. Large kernel matters – improve semantic segmentation by global convolutional network.

Robinson, C., Hohman, F., Dilkina, B., 2017a. A deep learning approach for population estimation from satellite imagery. In: Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities. Presented at the SIGSPATIAL'17: 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, Redondo Beach CA USA, pp. 47–54. http://dx.doi.org/10.1145/3149858.3149863.

Robinson, C., Hohman, F., Dilkina, B., 2017b. A deep learning approach for population estimation from satellite imagery. In: Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities. Presented at the SIGSPATIAL'17: 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, Redondo Beach CA USA, pp. 47–54. http://dx.doi.org/10.1145/3149858.3149863.

Schneider, A., Friedl, M.A., Potere, D., 2009. A new map of global urban extent from MODIS satellite data. Environ. Res. Lett. 4, 044003. http://dx.doi.org/10.1088/1748-9326/4/4/044003.

Smith, S.K., 1987. Tests of forecast accuracy and bias for county population projections. J. Am. Stat. Assoc. 82, 991–1003. http://dx.doi.org/10.1080/01621459.1987.10478528.

Sorichetta, A., Hornby, G.M., Stevens, F.R., Gaughan, A.E., Linard, C., Tatem, A.J., 2015. High-resolution gridded population datasets for latin america and the caribbean in 2010, 2015, and 2020. Sci. Data 2, 150045. http://dx.doi.org/10.1038/sdata.2015.45.

Stevens, F.R., Gaughan, A.E., Linard, C., Tatem, A.J., 2015. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. PLOS ONE 10, e0107042. http://dx.doi.org/10.1371/journal.pone.0107042.

Suganuma, M., Liu, X., Okatani, T., 2019. Attention-based adaptive selection of operations for image restoration in the presence of unknown combined distortions. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Long Beach, CA, USA, pp. 9031–9040. http://dx.doi.org/10.1109/CVPR.2019.00925.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A., 2016. Inception-v4, inception-ResNet and the impact of residual connections on learning.

Tatem, A.J., Adamo, S., Bharti, N., Burgert, C.R., Castro, M., Dorelien, A., Fink, G., Linard, C., John, M., Montana, L., Montgomery, M.R., Nelson, A., Noor, A.M., Pindolia, D., Yetman, G., Balk, D., 2012. Mapping populations at risk: improving spatial demographic data for infectious disease modeling and metric derivation. Popul. Health Metr. 10, 8. http://dx.doi.org/10.1186/1478-7954-10-8.

Tatem, A.J., Campiz, N., Gething, P.W., Snow, R.W., Linard, C., 2011. The effects of spatial population dataset choice on estimates of population at risk of disease. Popul. Health Metr. 9, 4. http://dx.doi.org/10.1186/1478-7954-9-4.

Thomson, C.N., Hardin, P., 2000. Remote sensing/GIS integration to identify potential low-income housing sites. Cities 17, 97–109. http://dx.doi.org/10.1016/S0264-2751(00)00005-6.

Tobler, W., Deichmann, U., Gottsegen, J., Maloy, K., 1997. World population in a grid of spherical quadrilaterals. Int. J. Popul. Geogr. 3, 203–225. http://dx.doi.org/10.1002/(SICI)1099-1220(199709)3:3<203::AID-IJPG68>3.0.CO;2-C.

Uzun, B., Cete, M., 2004. A model for solving informal settlement issues in developing countries. p. 8.

Wang, Y., Huang, C., Zhao, M., et al., 2020. Mapping the population density in mainland China using NPP/VIIRS and points-of-interest data based on a random forests model. Remote Sens. 12, 3645.

Xing, X., Huang, Z., Cheng, X., Zhu, D., Kang, C., Zhang, F., Liu, Y., 2020. Mapping human activity volumes through remote sensing imagery. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 5652–5668. http://dx.doi.org/10.1109/JSTARS.2020.3023730.

Ye, L., Rochan, M., Liu, Z., Wang, Y., 2019a. Cross-modal self-attention network for referring image segmentation.

Ye, T., Zhao, N., Yang, X., Ouyang, Z., Liu, X., Chen, Q., Hu, K., Yue, W., Qi, J., Li, Z., Jia, P., 2019b. Improved population mapping for China using remotely sensed and points-of-interest data within a random forests model. Sci. Total Environ. 658, 936–946. http://dx.doi.org/10.1016/j.scitotenv.2018.12.276.

Yin, G., Huang, Z., Yang, L., Ben-Elia, E., Xu, L., Scheuer, B., Liu, Y., 2023. How to quantify the travel ratio of urban public transport at a high spatial resolution? A novel computational framework with geospatial big data. Int. J. Appl. Earth Obs. Geoinform. 118, 103245. http://dx.doi.org/10.1016/j.jag.2023.103245.

Zhang, L., Peng, J., Liu, Y., Wu, J., 2017. Coupling ecosystem services supply and human ecological demand to identify landscape ecological security pattern: A case study in Beijing–Tianjin–Hebei region, China. Urban Ecosyst. 20, 701–714. http://dx.doi.org/10.1007/s11252-016-0629-y.

Zhang, X., Yu, W., Pun, M.-O., 2022. Multilevel deformable attention-aggregated networks for change detection in bitemporal remote sensing imagery. IEEE Trans. Geosci. Remote Sens. 60, 1–18. http://dx.doi.org/10.1109/TGRS.2022.3157721.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Honolulu, HI, pp. 6230–6239. http://dx.doi.org/10.1109/CVPR.2017.660.