Mapping Gridded Gross Domestic Product Distribution of China Using Deep Learning With Multiple Geospatial Big Data

Yuehong Chen, Guohao Wu, Yong Ge, *Senior Member, IEEE*, and Zekun Xu

***Abstract*—Timely gridded gross domestic product (GDP) data is a fundamental indicator in many applications. It is critical to characterize the complex relationship between GDP and its auxiliary information for accurately estimating gridded GDP. However, few knowledge is available about the performance of deep learning approaches for learning this complex relationship. This article develops a novel convolutional neural network based GDP downscaling approach (GDPnet) to transform the statistical GDP data into GDP grids by integrating various geospatial big data. An existing autoencoder-based downscaling approach (Resautonet) is employed to compare with GDPnet. The latest county-level GDP data of China and the multiple geospatial big data are adopted to generate the 1-km gridded GDP data in 2019. Due to the different related auxiliary data of each GDP sector, the two downscaling approaches are first separately built for each GDP sector and then the results are merged to the gridded total GDP data. Experimental results show that the two deep learning approaches had good predictive power with R^2 over 0.8, 0.9, and 0.92 for the three sectors tested by county-level GDP data. Meanwhile, the proposed GDPnet outperformed the existing Resautonet. The average R^2 of GDPnet was 0.034 higher than that of Resautonet in terms of county-level GDP test data. Furthermore, GDPnet had higher accuracy ($R^2 = 0.739$) than Resautonet ($R^2 = 0.704$) assessed by town-level GDP data. In addition, the proposed GDPnet is faster (about 78% running time) than the Resautonet. Hence, the proposed approach provides a valuable option for generating gridded GDP data.**

***Index Terms*—1-km gridded gross domestic product (GDP), China, deep learning, downscaling, geospatial big data, gross domestic product (GDP).**

I. Introduction

GROSS domestic product (GDP) is a monetary metric that measures the market value of all final goods and services within a given region during a period of time, often seasonally and annually [1]–[3]. GDP is the worldwide standard statistical indicator to measure the economic development condition in a country or region and it is of great significance to formulate policies in various aspects for regional development [1], [3]. GDP is an essential indicator of economic activity in assessing disaster risk, simulating climate change, reducing poverty, and identifying economic development inequality [4]–[8], most of which are related to the sustainable development goals published by the United Nations.

The traditional manner of collecting GDP data is through statistical tables in administrative units (e.g., province, city, and county) and they are spatially linked to the boundary of these administrative units [9]. However, this form of statistical GDP data only provides a single value of GDP for each irregular administrative unit; hence, it cannot represent the spatial variation of GDP within a unit and it is often difficult to integrate with gridded environmental variables derived from remote sensing images [10]–[12]. Thus, transforming the traditional statistical GDP data into gridded GDP data (also termed as GDP spatialization or GDP downscaling) is an important way to facilitate the integration of GDP data with other environmental variables [5], [13]–[17].

GDP downscaling is an ill-posed problem and it usually requires auxiliary data to reduce its uncertainty to enhance the accuracy of estimating gridded GDP data from the traditional statistical GDP data [14], [16]. Elvidge *et al.* [14] first found the strong correlation between GDP and night-time lights (NTL). Since then, NTL dataset has been one of the most widely used auxiliary variables in GDP downscaling. The Defense Meteorological Satellite Program's Operational Linescan System (DMSP/OLS) dataset is the first type of NTL data for GDP downscaling. For instance, Doll *et al.* [13] employed the DMSP/OLS data to produce the global gridded purchasing power parity GDP map at the spatial resolution of 1° and Sutton and Costanza [5] used it to generate the global gridded GDP map at a finer spatial resolution of 1-km. Recently, the visible infrared imaging radiometer suite (VIIRS) NTL data has proved its better performance than DMSP/OLS data in GDP downscaling because of its higher spatial resolution (i.e., 500 m) [18], [19]. Although NTL data has been widely applied to GDP downscaling, its shortcomings are obvious. For example, NTL data cannot be a proxy for the primary sector of

Manuscript received October 9, 2021; revised November 26, 2021 and December 30, 2021; accepted January 30, 2022. Date of publication February 7, 2022; date of current version February 24, 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFC1510601, in part by the Strategic Priority Research Program of Chinese Academy of Sciences under Grant XDA20030302, and in part by the National Natural Science Foundation of China under Grant 42071315. *(Corresponding author: Yong Ge.)*

Yuehong Chen, Guohao Wu, and Zekun Xu are with the College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China (e-mail: chenyh@lreis.ac.cn; wgh.hhu.edu.cn@hhu.edu.cn; xuzk@hhu.edu.cn).

Yong Ge is with the State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China (e-mail: gey@igsnrr.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3148448

GDP (e.g., agricultural activities) and only NTL data is hard to differentiate the secondary sector (e.g., industry activities) and the tertiary sector (service activities) of GDP [15], [16].

To handle the limitation of the only usage of NTL data, a large number of geospatial big data (e.g., land use/cover, road network, points of interest (POI), footprint of buildings, and human distribution data) have been adopted to combine with NTL data for GDP downscaling [15], [16], [18]–[22]. To deal with the absence of agricultural distribution by NTL data [20], [21], the land use/cover was integrated with NTL data to delineate the agricultural and nonagricultural distributions for estimating global 1-km GDP grids in 2015 by Wang *et al.* [19] and national gridded GDP data [15], [23]. The road network density was adopted by Murakami and Yamagata [22] to represent the industry activities and it was combined with land cover as auxiliary variables to downscale the national-level GDP to generate global GDP grids at the spatial resolution of 0.5° in 1980, 1990, 2000, and 2010. Recently, Chen *et al.* [16] applied different auxiliary data to model the relationship with each GDP sector and it has demonstrated that the combined usage of various auxiliary variables can effectively represent the distribution of different GDP sectors in generating the 1-km gridded GDP image for China in 2010.

Over the past few decades, a variety of methods have been applied to downscale the statistical GDP data. In the beginning, the simple linear regression was used to investigate the relationship between GDP and NTL data [5], [13]. Subsequently, several nonlinear algorithms were developed to characterize the nonlinear relationship between GDP and auxiliary variables [15], [23], [24]. Zhao *et al.* [24] found the nonlinear method (i.e., quadratic polynomial regression) outperformed a linear method and it obtained a high determinate coefficient ($R^2 = 0.92$) in South China. Recently, ensemble approaches (e.g., random forest regression) have been applied to GDP downscaling and they have achieved better performance than simple linear or nonlinear regression algorithms [16], [18], [19], [22]. Liang *et al.* [18] demonstrated random forest regression performed better than multivariate linear regression and it achieved the determinate coefficient ($R^2 = 0.77$) in Ningbo city of China. Chen *et al.* [16] used random forest regression to estimate the gridded GDP map in 2010 for China and the determinate coefficient ranges from 0.7 to 0.95. However, the performance of the traditional machine learning algorithms is limited to handle geospatial big data [25].

With the development of deep learning techniques, several studies have reported that deep learning approaches outperformed some traditional machine learning algorithms (e.g., XGBoost [26]) in a series of geoscience applications [27]–[29]. Because deep learning models take multiple layers to learn representations of data with multiple levels of abstraction to better learn the complex relationships, including linear and nonlinear relationships, than the traditional machine learning algorithms [25]. For downscaling geoscience variables, the autoencoder (AE) deep neural networks and convolutional neural networks (CNN) are the two typical and widely used deep learning techniques [30]. Li *et al.* [27] developed an AE algorithm, AE-based residual deep network (Resautonet), to downscale the aerosol optical depth observations from moderate resolution imaging spectroradiometer (MODIS). Although the Resautonet has proved its strength to regress and downscale the aerosol optical depth, its feasibility for downscaling other geoscience variables (e.g., GDP) is unknown. Meanwhile, CNN has been successfully applied to downscaling spatially 2-D data (e.g., remote sensing images) and it is seldom applied to downscaling spatially 1-D data, such as statistical GDP data [30], [31]. Compared with the traditional fully connected neurons between adjacent layers in AE networks, CNN is much easier to train and generalize much better than the fully connected networks as it uses the shared weights in a convolutional layer [25]. Until now, few knowledge is available for the performance of deep learning techniques in generating gridded GDP data. In addition, existing gridded GDP datasets for China are out of date and they cannot meet the timely requirement of several applications in China as it has a very fast economic growth speed [16], [20].

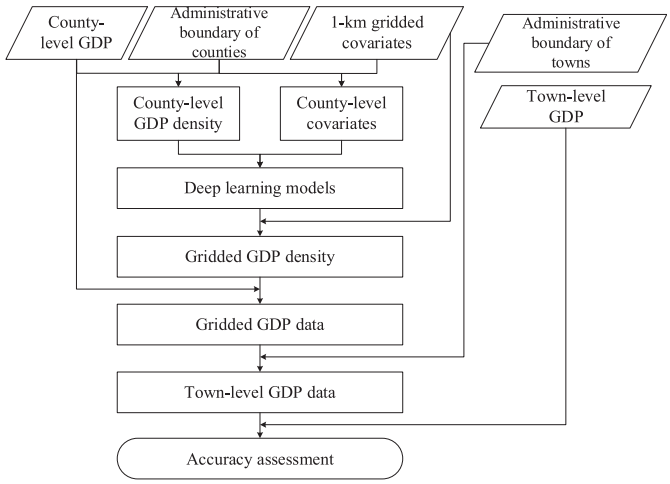


Fig. 1. Flowchart of GDP downscaling using deep learning.

To fill the above-mentioned research gaps, we first develop a novel GDP downscaling approach using 1-D CNN and residual connection (GDPnet). Then, both the proposed GDPnet and the existing Resautonet are employed to generate the gridded GDP data of China using the multiple geospatial big data and the latest county-level statistical GDP data from the 2020 yearbook. Finally, the accuracy of gridded GDP maps is evaluated for the two deep learning approaches.

The rest of this article is organized as follows. Section II describes the methodology. Section III and Section IV show the experimental results and discussion, respectively. Finally, Section V concludes this article.

II. Methodology

The flowchart of GDP downscaling is shown in Fig. 1, and there are five main steps as follows.

1) Collecting and preprocessing GDP data and covariates, including VIIRS NTL, POI, land cover, digital elevation model (DEM), OpenStreetMap, Tencent user positioning data, and centroid of girds. The county-level and town-level GDP data, including the annual primary sector of GDP (GDP_1), the annual secondary sector of GDP

(GDP_2), the annual tertiary sector of GDP (GDP_3), and the sum of the three sectors (i.e., total GDP), were spatially linked to the corresponding administrative boundaries and their coordinate system was transformed to the Albers equal-area conic projection. For GDP_1, it mainly includes the economic activities of farming, forestry, livestock, and fishery [15]. The manufacturing of goods and construction in the economy often belongs to GDP_2, while the service economic activities are related to GDP_3 [15]. The spatial distribution of the three sectors is different and their related covariates are different as well. Thus, it is needed to separately process each sector [15], [16]. The 1-km gridded covariates were preprocessed and reprojected to the Albers equal-area conic coordinate system as well. Additionally, town-level total GDP data was collected for accuracy assessment purpose.

2) Building deep learning models: We build a 1-D CNN-based GDP downscaling approach with residual connection (GDPnet) and fine tune the existing Resautonet for the GDP downscaling.
3) Training deep learning models: Because each GDP sector has different related covariates [15], [16], different combinations of covariates were prepared for each sector in training the two approaches. With the county-level covariates and the logarithm of the GDP density for each sector as inputs, the two approaches were separately implemented on each sector of GDP; hence, six models will be trained.
4) Predicting gridded GDP density using trained deep learning models: Taking 1-km gridded covariates as inputs, each trained GDP downscaling model was applied to predict the 1-km gridded density of each GDP sector. To ensure the coherency between the statistical GDP and the sum of GDP density grids within each county, the correction was implemented on the predicted gridded GDP densities to yield the gridded GDP data of each sector and the corrected gridded GDP data of three sectors were summed to the total gridded GDP data for each approach.
5) Assessing the accuracy of gridded GDP data: The town-level GDP data was adopted to assess the accuracy of gridded GDP data by comparing the statistical town-level GDP data from the yearbook and the corresponding town-level GDP data aggregated from the estimated gridded GDP data.

A. Preparing GDP and Auxiliary Data

The county-level GDP density for each sector was calculated by dividing the value of GDP in a county by the area of the corresponding county and the logarithm of each GDP sector density was adopted as the dependent variable in training the GDP downscaling models. Then, county-level covariates were aggregated by the Zonal Statistics tool in ArcGIS 10.6 from the 1-km gridded covariates to spatially align with the county-level GDP data and they were used as independent variables (i.e., auxiliary data) in training.

B. Deep Learning Models

The Resautonet, proposed by Li *et al.* [27], took advantage of AE deep neural network and residual connection to model the regression relationship between station-level $PM_{2.5}$ and other related covariates (e.g., aerosol optical depth). Resautonet is a mirrored structure with encoding and decoding layers. Resautonet contains input layer, three encoding layers, one central layer, three decoding layers, and output layer. The neural nodes of the input layer depend on the specific data and the output layer could be one or more nodes. The neural nodes of three encoding layers are 128, 64, and 32, respectively. The central layer has 16 nodes. The neural nodes of three decoding layers are 32, 64, and 128, respectively. The residual connection was implemented on two symmetrical encoding and decoding layers with the same nodes.

To take advantage of CNN, we developed the GDPnet using CNN and residual connection to model the regression relationship between GDP and its covariates. GDPnet employed the 1-D CNN to deal with the spatially 1-D GDP data. The architecture of GDPnet consists of input layer, encoding CNN layers, central CNN layer, decoding CNN layers, fully connected layer, and output layer, as shown in Fig. 2. Inspired by Resautonet, both encoding and decoding parts used three CNN layers. The three encoding CNN layers, respectively, used 128, 64, and 32 filters, whereas the three decoding CNN layers, respectively, employed 32, 64, and 128 filters. The number of filters in the central CNN layer is set to 16. To avoid the vanishing/exploding gradients problem [31], [32], the residual connects were also taken to any two symmetrical CNN layers with the same filters in the encoding and decoding parts of GDPnet. After the decoding part, a fully connected layer was adopted. The output layer was used to yield the prediction of GDP. The batch normalization and the rectified linear activation (ReLU) operations were performed after each hidden layer. The linear activation function was employed for the output layer. In addition, the kernel size of filters in CNN layers is set to an empirical value of seven.

C. Training

According to the designed GDPnet and existing Resautonet, the county-level covariates and the logarithm of GDP density were adopted to train the two approaches and we chose the loss function of mean square error for training as the following:

$$L\left(\theta_{w,b}\right) = \operatorname{argmin}\left\{\frac{1}{N}\sum_{i=1}^{N}\left|Y_i - f_{\theta_{w,b}}(X_i)\right|^2\right\} \quad (1)$$

where $\theta_{w,b}$ is the model parameters of weight w and bias b to be optimized for model f, Y_i is the logarithm of county-level GDP density for county i, $f_{\theta_{w,b}}(X_i)$ is the estimated logarithm of GDP density using the covariates X_i for county i, and N is the batch size.

In training, the adaptive moment estimation (Adam) was used to optimize the model. The batch size and the number of epochs were set to 100 and 500, respectively. For each sector, the GDP data was randomly split into two parts: 20% of counties for test and the rest for training and validation.

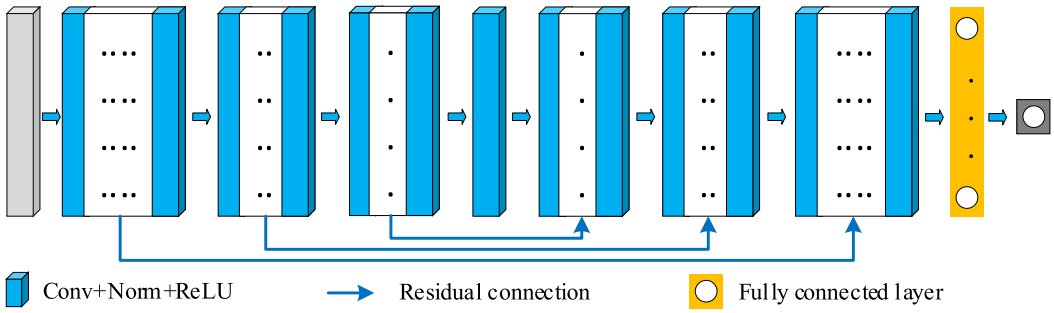


Fig. 2. Architecture of GDPnet.

D. Prediction

After the training of Resautonet and GDPnet for each sector of GDP, the 1-km gridded GDP density of each sector can be estimated by

$$g_k(j) = \mathrm{e}^{f_{\theta_{w,b}}(x_j)} \tag{2}$$

where $f_{\theta_{w,b}}(x_j)$ is the estimated logarithm of GDP density for grid j, x_j is the covariates for grid j, and $k \in \{1, 2, 3\}$ is the index of GDP sector.

To maintain the GDP coherency within each county, the gridded GDP density of each sector was corrected by multiplying the ratio between the actual GDP and the sum of predicted GDP grids within a county as

$$\mathrm{GDP}_k(j) = g_k(j) \frac{\mathrm{GDP}_k(i)}{\sum_{j \in i} g_k(j)} \tag{3}$$

where $\mathrm{GDP}_k(i)$ is the actual GDP within county i for sector k.

The total GDP for grid j can be calculated by

$$\mathrm{GDP}(j) = \sum_{k=1}^{3} \mathrm{GDP}_k(j) \ . \tag{4}$$

E. Accuracy Assessment Metrics

To evaluate the performance of the proposed GDPnet and the existing Resautonet in predicting gridded GDP data, the actual town-level GDP data was adopted to compare with the corresponding town-level GDP data aggregated from the estimated 1-km gridded GDP data. Three metrics of the mean absolute error (MAE), the root-mean-square error (RMSE), and the coefficient of determination (R^2) were measured for each 1-km gridded GDP data of each approach. Their formulas are expressed as

$$\mathrm{MAE} = \frac{1}{M} \sum_{i=1}^{M} \left| \mathrm{GDP}(i) - \widetilde{\mathrm{GDP}}(i) \right| \tag{5}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{M} \sum_{i=1}^{M} \left(\mathrm{GDP}(i) - \widetilde{\mathrm{GDP}}(i) \right)^2} \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{M} \left(\mathrm{GDP}(i) - \widetilde{\mathrm{GDP}}(i) \right)^2}{\sum_{i=1}^{M} \left(\mathrm{GDP}(i) - \overline{\mathrm{GDP}} \right)^2} \tag{7}$$

where $\mathrm{GDP}(i)$ is the actual GDP for town i, $\widetilde{\mathrm{GDP}}(i)$ is the estimated GDP for town i, $\overline{\mathrm{GDP}}$ is the mean of actual GDP for all towns, and M is the number of towns.

III. Experiments

A. Study Area and Data

The whole mainland China is selected as the study area. The economy of mainland China is the second in the world and it grows fast. Therefore, it is highly desired to generate the timely gridded GDP maps for its subsequent applications, such as risk assessment and poverty reduction.

Four types of datasets are used in this study. The first type is the county-level statistical GDP data from the 2020 yearbook, and it is considered as the dependent variable. The second type is the gridded auxiliary data and they are regarded as covariates (independent variables). The third type is the town-level statistical GDP data that is finer than the county-level data and it is employed for the accuracy assessment of the gridded GDP data. The last type is the county-level and town-level administrative boundary data, which are used to spatially link to the statistical GDP data.

1) County-Level Statistical GDP Data: The latest county-level statistical GDP data of China was collected to generate the gridded GDP data to meet the timely requirement of several applications. Four county-level indicators, GDP_1, GDP_2, GDP_3, and the annual total GDP, were first collected from the 2020 yearbook of China,[1] and the missing GDP data was then obtained from the 2020 yearbook of each province in mainland China. Finally, we obtained the GDP data for 2757 counties in mainland China, as shown in Fig. 3. Hong Kong, Macao, and Taiwan were excluded from this study because of the different usage of currency.

2) Gridded Auxiliary Data: Auxiliary datasets are crucial in accurately estimating gridded GDP data from the statistical GDP data. According to the previous studies [15], [16], we collected eight types of 1-km gridded auxiliary data that are related to GDP activities, as shown in Fig. 4 and Table I. Table I presents the correlation coefficients between each GDP sector and the used covariates in the sector.

[1][Online]. Available: https://data.cnki.net/Yearbook/Single/N2021050062

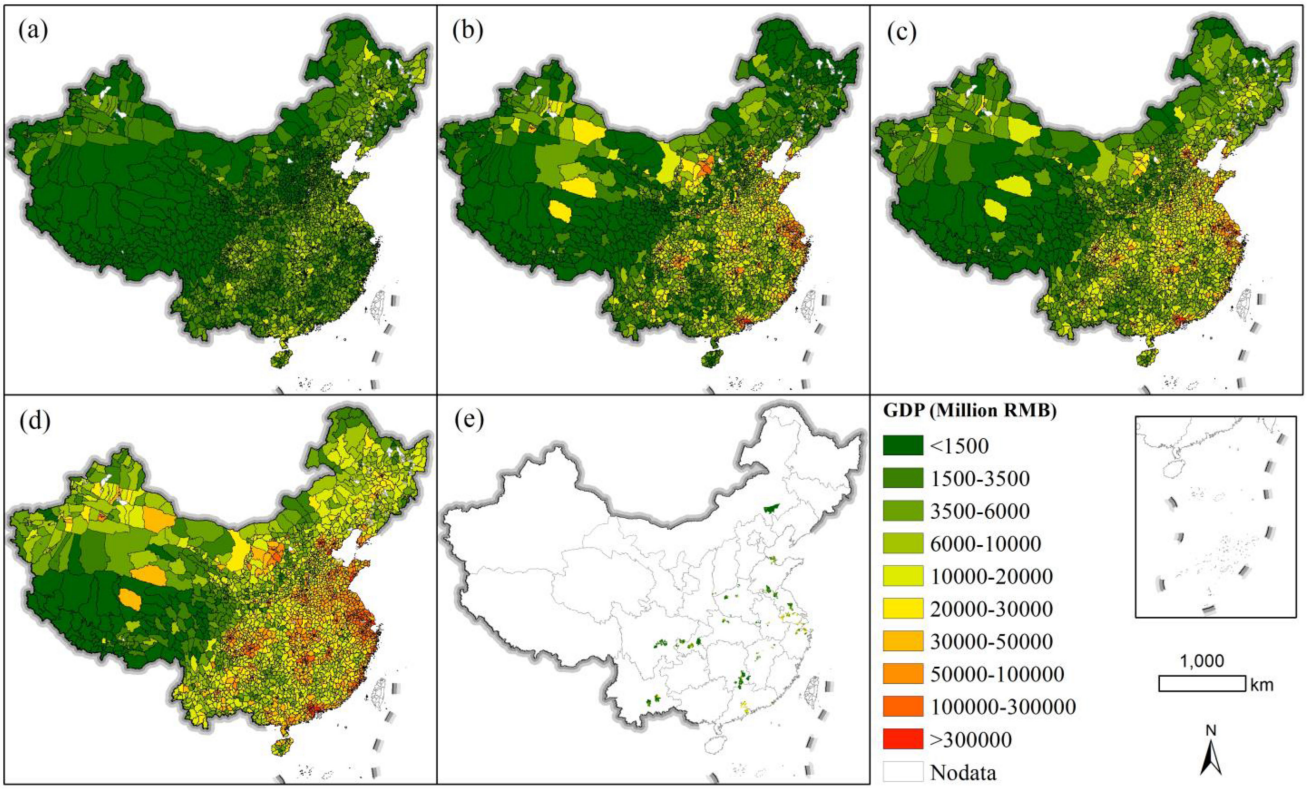


Fig. 3. County-level annual GDP data of mainland China. (a) Primary sector. (b) Secondary sector. (c) Tertiary sector. (d) Total GDP. (e) Total GDP of towns for accuracy assessment.

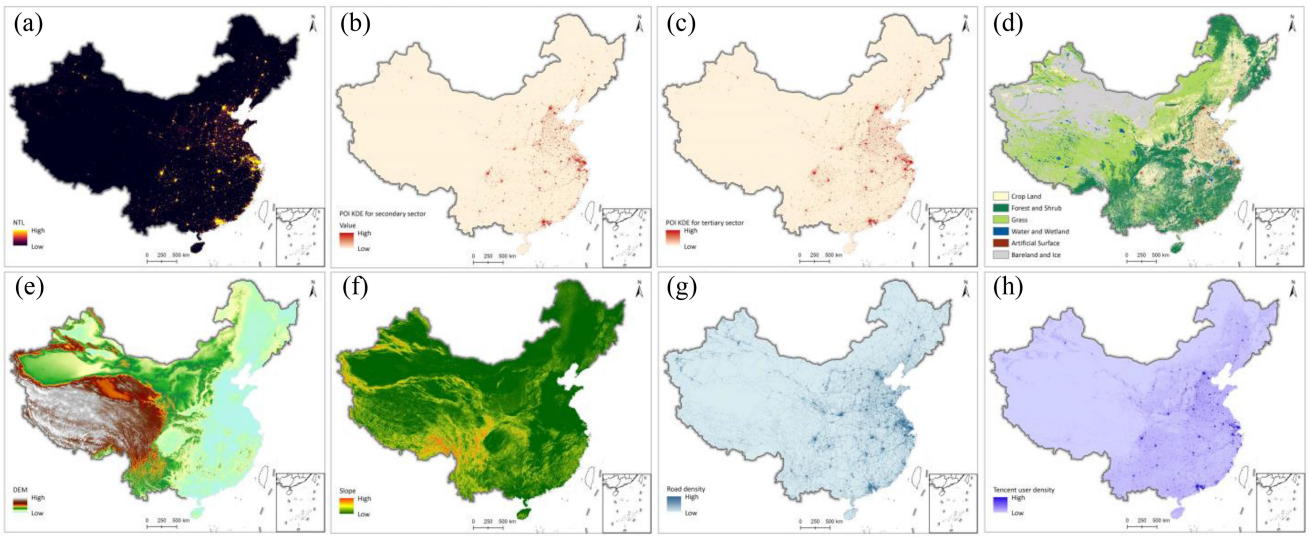


Fig. 4. 1-km gridded covariates of China. (a) NTL. (b) POI KDE for the secondary sector. (c) POI KDE for the tertiary sector. (d) Land cover. (e) DEM. (f) Slope. (g) Road density. (h) Tencent user density.

The annually composited VIIRS NTL image in 2019 was obtained from the website:[2] and it has implemented the saturation correction [33]. The original 500-m VIIRS NTL image was aggregated to 1-km VIIRS NTL image as one of the covariates. The correlation coefficients between GDP and aggregated NTL data for the secondary and tertiary sectors, as given in Table I, suggest that NTL data is highly correlated to the two sectors of GDP.

We obtained 50 347 282 POIs from one of the largest online maps.[3] They involve 5 070 471 POIs related to industry and

[2][Online]. Available: https://eogdata.mines.edu/nighttime_light/annual/v20/

[3][Online]. Available: https://amap.com/

TABLE I
COVARIATES FOR DOWNSCALING THE GDP OF CHINA

AUXILIARY DATA SOURCE	COVARIATES	CORRELATION TO GDP_1	CORRELATION TO GDP_2	CORRELATION TO GDP_3	TIME	USE IN SECTORS
VIIRS NTL data	VIIRS NTL image	-	0.59	0.67	2019	GDP_2 and GDP_3
POI data	KDE image related to GDP_2	-	0.79	-	2018	GDP_2
	KDE image related to GDP_3	-	-	0.82		GDP_3
Globeland30	Cropland proportion	0.49	-	-	2020	GDP_1
	Forest and Shrub proportion	0.07	-	-		GDP_1
	Grass proportion	-0.16	-	-		GDP_1
	Water and Wetland proportion	0.10	-	-		GDP_1
	Bareland and Ice proportion	-0.07	-	-		GDP_1
	Artificial surface proportion	-	0.52	0.57		GDP_2 and GDP_3
DEM ASTER GDEM	DEM	-0.23	-	-	-	GDP_1
	Slope	-0.18	-	-		GDP_1
OpenStreetMap	Road network density	-	0.64	0.81	2020	GDP_2 and GDP_3
Tencent user positioning data	Tencent user density	-	0.56	0.73	2019	
Centroid of girds	Longitude The square of longitude Latitude The square of latitude The product of longitude and latitude	-	-	-	-	GDP_1, GDP_2 and GDP_3

construction activities for the secondary sector of GDP and 45 276 811 POIs related to service activities for the tertiary sector of GDP. The count of POIs is highly correlated to the secondary and tertiary sectors in Table I. Referred in the study [16], we used the kernel density estimation (KDE) to generate the 1-km gridded POI density data for the secondary sector and the tertiary sector, respectively. The bandwidth of KDE was set from 500 to 5000 m with an interval of 100 m. According to the correlation coefficient between the GDP data and the estimated KDE data with each bandwidth, the optimal bandwidth of KDE was 1600 m for the secondary sector and 1200 m for the tertiary sector to generate the 1-km gridded POI density data.

The 30-m global land cover data in 2020, GlobeLand30,[4] was first resampled to 30.303 m for generating the 1-km land cover proportions using a window of 33×33 pixels. Then, the proportion of each land cover class within a 1-km grid was calculated. The six final land cover proportion images (i.e., Cropland, Forest and Shrub, Grass, Water and Wetland, Bareland and Ice, and Artificial surface) were achieved. The land cover of the Artificial surface is correlated to the secondary and tertiary sectors in Table I. For the land covers related to the primary sector, the Cropland has the highest correlation coefficient, followed by Water and Wetland and Forest, the Grass and Bareland and Ice are negatively correlated to the primary sector. Although a few land cover classes are not highly correlated to the primary sector, they are also employed to learn the nonlinear relationship in the two deep learning approaches in terms of previous studies [15], [16], [18], [22], [23]. To avoid abnormal values, the class of Water and Wetland was masked out for the secondary and tertiary sectors.

The 30-m DEM data, advanced spaceborne thermal emission and reflection radiometer global digital elevation model (ASTER GDEM), was collected from the website of the Earth Remote Sensing Data Analysis Center of Japan.[5] The 30-m DEM was first resampled to 30.303 m and then it was aggregated to 1-km DEM using a window of 33×33 pixels and the 1-km slope was further computed. Both DEM and slope are negatively correlated to the primary sector in Table I. Inspired by the articles presented in [16], [18], they are used for the primary sector.

The 1-km road network density image (i.e., road length within each grid) was calculated in ArcGIS 10.6 using the road network data from OpenStreetMap.[6] The road density is highly correlated to the secondary and tertiary sectors, as given in Table I.

The real-time Tencent user positioning density images at a spatial resolution of 0.01°[7] were collected from January 1 to June 12 in 2019 and all the density images were averaged to yield one final Tencent user density image. The averaged Tencent user density image was reprojected to 1-km final Tencent user density image with the Albers equal-area conic coordinate system. The 0.01° real-time Tencent user density was unavailable after June 12, 2019; thus, we only used the data from June 1 to June 12 in 2019. The Tencent density user positioning data, provided by one of the biggest companies in China, can effectively characterize a proxy of human activities for the secondary sector and tertiary sector of GDP [10]. It is highly correlated to the secondary and tertiary sectors in Table I.

Five location-related data, longitude, the square of longitude, latitude, the square of latitude, and the product of longitude and latitude, were calculated for each grid to account for the first law of geography (i.e., spatial autocorrelation) [34]. There are final 18 covariates derived from the 8 types of auxiliary data for downscaling the statistical GDP of mainland China in Table I.

For GDP_1, 12 covariates that are related to the primary sector of GDP were adopted and they involved five land cover proportions (i.e., Forest and Shrub, Grass, Water and Wetland, Bareland and Ice, and Cropland), DEM, slope, and five location-based variables (i.e., longitude, the square of longitude, latitude, the square of latitude, and the product of longitude and latitude). For GDP_2, ten covariates were used for the secondary sector of GDP

[4][Online]. Available: http://www.globallandcover.com/home.html
[5][Online]. Available: http://www.gdem.aster.ersdac.or.jp/search.jsp
[6][Online]. Available: http://download.geofabrik.de/index.html
[7][Online]. Available: https://heat.qq.com/

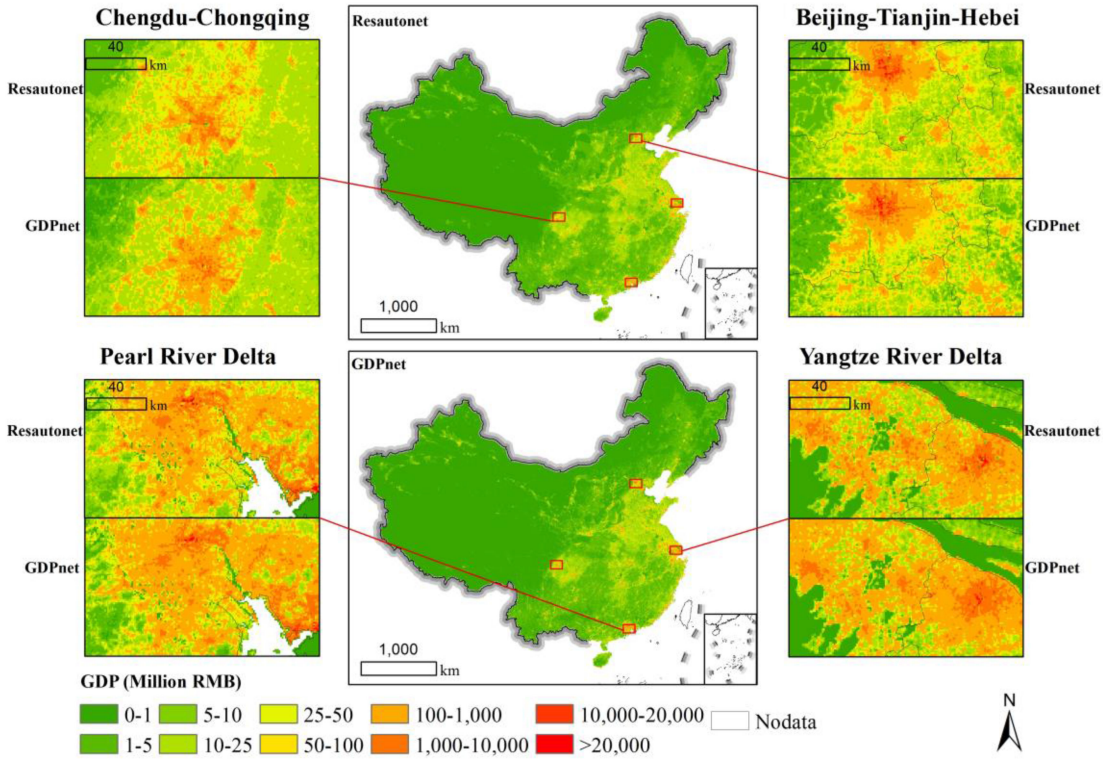


Fig. 5. 1-km gridded total GDP map of China.

and they contained one land cover proportion (i.e., Artificial surface), VIIRS NTL data, Tencent user density, POI KDE related to the secondary sector of GDP, road network density, and five location-based variables. For GDP_3, ten covariates were employed for the tertiary sector of GDP and they included one land cover proportion (i.e., Artificial surface), VIIRS NTL data, Tencent user density, POI KDE related to the tertiary sector of GDP, road network density, and five location-based variables.

3) Town-Level Statistical GDP Data: Due to the lack of ground truth GDP data within the 1-km grids, there are often two ways to evaluate the accuracy of gridded GDP data. The first way is to compare a part of county-level GDP data that is uninvolved in training GDP downscaling models with the corresponding county-level GDP data predicted by the trained model [16]. However, this way is better to test the effectiveness and generalization of GDP downscaling approaches. The second way is to compare the ground truth GDP data at finer scales (e.g., town-level GDP data) with the corresponding town-level GDP aggregated from the estimated gridded GDP data. The estimated town-level GDP can be calculated by the Zonal Statistics tool in ArcGIS. The second way is a popular way in the accuracy assessment of GDP downscaling as it provides the ground truth data at a finer scale than the scale of county-level GDP data. At present, town-level GDP data, the finest statistical GDP data, has the closet scale to 1-km gridded GDP data in China because town is the smallest administrative unit for GDP data. Town-level GDP data is often unavailable for the vast majority of towns because the Chinese government is not scheduled to release the town-level GDP data. Even though we have done our utmost to collect 791 town-level total GDP data from the 2020 yearbook, as shown in Fig. 3(e), to assess the accuracy of gridded total GDP data derived from the two deep learning approaches.

4) Administrative Boundary Data: The administrative boundary of counties in mainland China was collected from the National Catalogue Service for Geographic Information.[8] The administrative boundary of towns for accuracy assessment purpose was obtained from the Resource and Environment Science and Data Center.[9] They were adopted to spatially link to the statistical GDP data.

B. Gridded GDP Maps

The 1-km gridded total GDP maps in mainland China are displayed in Fig. 5. It shows that relatively high GDP grids mainly distribute in eastern China (e.g., Huanghuaihai Plain and eastern coastal region) and that the two deep learning approaches have basically similar spatial distribution of GDP grids in visual. Four representative zoom-in regions (i.e., Chengdu-Chongqing, Beijing-Tianjin-Hebei, Pearl River Delta, and Yangtze River Delta) are used to analyze the GDP spatial distribution in detail. The four regions are the largest metropolitan regions in China and it is worthy of analysis. It can be seen that city centers have higher GDP grids than suburbs and that the city centers of Beijing, Shanghai, Guangzhou, and Shenzhen present significantly higher GDP grids than other city centers. This

[8][Online]. Available: https://www.webmap.cn/main.do?method=index
[9][Online]. Available: https://www.resdc.cn/data.aspx?DATAID=203

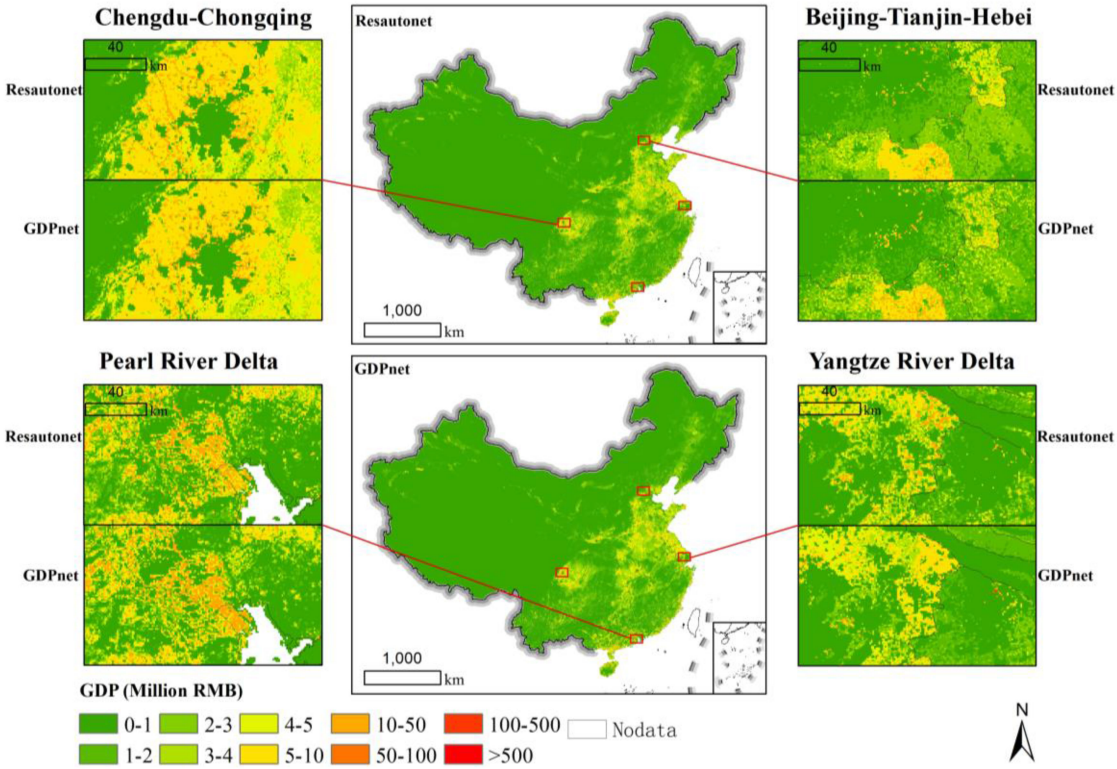


Fig. 6. 1-km gridded GDP map of China for the primary sector.

spatial distribution of GDP grids is intuitively consistent with the basic understanding of GDP in mainland China, indicating the reasonability of the two deep learning approaches.

The 1-km gridded GDP maps of the primary sector in mainland China are shown in Fig. 6. It presents that the two maps have a similar spatial distribution pattern that the agglomeration of relatively higher GDP grids mainly locates in eastern China and Sichuan Basin. But, GDP grids for the primary sector in eastern China have obviously lower values than those in the total GDP map in Fig. 5. Focusing on the four zoom-in regions, the grids in citer centers have lower values than those in urban suburbs and countryside, especially in Chengdu-Chongqing, Pearl River Delta, and Yangtze River Delta regions. The grids with low values in city centers and the grids with high values in suburbs for the primary sector of GDP are reasonable because the agricultural activities mainly distribute out of city centers.

The 1-km gridded GDP maps of the secondary sector in mainland China are shown in Fig. 7. It shows that eastern China has the clustered high GDP grids for the secondary sector and that the spatial pattern of the two maps generated by the two approaches is very close. The four zoom-in regions display that the GDP grids with high values mostly distribute in city centers and urban suburbs. When comparing the spatial distribution of GDP grids in the four zoom-in regions for the secondary sector, the GDP grids with high values in Yangtze River Delta region are more uniformly distributed than those in the other three zoom-in regions. There is an obvious difference between the two approaches' results is that GDPnet tends to allocate higher values of grids along with road network than Resautonet dose, especially in the Chengdu-Chongqing and Beijing-Tianjin-Hebei zoom-in regions.

TABLE II
PERFORMANCE OF THE TWO DEEP LEARNING APPROACHES

	MODEL	R^2	MAE	RMSE
GDP_1	Resautonet	0.800	0.006	0.013
	GDPnet	0.848	0.005	0.011
GDP_2	Resautonet	0.901	0.095	0.392
	GDPnet	0.909	0.089	0.375
GDP_3	Resautonet	0.918	0.285	1.472
	GDPnet	0.963	0.160	0.988

The 1-km gridded GDP maps of the tertiary sector in mainland China are displayed in Fig. 8. Again, the high GDP grids for the tertiary sector mostly distribute in eastern China and the two approaches generated comparable gridded GDP data in visual. City centers in the four zoom-in regions have significantly higher GDP grids than urban suburbs and countryside. Despite the unbalanced spatial distribution in the four regions, the GDP grids for the tertiary sector in Yangtze River Delta are more evenly distributed than the other three zoom-in regions.

C. Model Evaluation and Accuracy Assessment

To evaluate the performance of six trained models, the unused county-level GDP data in model training was used to calculate the three metrics of the two approaches for each GDP sector, as given in Table II. It shows that the proposed GDPnet performed better than the existing Resautonet for each GDP sector, as

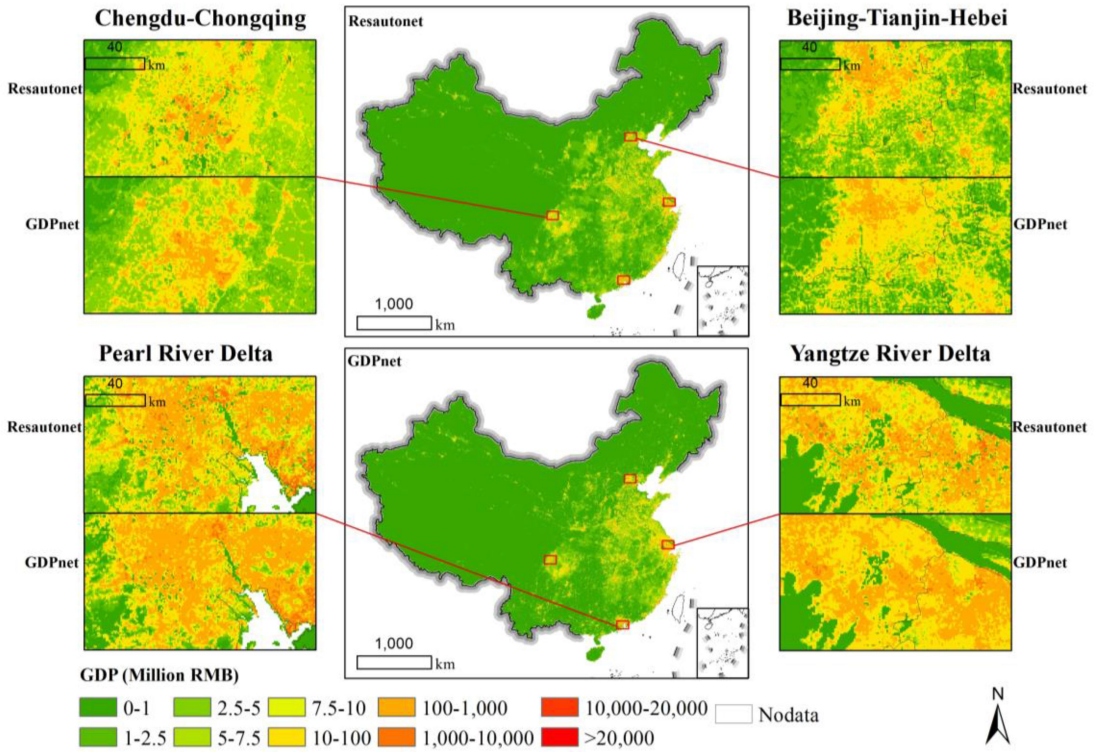


Fig. 7. 1-km gridded GDP map of China for the secondary sector.

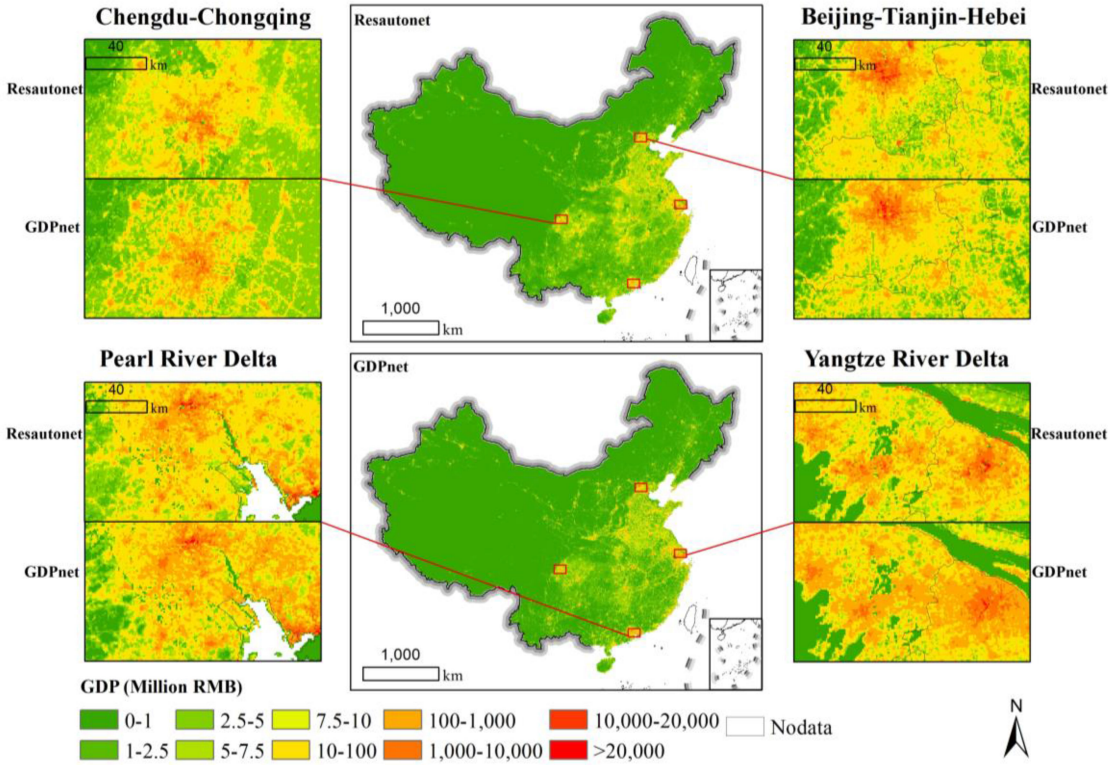
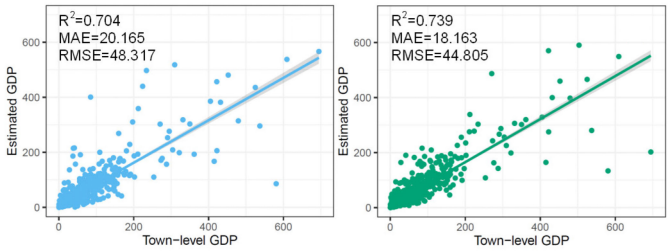
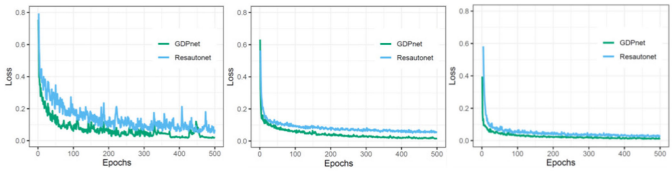


Fig. 8. 1-km gridded GDP map of China for the tertiary sector.



(a) (b)

Fig. 9. Accuracies of gridded GDP maps assessed by town-level total GDP data. (a) Resautonet. (b) GDPnet.



(a) (b) (c)

Fig. 10. Loss curves of GDPnet and Resautonet for each sector. (a) Primary sector. (b) Secondary sector. (c) Tertiary sector.

indicated by higher R^2 and lower MAE and RMSE for all the three sectors. Meanwhile, the three metrics present that the two approaches have the best performance in the tertiary sector, followed by the secondary sector, and the primary sector is the worst. Despite the difference performances in the three sectors by the two approaches, they performed well, as indicated by the R^2 over 0.8, 0.9, and 0.92 for the three sectors, respectively.

IV. Discussion

A. Intercomparison of Two Approaches

The 1-km gridded GDP maps were produced by the proposed GDPnet and the existing Resautonet. The two deep learning approaches were built on different architectures. GDPnet is based on the 1-D CNN, while Resautonet is based on the AE deep neural network. Their performances in generating gridded GDP maps are compared here. The quantitative metrics, as given in Table II, show that the proposed GDPnet outperformed Resautonet with higher R^2 and lower MAE and RMSE for the county-level GDP test data. Specifically, the average R^2 of GDPnet is 0.034 higher than that of Resautonet, while the average MAE and RMSE of GDPnet are 78% and 82% of those for Resautonet, respectively. At the same time, the accuracy assessment by town-level total GDP data, as shown in Fig. 9, also confirms the finding, and it shows that the R^2 of GDPnet is 0.035 higher that of Resautonet. The main reason may be that the CNN are easier to train and generalized much better than the fully connected networks [25]. Fig. 9 presents that both GDPnet and Resautonet performed better for towns with low total GDP than towns with high total GDP. There may be two main reasons. First, the aggregation error from gridded GDP to town-level GDP for towns with high GDP may be larger than that in towns with low GDP. Towns with high GDP primarily locate in the downtown area and they have a relatively small area, which easily results in larger errors due to the obviously inconsistent boundary between small towns and grids. Second, most of the counties locate out of the downtown area and they have relatively low GDP and the two approaches may learn the better complex relationship between GDP and covariates for the counties with low GDP than that of counties with high GDP. Moreover, Fig. 10 displays the loss curves of the two approaches for each sector in training process. Fig. 10 presents that the loss values of the two approaches decrease dramatically at the beginning and gradually converge after 300 epochs for the primary sector, 200 epochs for the secondary sector, and 100 epochs for the tertiary sector. It also shows that the proposed GDPnet has lower loss values than those of Resautonet for all the three sectors, which confirms the findings in Fig. 9. The loss values between GDPnet and Resautonet have the largest difference in the primary sector, followed by the secondary sector, and the tertiary sector is the least. Additionally, the loss curves of the two approaches for the primary sector fluctuate more than those of the other two sectors, suggesting that the stability of the two approaches in the primary sector is less than that in the other two sectors.

TABLE III
Maximum and Standard Deviation of the Gridded GDP Maps

	Maximum (Billion RMB)		Standard Deviation (Million RMB)	
	Resautonet	GDPnet	Resautonet	GDPnet
GDP_1	0.85	1.44	1.51	1.33
GDP_2	15.5	10.20	40.21	34.03
GDP_3	57.82	40.60	134.63	109.26
Total	57.92	40.71	144.15	123.51

B. Intercomparison of Three Sectors

Considering the different contribution of covariates for each GDP sector, different combination of covariates was separately prepared for training the two approaches for each sector. The performance of the two approaches was investigated for each sector. When comparing the four zoom-in regions, as shown in Figs. 6–8, the primary sector has less GDP grids with high value than the secondary and tertiary sectors, and the primary sector mainly distributes out of city centers (i.e., urban suburbs and countryside). The secondary and tertiary sectors mainly distribute in city centers and urban suburbs, and the GDP grids with high value in the two sectors are more than those in the primary sector. Meanwhile, the tertiary sector has relatively more GDP grids with high value in city centers than the secondary sector (see Figs. 7 and 8), which is basically consistent with reality as a large number of service-related activities locate in city centers. This phenomenon would be reasonable and consistent with the actual distribution of GDP in China. In addition, the four zoom-in regions, as shown in Figs. 7 and 8, display that GDPnet tends to allocate higher values of grids along with road network than Resautonet dose for the secondary and tertiary sectors, which suggest that the distribution by GDPnet is more reasonable than that by Resautonet for the two sectors as the economic activities in the secondary and tertiary sectors are more related to transportation. Table III presents the statistics

TABLE IV
TRAINING TIME(S) OF RESAUTONET AND GDPNET

	RESAUTONET	GDPNET
GDP_1	246	192
GDP_2	230	180
GDP_3	227	178
Total	703	550

(i.e., maximum and standard deviation) of the three sectors and the total GDP. It suggests that the existing Resautonet generated GDP grids with larger value range and variation than the proposed GDPnet did, as indicated by the larger standard deviation and maximum value by the Resautonet for the three sectors and the total GDP. The reason is likely that the GDPnet used the CNN layers to generate results with lower variation. The minimum of all gridded GDP maps is zero and the mean of two approaches is almost the same for each type of gridded GDP map due to the coherency manipulation in (3); thus, the minimum and mean of gridded GDP maps are excluded in Table III.

C. Comparison Against Existing GDP Downscaling Methods

Several previous GDP downscaling studies have proved that ensemble algorithms (e.g., random forest regression) performed better than simple regression methods in estimating gridded GDP data [16], [18], [24]. Chen *et al.* [16] used random forest regression to estimate the 1-km gridded GDP data for China in 2010 and they adopted tenfold cross validation to evaluate the accuracy using county-level GDP data. It is similar to the two deep learning approaches that random forest regression performed the best for the tertiary sector, followed by the secondary sector, and the primary sector has the worst accuracy. According to Table II tested by 20-fold county-level GDP data, the Resautonet has the comparable performance with the random forest regression in [16], but the proposed GDPnet has higher accuracy (e.g., R^2) than the random forest regression for each sector. Meanwhile, previous studies have proved that deep learning approaches have better performance than several ensemble algorithms (e.g., XGBoost [26]) through the multiple levels of data representation and abstraction [25]. Although the deep learning downscaling approaches have better performance, their computational efficiency is lower than the traditional ensemble algorithms.

D. Computational Efficiency

The structure depth of Resautonet and GDPnet is the same and they used the same residual connections. Resautonet used the fully connected layers, whereas GDPnet employed the 1-D convolutional layers. The neuron number of a fully connected layer in Resautonet is the same to the filter number of the corresponding convolutional layer in GDPnet. Thus, Resautonet and GDPnet have similar structure except the different types of layers. Their computational efficiency is worthy of comparison. The training time of the two deep learning approaches is given in Table IV. It indicates that the proposed GDPnet is more efficient than the existing Resautonet for each sector. The time of GDPnet is 78.05%, 78.26%, and 78.41% of Resautonet for the primary sector, secondary sector, and tertiary sector, respectively. The reason may be that the weight of CNN layers in GDPnet is faster and easier to train than that of the fully connected neurons between adjacent layers in Resautonet. Additionally, the secondary and tertiary sectors with the same number of covariates have almost same time in each approach, while the primary sector needs slightly more time than the other two sectors for the two approaches because the primary sector used two more covariates than the other two GDP sectors.

E. Limitations

Although multiple geospatial big data and deep learning approaches were made full use to produce the gridded GDP data in China, limitations of this study were inevitable. The different POIs were employed to separately characterize the secondary and tertiary sectors; however, POIs only give the quantity information rather than the area of each POI [16]. It limits the accuracy of gridded GDP for the secondary and tertiary sectors. Thus, potential geospatial big data that can represent the density and area of economic activities can be incorporated in the future. The uncertainty of transforming gridded data into county-level and town-level data is another limitation of this study. The training of deep learning approaches used the county-level covariates and the accuracy assessment process took the town-level GDP data, which were aggregated from the gridded data by the Zonal Statistics tool. The Zonal Statistics tool only calculates the value of an administrative unit using grids whose centroids locate within the administrative unit, which inevitably introduces errors by the inconsistent boundary between grids and administrative units. The finer units (e.g., towns) may introduce more errors.

V. CONCLUSION

This article presents a novel downscaling approach using deep learning, GDPnet, for estimating gridded GDP data from the traditional statistical GDP data by integrating a variety of geospatial big datasets. It aimed to take advantage of CNN and residual connection to characterize the complex relationship between GDP and auxiliary data. An existing AE-based downscaling approach, Resautonet, was compared. To consider the different related auxiliary data of each GDP sector, the proposed downscaling approach was separately built for each sector. The latest county-level GDP data of China from the 2020 yearbook was adopted to substantiate the proposed approach in generating 1-km gridded GDP data of China in 2019. Experimental results have demonstrated that the two downscaling approaches using deep learning had good predictive power with R^2 over 0.8, 0.9, and 0.92 for the three sectors tested by county-level GDP data and that the proposed GDPnet outperformed the existing Resautonet. The average R^2 of GDPnet was 0.034 higher than that of Resautonet in terms of county-level GDP test data, and GDPnet also had higher accuracy ($R^2 = 0.739$) than Resautonet ($R^2 = 0.704$) assessed by town-level GDP data. Meanwhile, GDPnet generated intuitively more reasonable distribution of GDP than Resautonet. In addition, the proposed GDP downscaling approach only needs about 78% running time of the AE-based downscaling method for the three sectors. Hence, the proposed CNN-based GDP downscaling approach is an effective option for estimating the gridded GDP distribution data.

REFERENCES

[1] T. Geiger, "Continuous national gross domestic product (GDP) time series for 195 countries: Past observations (1850–2005) harmonized with future projections according to the shared socio-economic pathways (2006–2100)," *Earth Syst. Sci. Data*, vol. 10, pp. 847–856, 2018.

[2] J. V. Henderson, A. Storeygard, and D. N. Weil, "Measuring economic growth from outer space," *Amer. Econ. Rev.*, vol. 102, pp. 994–1028, 2012.

[3] J. Sun, L. Di, Z. Sun, J. Wang, and Y. Wu, "Estimation of GDP using deep learning with NPP-VIIRS imagery and land cover data at the county level in CONUS," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1400–1415, Mar. 2020.

[4] Y. Song *et al.*, "Spatial and temporal variations of spatial population accessibility to public hospitals: A case study of rural–urban comparison," *GI Sci. Remote Sens.*, vol. 55, pp. 718–744, 2018.

[5] P. C. Sutton and R. Costanza, "Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation," *Ecological Econ.*, vol. 41, pp. 509–527, 2002.

[6] Y. Ge, Y. Yuan, S. Hu, Z. Ren, and Y. Wu, "Space–time variability analysis of poverty alleviation performance in China's poverty-stricken areas," *Spatial Statist.*, vol. 21, pp. 460–474, 2017.

[7] M. Oppenheimer *et al.*, "Emergent risks and key vulnerabilities," in *Climate Change 2014—Impacts, Adaptation and Vulnerability: Part A: Global and Sectoral Aspects*. Cambridge, U.K.: Cambridge Univ. Press, 2014, pp. 1046–1050.

[8] J. Yang *et al.*, "Projecting heat-related excess mortality under climate change scenarios in China," *Nature Commun.*, vol. 12, no. 1, Feb. 2021, Art. no. 1039.

[9] J. Chen and L. Li, "Regional economic activity derived from MODIS data: A comparison with DMSP/OLS and NPP/VIIRS nighttime light data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 3067–3077, Aug. 2019.

[10] Y. Chen, R. Zhang, Y. Ge, J. Yan, and Z. Xia, "Downscaling census data for gridded population mapping with geographically weighted area-to-point regression kriging," *IEEE Access*, vol. 7, pp. 149132–149141, 2019.

[11] S. Leyk *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*, vol. 11, pp. 1385–1409, 2019.

[12] X. Cao, J. Wang, J. Chen, and F. Shi, "Spatialization of electricity consumption of China using saturation-corrected DMSP-OLS data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 28, pp. 193–200, May 2014.

[13] C. N. H. Doll, J.-P. Muller, and C. D. Elvidge, "Night-time imagery as a tool for global mapping of socioeconomic parameters and greenhouse gas emissions," *AMBIO*, vol. 29, pp. 157–162, 2000.

[14] C. D. Elvidge, K. E. Baugh, E. A. Kihn, H. W. Kroehl, E. R. Davis, and C. W. Davis, "Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption," *Int. J. Remote Sens.*, vol. 18, pp. 1373–1379, Apr. 1997.

[15] Q. Chen, X. Hou, X. Zhang, and C. Ma, "Improved GDP spatialization approach by combining land-use data and night-time light data: A case study in China's continental coastal area," *Int. J. Remote Sens.*, vol. 37, pp. 4610–4622, 2016.

[16] Q. Chen *et al.*, "Mapping China's regional economic activity by integrating points-of-interest and remote sensing data with random forest," *Environ. Plan. B, Urban Analytics City Sci.*, vol. 48, no. 7, pp. 1876–1894, 2020.

[17] J. Yi, Y. Du, F. Liang, W. Tu, W. Qi, and Y. Ge, "Mapping human's digital footprints on the Tibetan plateau from multi-source geospatial big data," *Sci. Total Environ.*, vol. 711, Apr. 2020, Art. no. 134540.

[18] H. Liang, Z. Guo, J. Wu, and Z. Chen, "GDP spatialization in Ningbo city based on NPP/VIIRS night-time light and auxiliary data using random forest regression," *Adv. Space Res.*, vol. 65, pp. 481–493, 2020.

[19] X. Wang, P. C. Sutton, and B. Qi, "Global mapping of GDP at 1 km^2 using VIIRS nighttime satellite imagery," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 12, 2019, Art. no. 580.

[20] N. Zhao, Y. Liu, G. Cao, E. L. Samson, and J. Zhang, "Forecasting China's GDP at the pixel level using nighttime lights time series and population images," *GI Sci. Remote Sens.*, vol. 54, pp. 407–425, 2017.

[21] M. Kummu, M. Taka, and J. H. A. Guillaume, "Gridded global datasets for gross domestic product and human development index over 1990–2015," *Sci. Data*, vol. 5, no. 1, Feb. 2018, Art. no. 180004.

[22] D. Murakami and Y. Yamagata, "Estimation of gridded population and GDP scenarios with spatially explicit statistical downscaling," *Sustainability*, vol. 11, 2019, Art. no. 2106.

[23] E. Ustaoglu, R. Bovkır, and A. C. Aydınoglu, "Spatial distribution of GDP based on integrated NPS-VIIRS nighttime light and MODIS EVI data: A case study of Turkey," *Environ., Develop. Sustainability*, vol. 23, pp. 10309–10343, Jul. 2021.

[24] M. Zhao, W. Cheng, C. Zhou, M. Li, N. Wang, and Q. Liu, "GDP spatialization and economic differences in South China based on NPP-VIIRS nighttime light imagery," *Remote Sens.*, vol. 9, no. 7, 2017, Art. no. 673.

[25] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[26] L. Li, Y. Fang, J. Wu, J. Wang, and Y. Ge, "Encoder-decoder full residual deep networks for robust regression and spatiotemporal estimation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 9, pp. 4217–4230, Sep. 2021.

[27] L. Li *et al.*, "Spatiotemporal imputation of MAIAC AOD using deep learning with downscaling," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111584.

[28] M. Reichstein *et al.*, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, pp. 195–204, Feb. 2019.

[29] H. Shen, Y. Jiang, T. Li, Q. Cheng, C. Zeng, and L. Zhang, "Deep learning-based air temperature mapping by fusing remote sensing, station, simulation and socioeconomic data," *Remote Sens. Environ.*, vol. 240, Apr. 2020, Art. no. 111692.

[30] Q. Yuan *et al.*, "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, May 2020, Art. no. 111716.

[31] Y. Chen, K. Shi, Y. Ge, and Y. Zhou, "Spatiotemporal remote sensing image fusion using multiscale two-stream convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 112, Apr. 2021, doi: 10.1109/TGRS.2021.3069116.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[33] C. D. Elvidge, M. Zhizhin, T. Ghosh, F.-C. Hsu, and J. Taneja, "Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019," *Remote Sens.*, vol. 13, no. 5, 2021, Art. no. 922.

[34] H. J. Miller, "Tobler's First Law and spatial analysis," *Ann. Assoc. Amer. Geographers*, vol. 94, pp. 284–289, 2004.

Yuehong Chen received the B.S. degree in geographical information system from Hohai University, Nanjing, China, in 2010, and the M.Sc. and Ph.D. degrees in cartography and geographical information system from the State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, University of Chinese Academy of Sciences, Beijing, China, in 2013 and 2016, respectively.

He is an Associate Professor with the College of Hydrology and Water Resources, Hohai University. His current research interests include geospatial big data analysis and remote sensing image processing.

Guohao Wu received the B.S. degree in geographical information science in 2021 from Hohai University, Nanjing, China, where he is currently working toward the master's degree in cartography and geographical information system with the College of Hydrology and Water Resources.

His research interest focuses on geospatial big data analysis.

Yong Ge (Senior Member, IEEE) received the Ph.D. degree in cartography and geographical information system from the Chinese Academy of Sciences (CAS), Beijing, China, in 2001.

She is a Professor with the State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS. She has directed research in more than ten national projects. She is the author or coauthor of more than 80 scientific papers published in refereed journals, one book, and six chapters in books; she is the editor of one book, and she holds three granted patents in improving the accuracy of information extraction from remotely sensed imagery. Her research activity focuses on spatial data analysis and data quality assessment.

Dr. Ge has been involved in the organization of several international conferences and workshops. She is a member of the Theory and Methodology Committee of *Cartography and Geographic Information Society, International Association of Mathematical Geosciences*, and the Editorial Board of *Spatial Statistics* (Elsevier).

Zekun Xu is currently working toward the B.S. degree in geographical information science from Hohai University, Nanjing, China.

His research interest focuses on spatial analysis.