

Weakly Supervised Deep Learning for Fine-grained Socioeconomic Development Index Inference Based on Satellite Imagery

Zaishuo Xia¹, Yanbing Bai^{1*}, Bolin Zhang², Shubo Zeng¹, Zhe Liu¹, Ziyue Zhang¹ and Xuxin Mao³

¹Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, 100872, China.

²School of Finance, Renmin University of China, Beijing, 100872, China.

³National Institute of Economic and Social Research, London, SW1P 3HE, United Kingdom.

*Corresponding author(s). E-mail(s): ybbai@ruc.edu.cn;

Abstract

Acquiring timely and accurate social and economic development dynamic information is very important for government decision-making. However, existing methods for acquiring social and economic indicators mainly rely on a time-consuming and laborious economic census, which limits the quality and timeliness of the data. In recent years, satellite remote sensing data with high timeliness and coverage have provided us with the possibility of fast and high-precision insight into economic development indicators. Therefore, we developed a weakly supervised training framework, which uses a large amount of unlabeled satellite imagery data to estimate regional social and economic indicators, reduces the dependence on labeling, and improves the reliability of the model. Our model shows better performance than traditional supervised methods and has better interpretability, which provides new insights for monitoring the drivers behind the dynamic of social and economic indicators with satellite imagery big data. Our paper successfully applies this weakly supervised framework to estimate the social development index across China and aggregate it for the county-level development index. Additionally, we adjust the county-level index based on the corresponding administrative region's area. Our method shows better performance than

2 WS-DL for FG-SEDI Inference Based on Satellite Imagery

traditional supervised methods, which has been shown to achieve Pearson correlation coefficients above 0.8 with multiple economic indicators. Moreover, we have conducted a deeper discussion on interpretability, which indicates that our model is sensitive to roads and buildings but insensitive to farmlands. Codes and implementation details are made available at <https://github.com/xiafire/Economic-Development>.

Keywords: Cluster analysis, Computational Socioeconomics, Computer Vision, Economic and Social Statistics, High-resolution satellite imagery, Weakly Supervised Deep Learning

1 Introduction

Accurately measuring the level of economic development in a region is of significant importance. For the government, it can evaluate the effect of the policy implementation in the previous stage and can provide a foundation for policy formulation in the next step; for the companies, economic development will affect corporate development strategies such as plant setting; and for investors, discovering potential areas with sound economic growth can create higher returns in the future. However, in some developing countries, missing and outdated data is common (Wu and Ru, 2019). In China, the sheer scale and population present challenges in statistical work at the local level, including slow data updates and limitations in defining statistical boundaries based on administrative units (Gao and Fu, 2014; Liao, 2014; Zhang, 2015).

Utilizing large-scale and easily accessible remote sensing data can help alleviate these challenges and provide a more comprehensive and accurate assessment of economic development (Donaldson and Storeygard, 2016a; Ghosh et al, 2013). People are paying more and more attention to satellite imagery to measure specific social development indicators (Donaldson and Storeygard, 2016b; Burke et al, 2021). Compared with manual statistical data, satellite data has the advantages of low cost, fast update speed, and more comprehensive data, making it more convenient to use. For instance, the utilization of high-resolution satellite imagery has been instrumental in analyzing air pollution and land use patterns (Elvidge et al, 1997b, 2001; Welch, 1980; Sutton and Costanza, 2002). Deep convolutional neural networks have also been utilized to extract information from these images, leading to more precise and accurate results (Zhao et al, 2018; Yuan et al, 2020).

So far, the research on applying deep convolution neural networks and satellite imagery to obtain social and economic indicators has yielded promising outcomes (Burke et al, 2021; Pandey et al, 2018). Initially, most studies in this domain have utilized supervised learning, which requires labeled ground truth data for model training (Yeh et al, 2020a). Nonetheless, these approaches suffer from two key limitations. Firstly, most classification and regression models require labeled data for training, which may not be readily available for

many socioeconomic indicators (Bjorck et al, 2021). Secondly, deep convolution neural networks are widely perceived as a black-box process, raising concerns regarding the lack of interpretability (Haut et al, 2021).

In response to the above problems, this paper utilizes a weakly supervised learning framework for estimating fine-grained social development indices using satellite images and nightlight imagery based on the solutions proposed in Han et al (2020) and Weakly supervised learning can use limited, noisy or inaccurate data to train model parameters. In this paper, we use inexact supervision(Zhou, 2017), a form of weakly supervised learning, to estimate fine-grained social development indices. More specifically, coarse nightlight intensity data are used as proxy variables to train the model. The framework mainly consists of four steps: (1) Feature vector extraction; (2) Feature vector clustering; (3) Partial Order Graph(POG) generation; (4) Index assignment. This framework breaks through the requirements of previous methods for labeling data sets, and at the same time, the meaning of each step is more clear, which solves the "black box" nature of deep neural networks to a certain extent. This paper will use this framework to estimate the social development index in China, and analyze the practical significance of the predicted index. Compared with previous studies, this paper has the following innovations:

- We utilize a novel weakly supervised framework for evaluating regional social development indicators, which reduces the dependence on labeled data and utilizes massive large-scale remote sensing images more fully. Our method uses nightlight signals, without the need for manual annotation, to train the deep learning model. This method not only reduces the cost of data collection but also enhances the coverage and accuracy of the data.
- Due to the use of weakly supervised methods, this paper breaks through the limitations of administrative divisions on forecasting units, using a large amount of satellite images covering the entire China, and obtains more fine-grained social development indicators than county-level administrative divisions, which can provide forward-looking information before relevant government statistics are carried out;
- We conducted a thorough analysis of the interpretability of our model, providing insights into the key features in remote sensing images that reflect social development and the economic indicators that are suitable for predicting socioeconomic indicators. With China's strong spatial heterogeneity and standardized government statistical standards, this not only enhances our understanding of the connection between remote sensing imagery and social development, but also provides a valuable example for future big data applications in this field.

2 Related Works

Satellite imagery has emerged as a powerful tool for predicting economic activity in recent years. Early research in this field focused on extracting low-dimensional features (e.g., the city's building area and the intensity of

4 WS-DL for FG-SEDI Inference Based on Satellite Imagery

nighttime lighting) from satellite images to estimate social and economic indicators. In the late 1960s, [Tobler \(1969\)](#) used LANDSAT satellite images to verify the relationship between a city's building area and its population. The study found that the settlement size coefficient varied across different regions and times but generally followed the expected trend, demonstrating the value of satellite data in urban geography research. [Welch \(1980\)](#) utilized images from the DMSP satellite to measure electricity consumption in US cities and conducted correlation analyses with population and urban area. Additionally, it used microdensitometry to create brightness profiles of cities and compared them across different types and sizes of cities. The research demonstrated the potential and innovation of utilizing nightlight data in urban research, pioneering the use of nighttime light data to analyze urban energy utilization patterns. There has been a proliferation of research on the relationship between nighttime lighting and various economic indicators, such as GDP, carbon emissions, and urbanization, at different spatial and temporal scales ([Elvidge et al, 2001, 1997a, 1999](#)). By analyzing nighttime light satellite imagery data, [Doll et al \(2006\)](#) examine regional economic productivity across multiple subnational levels in 11 EU countries and the United States. The findings of this research reveal a strong correlation between nighttime light data and regional economic productivity, indicating that nighttime light intensity can serve as an indicator to measure regional economic activity.

Predicting social and economic indicators solely through extracting low-dimensional features from satellite images poses several challenges. Firstly, the relationship between the extracted features and the targeted indicators may not be straightforward [Bluhm and McCord \(2022\); Mellander et al \(2015\)](#). In fact, nightlight intensity can be affected by many factors, such as lighting technology, urban design, and cultural habits, which cannot directly reflect economic activity. Additionally, the use of OLS or GWR limits the complexity of the regression functions, which may not capture all the potential variables and relationships in the data [Charlton et al \(2009\); Bagan and Yamagata \(2015\); Li et al \(2020\)](#). Moreover, the low-dimensional features extracted are susceptible to noise, and can not effectively capture fine-grained spatial heterogeneity, which can be crucial for accurately predicting social and economic indicators [Wang et al \(2019\)](#).

Researchers have used convolutional neural networks to extract high-dimensional features from satellite imagery that are linked to economic activities, such as vehicles ([Minetto et al, 2020](#)), buildings,roads ([Ackermann et al, 2020; Alshehhi et al, 2017; Buslaev et al, 2018; Minetto et al, 2020](#)) and markets ([Baragwanath et al, 2021](#)). [Engstrom et al \(2017\)](#) explores how convolutional neural networks (CNNs) can be used to extract features from high-resolution satellite images of Sri Lanka and estimate poverty rates and average log consumption. The paper compares the performance of CNNs with traditional methods that use handcrafted features, finding that CNNs outperform traditional methods in terms of predictive accuracy. [Yeh et al \(2020b\)](#) uses ResNet-18 CNN models to predict household consumption expenditure

and asset wealth from high-resolution satellite images of Nigeria, Tanzania, Uganda and Malawi. The paper also explores how different types of satellite data (optical, infrared and nightlight) can improve the predictions. Liu et al (2021) proposed a method to infer the GDP of Chinese counties using nighttime light data, utilizing a convolutional neural network with attention mechanism on daytime satellite images.

Although deep learning techniques have shown great potential in extracting features from remote sensing images for predicting socioeconomic indicators, the challenges of training these models with limited labeled data still exist. One approach is to use transfer learning, which allows the model to leverage pre-trained weights on a large dataset to improve performance on a smaller dataset with limited labeled data (Bengio, 2012). Jean et al (2016) first trained a CNN on high-resolution satellite images of the United States and then applied the learned features to low-resolution satellite images of Africa, resulting in improved accuracy and efficiency in predicting poverty levels. These approaches demonstrate the potential of deep learning and transfer learning in leveraging large-scale satellite imagery to better understand and predict economic activities. Another approach is to use weakly supervised learning, which leverages open-source remote sensing data in an interpretable manner to estimate social and economic development indicators (Zhou, 2017).

3 Data Collection

The data used in this paper include satellite remote sensing images, nighttime light data, and representative socioeconomic statistical indicators. We match these spatial data with statistical data through a geographic fence matching algorithm, which is described below:

3.1 Satellite Remote Sensing Data

The ArcGIS website provides satellite remote sensing imagery at different levels covering the globe. These images are divided into 24 levels according to the resolution from low to high, corresponding to different choices of zoom-level from 0 to 23, and the actual geographic range covered is correspondingly different. Zoom level is a term used to describe the level of spatial resolution, whose value corresponds to the number of subdivisions made to the earth's surface. For example, a zoom level of 0 would show the entire earth as a single texture, while a zoom level of t would divide the earth into 2^t parts. We take the image with the highest resolution of 0.6m provided by the website as an example. Currently, the corresponding zoomlevel=18, and the size of each image is 256×256 pixels. From this calculation, each image covers the actual range of $150m \times 150m$. Due to the huge geographical scope of China and the large number of counties to be studied, this paper comprehensively considers the training difficulty and feature extraction performance and finally chooses to download the image with zoomlevel=12, and each image covers the actual

6 *WS-DL for FG-SEDI Inference Based on Satellite Imagery*

geographical range of 9.7km × 9.7km. A total of 131,952 satellite images have been obtained in this study.

According to the high-definition global nighttime light map released by EOG (Earth Observation Group), this paper selects the annual composite map of 2020 under the VCM version. This version has removed the effects of stray light, temporary light, and background light and more accurately reflects the traces of human economic activity in the year. For each latitude and longitude coordinate loc_i of the satellite images, this paper will calculate the average d_i of light intensities with the corresponding coordinate as the center, and d_i can be regarded as an approximation of economic activities in the coverage area of the corresponding satellite image x_i . In this paper, the Gaussian mixture model (GMM) algorithm ([Biernacki et al, 2000](#)) is used to divide the light intensity into three categories: low intensity, medium intensity, and high intensity. However, due to the high proportion of low-level samples, some samples with low nightlight intensity levels are dropped to maintain data balance. The corresponding relationship between the light intensity d_i and the light intensity level D_i is as follows:

$$D_i = \begin{cases} 0 & d_i \leq 0.02 \\ 1 & 0.02 < d_i \leq 3 \\ 2 & d_i > 3 \end{cases} \quad (1)$$

In the formula, 0 represents low intensity, 1 represents medium intensity, and 2 represents high intensity. And the sample result is shown in Fig. 1.

Since low-intensity images often correspond to areas such as rivers, mountains, and grasslands, this paper samples low-intensity images to balance the data distribution and enhance the model's efficiency to learn features related to socioeconomic activities. Finally, the distribution of light intensity levels corresponding to the satellite images in 2020 is shown in Table 1.

Table 1 Remote Sensing Data Light Intensity Level Distribution

Light intensity level	Low intensity(0)	medium intensity(1)	high intensity(2)
Ratio(%)	20.0	55.6	24.4

3.2 Socioeconomic Indicators

This paper selects a series of critical socioeconomic indicators to measure the validity of the model output. We refer to some articles that study the Chinese economy with certain indicators, which are labeled in Table 2. Since it is difficult to obtain more fine-grained statistical data covering the whole country, this paper selects the government statistical data of the smallest, county-level, unit. After considering the integrity and socioeconomic significance of the data, this paper selects the following (Table 2) county-level indicators and shows corresponding abbreviations in 2020.

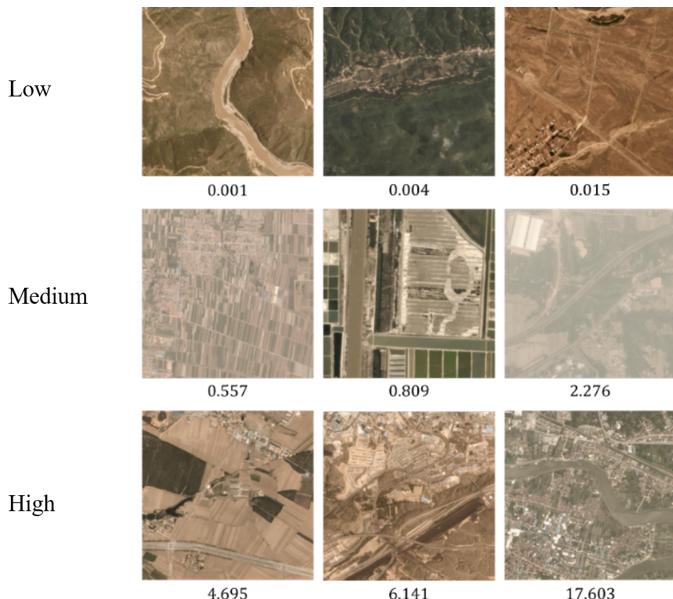


Fig. 1 Geographic satellite images and corresponding nightlight intensity levels. The above figure is an example of geographic satellite pictures with low, medium, and high levels of nighttime light intensity, and the actual value of nighttime light intensity is below each figure.

Table 2 Socioeconomic Indicators and corresponding abbreviations

Indicator	Abbreviation
Household registration population	Population (Chen and Wu, 2014; Wang and Cindy, 2009)
Gross Regional Product	Gross Product (Li et al., 2014; Ru et al., 2014; Liu et al., 2021)
The added value of the primary industry	Primary Industry (Chen et al., 2013; Jia et al., 2011; Xie and Cai, 2015; Lv et al., 2012)
The added value of the secondary industry	Secondary Industry (Chen et al., 2013; Jia et al., 2011; Xie and Cai, 2015; Lv et al., 2012)
number of beds in medical and health institutions	Beds

3.3 Data Matching

In order to discuss the significance of the social development index predicted by the model, it is necessary to compare the index with existing statistical data, which requires matching remote sensing data with county-level administrative units. Since the coverage area of a single satellite image is much smaller than that of a county-level administrative unit, this paper believes that satellite images and their corresponding nighttime light intensity belong to the county where their center points fall. This paper uses a geographic fence algorithm to match center points with county-level administrative units. The data matching process is shown in Fig. 2.

4 Weakly Supervised Learning Framework

This research intends to directly learn the index representing the degree of social development from remote sensing data in the absence of labeled data. We also try to use easy-to-obtain open-source datasets and pre-trained model

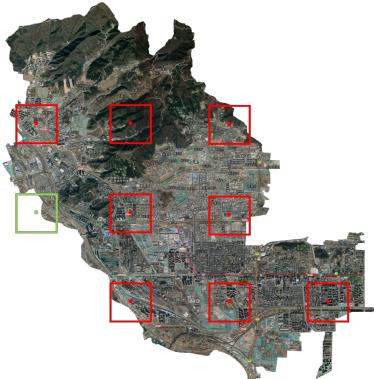


Fig. 2 Map of satellite image matching of county administrative unit. The above is the satellite photo of Shijingshan District <http://www.bigemap.com/source/satel-455.html>, where the red and green boxes represent the acquired satellite images, and the red boxes represent the satellite images belonging to Shijingshan District while green boxes represent not. For the convenience of expression, the area ratio between the satellite image and the administrative unit is enlarged.

parameters in this process so that the method has a strong ability of transference. The core idea of this approach is to extract economic development indicators from remote sensing images, which can be seen as a task of feature extraction from images. Unlike other feature extraction tasks, economic development indicators have a measurable property that reflects the level of economic development. Therefore, a weakly supervised signal is established by utilizing the relative level of economic development among regions for weakly supervised training. Firstly, a supervised signal is established by selecting nighttime light as an indicator to compare the prosperity level between regions, and an ordered graph is output as the supervision signal. Then, supervised training is performed using a metric learning method and a learn-to-rank loss function. Inspired by the methods of Jean et al (2016), Bai et al (2022) and Han et al (2020), this paper will complete the learning of the weakly supervised social development index in four steps:

- A convolutional neural network extracts high-dimensional feature vectors from the corresponding satellite image.
- All high-dimensional feature vectors are divided into several clusters by the clustering algorithm after dimensionality reduction.
- Based on the nighttime light intensity, the clusters are sorted from low to high to form a directed graph, which can be considered to represent the order of social development in the corresponding area of each cluster.
- Use another convolutional neural network to assign a value to the picture according to the sorting information in the directed graph and obtain the social development index of the corresponding area.

The above steps can be described in the following flow chart (Fig. 3).

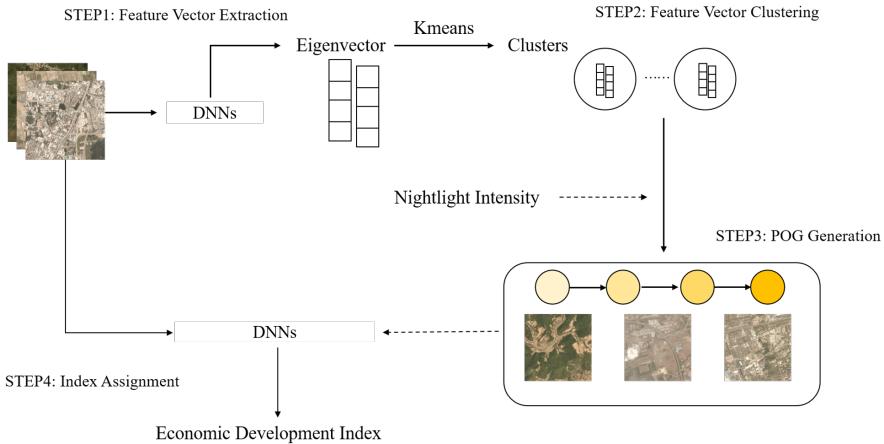


Fig. 3 Weakly supervised social development index learning process. The method consists of four steps: (1) Feature vector extraction; (2) Feature vector clustering; (3) POG generation; (4) Index assignment. Among them, the yellow circles in steps (3) and (4) represent clusters formed by feature vectors (or satellite images corresponding to the vectors), and the darker the color, the higher the degree of social development in the area covered by the graphics in the cluster.

4.1 Extract High-dimensional Feature Vectors

First, this paper hopes to extract high-dimensional feature vectors related to economic activities and social development from remote sensing data. This paper uses the ResNet-18 network structure pre-trained on ImageNet to accomplish this task and adds an attention layer before the last two convolutional layers to improve feature extraction performance. In this paper, the Adam iterative optimization method is used, and the loss function is as follows:

$$L_{\text{extract}} = -\frac{1}{|B|} \sum_{i \in B} \sum_{j \in \{0,1,2\}} y_j^i \log \hat{y}_j^i \quad (2)$$

In the function, B represents the subscript collection of the current batch of pictures, y_j^i is the actual probability that the nighttime light intensity level of the i -th picture x_i is j , and \hat{y}_j^i is the probability that the ResNet-18 network prediction x_i is the corresponding light intensity level j .

After the model is trained to converge, the last fully connected layer is removed, and the high-dimensional feature vector $G \in \mathcal{R}^{1280}$ extracted from the remote sensing data can be obtained.

4.2 Feature Vector Clustering

Next, this paper performs dimensionality reduction and clustering processing on the acquired high-dimensional information. First, retain more than 80% of the original data. For the characteristics of low, medium, and high development levels, this paper uses PCA method to reduce the original 1280-dimensional

vector to 29, 12, and 37 dimensions, respectively. Then, according to the similarity of features, the K-means method is used to hierarchically cluster the features of low, medium and high development levels, and further cluster them into 5, 5, and 5 classes, respectively. In this way, this paper obtains 15 picture clusters divided by all pictures. It can be considered that the picture characteristics in the clusters are relatively similar, and the regions represented by them have relatively close development levels.

Fig. 4 shows example cluster images, composed of 5 urban, 5 rural, and 5 uninhabited cluster. Each cluster exhibits consistent semantic features in its image patterns. Specially, cluster 0 contains a large number of deserts, cluster 7 has many farmlands, and cluster 14 has densely packed houses, representing high economic levels.s.

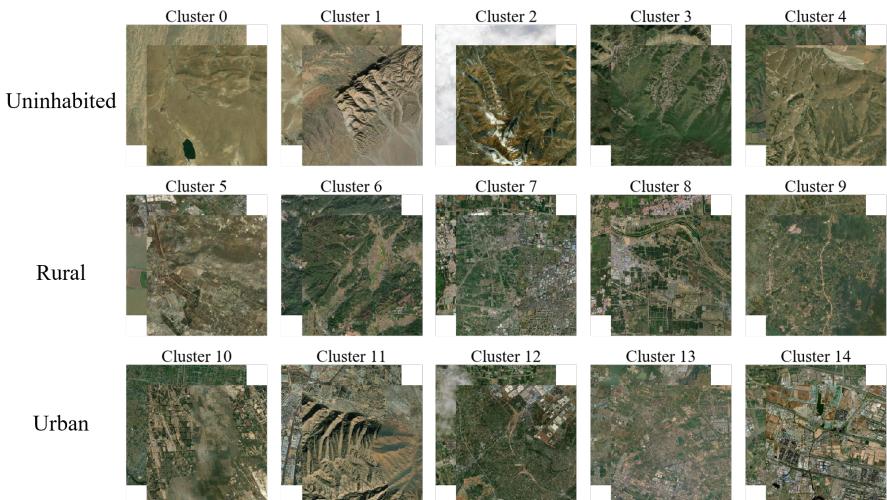


Fig. 4 Satellite Imagery Clustering Results based on Kmeans Algorithm

4.3 POG Generation

It can be considered that the social development levels of the corresponding regions of the pictures in a cluster are similar. This paper hopes to rank the social development levels of the corresponding regions of different clusters and generate a directed graph of the clusters. The basic idea is that if the social development levels of the corresponding regions of the two clusters are significantly different, an edge is generated from a low-level cluster to a high-level cluster.

Since the accurate social development level index is unknown, this paper still uses nighttime light intensity as a proxy variable for the intensity and scale of economic activity. This paper calculates the average of light intensity at night of all pictures in each cluster. A two-sample t-test with a significant

difference of 0.01 is used to judge whether there is a significant difference between the averages of light intensity. If the test is significant, an edge is generated from the low-average cluster to the high-average cluster. Pairwise tests are performed on all clusters, and finally, a directed graph representing the relative level of social development is obtained. The result of POG can be seen in Fig. 5.

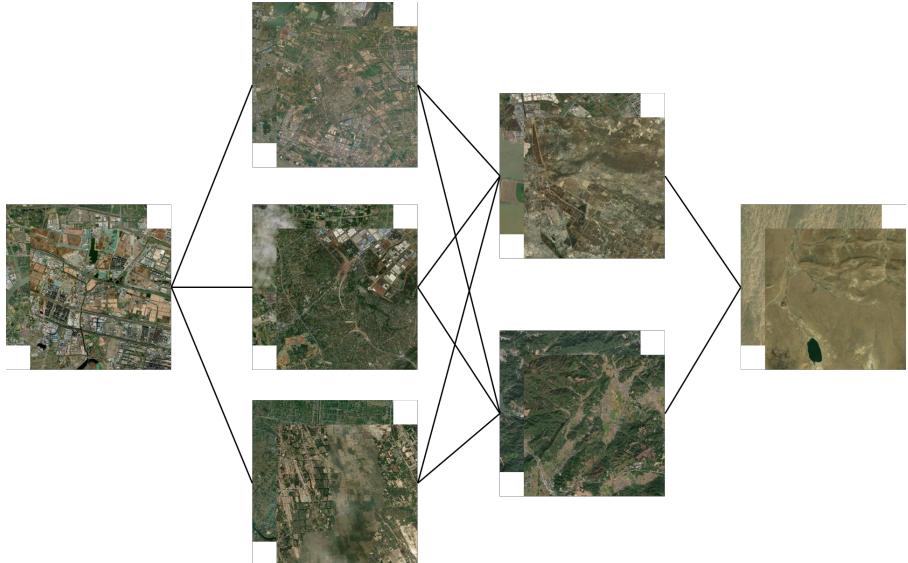


Fig. 5 Example Satellite Imagery Ranking Results based on Partial Order Graphs Algorithm. Clusters are sorted from rural to urban, from left to right.

4.4 Index Assignment

In this paper, deep learning model is used to assign the image value as the social development index of the area covered by the image. The value range of this index is [0, 1]. Since the directed graph generated in section 4.3 contains the relative development level of different regions, the assignment process mainly refers to the sorting information in the directed graph.

For a path $P_j = \{C_{j1}, C_{j2}, \dots, C_{jm}\}$ in a directed graph, we randomly select n_s pictures from each cluster on the path, and assign values to these $n_s \times m$ pictures, and obtain a total of n_s groups of value vectors $s_1^j, s_2^j, \dots, s_{n_s}^j, s_i^j \in R^m$, where:

$$s_i^j = \left(f\left(x_i^{j1}\right), f\left(x_i^{j2}\right), \dots, f\left(x_i^{jm}\right) \right) \quad (3)$$

In the function, x_i^{jk} represents the i th picture extracted from C_{jk} .

Next, we calculate the rank of each element in $s_i^j, 1 \leq i \leq n_s$ within the vector. For a vector $s = (s_1, \dots, s_n)$, the following function is used to record the relative order of any two elements s_w, s_v in the vector:

$$\sigma_\lambda(s_w, s_v) = \frac{1}{1 + s^{-\lambda}(s_v - s_w)}, \lambda > 0 \quad (4)$$

The function has a positive parameter λ , whose value range is $(0, 1)$. When s_w is much larger than s_v , the value of $\sigma_\lambda(s_w, s_v)$ is infinitely close to 0; on the contrary, when s_w is much smaller than s_v , the value of $\sigma_\lambda(s_w, s_v)$ is infinitely close to 1. Compared with direct sorting and recording the ranking of natural numbers, this function can more clearly reflect the difference in the absolute value of each element in the vector. For the value vector s_i^j , the rank of the w -th element in the entire vector is:

$$r(s_i^j, w) = \sum_{v=1, v \neq w}^m \sigma_\lambda(s_w, s_v) \quad (5)$$

Now, we get the ranking vector s_i^j of $rank(s_i^j) = (r(s_i^j, 1), \dots, r(s_i^j, m))$. According to the ordering information of the directed graph, the true ranking vector in P_j should be $r_j = (1, 2, \dots, m)$. Comparing $rank(s_i^j)$ with r_j , the loss function is:

$$L_s = \frac{1}{n_s} \sum_{j=1}^{n_s} \| rank(s_i^j) - r_j \|_2^2 \quad (6)$$

In the function, L_s reflects the ability of the assignment function to distinguish images in different clusters. Still, at the same time, we need to control the scores of images in the same cluster as close as possible. In this paper, the following regular terms are used to control intra-cluster differences:

$$L_{\text{var}} = \frac{1}{|P_j|} \sum_{C_{jk} \in P_j} \text{Var}(f(X_{C_{jk}})) \quad (7)$$

In the function, $X_{C_{jk}}$ represents the current batch of images in C_{jk} , and $\text{Var}(\cdot)$ calculates the variance of the current batch of image values.

Considering the assignment loss function and the regular term, the final loss function in the assignment stage is:

$$L_{\text{score}} = L_s + \alpha L_{\text{var}} \quad (8)$$

In the function, α and λ in the ranking function are hyperparameters, and n_s is the number of training images per batch.

In the following experiments, this paper uses the network pre-trained on labeled satellite images provided in Han et al (2020) to assign image values.

5 EXPERIMENTS

To assess the effectiveness of the model, this section will compare the correlation coefficients (Pearson and Spearman) between the economic index and model output across the countries.

5.1 Method

We compare our methods with other baselines including Autoencoder, PCA, and Supervised training.

- **Ours.** We used the classic ResNet-18 (He et al, 2016) as the network structure, set the clusters size in section 4.3 to 15, training 50 epochs at a learning rate of 1e-4.
- **Autoencoder** (Bourlard and Kamp, 1988) is a neural network architecture that aims to reconstruct its inputs from a reduced representation in the bottleneck layer. In our case, the autoencoder is trained on a large dataset of unlabeled input satellite images to learn the underlying structure of the data.
- **Principal Component Analysis** is a dimension reduction technique that aims to find the most important underlying structure in a dataset. We use PCA to extract k principal components from the satellite imagery, and then use linear regression to fit the nighttime light data. We choose k=10.
- **Supervised learning:** In this method, we use nightlight data as the label and satellite images as the input. The model is trained to learn the relationship between the satellite imagery and the nightlight intensity to perform the task of community detection. ResNet18 is trained for 50 epochs at a learning rate of 1e-4, which is the same as in Ours.

5.2 Economic Index

The economic index we used includes Gross Domestic Product (GDP), primary industry, secondary industry, population, and the number of hospital and clinic beds.

- **GDP** is a measure of a country's overall economic output and is a widely used indicator of economic health.
- **Primary industry** is defined as the sector of the economy that involves the extraction and production of raw materials, such as agriculture and mining.
- **Secondary industry** is the sector of the economy that involves the processing and manufacturing of goods, such as manufacturing and construction.

- **Population** is a key factor in understanding a region's economic potential, as a large population can provide a larger labor force and consumer base.
- **The number of hospital and clinic beds** is also considered as it reflects the availability of medical resources and healthcare facilities in a region.

5.3 Evaluation Matrix

To evaluate the correlation between the model output and economic index, we obtain the Pearson correlation coefficient and Spearman rank correlation coefficient, which are all measures of the relationship between two variables. However, Pearson correlation measures the linear relationship between the two variables and Spearman correlation measures the monotonic relationship.

- **Pearson Correlation Coefficient** is a statistical measure that quantifies the linear relationship between two continuous variables. This measure assumes that the variables are normally distributed and that their relationship is linear. To account for the assumption of normality in the Pearson correlation, the social development index was logarithmically transformed before analysis. The formula for Pearson correlation is:

$$\rho_{X,Y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

- **Spearman Rank Correlation Coefficient** is a nonparametric measure of the association between two variables. Unlike Pearson's correlation coefficient, it is based on the ranks of the data rather than the actual values, and it can handle non-linear relationships and is not affected by outliers. Spearman rank correlation was calculated using the raw data without transformation. The formula for Spearman rank correlation is:

$$\rho_S = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

where d_i is the difference between the rank of X_i and the rank of Y_i .

5.4 Performance Analysis

We present spearman rank corelation coefficient in Table 3. The results show that the ours method outperforms the PCA, AutoEncoder, and Supervised methods in terms of the spearman correlation coefficient, except for the population indicators. Table 4 presents the Pearson correlation coefficients. As various economic development indicators often exhibit highly skewed distributions, taking the logarithm can make the indicators more consistent with the normal distribution assumption, thereby making the correlation coefficients more accurate and reliable. As a result, we have taken the logarithm of various economic development indicators.

It is noteworthy that the Pearson correlation coefficient between nightlights and other economic indicators is low, which lead to poor performance in Supervised learning. However, since our method only uses nightlights to measure the relative level of regional economic development, it still achieved a good Pearson correlation with the economic indicators.

It can be obtained that our method has achieved best results in all indicators, which demonstrates that the economic development index generated by our method is effective and can well measure the relative level of economic development among different regions.

Table 3 Spearman rank corelation coefficient results. The abbreviations used in this figure represent the following economic indices: Gross Domestic Product (GDP), Primary Industry (PI), Secondary Industry (SI), Population (Pop), and Hospital Beds (HB).

	Economic Index	GDP	PI	SI	Pop	HB
Economic Index	GDP	1.	0.818233	0.939339	0.827765	0.844170
	PI	0.818233	1.	0.663937	0.831202	0.796488
	SI	0.939339	0.663937	1.	0.696805	0.719106
	Pop	0.827765	0.831202	0.696805	1.	0.943342
	HB	0.844170	0.796488	0.719106	0.943342	1.
	Nightlights	0.392843	0.414653	0.324825	0.380516	0.371398
Method	PCA	0.382493	0.416577	0.305940	0.463555	0.428463
	AutoEncoder	0.076812	0.063600	0.080957	0.077375	0.069411
	Supervised	0.502993	0.494340	0.451617	0.524664	0.497071
	Ours	0.540359	0.529399	0.475799	0.516072	0.500987

Table 4 Pearson corelation coefficient results. We have taken the logarithm of various economic development indicators. The abbreviations used in this figure represent the following economic indices: Gross Domestic Product (GDP), Primary Industry (PI), Secondary Industry (SI), Population (Pop), and Hospital Beds (HB).

	Economic Index	GDP	PI	SI	Pop	HB
Economic Index	GDP	0.715014	0.467647	0.691549	0.542645	0.573387
	PI	0.679708	0.821352	0.550101	0.699872	0.663789
	SI	0.646074	0.357086	0.670269	0.436029	0.473940
	Pop	0.693514	0.664699	0.612710	0.833260	0.785243
	HB	0.707086	0.614115	0.629403	0.771466	0.829593
	Nightlights	0.189148	0.094344	0.178024	0.151712	0.161095
Method	PCA	0.311397	0.339096	0.251604	0.388912	0.363358
	AutoEncoder	0.071693	0.056554	0.076407	0.069084	0.063187
	Supervised	0.302331	0.204368	0.284477	0.262638	0.265018
	Ours	0.544141	0.514694	0.485231	0.503338	0.499468

5.5 Performance Comparison of Different Network Architectures

In this section, the performance of different network architectures, including ResNet (He et al, 2016), VGG (Simonyan and Zisserman, 2014), and Vision Transformer (Dosovitskiy et al, 2020) are evaluated using our proposed method. The experiments are conducted with the same hyperparameters (i.e., 50 epochs and a learning rate of 1e-4).

Table 5 The performance of different network architectures, evaluated using the same methods as described in the section 5.4.

Metric	network	GDP	PI	SI	Pop	HB
Pearson	ResNet-18	0.544141	0.514694	0.485231	0.503338	0.499468
	ViT	0.590918	0.515851	0.545173	0.499958	0.502207
	VGG-16	0.385825	0.391049	0.336382	0.471334	0.424388
Spearman	ResNet-18	0.540359	0.529399	0.475799	0.516072	0.500987
	ViT	0.549522	0.515488	0.510706	0.477045	0.472732
	VGG-16	0.397163	0.403800	0.333956	0.472933	0.425199

ViT performs better than ResNet-18 and VGG-16 in terms of Pearson correlation coefficient for all the indicators, while ResNet-18 and ViT perform similarly in terms of Spearman correlation coefficient. Moreover, VGG-16 shows the lowest correlation coefficients for all the indicators. While there are slight differences in performance among different network architectures, they are all effective, demonstrating the applicability of our method.

6 DISCUSSIONS

In this section, the significance of the social development index will be explored. Moreover, we apply our method to a case study of regional development imbalances in China.

6.1 Implementation Details

Considering that solid spatial heterogeneity exists in China, we divide China into seven regions: East, Central, South, North, Northeast, Northwest, and Southwest, for the convenience of the research. Table 6 shows the corresponding included provinces in each area. And due to certain reasons, data for Taiwan Province were not available. As for model training, We train all the regions as a whole to gain the relative county-level index across the country.

In generating the directed graph, this paper ensures that the clusters with lower light intensity levels are located up of the clusters with higher light intensity levels. The specific method generates the directed graphs of each light intensity level cluster and then sequentially splice the directed graphs.

Table 6 Corresponding included provinces in each area

Area	Included provinces
East	Shandong, Jiangsu, Zhejiang, Anhui, Jiangxi, Fujian, Shanghai
Central	Henan, Hunan
South	Guangdong, Guangxi Zhuang Autonomous Region, Hainan, Hong Kong , Macao , Taiwan Province(data missing)
North	Shanxi, Hebei, Beijing, Tianjin, Inner Mongolia Autonomous Region
Northeast	Jilin, Heilongjiang, Liaoning
Northwest	Qinghai, Ningxia Hui Autonomous Region, Shaanxi, Gansu, Xinjiang Uygur Autonomous Region
Southwest	Guizhou, Yunnan, Chongqing, Sichuan, Tibet Autonomous Region

6.2 Model Interpretation using SHAP Values

Using SHAP (Shapley Additive Explanations), a technique for analyzing complex models through simpler surrogate models for each test instance (Lundberg and Lee, 2017), we formed feature associations with changes in the social development index in the images. The resulting hotspots represent important features that the model has learned.

In SHAP, red coloring indicates the positive association with social development index, and blue coloring indicates the negative association with social development index, as shown in Fig. 6.

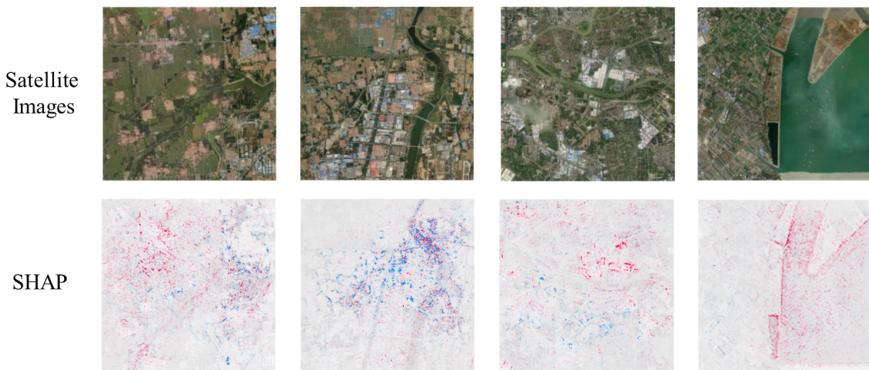


Fig. 6 A set of satellite images and their corresponding SHAP samples. Red coloring indicates the positive association with social development index, and blue coloring indicates the negative association with social development index.

With SHAP, the model was able to associate common features with county-level social development index. We broadly divided the instances SHAP spotted into four main categories:

1. Roads;
2. Buildings;
3. Farmlands;
4. Landscapes.

First, highway, railway, and waterway represented the infrastructures of transportation. Here we typically defined roads as their particular parts across the human settlements. We noticed that the model was the most sensitive to

roads, only presenting differences in reference to the positive-negative direction. For main roads in cities, which could be easily distinguished from the images, they were more likely to be positively associated with social development index. As for lanes in suburbs and villages where there are not so much surroundings, negative associations appear. There was an old saying, "If you want to be rich, you must first build roads." Highway, railway, and waterway provided huge facilitation in transportation for cities, helping turn resources into values. But a village with less productivity was temporarily under-developed, while roads took over money and grounds in practice.

Next, buildings had similar patterns with roads, that they had positive effect in dense area and had negative effect in sparse area on social development index. Besides, the model was sensitive to those buildings with light-colored roofs, which were significant in deep-colored backgrounds. Notably, buildings near by the water, such as docks, were positively associated with social development index. We thought buildings and roads had some mutual reinforcing effects, thus that either of those was sparse would impact the other.

The model was insensitive to farmlands, especially when they were contiguous with landscapes. That means the model could not differentiate green farmland from green landscape, even though the former was distinctly rectangular. We thought the model was more sensitive to color than shape in images, since some light-colored features in landscapes, such as snow and cloud, could be marked by SHAP.

6.3 The Significance of the Social Development Index

To further explore the meaning of the model output results, this section will compare the correlation coefficient between the social development indices of the country and different regions and the statistical indicators that reflect information such as population, production, and medical conditions. Since the labeled dataset is of county-level granularity, this paper takes the average of all grid-grained indices in the same county-level unit as the county's social development index s_c . The correlation coefficients of social development index, and socioeconomic indicators of 5 county-levels in different regions are shown in Fig. 7.

In different regions, the social development index exhibit the highest correlation in the western regions (including the southwest and northwest regions), reaching 0.8, while in other regions, the social development index shows relatively consistent values between 0.3 and 0.4. This consistency is observed in the five economic indicators. One possible reason is that in the west regions, the land is sparsely populated and economic development is more concentrated. Furthermore, when calculated by region, the pearson correlation coefficient does not always outperform the spearman correlation coefficient, indicating that both pearson correlation coefficient and spearman correlation coefficient need to be considered.

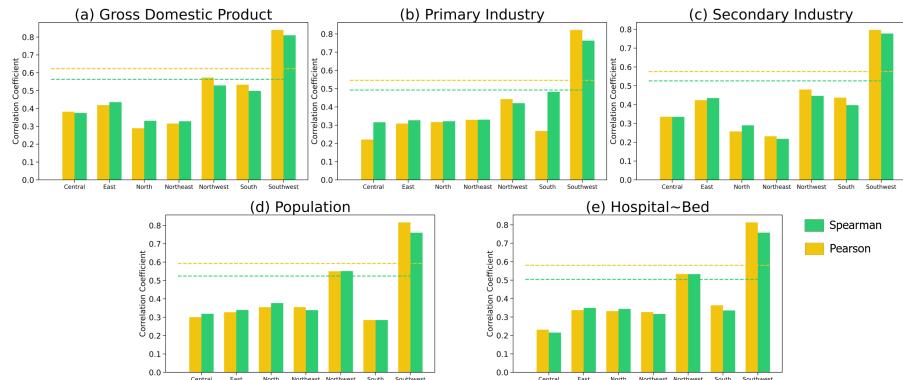


Fig. 7 County-level social development index and various economic indicators, correlation coefficient chart calculated by region. The horizontal line in each subgraph represents the evaluation result of the corresponding value.

6.4 Adjusted Social Development Index

This paper believes that indicators such as population and GDP are affected by the intensity of economic activities and related to the scale of economic activities. The practice of taking the average value of all grid-grained social development indexes in the same county-level unit ignores the relationship between the area of the county-level unit and the scale of economic activity. In this regard, this paper uses the following methods to revise the county-level social development index:

$$s_{c_adj} = \text{Scale}(s_c \times \text{area}) = \text{Scale}\left(\frac{1}{|\mathcal{J}|} \sum_{i \in \mathcal{J}} s_i \times \text{area}\right) \quad (9)$$

In the function, \mathcal{J} refers to the subscript set of all grids in the county-level unit, s_i refers to the grid-grained social development index, area is the county-level unit area, s_c is the county-level social development index calculated by the average method, and s_{c_adj} is the county-level social development index adjusted by area, and $\text{Scale}(\cdot)$ is a linear normalization function. The following is the Spearman correlation coefficient between before and after adjustment of population, regional GDP, the added value of the primary industry, added value of the secondary industry, number of beds in medical institutions and county-level social development index:

As can be seen from Fig. 8, there is a significant increase in the correlation coefficients after adjustment for both Pearson and Spearman coefficients. Specifically, the values increase from around 0.6 to above 0.8 after adjustment for all five economic development indicators.

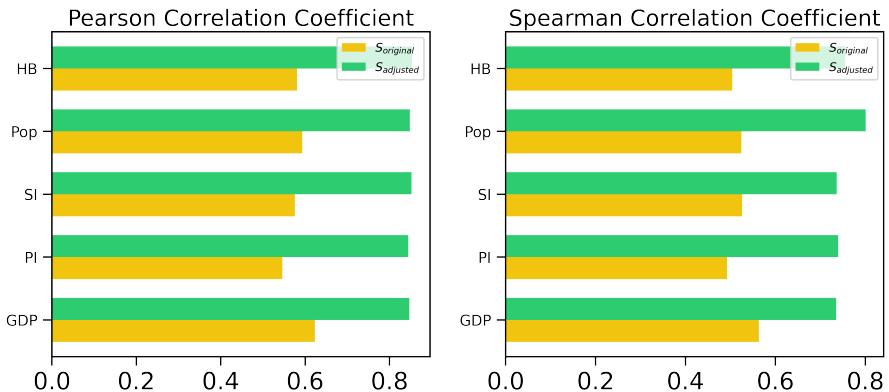


Fig. 8 The correlation coefficient between the county-level social development index and each index before and after adjustment.

7 Conclusion

7.1 Research Results

This paper proposes an weakly supervised method for learning social development index from open-source remote sensing data. The method consists of four steps: (1) Feature vector extraction; (2) Feature vector clustering; (3) POG generation; (4) Index assignment. The method in this paper breaks through the limitation of label datasets and theoretically can estimate the social development index of any granularity and any region. This method performs better than traditional methods and is effective and robust for different network architectures.

In order to further explore the meaning and application scope of the index output by the deep learning model, this paper calculates the county-level social development index for the whole of China and seven regions. After considering the influence of the administrative division area on the scale of economic activities, this paper revised the county-level social development index according to the area of county-level administrative units and found that the revised value is reasonable, which have a huge improvement.

7.2 Reflection and Outlook

In addition to the above results, this paper still has shortcomings. First, the severe lack of satellite images in this paper exists. The actual area coverage is too low to hinder exploring the relationship between the social development index learned from remote sensing images and actual economic activities. Second, this paper mainly uses the 2020 data as an example to explore the meaning and application scope of the social development index without considering the temporal heterogeneity. Next, this paper plans to supplement satellite images, estimate the social development index in a different year, and compare the proxy effect of the index.

Acknowledgments This research was supported by Public Computing Cloud, Renmin University of China.

Declarations

- Funding: This work was jointly supported by National Natural Science Foundation of China (NSFC) under grants 62206301; Major Project of the MOE(China)National Key Research Bases for Humanities and Social Sciences(22JJD910003); fund for building world-class universities (disciplines) of Renmin University of China. Project No.KYGJA2022001; and the Wine Group's research grant No.09202188.
- Conflict of interest/Competing interests: The authors have declared that no conflict of interest/competing interests exist.
- Ethics approval: Not applicable.
- Consent to participate: The authors have agreed to participate.
- Consent for publication: The authors have agreed to publication.
- Data availability: All data in the experiment is authoritative and available. Data are made available at:<https://www.kaggle.com/datasets/hhyyyyy/daytime-image-zl12>. Codes and implementation details are made available at:<https://github.com/xiafire/Economic-Development>.
- Authors' contributions: Zaishuo Xia conducted the experiments, while Shubo Zeng, Bolin Zhang, and Zaishuo Xia wrote the main manuscript text. Zhe Liu contributed to the interpretability section of the model, and Ziyue Zhang prepared the figures. Xuxin Mao provided guidance throughout the project. Yanbing Bai was involved in conceptualization, methodology, resource acquisition, supervision, project administration, and funding acquisition.

Supplementary Material

economicdata.xlsx: Socioeconomic data used in Section 3.2.

train.ipynb: Code for training model by weakly supervised framework.

evaluation.ipynb: Code for evaluate model with correlation coefficient.

score_satellite.py: Python module containing helper functions required for train.ipynb.

References

Ackermann K, Chernikov A, Anantharama N, et al (2020) Object recognition for economic development from daytime satellite imagery. arXiv preprint arXiv:200905455

Alshehhi R, Marpu PR, Woon WL, et al (2017) Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks. ISPRS Journal of Photogrammetry and Remote Sensing 130:139–149

- Bagan H, Yamagata Y (2015) Analysis of urban growth and estimating population density using satellite images of nighttime lights and land-use and population data. *GIScience & Remote Sensing* 52(6):765–780
- Bai Y, Wu Y, Hao Y, et al (2022) Using satellite imagery and deep learning to measure the economic development level. In: ESCoE Conference on Economic Measurement 2022, ESCoE
- Baragwanath K, Goldblatt R, Hanson G, et al (2021) Detecting urban markets with satellite imagery: An application to india. *Journal of Urban Economics* 125:103,173
- Bengio Y (2012) Deep learning of representations for unsupervised and transfer learning. In: Proceedings of ICML workshop on unsupervised and transfer learning, JMLR Workshop and Conference Proceedings, pp 17–36
- Biernacki C, Celeux G, Govaert G (2000) Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE transactions on pattern analysis and machine intelligence* 22(7):719–725
- Bjorck J, Rappazzo BH, Shi Q, et al (2021) Accelerating ecological sciences from above: Spatial contrastive learning for remote sensing. *Proceedings of the AAAI Conference on Artificial Intelligence* 35(17):14,711–14,720. <https://doi.org/10.1609/aaai.v35i17.17728>, URL <https://ojs.aaai.org/index.php/AAAI/article/view/17728>
- Bluhm R, McCord GC (2022) What can we learn from nighttime lights for small geographies? measurement errors and heterogeneous elasticities. *Remote Sensing* 14(5):1190
- Bourlard H, Kamp Y (1988) Auto-association by multilayer perceptrons and singular value decomposition. *Biological cybernetics* 59(4-5):291–294
- Burke M, Driscoll A, Lobell DB, et al (2021) Using satellite imagery to understand and promote sustainable development. *Science* 371(6535):eabe8628
- Buslaev A, Seferbekov S, Iglovikov V, et al (2018) Fully convolutional network for automatic road extraction from satellite imagery. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 207–210
- Charlton M, Fotheringham S, Brunsdon C (2009) Geographically weighted regression. White paper National Centre for Geocomputation National University of Ireland Maynooth 2
- Chen D, Peng B, Xiong J (2013) Study on the coupling characteristics of the ecological economic system around the dongting lake. *Scientia Geographica*

Sinica 33(11):1338–1346

- Chen W, Wu Q (2014) Economic efficiency of urban construction land in the yangtze river delta and its influencing. *Economic Geography* 34(9):142–149
- Doll CN, Muller JP, Morley JG (2006) Mapping regional economic activity from night-time light satellite imagery. *Ecological Economics* 57(1):75–92
- Donaldson D, Storeygard A (2016a) The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives* 30(4):171–98
- Donaldson D, Storeygard A (2016b) The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives* 30(4):171–198
- Dosovitskiy A, Beyer L, Kolesnikov A, et al (2020) An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929
- Elvidge CD, Baugh KE, Kihn EA, et al (1997a) Mapping city lights with night-time data from the dmsp operational linescan system. *Photogrammetric Engineering and Remote Sensing* 63(6):727–734
- Elvidge CD, Baugh KE, Kihn EA, et al (1997b) Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing* 18(6):1373–1379
- Elvidge CD, Baugh KE, Dietz JB, et al (1999) Radiance calibration of dmsp-ols low-light imaging data of human settlements. *Remote Sensing of Environment* 68(1):77–88
- Elvidge CD, Imhoff ML, Baugh KE, et al (2001) Night-time lights of the world: 1994–1995. *ISPRS Journal of Photogrammetry and Remote Sensing* 56(2):81–99
- Engstrom R, Hersh JS, Newhouse DL (2017) Poverty from space: using high-resolution satellite imagery for estimating economic well-being. *World Bank Policy Research Working Paper* (8284)
- Gao M, Fu H (2014) An overall analysis of china's current gdp accounting system. *Economic Theory and Business Management* 9:5–14
- Ghosh T, Anderson SJ, Elvidge CD, et al (2013) Using nighttime satellite imagery as a proxy measure of human well-being. *Sustainability* 5(12):4988–5019
- Han S, Ahn D, Park S, et al (2020) Learning to score economic development from satellite imagery. In: *Proceedings of the 26th ACM SIGKDD*

- International Conference on Knowledge Discovery & Data Mining, pp 2970–2979
- Haut JM, Paoletti ME, Moreno-Álvarez S, et al (2021) Distributed deep learning for remote sensing data interpretation. *Proceedings of the IEEE* 109(8):1320–1349
- He K, Zhang X, Ren S, et al (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 770–778
- Jean N, Burke M, Xie M, et al (2016) Combining satellite imagery and machine learning to predict poverty. *Science* 353(6301):790–794
- Jia J, Guo Q, Ning J (2011) Fiscal decentralization, government governance structure and county-level financial difficulties. *Management World* 1:30–39
- Li P, Chen W, Sun W (2014) Spatial differentiation and influencing factors of multi-functionality in rural areas in economically developed areas: Taking jiangsu province as an example. *Acta Geographica Sinica* 69(6):797–807
- Li Y, Cheng G, Yang J, et al (2020) A fine simulation of regional economic spatial pattern based on noctilucent remote sensing data: A case study of henan province. *Regional Research and Development* 39(4):41–47
- Liao X (2014) Reflections on county-level gdp accounting. *Modern Economic Information* 10:463+465
- Liu H, He X, Bai Y, et al (2021) Nightlight as a proxy of economic indicators: Fine-grained gdp inference around chinese mainland via attention-augmented cnn from daytime satellite imagery. *Remote Sensing* 13(11):2067
- Lundberg SM, Lee S (2017) A unified approach to interpreting model predictions. CoRR abs/1705.07874. URL <http://arxiv.org/abs/1705.07874>, <https://arxiv.org/abs/1705.07874>
- Lv Y, Zhang L, Wang Z (2012) Analysis of regional differences in china's rural financial development. *Statistics and Decision* 19:111–115
- Mellander C, Lobo J, Stolarick K, et al (2015) Night-time light data: A good proxy measure for economic activity? *PloS one* 10(10):e0139,779
- Minetto R, Segundo MP, Rotich G, et al (2020) Measuring human and economic activity from satellite imagery to support city-scale decision-making during covid-19 pandemic. *IEEE Transactions on Big Data* 7(1):56–68
- Pandey S, Agarwal T, C. Krishnan N (2018) Multi-task deep learning for predicting poverty from satellite images. *Proceedings of the AAAI Conference*

- on Artificial Intelligence 32(1). <https://doi.org/10.1609/aaai.v32i1.11416>, URL <https://ojs.aaai.org/index.php/AAAI/article/view/11416>
- Ru L, Miao C, Wang H (2014) Research on the level and spatial pattern of financial agglomeration in my country's central cities. *Economic Geography* 34(2):58–66
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556
- Sutton PC, Costanza R (2002) Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation. *Ecological economics* 41(3):509–527
- Tobler WR (1969) Satellite confirmation of settlement size coefficients. *Area* 1(3):30–34
- Wang H, Cindy F (2009) Analysis of inter-provincial population migration and reasons. *Population Journal* 5:50–53
- Wang J, Kuffer M, Pfeffer K (2019) The role of spatial heterogeneity in detecting urban slums. *Computers, Environment and Urban Systems* 73:95–107. <https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2018.08.007>, URL <https://www.sciencedirect.com/science/article/pii/S0198971518301947>
- Welch R (1980) Monitoring urban population and energy utilization patterns from satellite data. *Remote sensing of Environment* 9(1):1–9
- Wu Z, Ru Z (2019) Make full use of night light data to analyze the reliability of regional economic development. *Marketing Industry* 30:99–100
- Xie S, Cai H (2015) Research on the coupling coordination degree between logistics industry and regional economy in the yangtze river delta. *Journal of Jiangxi University of Finance and Economics* 5:20–27
- Yeh C, Perez A, Driscoll A, et al (2020a) Using publicly available satellite imagery and deep learning to understand economic well-being in africa. *Nature communications* 11(1):1–11
- Yeh C, Perez A, Driscoll A, et al (2020b) Using publicly available satellite imagery and deep learning to understand economic well-being in africa. *Nature communications* 11(1):2583
- Yuan Q, Shen H, Li T, et al (2020) Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment* 241:111,716

26 *WS-DL for FG-SEDI Inference Based on Satellite Imagery*

Zhang M (2015) County-level gdp accounting work. *Statistics and Management* 9:6–7

Zhao K, Kang J, Jung J, et al (2018) Building extraction from satellite images using mask r-cnn with building boundary regularization. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 247–251

Zhou ZH (2017) A brief introduction to weakly supervised learning. *National Science Review* 5(1):44–53. <https://doi.org/10.1093/nsr/nwx106>, URL <https://doi.org/10.1093/nsr/nwx106>