

# Combining Residual Neural Networks and Feature Pyramid Networks to Estimate Poverty Using Multisource Remote Sensing Data

Yumin Tan<sup>ID</sup>, Peng Wu<sup>ID</sup>, Guanhua Zhou, Yunxin Li, and Bingxin Bai

**Abstract**—Reliable poverty data are critical for regional economic analysis and policy making, especially considering that economic inequality and sustainable development are widespread social concerns. This article proposes a multitask learning model combining deep residual neural networks and feature pyramid networks to estimate poverty level from multiple sources including the night-time light data, Landsat 8 imagery, and spectral index data. We first train the multitask learning model using the multisource data in Chongqing, China and then estimate the representative economic indicators in the study area. The model is evaluated with the Pearson correlation coefficient of the actual and estimated economic indicators. The result shows that the proposed model outperforms other models with the Pearson correlation coefficient up to 0.87 in the annual estimates of economic indicators between 2013 and 2017. As all the data used in this article are publicly available, the proposed model can be used to estimate the economic indicators in other regions as well.

**Index Terms**—Economic indicators, multitask learning model, neural network, night-time light, poverty.

## I. INTRODUCTION

POVERTY has always been the focus of attention in all countries around the world [1]. World Bank report of 2018 states that 10% of the world's population still lives below the international poverty line of \$1.9 a day [2]. Estimation and monitoring of regional poverty level are very important for researchers and policymakers to understand the extent of poverty and to formulate strategies for poverty eradication. The traditional method to measure the poverty level is mainly through the statistical data, mostly economic indicators including income, expenditure, and household assets. [3], [4]. However, the collection of these data requires a lot of labor, material, and financial resources, which greatly limits the data collection capacity in many countries, especially in developing countries

[5]. One possible solution is remote sensing data, which are widely used in studies such as regional economic conditions and environmental health and population distribution due to their free availability, broad coverage, and timeliness.

Currently, environmental and telephonic data have been constantly adopted in regional poverty research. Environmental data include remote sensing data and publicly available measurement data. The two main sources of remote sensing data in poverty estimation are night-time light data and high-resolution satellite imagery. Night-time light data, the artificial light data recording human settlements, have been used in many applications including population mapping [6]–[8], GDP estimation [9]–[11], electricity consumption [12]–[14], Ecosystem services product [15] and carbon dioxide (CO<sub>2</sub>) emissions [16]–[18]. The National Oceanic and Atmospheric Administration (NOAA) provides annual night-time light data of the whole world. There are two different types of data: DMSP-OLS data with a resolution of 30 arc-s (about 1000 m) acquired by the Defense Meteorological Satellite Program's Operational Line Scan System (DMSP-OLS) [19]–[21], and NPP-VIRS data with a resolution of 15 arc-s (about 500 m) from the Visible Infrared Imaging Radiometer Suite (VIIRS) Day-Night Band carried by the Suomi National Polar-orbiting Partnership (NPP) Satellite [22]–[24]. The NPP-VIRS data have already been calibrated and, thus, are more reliable in mapping temporal changes of lighting than the DMSP-OLS data [25]. Based on environmental data for poverty estimation, Dasgupta *et al.* [26] used regression model to analyze the daily consumption expenditure in four different regions and concluded that the daily consumption expenditure exhibits substantial variations across countries. Vista and Murayama [27] applied linear regression to estimate per capita income in the Philippines and found that the spatial variation in poverty is mainly caused by disparities of access to road infrastructure. Watmough *et al.* [28] discovered that satellite sensor data can be used to calculate the rural welfare in large areas of developing countries and estimated the relative welfare in India through random forest model. For telephonic data, Pokhriyal *et al.* [29] studied the relationship between user behavior statistics collected from mobile communications and poverty index, and concluded that telephonic data can predict poverty index at any spatial resolution. Blumenstock *et al.* [30] accurately reconstructed a country's wealth distribution from mobile phone data and inferred the distribution of asset in a small area with a few households. Multisource data, rather than single

Manuscript received October 2, 2019; revised December 12, 2019; accepted January 7, 2020. Date of publication January 27, 2020; date of current version February 13, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 41971320, and in part by the Project for Follow-up Work in Three Gorges (2017HXNL-01). (Corresponding author: Peng Wu.)

Y. Tan, P. Wu, Y. Li, and B. Bai are with the School of Transportation Science and Engineering, Beihang University, Beijing 100191, China (e-mail: tanym@buaa.edu.cn; zy1713401@buaa.edu.cn; liyunxin@buaa.edu.cn; baibx@buaa.edu.cn).

G. Zhou is with the School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China (e-mail: zhouguanhua@buaa.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.2968468

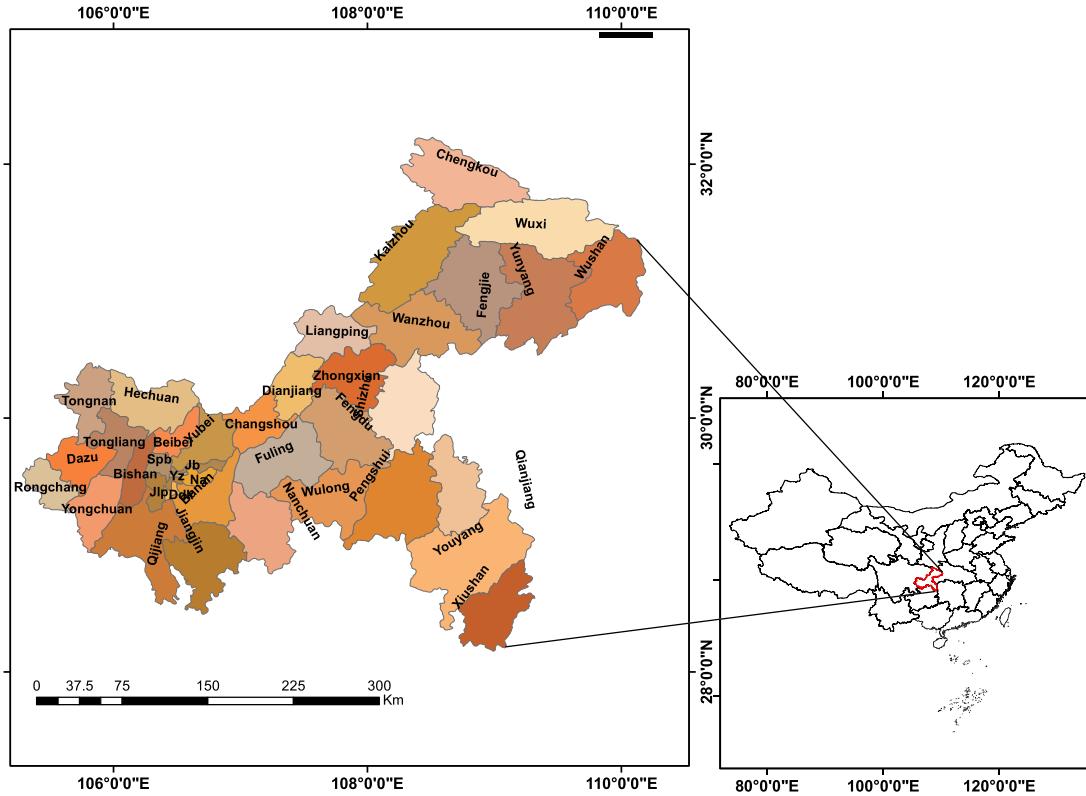


Fig. 1. Study area. In the map, Spb, Yz, Jlp, Ddk, Na, and Jb refer to Shapingba, Yuzhong, Jiulongpo, Dadukou, Nanan, and Jiangbei, respectively.

source data, are also applied in a number of research studies to estimate poverty [31], [32].

Recently, a new method of combining satellite imagery and machine learning to estimate poverty has been proposed [33]. It employs convolutional neural network to extract features from day-time satellite imagery, which can explain up to 75% of the variation in Rwandan economic livelihood indicators. In 2015, He *et al.* [34] introduced a new network structure, residual neural network (ResNet), that won the ImageNet Large Scale Visual Recognition Challenge & Microsoft Common Objects in Context competitions. In their paper [34], they put forward the idea of deep residual learning that can implement the training of deep network better. In 2017, Lin *et al.* [35] developed another new network structure, feature pyramid network (FPN), that solves the multiscale problem in object detection. Through simple network connection change, it greatly improves the performance of small object detection without increasing the calculation amount of the original model [35]. Based on recent advances in deep learning [34]–[36], we propose a deep learning model that combines ResNet-50 and FPN for multitask learning. Multitask learning has been shown to improve the generalization of machine learning models [37]. Normalized Difference Vegetation Index (NDVI) [38], [39] and Modified Normalized Difference Water Index (MNDWI) [40] are chosen to reflect vegetation cover condition and water resources in the region. The night-time light data and Normalized Difference Built Index (NDBI) [41] are adopted to reflect regional economic activity and urban condition, respectively. The input of the proposed model is a satellite image, and the output is the prediction

class of each classification task. The prediction tasks are to predict NDVI, MNDWI, nightlight intensity (NI), and NDBI. All these prediction tasks are equally important. Once the model is trained, image features are extracted from the global average pooling layer of the multitask learning model, and then a cross-validation linear model is established to estimate the economic indicators. Pearson correlation coefficient ( $R^2$ ) is selected to quantitatively evaluate the proposed model. In order to obtain the best prediction results, we conducted multiple experiments. The remainder of this article is organized as follows. Section II introduces the study area and dataset. Section III discusses the flowchart to estimate the economic indicators, the construction of the multitask learning model, and model evaluation. Section IV presents the  $R^2$  results of the four economic indicators, visualizes the features extracted from the model, and also explores the relationship between per capita GDP (PCGDP) and the national poverty-stricken counties (NPCs) in Chongqing. Section V compares the estimation results of the economic indicators from our model with those from other two methods, states the shortcomings of this article, and discusses the way to improve them. Finally, Section VI concludes this article with a summary.

## II. STUDY AREA AND DATASETS

### A. Study Area

Chongqing, China (see Fig. 1) is the only municipality in the central and western regions that is directly administrated by the central government of the People's Republic of China. It is an

TABLE I  
SPECIFIC DATA IN THIS STUDY

Data	Spatial resolution	Temporal resolution	Timeline
Landsat 8 images	30m	1 year	2013-2017
Spectral index data (NDVI, MNDWI, NDBI)	30m	1 year	2013-2017
Nighttime light data	500m	1 year	2013-2017
PCGDP and GPBE data	District or County	1 year	2013-2017
PCDIRR and PCLERR data	District or County	1 year	2014-2017

important junction of China's "One Belt and One Road" strategy and consists of 26 districts, 8 counties, and 4 autonomous counties. On March 19, 2012, the State Council Leading Group for Poverty Alleviation and Development announced a list of 665 key counties for poverty alleviation and development. These counties are designated as NPCs. The Chinese government has been implementing a targeted poverty alleviation strategy since 2013 in order to alleviate poverty and improve people's living standards in poverty-stricken areas [42]. Between 2013 and 2017, the number of NPCs in Chongqing was reduced from 14 to 6 (Chengkou, Fengjie, Wuxi, Shizhu, Youyang, and Pengshui), and the GDP of Chongqing exceeded 2 trillion yuan in 2018, making it the fifth-largest city of China in terms of GDP after Shanghai, Beijing, Shenzhen, and Guangzhou. Despite the rapid increase of total GDP in Chongqing, there are substantial economic imbalances across space, with a large disparity of GDP among counties. Chongqing area presents many mountains and rivers, rugged roads and poor natural conditions. A number of counties are still in poverty owing to these natural factors. Most of the poor counties concentrate in the southeast and northeast regions. Therefore, targeting poverty in Chongqing is of significance in understanding the situation of the poor and helping researchers and policy makers make effective decisions on poverty alleviation.

### B. Data Overview

The data in this article include Landsat 8 images, spectral index data (NDVI, MNDWI, and NDBI), night-time light data, and statistical yearbook data (see Table I)

**Landsat 8 images:** Landsat series satellites have been jointly managed by NASA and the United States Geological Survey since 1972. Landsat 8 is the latest of this series that provides medium resolution (30 m) imagery in nine surface reflectance bands and two thermal bands: Coastal, blue, red, green, near infrared, short wave infrared 1, short wave infrared 2, panchromatic, cirrus, thermal infrared 1, and thermal infrared 2 [43], [44]. The satellite carries two sensors, the land imager (OLI) and the thermal infrared sensor, which collect data in every 16 days. Problems such as shadow, clouds, and snow often appear in Landsat 8 satellite imagery [45]. As a planetary-scale platform for earth science data and analysis, Google Earth Engine has a wide range of applications in forest cover change detection [46], animal habitat monitoring [47], land use classification [48], and water resources research [49]. In this article, Google Earth Engine was adopted to collect Landsat 8 Collection 1 Tier 1 raw scenes of various districts and counties in Chongqing from 2013

to 2017. Landsat SimpleComposite tool was used to synthesize images of various districts and counties in each year. Then, the acquired satellite images were divided into image blocks (tiles) of  $128 \times 128$  pixels. As the vector boundaries of the various districts and counties are irregular, all images occupying blank regions were discarded from the dataset.

**Night-time light data:** The night-time light data were acquired by the Suomi NPP satellite. We downloaded annual and monthly NPP-VIRS data from the NOAA website for the years from 2013 to 2017. Annual data have been preprocessed to remove temporal lights and background (nonlight) values while monthly data are not filtered to screen out auroras, fires, ships, and other temporal lights. First, the monthly NPP-VIRS data from 2013 to 2017 were averaged to obtain the annual NPP-VIRS data for each of the five years. After that, the district and county vector boundaries were applied to clip the annual night-time light data from 2013 to 2017. Finally, the night-time light images for each district and county were divided into small tiles for training as we did for the Landsat 8 images. The calculated total night-time light intensity value of each night-time light image tile represents the NI of the corresponding Landsat 8 image tile. From the histogram of NI in training set (all the night-time light image tiles), the categories of NI were obtained by fitting the Gaussian mixture model to the relative frequency of NI.

**Spectral index data (NDVI, MNDWI, and NDBI):** The normalizedDifference tool in Google Earth Engine was used to generate NDVI, MNDWI, and NDBI data between 2013 and 2017 from the previously acquired Landsat 8 images. The average value of the three index data corresponding to the same area as the Landsat 8 image was calculated to reflect the vegetation, construction, and water condition of the region, respectively. In subsequent model classification learning, the classification categories of NDVI, MNDWI, and NDBI were obtained using the Gaussian fitting model and the categories of night-time light intensity.

**Statistical yearbook data:** In the statistical yearbook, we collected economic survey data including PCGDP, general public budgetary expenditure (GPBE), per capita living expenditure of rural residents (PCLERR), and per capita disposable income of rural residents (PCDIRR). PCGDP and GPBE data were collected from 2013 to 2017, while PCDIRR and PCLERR were collected from 2014 to 2017 due to the absence of 2013 data. Among them, the PCGDP indicator, which is highly related to poverty level, is taken as the basis by the Poverty Alleviation and Development Group of the State Council of China and combined with other factors to identify 11 concentrated areas with special

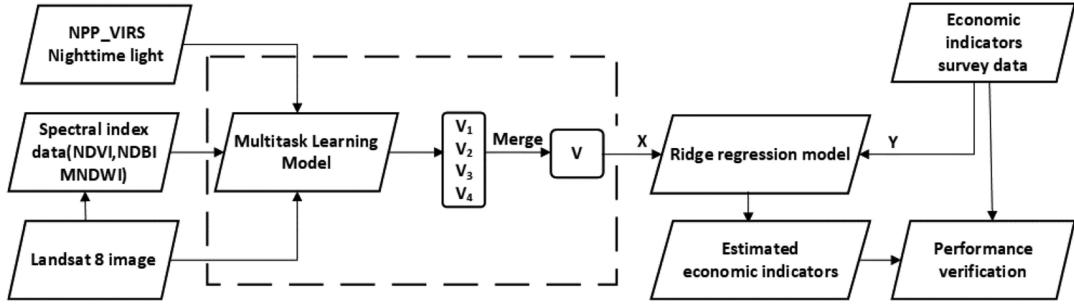


Fig. 2. Flowchart to estimate economic indicators with the proposed multitask learning model.

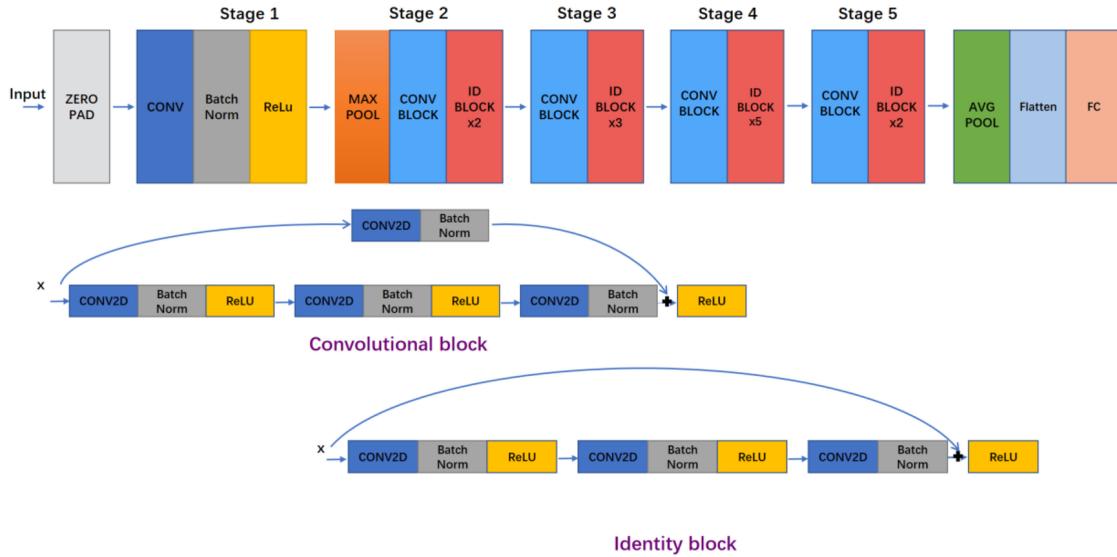


Fig. 3. ResNet-50 network architectures. It includes convolution block and identity block.

difficulties. Therefore, in this article, we focus on the PCGDP economic indicator. All data were obtained from the statistical yearbook of Chongqing website.

### III. METHODS

A flowchart to estimate the economic indicators with the multitask learning model is illustrated in Fig. 2. First, the weights of all the filters of the model are initialized using the pretrained weights on the ImageNet dataset. Second, the multitask learning model is trained to classify the four types of data (night-time light, NDVI, NDBI, and MNDWI), and then extracts various features from the four different classification tasks. These features are subsequently combined into a single feature as the final image feature extracted from the multitask learning model, which is then used to train the ridge regression model to estimate the economic indicators for each district and county, together with the economic indicator data from the statistical yearbook.

#### A. Multitask Learning Model

ResNet-50 (see Fig. 3) first introduced by He *et al.* [34], replaces each 2-layer block in the 34-layer net with the 3-layer bottleneck block and results in higher accuracy than the 34

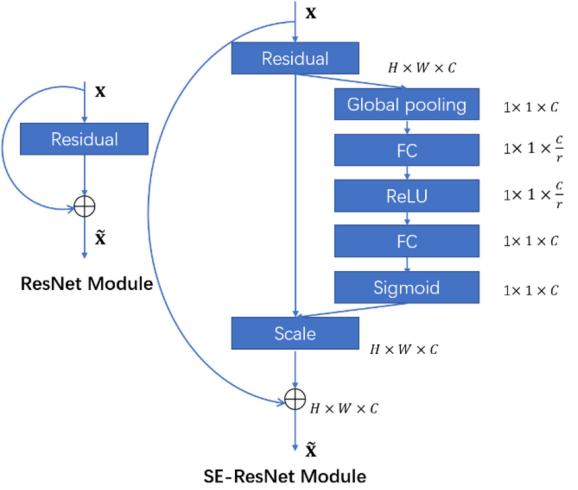


Fig. 4. Original residual module is on the left and the modified SE-ResNet module is on the right.

layers. The whole bottom-up network architecture is divided into five stages, i.e., Stage1, Stage2, Stage3, Stage4, and Stage5, respectively. Residual learning [50] is adopted to avoid the

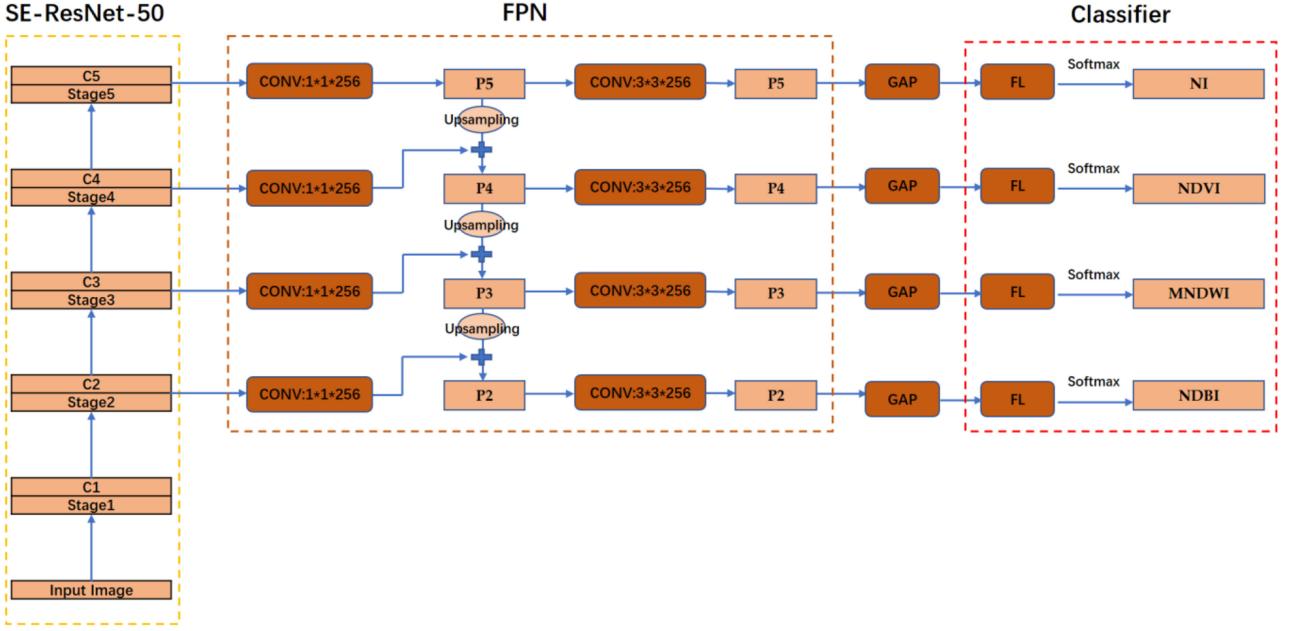


Fig. 5. Network architectures of the multitask learning model.

degradation problem (with the increase of network depth, the accuracy tends to be saturated, and then degraded rapidly). The feature activations output from each stage's last residual block were denoted as {C2, C3, C4, and C5} for Stage2, Stage3, Stage4, and Stage5, respectively. Squeeze-and-excitation block [51] is integrated into the ResNet-50 network architecture. In the block, the channel-wise attention can help the ResNet-50 to find more critical feature channels. The details of this block are shown in Fig. 4 where the ResNet-50 original residual learning block is displayed on the left, and the modified residual learning block SE-ResNet is on the right with the squeeze-and-excitation block being integrated into the ResNet residual learning block.

Adopting the methods mentioned in [35] to build FPN, we obtain the feature map collections {P2, P3, P4, and P5}. The final set of feature maps {P2, P3, P4, and P5} have the same size with {C2, C3, C4, and C5}, respectively. In this article, the feature map P5 is used to classify NI, and the feature maps P2, P3, and P4 that integrate low-level features with high-level features are used to classify NDVI, MNDWI, and NDBI, respectively. The classifier is a two-way fully connected layer behind the global average pooling layer, with 1024 neurons and a softmax activation layer. Fig. 5 shows the network architecture of the multitask learning model composed of SE-ResNet-50, FPN, and the classifier in which GAP and FL represent the global average pooling layer and fully connected layer, respectively.

In terms of the algorithm, the loss functions for the multitask learning model are

$$L_{\text{multitask}} = \sum_{t \in \text{Tasks}} L_t \quad (1)$$

$$L_t = -\frac{1}{n} \sum_{i=1}^n (y_{ij} \log \hat{y}_{ij} + (1 - y_{ij}) \log (1 - \hat{y}_{ij})) \quad (2)$$

where  $L_{\text{multitask}}$  represents the loss of the multitasking model and is the sum of each classified task,  $L_t$  is cross-entropy loss used for each classifier in the classification term,  $n$  is the batch size,  $y_{ij} \in [0, 1]$  denotes the  $j$ th dimension of the ground-truth class label vector for the training image  $I$ , and  $\hat{y}_{ij}$  is the output of the softmax layer. In subsequent sections, when we mention extracted features, the extracted features refer to the output of the global average pooling layer.

### B. Model Performance Evaluation

Once the multitask learning model is trained, we extract the aforementioned layer of each classifier and use it as the feature representation of the input image. Because the vector boundaries of the counties are different from each other, the number of feature vectors for each district also varies. The small image blocks are employed to obtain the features that are then averaged into one feature. The same procedure is performed for each task, and four different 256-dimensional features are generated. These four 256-dimensional features are subsequently merged into a 1024-dimensional feature to represent the image feature extracted from the total satellite images of each region and county. With these features as input, we train a linear ridge regression model to estimate the economic indicators. In the regression model, the 10-fold cross validation method is adopted to predict the economic indicators, where the data are randomly divided into 10 folds, 9 of which are used for each training and the remaining fold is for evaluation. Pearson correlation coefficients ( $R^2$ ) are calculated to evaluate the performance of the proposed multitask learning model based on the actual economic indicators data in the statistical yearbook and the economic indicators data predicted by the ridge regression model. The actual economic indicators are log transformed.

TABLE II  
ACCURACY OF THE TRAINING AND TEST SET FOR THE  
FOUR CLASSIFICATION TASKS

Classification tasks	Training Accuracy (%)	Test Accuracy (%)
NI	99.66	96.88
NDVI	99.00	96.88
MNDWI	99.00	96.88
NDBI	98.97	93.75

Previous studies emphasize only on short-term effects of poverty alleviation and the benefits of economic growth in different regions of China [52], [53]. By contrast, this article makes a comprehensive evaluation of the long-term impact of poverty alleviation in Chongqing by estimating the economic indicators from 2013 to 2017. Our work also systematically and comprehensively assesses the economic development of Chongqing NPCs, which have adopted the targeted poverty alleviation strategy implemented by the Chinese government since 2013.

#### IV. RESULTS

##### A. Estimation of Economic Indicators

The economic indicators are estimated with the features extracted from satellite images for each year from 2013 to 2017. Before model training, all filter weights of ResNet-50 have been initialized with the pre-trained weights on the ImageNet dataset. During the model training process, there are 39 145 satellite images, 70% of which are used for training and 30% for testing. The models select a batch size of 32, a learning rate of 0.01 and 30 000 iterations. We additionally lower the learning rate by 0.1 after 30 epochs of training as this is where the model plateaus, with momentum and weight decay, being set to 0.9 and 0.0001, respectively. Table II reports the accuracy of the training and test sets for each classified data. The training accuracy and test accuracy of NI, NDVI, MNDWI, and NDBI are all over 90%.

We further estimate PCGDP, GPBE, PCLERR, and PCDIRR, but focus on PCGDP, which is highly related to poverty. Fig. 6 shows the line charts of predicted PCGDP and actual PCGDP for each year. It can be seen that the predicted PCGDP is close to the actual PCGDP. In order to evaluate the performance of our model better, we train the ridge regression model with 20 trials for 10-fold cross validation, and then use the average  $R^2$  to evaluate our multitask learning model. Table III shows the  $R^2$ s of the four economic indicators, where the average  $R^2$  value is more than 200 folds (20 trials, 10 folds each). In each fold of cross validation, different regularization constants are used to train the ridge regression model, and the economic indicators are subsequently estimated with the test set. The  $R^2$  of the actual economic indicators and the estimated economic indicators are subsequently calculated. Due to the data unavailability of PCLERR and PCIR, the  $R^2$ s of the two economic indicators are not obtained for the entire period of 2013–2017. The single-year  $R^2$  of PCGDP is constantly higher than GPBE between 2013 and 2017, and the single-year  $R^2$  of PCDIRR is consistently higher than PDLERR between 2014 and 2017.

Excluding 2013, the best single-year  $R^2$  among all the four economic indicators is PCDIRR in 2017 (0.87). The images for the periods of 2013–2014 and 2014–2017 are also used to estimate PCGDP and GPBE, and PCDIRR and PCLERR, respectively. The best indicator is PCGDP with a  $R^2$  of 0.8. PCGDP and GPBE estimated with the image for the period 2013–2017 have higher  $R^2$ s (0.8 and 0.69) than those estimated with their corresponding yearly images, while PCDIRR and PCLERR estimated with the image for the period 2014–2017 exhibit lower  $R^2$ s (0.75 and 0.57) than those estimated with their corresponding yearly images.

To test whether the multitask learning model trained by Chongqing data can be applied to other regions, we choose Guizhou Province for a case study, which is geographically close to Chongqing and with low economic development. To predict PCDIRR of each county in Guizhou Province in 2016, the multitask learning model is trained with the remote sensing images of the province to acquire image characteristics. The actual PCDIRR economic indicators are obtained from the statistical yearbook of Guizhou Province. A scatterplot between the actual PCDIRR and the predicted PCDIRR is drawn (see Fig. 7) with a  $R^2$  of 0.61. This case study indicates that the multitask learning model is universal and can be used to estimate the economic indicators in other regions.

Due to the absence of data for PCDIRR and PCLERR, the  $R^2$ s of these two indicators are not available in 2013 and the period of 2013–2017. Similarly,  $R^2$ s of PCGDP and GPBE are not available in the period of 2014–2017.

In addition, in order to ensure the statistical significance of our results, we randomly redistribute the satellite images of districts and counties to other districts and counties, and train the ridge regression with these randomly distributed images to estimate economic indicators and calculate  $R^2$ . It should be noted that we repeat the experiments 1000 times using images of all available years, in which the  $R^2$ s (see Fig. 8) are randomly distributed for each of the 1000 experiment results. It shows that the  $R^2$ s resulting from the original satellite images are much higher than those from the randomly distributed satellite images. To test the performance of our model in estimating economic indicators in different years, we use the yearly and 2014–2017 images to obtain features and estimate the economic indicators in other single years and the period of 2014–2017. We train the ridge regression with 50 trials for 10-fold cross validation, and take the average of all  $R^2$ s for evaluation (see Fig. 9). It seems that the model performs better in estimating the same year's indicators than other years', but still with acceptable results for other years. In addition, the image of 2014–2017 produces much more accurate results than the yearly images. These results indicate that our model is able to predict economic indicators in other years. In particular, the model can achieve higher performance in estimating economic indicators using images of the entire period of the years.

##### B. Visualization of the Extracted Features and PCGDP Map

The feature vectors extracted from the training images by the deep-learning model are high-dimensional. In order to evaluate whether the features obtained with our model are

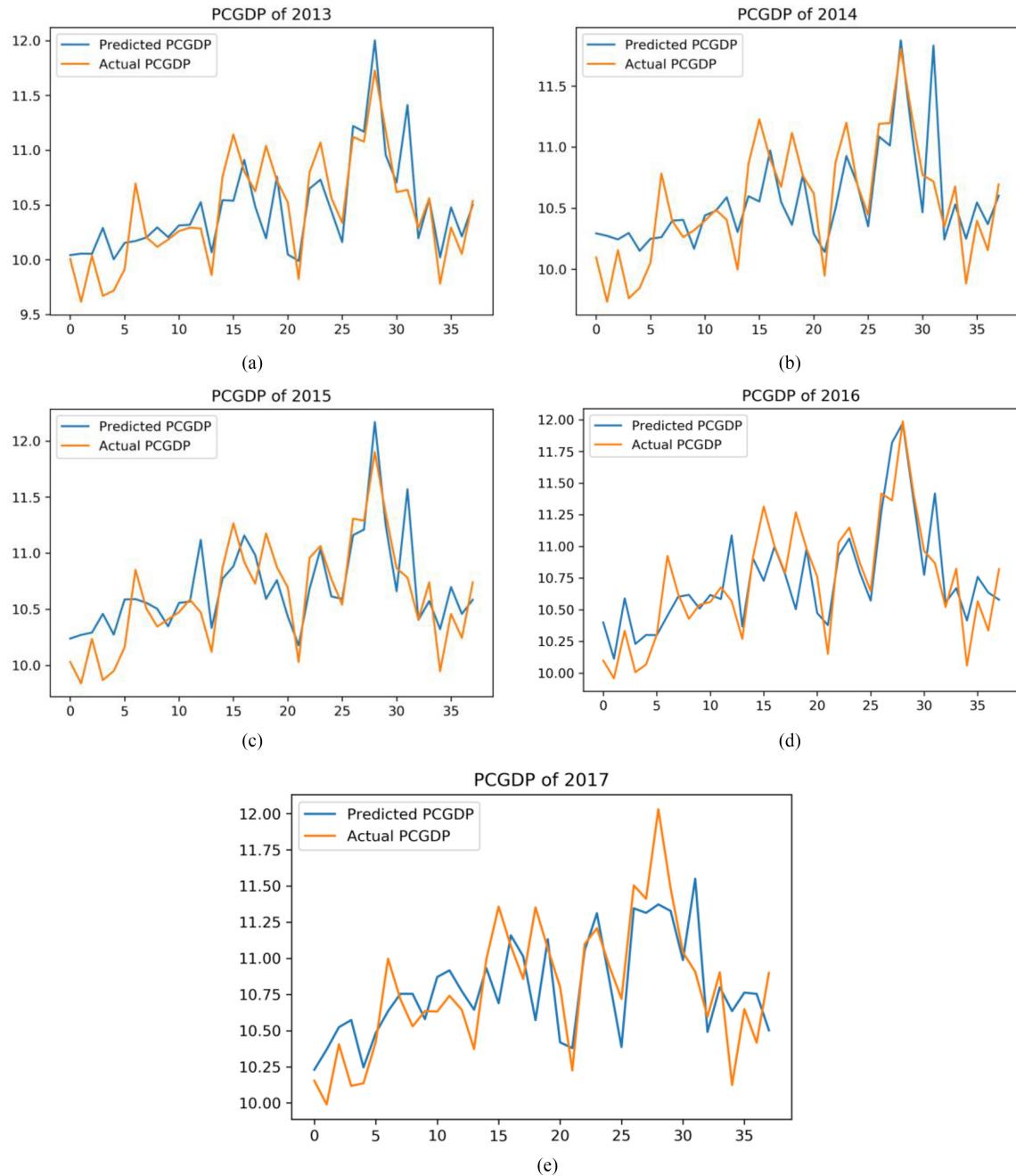


Fig. 6. Line charts of predicted PCGDP and actual PCGDP in (a) 2013. (b) 2014. (c) 2015. (d) 2016. (e) 2017. The x-axis represents different districts and counties in Chongqing, and the y-axis represents the value of PCGDP in Chinese yuan.

TABLE III  
 $R^2$  OF PCGDP, GPBE, PCDIRR, AND PCLERR

Economic indicators	2013	2014	2015	2016	2017	2013-2017	2014-2017
PCGDP	0.71	0.70	0.69	0.72	0.69	0.8	—
GPBE	0.65	0.62	0.61	0.58	0.62	0.69	—
PCDIRR	—	0.86	0.83	0.84	0.87	—	0.75
PCLERR	—	0.76	0.72	0.73	0.69	—	0.57

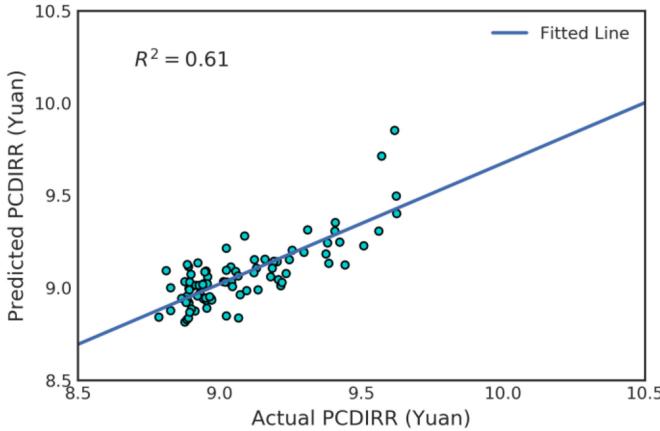


Fig. 7. Scatterplot between the actual PCDIRR and the predicted PCDIRR in Guizhou Province. The blue line is the best fit line.

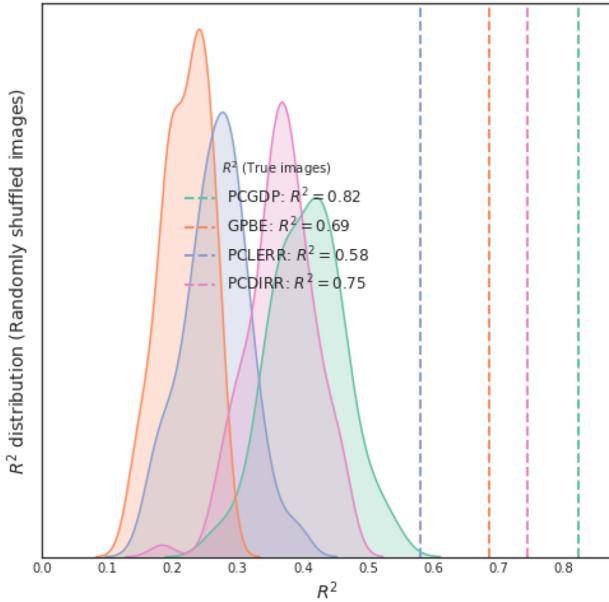


Fig. 8.  $R^2$  (dashed line) obtained from correctly allocated satellite images and  $R^2$  (kernel density estimation map, 1000 results distribution) obtained from randomly distributed satellite images. The  $R^2$  is estimated for the four economic indicators, respectively.

related to regional economic conditions, the high-dimensional feature vectors are mapped to 2-D space. The final feature layer is derived by averaging all the feature layers acquired from the deep learning model. Some landscape features related to regional economic conditions are identified by the model (see Fig. 10). Although there is no direct supervision, the model is able to automatically detect semantically meaningful features such as urban areas, nonurban areas, water, and roads. This shows that the model can recognize regions of interest related to economic status by highlighting the cluttered background of the image. PCGDP is a key indicator in classifying NPCs by the state Council Leading Group Office of Poverty Alleviation and

Development. Therefore, we produce the annual maps of the estimated PCGDP from 2013 to 2017. These PCGDP maps are then divided into five categories using Jenks Natural Breaks [54] classification method. By comparing the PCGDP maps with the Chongqing NPC distribution map published by state Council Leading Group Office of Poverty Alleviation and Development, we explore the relationship between PCGDP and NPCs.

Fig. 11(a)–(c) shows the spatial distribution of the estimated PCGDP in 2013, 2014, and 2015, respectively, while Fig. 11(d) is the spatial distribution of the NPCs for the period of 2013–2015 during which the number of NPCs remained unchanged. There are 14 NPCs during this period, most of which are located in the southeast and northeast of Chongqing. In Fig. 11(a)–(c), the lowest PCGDP category contains 8, 13, and 8 counties in 2013, 2014, and 2015, respectively, among which 8, 12, and 8 are NPCs. Fig. 12(a) and (b) displays the spatial distribution of the estimated PCGDP and the NPCs in 2016, and Fig. 12(c) and (d) presents the spatial distribution of the estimated PCGDP and the NPCs in 2017. Four of the counties in the lowest PCGDP category in Fig. 12(a) belong to the NPCs in Fig. 12(b). In 2016, five NPCs, Wanzhou, Qianjiang, Wulong, Fengdu, and Xiushan, achieved poverty alleviation and became non-NPCs. There are six counties in the lowest PCGDP category in Fig. 12(c), three of which belong to the NPCs in Fig. 12(d). In 2017, three NPCs achieved poverty alleviation, namely Yunyang, Kaizhou, and Wushan. Most of the counties in the lowest and second lowest PCGDP categories are NPCs.

## V. DISCUSSION

In Section IV-A, we estimate four economic indicators using the yearly image from 2013 to 2017, of which the most accurate is PCDIRR with all its  $R^2$  results above 0.8, and the least accurate is GPBE. Among these indicators, we focus on the economic indicator of PCGDP. From 2013 to 2017, the  $R^2$ s between actual PCGDP and estimated PCGDP are 0.71, 0.70, 0.69, 0.72, and 0.69, respectively, which echoes the results of previous research [32], [55]. It shows that although the economic indicators estimated by our model are different from those estimated by previous studies, the model is applicable to the estimation of economic indicators in Chongqing. Furthermore, the image of 2013–2017 is adopted to estimate the economic indicators of PCGDP and GPBE, resulting in  $R^2$ s of 0.8 and 0.69, respectively. These  $R^2$ s are higher than the  $R^2$ s calculated from the indicators estimated with yearly images. The image of 2014–2017 is used to estimate PCDIRR and PCLERR, too, with resulting  $R^2$ s lower than corresponding yearly image results. We also test the use of a single-year image to estimate the economic indicators in other years. The results are very close to those estimated with the image of the same year, which implies that our model can use the image of one year to estimate the economic indicators in other years. In Section IV-B, the features extracted from the model are visualized. The visualization indicates that our model is able to automatically identify the features related to economic conditions. In addition, we map the spatial distributions of PCGDP from 2013 to 2017 and compare them with the

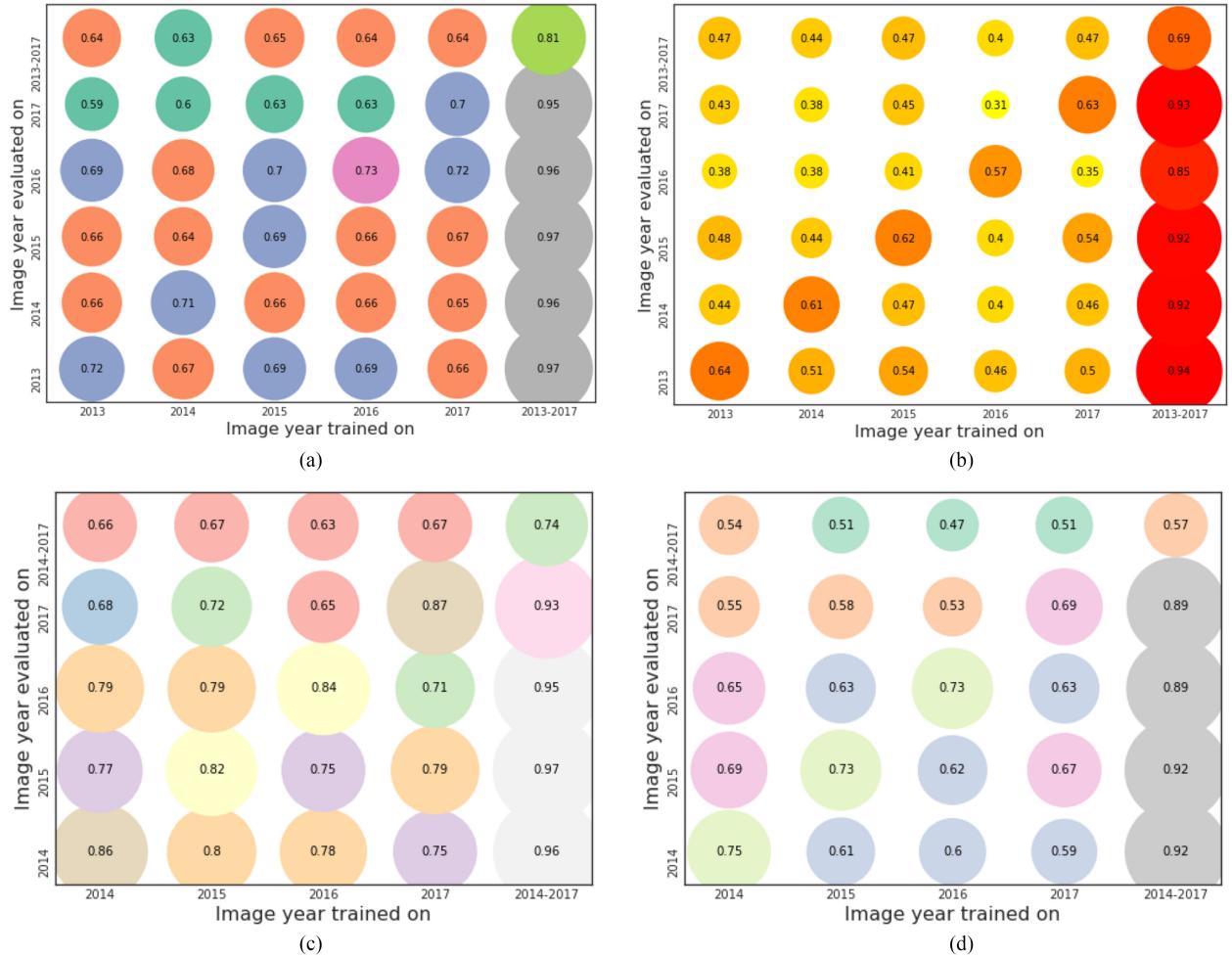


Fig. 9. Cross-validated  $R^2$  values of the four economic indicators predicted from other single years and the period of 2014–2017. (a) PCGDP. (b) GPBE. (c) PCDIRR. (d) PCLERR. Reported  $R^2$  values are averaged over 500 folds (50 trials, 10 folds each).

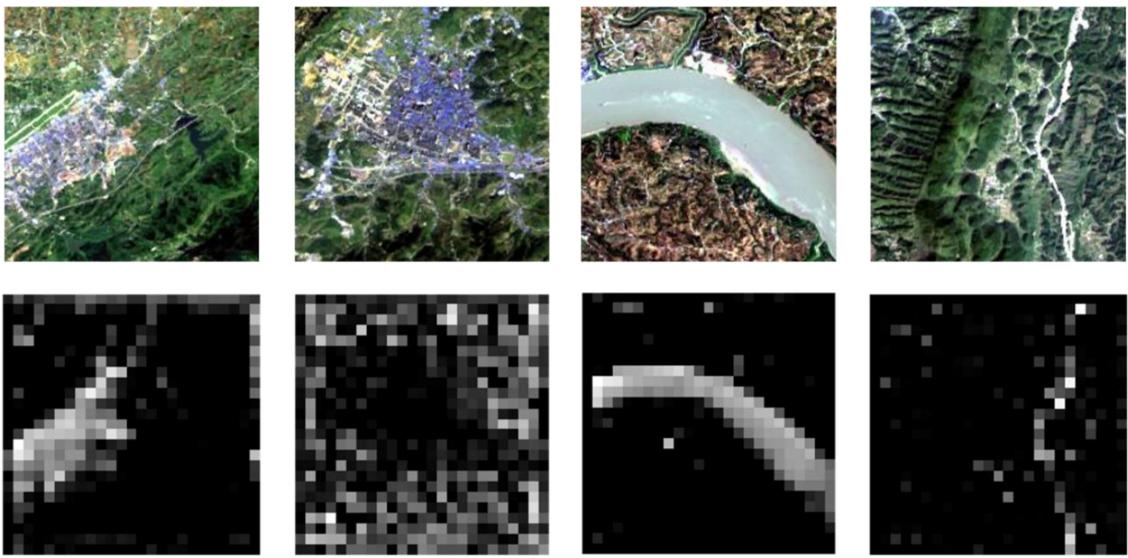


Fig. 10. Visualization of the extracted features from the model. The first row shows four Landsat 8 images with different object categories, which are urban areas, nonurban areas, water, and roads. The second row presents filtered activation maps displaying the features that the model learns from the corresponding Landsat 8 image in the first row.

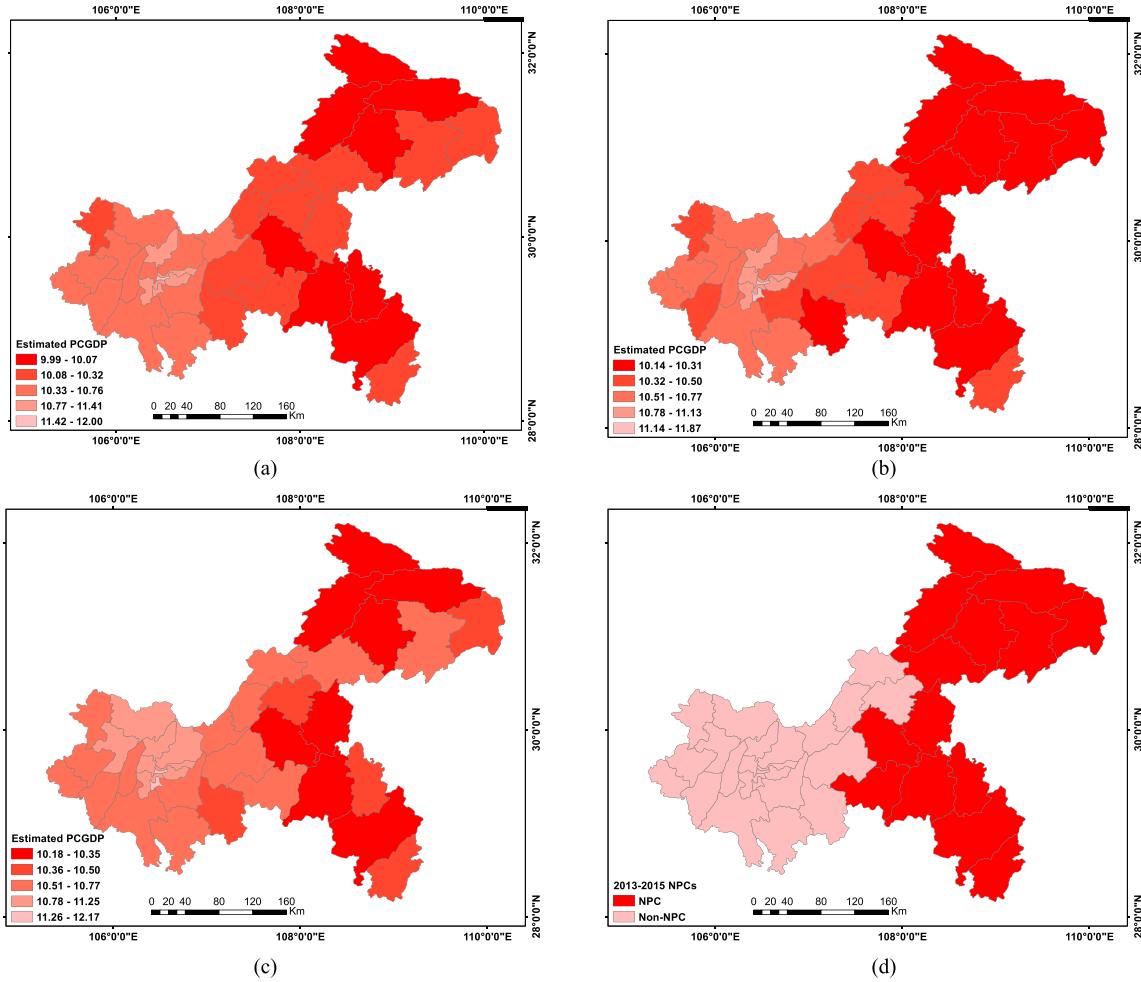


Fig. 11. Comparison between the estimated PCGDP and NPCs. (a) Spatial distribution of the estimated PCGDP at county level in 2013. (b) Spatial distribution of the estimated PCGDP in 2014. (c) Spatial distribution of the estimated PCGDP in 2015. (d) Spatial distribution of the NPCs for the period of 2013–2015, during which the number of NPCs remained unchanged.

spatial distributions of NPCs in Chongqing during this period. It is found that most of the counties with low PCGDP are NPCs. This further confirms that PCGDP is a key indicator to divide NPCs.

In order to test whether our multitask learning model is better than the direct use of night-time light data and spectral index data in estimating economic indicators, i.e., PCGDP, GPBE, PCDIRR, and PCLERR, we compare the results of our multitask learning model with the outcomes from two other models. The first model uses night-time light data to estimate economic indicators by linear regression. The second model adopts both night-time light data and spectral index data to estimate the log-transformed economic indicators with linear regression. Table IV lists the  $R^2$ 's of the four economic indicators estimated with night-time light data. All results are lower than those obtained by our model. Table V shows the  $R^2$ 's of the four economic indicators estimated with both night-time light data and spectral index data. The results are improved greatly compared with Table IV. However, they are still lower than our model. These experiments indicate that the multitask learning model based on deep learning can act as a baseline to analyze

TABLE IV  
 $R^2$  OF THE ECONOMIC INDICATORS ESTIMATED FROM NIGHT-TIME LIGHT DATA WITH LINEAR REGRESSION MODEL FROM 2013 TO 2017

Economic indicators	2013	2014	2015	2016	2017
PCGDP	0.26	0.24	0.23	0.21	0.2
GPBE	0.3	0.29	0.3	0.28	0.3
PCDIRR	—	0.38	0.36	0.36	0.34
PCLERR	—	0.22	0.26	0.25	0.22

TABLE V  
 $R^2$  OF THE ECONOMIC INDICATORS ESTIMATED FROM NIGHT-TIME LIGHT DATA AND SPECTRAL INDEX DATA WITH LINEAR REGRESSION MODEL FROM 2013 TO 2017

Economic indicators	2013	2014	2015	2016	2017
PCGDP	0.69	0.67	0.71	0.72	0.69
GPBE	0.62	0.56	0.59	0.41	0.55
PCDIRR	—	0.74	0.76	0.74	0.71
PCLERR	—	0.66	0.47	0.61	0.53

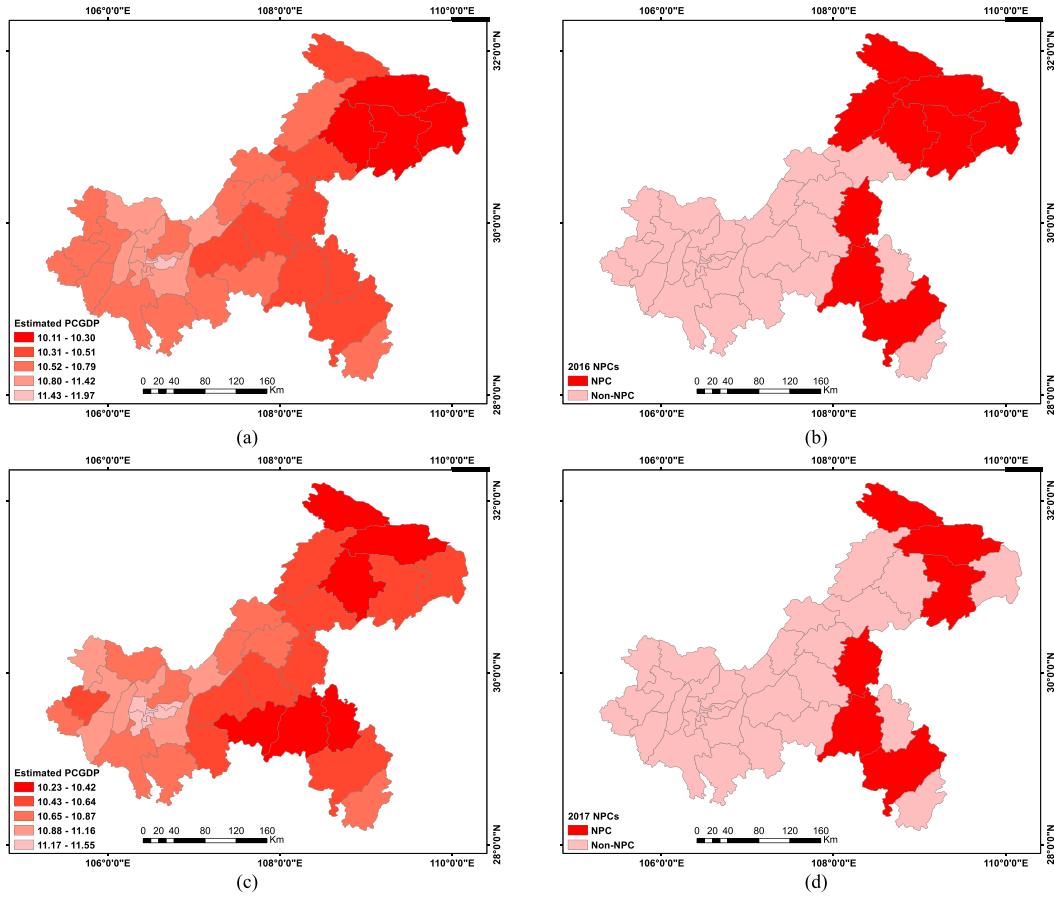


Fig. 12. Comparison between the estimated PCGDP and NPCs. (a) Spatial distribution of the estimated PCGDP at county level in 2016. (b) Spatial distribution of the NPCs in 2016. (c) Spatial distribution of the estimated PCGDP at county level in 2017. (d) Spatial distribution of the NPCs in 2017.

regional economic indicators and introducing spectral index data could better reflect the characteristics of regional economic conditions.

There are some shortcomings in this article. First, because of the different vector boundaries of the districts and counties, the clipped images cannot include all the districts and counties, which affect the extraction of the characteristics for each district and county and result in the decline of the estimation accuracy of economic indicators. Second, due to the unavailability of the economic indicators, it is impossible to compare the results of all the economic indicators from 2013 to 2017, as well as to identify the economic indicators that are better estimated by the model. As more satellite image data are provided and the training and test accuracies of the model are improved, higher estimation accuracy of economic indicators can be achieved.

The economic indicators of PCDIRR and PCLERR lack data in 2013, so the  $R^2$ 's of the two indicators in 2013 are not available and the corresponding results in Tables IV and V are empty.

## VI. CONCLUSION

In this article, we propose a deep-learning framework that combines ResNet-50 and FPN for multitask learning from multiple data sources including night-time light data and spectral

index data. A series of experiments are performed to evaluate the proposed model with Chongqing, China as the study area. Four economic indicators, PCGDP, GPBE, PCLERR, and PCDIRR, are estimated with the model, respectively. Among them, the  $R^2$  of PCGDP, the economic indicator we focus on, is above 0.68 annually from 2013 to 2017, and reaches 0.8 for the period of 2013–2017. Using the images for the period of 2013–2017 and 2014–2017, the indicator with the highest  $R^2$  is PCGDP (0.8), while the  $R^2$ 's of PCLERR and PCDIRR are 0.75 and 0.57, respectively. One of the reasons for the decline of model estimation accuracy is the difficulty to identify living condition change and income change of rural residents from long-time series of remote sensing images. Our model produces higher accuracy than the direct use of night-time light data and the combination of night-time light data and spectral index data to estimate economic indicators. We also compare the distribution maps of PCGDP with the distribution maps of Chongqing's NPCs from 2013 to 2017. Most of the counties with low PCGDP are NPCs, which is consistent with the fact that PCGDP is one of the basic indicators for the Poverty Alleviation and Development Group of the State Council of China to classify NPCs. Based on these analyses, we conclude our article with the following points.

- 1) Landsat 8 imagery can be used to analyze regional economic indicators.

- 2) Spectral index data and night-time light data can provide more information for regional economic analysis.
- 3) The multitask learning model produces relatively high accuracy in estimating economic indicators from the above multiple data sources.
- 4) All data in this article are open and globally available. Therefore, the proposed model can be applied to similar studies in other regions.

#### ACKNOWLEDGMENT

This work was done on Tensorflow and the keras API. The Landsat 8 images were collected from the Google Earth Engine.

#### REFERENCES

- [1] M. Jerven, "Comparability of GDP estimates in sub-Saharan Africa: The effect of revisions in sources and methods since structural adjustment," *Rev. Income Wealth*, vol. 59, no. S1, pp. S16–S36, Oct. 2013.
- [2] World Bank, "Decline of global extreme poverty continues but has slowed." [Online]. Available: <https://www.worldbank.org/en/news/press-release/2018/09/19/decline-of-global-extreme-poverty-continues-but-has-slowed-world-bank>. [Accessed on: Aug. 22, 2019].
- [3] S. Carvalho and H. White, "Combining the quantitative and qualitative approaches to poverty measurement and analysis," Washinton, DC, USA: The World Bank, 1997.
- [4] ECG-ISSER, "Ghana Socioeconomic Panel Study Survey: 2009-2010. Distributed by The World Bank. Study ID: GHA\_2009\_GSPS\_v01\_M."
- [5] F. H. G. Ferreira *et al.*, "A global count of the extreme poor in 2012: Data issues, methodology and initial results," *J. Econ. Inequality*, vol. 14, no. 2, pp. 141–172, 2016.
- [6] P. Sutton, D. Roberts, C. Elvidge, and K. Baugh, "Census from heaven: An estimate of the global human population using night-time satellite imagery," *Int. J. Remote Sens.*, vol. 22, no. 16, pp. 3061–3076, Nov. 2001.
- [7] X. Yang, W. Yue, and D. Gao, "Spatial improvement of human population distribution based on multi-sensor remote-sensing data: An input for exposure assessment," *Int. J. Remote Sens.*, vol. 34, no. 15, pp. 5569–5583, 2013.
- [8] P. C. Sutton, D. A. Roberts, C. D. Elvidge, and H. Melj, "A comparison of nighttime satellite imagery and population density for the continental united states," *Photogramm. Eng. Remote Sens.*, vol. 63, 1997, Art. no. 1303.
- [9] A. Marx and M. Ziegler Rogers, "Analysis of Panamanian DMSP/OLS nightlights corroborates suspicions of inaccurate fiscal data: A natural experiment examining the accuracy of GDP data," *Remote Sens. Appl. Soc. Environ.*, vol. 8, pp. 99–104, 2017.
- [10] 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," *GISci. Remote Sens.*, vol. 54, no. 3, pp. 407–425, May 2017.
- [11] K. Shi *et al.*, "Evaluating the ability of NPP-VIIRS nighttime light data to estimate the gross domestic product and the electric power consumption of China at multiple scales: A comparison with DMSP-OLS data," *Remote Sens.*, vol. 6, no. 2, pp. 1705–1724, 2014.
- [12] K. Shi, B. Yu, C. Huang, J. Wu, and X. Sun, "Exploring spatiotemporal patterns of electric power consumption in countries along the Belt and Road," *Energy*, vol. 150, pp. 847–859, May 2018.
- [13] K. Shi *et al.*, "Evaluating spatiotemporal patterns of urban electricity consumption within different spatial boundaries: A case study of Chongqing, China," *Energy*, vol. 167, pp. 641–653, Jan. 2019.
- [14] A. C. Townsend and D. A. Bruce, "The use of night-time lights satellite imagery as a measure of Australia's regional electricity consumption and population distribution," *Int. J. Remote Sens.*, vol. 31, no. 16, pp. 4459–4480, 2010.
- [15] 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," *Ecol. Econ.*, vol. 41, no. 3, pp. 509–527, 2002.
- [16] Y. Wang and G. Li, "Mapping urban CO<sub>2</sub> emissions using DMSP/OLS 'city lights' satellite data in China," *Environ. Plan. A*, vol. 49, no. 2, pp. 248–251, 2017.
- [17] X. Cui, Y. Lei, F. Zhang, X. Zhang, and F. Wu, "Mapping spatiotemporal variations of CO<sub>2</sub> (carbon dioxide) emissions using nighttime light data in Guangdong Province," *Phys. Chem. Earth*, vol. 110, pp. 89–98, Apr. 2019.
- [18] J. Ou, X. Liu, X. Li, M. Li, and W. Li, "Evaluation of NPP-VIIRS nighttime light data for mapping global fossil fuel combustion CO<sub>2</sub> emissions: A comparison with DMSP-OLS nighttime light data," *PLoS One*, vol. 10, no. 9, Sep. 2015, doi: [10.1371/journal.pone.0138310](https://doi.org/10.1371/journal.pone.0138310).
- [19] C. Li, G. Li, Y. Zhu, Y. Ge, H. te Kung, and Y. Wu, "A likelihood-based spatial statistical transformation model (LBSSTM) of regional economic development using DMSP/OLS time-series nighttime light imagery," *Spatial Statist.*, vol. 21, pp. 421–439, 2017.
- [20] B. Yu *et al.*, "Object-based spatial cluster analysis of urban landscape pattern using nighttime light satellite images: A case study of China," *Int. J. Geograph. Inf. Sci.*, vol. 28, no. 11, pp. 2328–2355, 2014.
- [21] C. D. Elvidge, K. E. Baugh, M. Zhizhin, and F.-C. Hsu, "Why VIIRS data are superior to DMSP for mapping nighttime lights," *Proc. Asia-Pac. Adv. Netw.*, vol. 35, 2013, Art. no. 62.
- [22] X. Li, H. Xu, X. Chen, and C. Li, "Potential of NPP-VIIRS nighttime light imagery for modeling the regional economy of China," *Remote Sens.*, vol. 5, no. 6, pp. 3057–3081, 2013.
- [23] C. D. Elvidge, M. Zhizhin, F. C. Hsu, and K. E. Baugh, "VIIRS nightfire: Satellite pyrometry at night," *Remote Sens.*, vol. 5, no. 9, pp. 4423–4449, 2013.
- [24] C. D. Elvidge, M. Zhizhin, F.-C. Hsu, and K. Baugh, "What is so great about nighttime VIIRS data for the detection and characterization of combustion sources?" *Proc. Asia-Pac. Adv. Netw.*, vol. 35, 2013, Art. no. 33.
- [25] K. Baugh, F.-C. Hsu, C. D. Elvidge, and M. Zhizhin, "Nighttime lights compositing using the VIIRS day-night band: Preliminary results," *Proc. Asia-Pac. Adv. Netw.*, vol. 35, 2013, Art. no. 70.
- [26] S. Dasgupta, U. Deichmann, C. Meisner, and D. Wheeler, "Where is the poverty-environment nexus? Evidence from Cambodia, Lao PDR, and Vietnam," *World Dev.*, vol. 33, no. 4, pp. 617–638, 2005.
- [27] B. M. Vista and Y. Murayama, "Spatial determinants of poverty using GIS-based mapping," in *Spatial Analysis and Modeling in Geographical Transformation Process*. New York, NY, USA: Springer-Verlag, 2011.
- [28] G. R. Watmough, P. M. Atkinson, A. Saikia, and C. W. Hutton, "Understanding the evidence base for poverty-environment relationships using remotely sensed satellite data: An example from Assam, India," *World Dev.*, vol. 78, pp. 188–203, Feb. 2016.
- [29] N. Pokhriyal, W. Dong, and V. Govindaraju, Virtual networks and poverty analysis in Senegal, Jun. 2015. [Online]. Available: <https://arxiv.org/abs/1506.03401>.
- [30] J. Blumenstock, G. Cadamuro, and R. On, "Supplementary materials for predicting poverty and wealth from mobile phone metadata," *Science*, vol. 350, no. 6264, pp. 1073–1076, 2015.
- [31] N. Pokhriyal and D. C. Jacques, "Combining disparate data sources for improved poverty prediction and mapping," *Proc. Natl. Acad. Sci.*, vol. 114, no. 46, pp. E9783–E9792, 2017.
- [32] X. Zhao *et al.*, "Estimation of poverty using random forest regression with multi-source data: A case study in Bangladesh," *Remote Sens.*, vol. 11, no. 4, 2019, Art. no. 375.
- [33] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, vol. 353, no. 6301, pp. 790–794, 2016.
- [34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [35] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 936–944.
- [36] A. Perez, S. Ganguli, S. Ermon, G. Azzari, M. Burke, and D. Lobell, "Semi-supervised multitask learning on multispectral satellite images using wasserstein generative adversarial networks (GANs) for predicting poverty, 2019. [Online]. Available: <https://arxiv.org/abs/1902.11110>.
- [37] R. Caruana, T. Mitchell, H. Simon, and D. Pomerleau, "Multitask learning Rich Caruana 23 September1997," School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA, 1997.
- [38] R. E. Kennedy, Z. Yang, and W. B. Cohen, "Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrend - Temporal segmentation algorithms," *Remote Sens. Environ.*, vol. 114, no. 12, pp. 2897–2910, 2010.
- [39] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.*, vol. 8, no. 2, pp. 127–150, 1979.

- [40] K. Rokni, A. Ahmad, A. Selamat, and S. Hazini, "Water feature extraction and change detection using multitemporal landsat imagery," *Remote Sens.*, vol. 6, no. 5, pp. 4173–4189, 2014.
- [41] H. Xu, "Extraction of urban built-up land features from Landsat imagery using a thematic-oriented index combination technique," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 12, pp. 1381–1391, 2013.
- [42] Z. Qi and Z. Tao, "Advances and practices for targeted poverty alleviation, *China Econ. Trans.*, vol. 1, pp. 134–141, 2018.
- [43] D. P. Roy *et al.*, "Landsat-8: Science and product vision for terrestrial global change research," *Remote Sens. Environ.*, vol. 145, pp. 154–172, Apr. 2014.
- [44] G. Rongali, A. K. Keshari, A. K. Gosain, and R. Khosa, "Split-Window algorithm for retrieval of land surface temperature using landsat 8 thermal infrared data," *J. Geovisualization Spat. Anal.*, vol. 2, no. 2, 2018, Art. no. 14.
- [45] A. K. Whitcraft, E. F. Vermote, I. Becker-Reshef, and C. O. Justice, "Cloud cover throughout the agricultural growing season: Impacts on passive optical earth observations," *Remote Sens. Environ.*, vol. 156, pp. 438–447, Jan. 2015.
- [46] J. A. Fortin, J. A. Cardille, and E. Perez, "Multi-sensor detection of forest-cover change across 45 years in Mato Grosso, Brazil," *Remote Sens. Environ.*, vol. 2, 2019, Art. no. 111266.
- [47] A. R. Joshi *et al.*, "Tracking changes and preventing loss in critical tiger habitat," *Sci. Adv.*, vol. 2, no. 4, Apr. 2016, Art. no. e1501675.
- [48] Y. Xie, T. J. Lark, J. F. Brown, and H. K. Gibbs, "Mapping irrigated cropland extent across the conterminous United States at 30 m resolution using a semi-automatic training approach on Google Earth Engine," *ISPRS J. Photogramm. Remote Sens.*, vol. 155, pp. 136–149, Sep. 2019.
- [49] Y. Zhou *et al.*, "Continuous monitoring of lake dynamics on the Mongolian plateau using all available Landsat imagery and Google Earth engine," *Sci. Total Environ.*, vol. 689, pp. 366–380, Nov. 2019.
- [50] M. Wang, X. Zhang, X. Niu, F. Wang, and X. Zhang, "Scene classification of high-resolution remotely sensed image based on ResNet," *J. Geovis. Spat. Anal.*, vol. 3, no. 2, pp. 1–9, 2019.
- [51] J. Hu, "Squeeze-and-excitation networks, May 2019. [Online]. Available: <https://arxiv.org/abs/1709.01507>.
- [52] 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, Aug. 2017.
- [53] Y. Liu, J. Liu, and Y. Zhou, "Spatio-temporal patterns of rural poverty in China and targeted poverty alleviation strategies," *J. Rural Stud.*, vol. 52, pp. 66–75, May 2017.
- [54] C. A. Brewer and L. Pickle, "Evaluation of methods for classifying epidemiological data on choropleth maps in series," *Ann. Assoc. Amer. Geographers*, vol. 92, no. 4, pp. 662–681, 2002.
- [55] N. Jean, M. Burke, M. Xie, W. M. Davis, D. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, vol. 353, no. 6301, pp. 790–794, 2016.

**Yumin Tan** received the M.S. degree in geodesy and surveying engineering from Tsinghua University, Beijing, China, in 2001, and the Ph.D. degree in cartography and geographic information systems from Chinese Academy of Sciences, Beijing, China, in 2004.

Currently, she is an Associate Professor with the School of Transportation Science and Engineering, Beihang University, Beijing, China. Her research interests include GIS, remote sensing image information extraction, and regional integrated applications.

**Peng Wu** is currently working toward the M.S. degree in architectural and civil engineering from Beihang University, Beijing, China.

His research interests include machine learning and remote sensing using night-time light data.

**Guanhua Zhou** received the M.S. degree in geophysics from Ocean University of China, Qingdao, China, in 2003, and the Ph.D. degree in cartography and geographic information system from Institute of Remote Sensing Applications, Chinese Academy of Sciences, China, 2007.

He is currently a Lecturer with the School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing, China. His research interests include environmental remote sensing, remote sensing applications, and polarization remote sensing.

**Yunxin Li** is currently working toward the M.S. degree in civil engineering at Beihang University, Beijing, China.

His research interests include point cloud data processing and UAV image processing.

**Bingxin Bai** received the M.S. degree in architectural and civil engineering from Beihang University, Beijing, China, and is currently working toward the Ph.D. degree in highway and railway engineering from Beihang University, Beijing, China.

Her research interests include GIS, remote sensing image information extraction, regional integrated applications, and remote Sensing data fusion.