DOI: 10.1111/jiec.13356

RESEARCH ARTICLE

JOURNAL OF INDUSTRIAL ECOLOGY WILEY

High-resolution quantification of building stock using multi-source remote sensing imagery and deep learning

Yi Bao[1,2] | Zhou Huang[1,2] | Han Wang[1,2] | Ganmin Yin[1,2] | Xiao Zhou[1,2] | Yong Gao[1,2]

[1]Institute of Remote Sensing and Geographical Information Systems, School of Earth and Space Sciences, Peking University, Beijing, China

[2]Beijing Key Lab of Spatial Information Integration & Its Applications, Peking University, Beijing, China

Correspondence
Zhou Huang, Room 413, Remote Sensing Building, Institute of Remote Sensing and Geographic Information Systems, Peking University, Beijing 100871, China.
Email: huangzhou@pku.edu.cn

Editor Managing Review: Gang Liu

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 42271471, 42201454, 41971331, 41830645; China Postdoctoral Science Foundation, Grant/Award Number: 2022M710193

Abstract

In recent decades, urbanization has led to an increase in building material stock. The high-resolution quantification of building stock is essential to understand the spatial concentration of materials, urban mining potential, and sustainable urban development. Current approaches rely excessively on statistics or survey data, both of which are unavailable for most cities, particularly in underdeveloped areas. This study proposes an end-to-end deep-learning model based on multi-source remote sensing data, enabling the reliable estimation of building stock. Ground-detail features extracted from optical remote sensing (ORS) and spatiotemporal features extracted from nighttime light (NTL) data are fused and incorporated into the model to improve accuracy. We also compare the performance of our feature-fusion model with that of an ORS-only regression model and traditional NTL regression for Beijing. The proposed model yields the best building-stock estimation, with a Spearman's rank correlation coefficient of 0.69, weighted root mean square error of 0.58, and total error in the test set below 14%. Using gradient-weighted class activation mapping, we further investigate the relationship between ORS features and building-stock estimation. Our model exhibits reliable predictive capability and illustrates the tremendous value of the physical environment for estimating building stock. This research illustrates the significant potential of ORS and deep learning for stock estimation. Large-scale, long-term building-stock investigations could also benefit from the end-to-end predictability and the data availability of the model.

KEYWORDS
building stock, deep learning, high-resolution quantification, industrial ecology, multi-source remote sensing imagery

1 | INTRODUCTION

The manufactured material stock that forms the built environment (i.e., buildings and infrastructure) is essential for shaping the use of materials and energy (Krausmann et al., 2017; Lanau et al., 2019; Mao et al., 2020). As urbanization has accelerated, global socioeconomic material stocks have increased 23-fold over the past century. More than half of the resources are used for urban expansion and infrastructure renewal, while the recycling rate is only 12% (Krausmann et al., 2017). As an essential foundation for urban development, buildings provide crucial services, such as housing and transportation, to the society (Han et al., 2018; Lanau et al., 2019; Mao et al., 2020). However, the construction, maintenance, and end-of-life management of material stocks are major sources of environmental impact (Mao et al., 2020; Peled & Fishman, 2021). The primary factor

driving the demand for bulk materials is the construction of buildings (Deetman et al., 2020). Simultaneously, the building sector is responsible for 36% of the greenhouse gas emissions (Langevin et al., 2019) and 31% of the energy consumption (Li et al., 2019). Therefore, estimating the material stocks accumulated in a building will improve the management of urban resources and promote a circular economy.

Quantifying in-use material stocks accumulated in buildings is the cornerstone of the socioeconomic metabolism, urban mining, and urban sustainability research (Aldebei & Dombi, 2021; Han et al., 2018; Peled & Fishman, 2021; Bao et al., 2023). A growing body of literature recognizes the importance of material stocks, and studies have proposed various estimation methods (Han et al., 2018; Liang et al., 2014; Mao et al., 2020; Müller, 2006). However, most methods (e.g., "top-down" and "bottom-up" methods) rely excessively on statistical and survey data, making it extremely difficult to achieve a balance between large scale and high spatial resolution.

Nonetheless, with the development of the internet and satellites, data sources, diversity, and accuracy have advanced and evolved, dramatically changing how people acquire data (Runting et al., 2020). Compared to traditional statistics, satellite data have wider coverage, faster updates, and higher resolution (Ferreira et al., 2020). Nighttime light (NTL), as a representative type of remote sensing data, is increasingly used to calculate socioeconomic indicators such as population (Yu et al., 2019), gross domestic product (Han et al., 2022), energy consumption (Xiao et al., 2018), and CO_2 emissions (Chen et al., 2020). In recent years, many studies have revealed a strong correlation between NTL radiation intensity and material stocks on a city or national scale (Hattori et al., 2014; Liang et al., 2014; Yu et al., 2018). Another approach is calculating the volume of a building using optical remote sensing (ORS) (Haberl et al., 2021) or NTL data (Peled & Fishman, 2021) and then estimating building stock by collecting material composition indicators (MCIs) for different types of buildings. Despite the increasing research emphasis on remote sensing data applications for stock estimation, their potential is still underexploited. For example, the textural features of ORS can reflect ground conditions and effectively identify buildings (Cao & Huang, 2021; Shao et al., 2020; Wang et al., 2021) and roads (Das et al., 2011; Lv et al., 2017). However, most extant studies ignore the critical role of ORS characteristics during stock estimation.

This study proposes an end-to-end high-resolution stock estimation model based on deep learning and multi-source remote sensing imagery to address the limitations of previous research. The proposed model can perform fast grid-level estimations based on publicly available remote sensing datasets. We extract the ORS features using a deep-learning model and combine them with the spatiotemporal features of NTL data to obtain fused features of multi-source remote sensing. These fused features are then used as the input to machine learning to yield a grid-scale building stock. We validate the proposed model using a Beijing dataset (Mao et al., 2020); the results reveal significant improvement in accuracy compared with using only NTL regression (Yu et al., 2018). Given the wide availability and accessibility of the data used, the model can be extended to new regions and benefit a wider area than the proposed location. In addition, it is possible to understand the factors that affect the accuracy of building-stock estimation using visualization deep-learning techniques.

2 | RELATED WORKS

2.1 | Methods for estimating building material stock

Stock estimation methods are generally categorized into the top-down approach, bottom-up approach, and remote sensing (Lanau et al., 2019). Each model has advantages and limitations, depending on its application and data availability. The top-down approach is based on the mass-balance principle, which estimates the net increase in stock from the difference between inflows and outflows of the material within the system (Müller et al., 2014). This approach is suitable for obtaining an overview of national-scale material stock over long periods (Hu et al., 2010; Huang et al., 2013; Müller et al., 2010; Sartori et al., 2008). However, the unavailability of statistical data makes it impossible to calculate high spatial resolution material stock within cities.

By contrast, the bottom-up approach enables fine-grained stock estimation by collecting cadastral-level building and infrastructure physical measurements, and the corresponding MCIs. However, as the bottom-up approach is labor-intensive (Lanau et al., 2019), the scope of existing studies is typically limited to the city level or smaller geographical scales (Gontia et al., 2018; Guo et al., 2020; Oezdemir et al., 2017; Tanikawa & Hashimoto, 2009) and a specific year (Mao et al., 2020). Simultaneously, due to the vast amount of the data required, most studies have focused on developed cities with detailed statistics, such as Vienna (Kleemann et al., 2017) and Philadelphia (Marcellus-Zamora et al., 2016).

Remote sensing is an excellent complement to statistics and can be applied in regions where statistical data are lacking. Previous studies have established linear regression relationships between NTL radiation intensity and building material stocks in the study area (Hattori et al., 2014; Huang et al., 2013; Takahashi et al., 2010). An NTL-based approach can estimate material stock on a large scale and over a long period at a low cost. However, stock estimates derived from NTL ignore non-illuminated buildings or misinterpret illuminated structures (Haberl et al., 2021). Background noise, radiation saturation, and other defects in NTL data also reduce the accuracy of the quantitative analyses (Levin et al., 2020). Meanwhile, due to scale effects, the accuracy of high-resolution stock estimation within cities can also be affected by directly applying a regression relationship derived at the national/city level (Yu et al., 2018, 2019). ORS and radar have higher resolutions compared to NTL and have recently been applied to high-resolution stock estimates (Haberl et al., 2021; Schandl et al., 2020). However, most studies only provide indirect estimates of

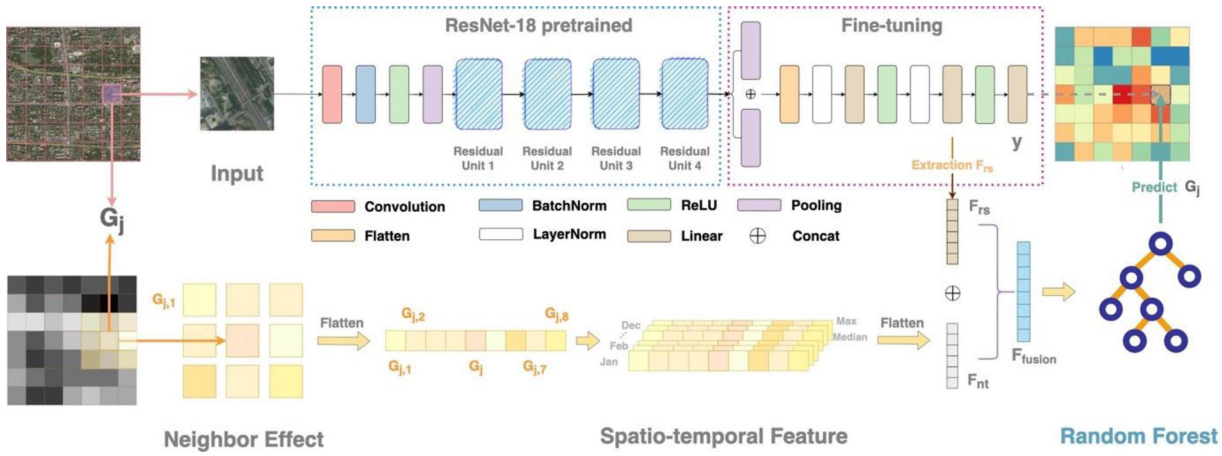

FIGURE 1 The framework of building-stock estimation through multi-source remote sensing and machine learning.

building stock using data on building volume and collected MCIs (Haberl et al., 2021; Peled & Fishman, 2021). Indirect methods require statistical data and introduce more uncertainty compared to end-to-end estimation methods.

2.2 | High-resolution socioeconomic estimation through deep learning

Since the emergence of deep learning, remote sensing feature extraction methods have become well established and are commonly used to predict socioeconomic variables (Abitbol & Karsai, 2020; Jean et al., 2016; Yeh et al., 2020; Huang et al., 2020; Feng et al., 2021). Xing et al. (2020) construct an end-to-end deep-learning model to estimate high-resolution human activity volumes using ORS images and the ResNet model (Neighbor-ResNet) with spatial neighbor augmentation. In addition, the combination of NTL (nighttime features) and ORS (daytime features) provides a better representation of socioeconomic development compared to a single remote sensing data source, being validated for high-resolution poverty prediction in Africa (Jean et al., 2016; Yeh et al., 2020). Previous research has provided a solid foundation for end-to-end building-stock estimation. Buildings have unique ground indicators on ORS images, and deep learning can be used to extract features and establish their mapping to material stock. Data-driven approaches combined with remote sensing can more easily be applied to a more extensive study area compared to traditional methods. Automatic feature-selection based on deep-learning methods minimizes human labor and establishes end-to-end mapping from remote sensing images to building stock.

3 | METHODS AND DATA

3.1 | Model framework and methods

The overall structure of the end-to-end building-stock estimation is shown in Figure 1. The model consists of four stages: input unit unification, feature extraction, feature fusion, and model prediction. (1) We divide the ORS images according to the grid size of NTL data. The first stage ensures spatial cell uniformity of the input multi-source remote sensing images and labels. (2) The model operates in two parallel branches for the feature extraction stage: ORS and NTL. For the ORS branch, the deep residual network (ResNet) is utilized as the fundamental model to extract the high-level features (denoted as F_{rs}) of ORS images. For the NTL branch, we construct the spatial-temporal neighborhood features (denoted as F_{nt}) of NTL images. (3) In the feature-fusion stage, we concatenate the ORS feature F_{rs} and NTL feature F_{nt} as fusion feature F_{fusion}. (4) Finally, the fusion feature F_{fusion} is fed into the random-forest model to predict the building stock of the corresponding grid.

3.1.1 | Optical remote sensing feature extraction

With the development of deep learning, convolutional neural networks (CNN) are steadily displacing conventional techniques and are increasingly used for feature extraction in high-resolution remote sensing (Ball et al., 2017; Yuan et al., 2020; Ye et al., 2022). ResNet is a novel CNN-based model that utilizes residual blocks to increase the depth of traditional CNN and avoid the degradation problem of deep networks (He et al., 2016).

A substantial body of research has demonstrated that ResNet can effectively extract remote sensing features (Wang et al., 2019; Xu & Chen, 2019; Zhu et al., 2021).

Training a new deep convolutional network from scratch requires a large amount of sample data to achieve higher accuracy. To reduce training costs, the feature extraction by deep learning in this study is divided into two main steps: pre-training and fine-tuning, as detailed in Supporting Information Figure S1. In the first stage, we initialize ResNet-18 with pre-trained weights on ImageNet (Deng et al., 2009). This initialization strategy promotes fast convergence, improves accuracy, and reduces training time (Sun et al., 2022; Xing et al., 2020). Although the pre-trained model can extract high-level features from the input images, the model needs to be fine-tuned for specific prediction tasks using the grid building-stock dataset that we created. In addition, layer normalization is used in the fine-tuning step to improve the training speed and accuracy of the model. The final 256-dimensional fully connected layer represents the output of the model and the feature F_{rs} of the input ORS imagery.

3.1.2 | Nighttime light feature extraction

Previous studies have not fully explored the spatiotemporal neighborhood characteristics of NTL. Seasonal variations considerably affect NTL radiation intensity in the same region. Compared with the single-band NTL image, the NTL time series contain richer information and imply building functions and human activity intensities more accurately (Stathakis & Baltas, 2018; Yuan et al., 2019; Zheng et al., 2021). These indicators are critical for predicting the building material stocks (Mao et al., 2020), so the NTL time series would be a better choice for feature extraction. Meanwhile, the features expressed in a single NTL radiation intensity value are limited because of saturation and blooming effects (Levin et al., 2020; Peled & Fishman, 2021). For most geographical phenomena, neighboring characteristics are essential for determining central targets (Xing et al., 2020; Yin et al., 2022). Therefore, we build the spatiotemporal features of NTL data using the following steps:

The NTL radiation intensity for grid j in the tth month is G_j^t. The radiation intensities of the eight grids adjacent to grid j are $\{G_{j,1}^t, G_{j,2}^t, \ldots G_{j,8}^t\}$, which represent the neighbor features of grid j in the tth month. In addition, we calculate the median and maximum values for each grid and its neighbors over 12 months, denoted by $\{G_j^{median}, G_{j,1}^{median}, G_{j,2}^{median}, \ldots G_{j,8}^{median}\}$ and $\{G_j^{max}, G_{j,1}^{max}, G_{j,2}^{max}, \ldots G_{j,8}^{max}\}$, respectively. For missing values in the neighbors, we use 0 for padding. The acquired 126-dimensional feature F_{nt} is then used as the NTL spatiotemporal feature of grid j.

3.1.3 | Feature fusion and model prediction

NTL and ORS data reflect nighttime and daytime features, respectively, and integrating these features often provides a complete representation of surface conditions (Jean et al., 2016; Yeh et al., 2020). Fusion feature F_{fusion}, which contains both nighttime and daytime features, is used as the random-forest regression input to estimate the grid-scale building stock. The random forest is a widely used machine learning model that adopts ensemble learning to achieve high prediction accuracy and good model robustness (Ho, 1998). In addition, it can evaluate the importance of input features, investigate the impact of various features on stock prediction, and offer guidance for the model's feature selection.

3.2 | Model evaluation

We evaluate prediction accuracy using Spearman's rank correlation coefficient (ρ_s), root mean square error (RMSE), and mean absolute percentage error (MAPE), which represent the rank-fitting performance, absolute error, and percentage error, respectively. Due to spatial heterogeneity, the distribution of grid-scale stock values ranges from hundreds to millions of (metric) tons. We logarithmically transform the raw stock values when assessing the RMSE performance. In addition, considering the heavy-tailed distribution of the data, a small number of outliers could have a marked impact on the RMSE and MAPE results. We thus assign weights to different data according to the distribution frequency of the original data and calculate the weighted RMSE (denoted by wRMSE) and weighted MAPE (indicated by wMAPE). The formulas for calculating these three indicators are as follows:

$$\rho_s = 1 - \frac{6\sum d_i^2}{n\left(n^2-1\right)}, \tag{1}$$

$$\mathrm{wRMSE}(y,\hat{y}) = \sqrt{\frac{1}{n}\sum_{i=0}^{n} w_i (y_i - \hat{y}_i)^2}, \tag{2}$$

$$\mathrm{wMAPE}(y,\hat{y}) = \frac{1}{n}\sum_{i=0}^{n} w_i \frac{|y_i - \hat{y}_i|}{y_i}. \tag{3}$$

In Equation (1), d_i denotes the rank difference between the actual and predicted values and n indicates the number of observations. In Equations (2) and (3), y_i represents the ith actual value, $\hat{y}_i$ refers to the predicted value of y_i, w_i is the corresponding weight of y_i, and $\sum w_i = 1$.

3.3 | Model comparison

Previous studies have demonstrated a strong linear correlation between the radiation intensity indicated by NTL data and material stock at large spatial scales, such as at the city or country level (Hattori et al., 2014; Liang et al., 2014; Peled & Fishman, 2021; Takahashi et al., 2010; Vilaysouk et al., 2021). City-scale regression models are also often applied to high-resolution grids within cities (Peled & Fishman, 2021; Yu et al., 2018, 2019). The detailed process of the algorithm is as follows.

Consider n cities whose total building stock and NTL radiation intensity are $[S_1, S_2, \ldots S_n]$ and $[N_1, N_2, \ldots N_n]$, respectively. The linear regression equation for S_i and N_i is as follows:

$$S_i = a{*}N_i + b. \tag{4}$$

Parameters a and b can be obtained by a least squares regression, which is then applied to the relationship between NTL and stock at the grid scale within cities. For a city containing m grids, the NTL radiation intensity of each grid is denoted as $\{G_1, G_2, \ldots G_m\}$. Therefore, the stock corresponding to grid j is expressed as follows (assuming each grid is subject to the same bias):

$$y_{ntl}\ (j) = a{*}G_j + \frac{b}{m}. \tag{5}$$

y_{ntl} denotes the building stock of grid j obtained by linear regression. The results of building stock y_{ntl} obtained by the NTL regression are the baseline for comparison with our proposed model.

3.4 | Model interpretation

Although random forests can capture the importance of input features to some extent, it is not possible to locate the mapping of these features within the original ORS imagery using F_{rs} alone because of the black-box nature of deep learning. We also cannot precisely understand which surface conditions in ORS imagery will significantly affect building-stock estimation.

To address this issue, we deploy gradient-weighted class activation mapping (Grad-CAM) (Selvaraju et al., 2017) to explore the visual features of the model. This approach can determine the feature extraction pattern of CNN-based models and has been widely used for various purposes, such as Alzheimer's disease prediction (Tang et al., 2019), plant disease classification (Cheng et al., 2022), and remote sensing image interpretation (Abitbol & Karsai, 2020; Xing et al., 2020).

Crad-CAM follows the gradient flow of CNN, back-propagating to the convolutional layer and generating an activation map. The places that contribute most to the predicted values are represented by a heat map on the activation map. We utilize the random-forest method to rank the importance of $\boldsymbol{F}_{rs}$; one out of 256-dimensional features that contribute most to the prediction is used to generate an activation map. The overlay of the generated activation map and the original ORS imagery enable visualization of the features of the ground indicators that contribute most to the building-stock prediction. The results enable us to understand how high-resolution remote sensing features enable prediction of building stock and what characteristics the model emphasizes.

3.5 | Study area and data description

Beijing is selected as the study area for this research. Beijing, which is the capital of China, is one of the largest cities worldwide, with a population of 21.7 million and over 3621 megatons (140 tons/cap) of construction materials incorporated within its built environment (Mao et al., 2020). Although Beijing has a total area of over 16,000 km^2, most buildings are concentrated within the 5th Ring Road, which covers an area of approximately 100 km^2. We thus investigate the building stock at high resolution in Beijing within the 5th Ring Road to explore the synergistic use of NTL and ORS data for building-stock estimation. The detailed spatial distribution of the study area is shown in Figure S2.

Table 1 lists the four primary data sources and their applications. ORS images are obtained using the open Application Programming Interface (API) of Google Maps. We acquire 1.2 m resolution ORS images of Beijing in 2018, containing three multispectral bands (red, green, and blue; RGB). True color ORS images with RGB channels can accurately reflect the detailed living environment and are consistent with human visual perception of the natural world (Xing et al., 2020). NTL data are gathered from the Suomi National Polar-Orbiting Partnership's Visible Infrared Imaging

TABLE 1 Data descriptions, applications, and sources in this research

Data description	Application	Sources
1.2 m resolution satellite imagery of Beijing	To obtain features of ORS imagery	Google Maps
NTL data of 50 Chinese cities in 2018	To obtain NTL features in the study area and establish a linear relationship between building stock and NTL	VIIRS-DNB
Detailed building stock in Beijing	To train and verify the proposed model	Mao et al. (2020)
Total building stocks in 50 Chinese cities	To establish a linear relationship between building stock and NTL	Figshare
Beijing land use dataset	To explore the relationship between land use and prediction error	Gong et al. (2020)

Abbreviations: NTL,: nighttime light; ORS, optical remote sensing; VIIRS-DNB, Visible Infrared Imaging Radiometer Suite Day/Night Band.

TABLE 2 Model comparison

Model	ρ_s	RMSE	wRMSE	MAPE	wMAPE	Total error (%)
NTL regression	0.432	1.111	0.840	2.708	0.656	44.1
ORS regression	0.681	0.856	0.588	1.362	0.430	16.0
Proposed model	**0.688**	**0.828**	**0.578**	**1.227**	**0.414**	**13.4**

Abbreviations: MAPE, mean absolute percentage error; NTL, nighttime light; ORS, optical remote sensing; RMSE, root mean square error; wMAPE, weighted MAPE; wRMSE, weighted RMSE.

Radiometer Suite Day/Night Band, which has a spatial resolution of 15 arcsec (approximately 350 m in Beijing). All data are preprocessed by the National Oceanic and Atmospheric Administration to remove outliers and background lighting, resulting in a monthly average NTL for 50 Chinese cities in 2018. The total building stock in the 50 Chinese cities is obtained from the data released on Figshare (Bao, 2022). We establish a linear regression model using the building stock and NTL radiation intensity of these 50 cities for comparison with the proposed model. Finally, we utilize the results of the high-resolution building stock within the 5th Ring Road of Beijing (Mao et al., 2020) as the model's labeling for training and validation.

3.6 | Data preprocessing

We vectorize the NTL raster data to obtain a total of 4179 grids within the study area. Then, the 4179 grids are divided into the training and test sets in a ratio of 8:2. Given the inconsistent resolution of the NTL and ORS datasets, we use a vectorized grid as the fundamental research unit. The ORS imagery, land use datasets, and labels are cropped to the same size as NTL data to unify the research unit. In addition, we preprocess the ORS input images using random flips, rotations, shifts, and normalization during training to improve the generalization ability of the model.

4 | RESULTS AND DISCUSSION

4.1 | Model performance

Using the data and methodology mentioned above, we implement the model on an Ubuntu platform utilizing four GeForce RTX TITAN GPUs and the PyTorch framework. We initialize the model using ResNet18 pre-trained with ImageNet by the torchvision library. *MSELoss and Adam* are chosen as the loss function and the optimizer for backpropagation learning and updating weights. The hyperparameters are tuned empirically based on the model's performance in the test sets in the study area. The learning rate and batch size in this experiment are 0.001 and 64, respectively.

4.1.1 | Overall performance

We compare the performance of the NTL linear regression, ORS deep-learning regression, and proposed hybrid model. A comparison of the results of the three models for each indicator is presented in Table 2. The NTL has an excellent linear relationship with building stock at the urban scale (see Figure S3 for details). However, the NTL linear regression approach yields the worst results when the urban-scale parameters are applied to a high-resolution grid within the city. This finding illustrates that urban-scale distribution patterns cannot be directly extended to the inner-city

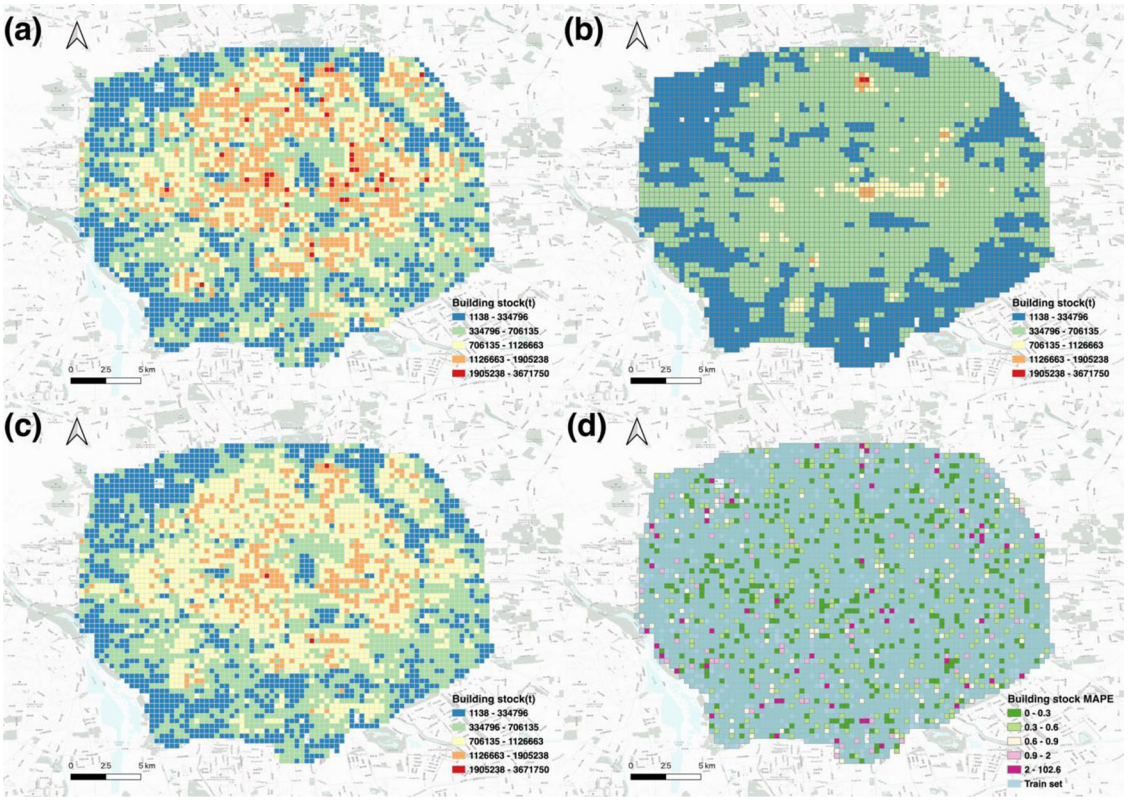

FIGURE 2 Prediction results at the grid scale. (a) The reference building stock. (b,c) The prediction results of the nighttime light regression and the proposed model, respectively. (d) The spatial distribution of the mean absolute percentage error for the proposed model. MAPE, mean absolute percentage error.

high-resolution grid because of scale effects and nonlinear relationships. The results estimated by the ORS regression are better than those obtained by the NTL regression, indicating that the features extracted by deep learning can better reflect building stock. However, ORS alone can capture only daytime features, and the role of nighttime lighting as a good proxy for building material stock should not be ignored. Our proposed model combines the advantages of NTL and ORS data to achieve the best prediction results.

With the multi-source remote sensing image features extracted by deep learning, we estimate the building stock at high resolution with relatively good accuracy without using any statistical or survey data. The wMAPE result for the test set is 41.4%, and the wRMSE is 0.578, which is the best performance among the three comparison models. The Spearman correlation coefficient, ρ_s, is 0.688, revealing a significant rank correlation between the predicted and observed values. In addition, we estimate an overall error of 13.4% for the test set, which is significantly lower than the 44% for the NTL regression.

4.1.2 | Spatial distribution of prediction

We plot the reference and prediction building stock onto a map to investigate the spatial distribution of errors and better understand the characteristics and performance of the proposed model. Figure 2 shows that the proposed model performs significantly better compared to the NTL regression; the MAPE of the former is less than 0.3 in most areas. By comparing Figure 2a,c, we note that our proposed model remains deficient at predicting extreme values, although the predictions are accurate in most cases. Simultaneously, since the land type where the buildings are located can substantially affect building-stock distribution, we further explore the relationship among land use type, building stock, and prediction MAPE.

The MAPE for different land use types also varies greatly. Figure S4 shows the distribution of MAPE box plots for 11 land use types in the test set. Figure 3 illustrates the distribution of the reference and predicted building stock, with larger points denoting larger MAPE and the color of the scatter representing the various land use types. The model's prediction error is significantly larger for the "Park" type than for the other land use types, probably due to the overestimation of many historical buildings in Beijing. Most grids' MAPE prediction for residential buildings is small, but the model still overestimates the residential building stock in a few areas. This may be due to the different ORS texture characteristics of residential buildings in different regions (e.g., city center and city periphery).

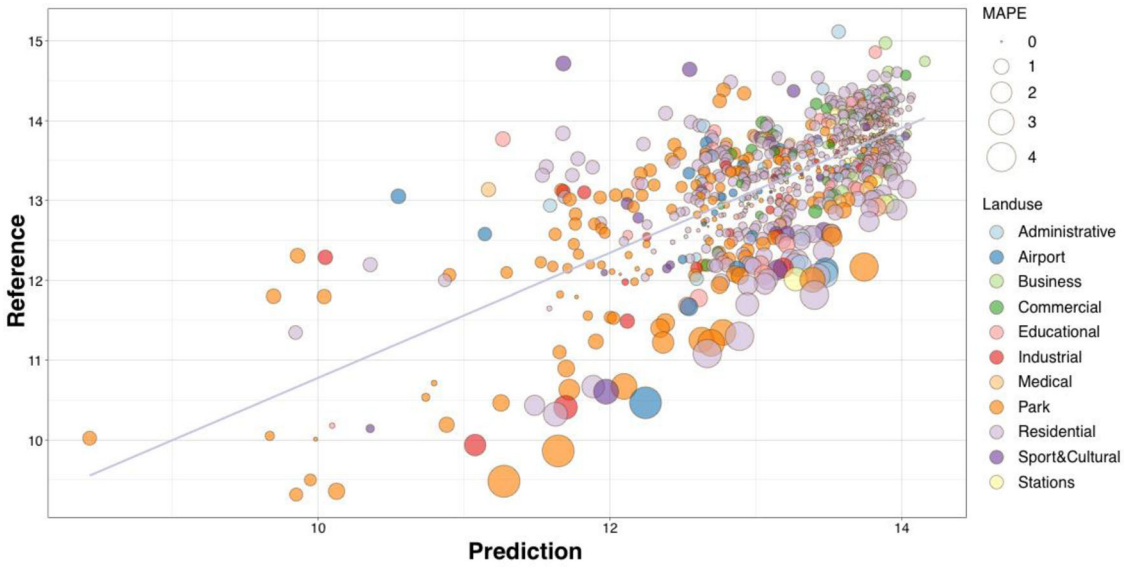

FIGURE 3 Scatterplot of reference and predicted values. MAPE, mean absolute percentage error. Data underlying this figure can be found in Supporting Information S2.

4.2 | Model interpretation

We calculate the importance of feature fusion for the random-forest model using the SHAP Python library developed previously (Lundberg & Lee, 2017). The plot is organized according to the absolute mean of the SHAP values for each feature in F_{fusion}, which serves as a proxy for the relative importance of the feature (Figure S5). The results show that the 227th dimensional feature of the ORS has the greatest influence on the prediction results of the model. In addition, the top 10 features of importance are all from ORS, which indicates that the ORS features extracted by deep learning have a decisive influence on the prediction of building stock. Based on the summation of importance calculated by the random forest, the ORS feature F_{rs} contributes approximately 85% to the final prediction, whereas NTL explains approximately 15% of the prediction. The contribution of NTL to the results is not as significant as that of ORS; nevertheless, the role of nighttime features as a complement to RGB remote sensing image features is not negligible. The results in Table 2 also demonstrate that the model with NTL inclusion is significantly better than that with the ORS regression.

We further visualize the ORS's 227th dimensional feature using Grad-CAM to locate the ground details retrieved by the proposed model. Figure 4 shows the informative regions of four randomly selected ORS images in the test set. The obtained heat map highlights specific ground objects and quantifies the relative contribution of ORS pixels to the prediction (Xing et al., 2020). Buildings are emphasized in all selected ORS, which indicates that the model makes decisions based on meaningful regions in the ORS images rather than focusing on random parts. For example, our model accurately captures only two buildings in Figure 4a (as shown in Figure 4e), which means that the model has an excellent ability to capture building features. When the building area covers most of the input area (e.g., Figure 4b–d), higher buildings are highlighted, indicating that taller buildings contribute more to the prediction results.

5 | CONCLUSIONS

Studies of high-resolution building material stock are essential to understand the environmental and socioeconomic consequences of construction. Although much effort has been expended to estimate building stock at high spatial resolution, existing methods still require a large amount of statistical data and human labor involvement. This study proposes a deep-learning-based model for building-stock estimation using publicly available multi-source remote sensing data. The model is implemented for Beijing, China, and the results indicate that the overall error in the test set is 13.4%. A comparison between our feature-fusion model, the ORS regression, and traditional NTL regression reveals that the proposed model achieves the best results for all six metrics. Our model can be widely applied to building-stock estimation with high spatial resolution. In our model, which is an end-to-end model, we reduce the human involvement in feature extraction and data collection significantly, enabling accurate and efficient high-resolution inventory estimation. Simultaneously, the availability of remote sensing data makes it possible to map the built environment stock over a long time series and at a global scale. The model is likely a credible building-stock estimator with excellent generalization capabilities over large unsampled areas, particularly for low-income nations and regions. Although only total building material stocks of Beijing are estimated in this study,

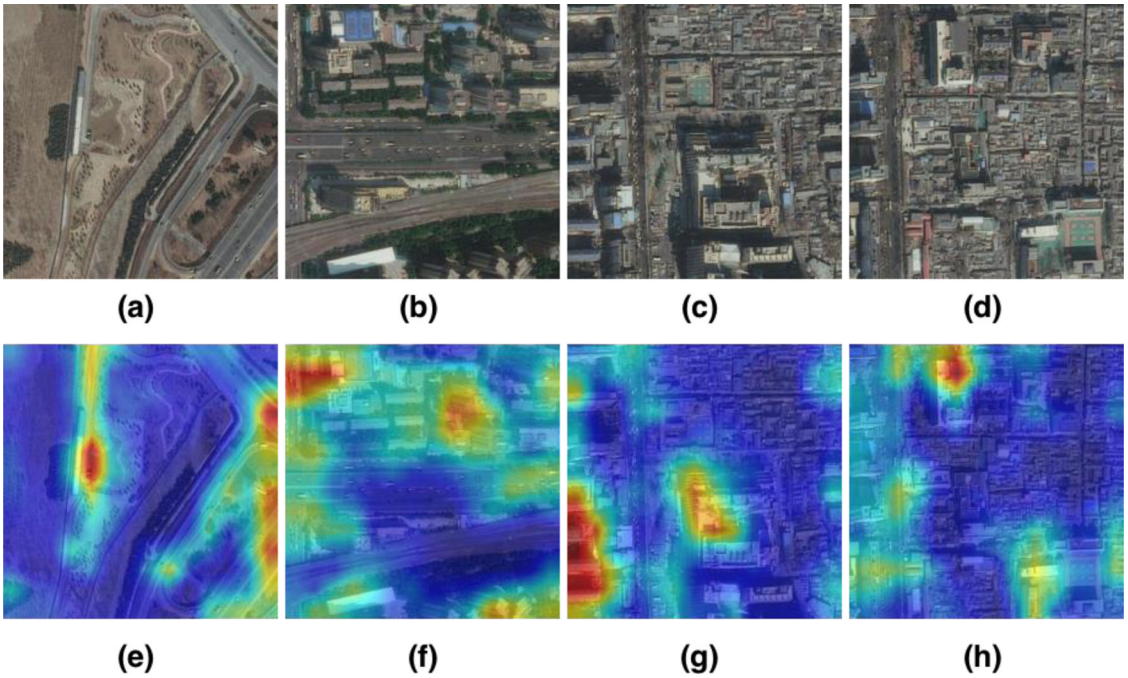

FIGURE 4 Feature importance map in optical remote sensing (ORS) imagery based on gradient-weighted class activation mapping. The heat map represents areas essential for the final decision. Heat maps that correlate to the ORS imagery are displayed in the bottom rows. The subfigures a, b, c, d in the first row are the four randomly selected ORS images, and the subfigures e,f,g,h represent the heat maps of a,b,c,d, respectively.

the model is also applicative for predicting building material stocks with individual categories. Because the features of multi-source remote sensing imagery also imply different building functions, the prediction model of typed building material stocks can be trained when the stock samples with categories are available.

This study has some limitations that need to be acknowledged. As a data-driven deep-learning method, the quality and quantity of the sample data significantly affect the performance of the model. The predictive and generalization abilities of the model improve as the sample size increases. However, owing to the lack of data, China's high-resolution building-stock labels are currently limited to a few major cities, such as Beijing and Shanghai. Moreover, the different layouts and unique appearances of cities in different regions and development levels would also affect the migration of the model. These disadvantages can be addressed by adding training data to enhance generalizability and prediction accuracy. Building height is one of the critical parameters of building stock; however, limited information regarding building height is provided in RGB-based multispectral remote sensing images. We are currently investigating the incorporation of additional remote sensing data sources into the model (e.g., synthetic aperture radar) to improve the expressiveness of the model and obtain more accurate stock estimation results.

ACKNOWLEDGMENTS

We appreciate the detailed comments from the editor and the anonymous reviewers.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request. The training and testing data were obtained from publicly available data from the literature and relevant websites.

ORCID

Zhou Huang https://orcid.org/0000-0002-1255-1913

REFERENCES

Abitbol, J. L., & Karsai, M. (2020). Interpretable socioeconomic status inference from aerial imagery through urban patterns. *Nature Machine Intelligence*, *2*(11), 684–692. https://doi.org/10.1038/s42256-020-00243-5

Aldebei, F., & Dombi, M. (2021). Mining the built environment: Telling the story of urban mining. *Buildings*, *11*(9), 388. https://doi.org/10.3390/buildings11090388

Ball, J. E., Anderson, D. T., & Chan, Sr, C. S. (2017). Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. *Journal of Applied Remote Sensing*, *11*(4), 042609. https://doi.org/10.1117/1.JRS.11.042609

Bao, Y. (2022). *High-resolution built environment stocks within cities in China*. https://doi.org/10.6084/m9.figshare.19387439

Bao, Y., Huang, Z., Li, L., Wang, H., Lin, J., & Liu, G. (2023). Evaluating the human use efficiency of urban built environment and their coordinated development in a spatially refined manner. *Resources, Conservation and Recycling*, *189*, 106723. https://doi.org/10.1016/j.resconrec.2022.106723

Cao, Y., & Huang, X. (2021). A deep learning method for building height estimation using high-resolution multi-view imagery over urban areas: A case study of 42 Chinese cities. *Remote Sensing of Environment*, *264*, 112590. https://doi.org/10.1016/j.rse.2021.112590

Chen, H., Zhang, X., Wu, R., & Cai, T. (2020). Revisiting the environmental Kuznets curve for city-level CO_2 emissions: Based on corrected NPP-VIIRS nighttime light data in China. *Journal of Cleaner Production*, *268*, 121575. https://doi.org/10.1016/j.jclepro.2020.121575

Cheng, X., Doosthosseini, A., & Kunkel, J. (2022). Improve the deep learning models in forestry based on explanations and expertise. *Frontiers in Plant Science*, *13*, 902105.

Das, S., Mirnalinee, T., & Varghese, K. (2011). Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(10), 3906–3931.

Deetman, S., Marinova, S., van der Voet, E., van Vuuren, D. P., Edelenbosch, O., & Heijungs, R. (2020). Modelling global material stocks and flows for residential and service sector buildings towards 2050. *Journal of Cleaner Production*, *245*, 118658. https://doi.org/10.1016/j.jclepro.2019.118658

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Feng, Y., Huang, Z., Wang, Y., Wan, L., Liu, Y., Zhang, Y., & Shan, X. (2021). An SOE-based learning framework using multisource big data for identifying urban functional ones. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 7336–7348. https://doi.org/10.1109/jstars.2021.3091848

Ferreira, B., Iten, M., & Silva, R. G. (2020). Monitoring sustainable development by means of earth observation data and machine learning: A review. *Environmental Sciences Europe*, *32*(1), 120. https://doi.org/10.1186/s12302-020-00397-4

Gong, P., Chen, B., Li, X., Liu, H., Wang, J., Bai, Y., Chen, J., Chen, X., Fang, L., Feng, S., Feng, Y., Gong, Y., Gu, H., Huang, H., Huang, X., Jiao, H., Kang, Y., Lei, G., Li, A., … Xu, B. (2020). Mapping essential urban land use categories in China (EULUC-China): Preliminary results for 2018. *Science Bulletin*, *65*(3), 182–187. https://doi.org/10.1016/j.scib.2019.12.007

Gontia, P., Nägeli, C., Rosado, L., Kalmykova, Y., & Österbring, M. (2018). Material-intensity database of residential buildings: A case-study of Sweden in the international context. *Resources, Conservation and Recycling*, *130*, 228–239. https://doi.org/10.1016/j.resconrec.2017.11.022

Guo, J., Fishman, T., Wang, Y., Miatto, A., Wuyts, W., Zheng, L., Wang, H., & Tanikawa, H. (2020). Urban development and sustainability challenges chronicled by a century of construction material flows and stocks in Tiexi, China. *Journal of Industrial Ecology*, *25*(1), 162–175. https://doi.org/10.1111/jiec.13054

Haberl, H., Wiedenhofer, D., Schug, F., Frantz, D., Virág, D., Plutzar, C., Gruhler, K., Lederer, J., Schiller, G., Fishman, T., Lanau, M., Gattringer, A., Kemper, T., Liu, G., Tanikawa, H., van der Linden, S., & Hostert, P. (2021). High-resolution maps of material stocks in buildings and infrastructures in Austria and Germany. *Environmental Science & Technology*, *55*(5), 3368–3379. https://doi.org/10.1021/acs.est.0c05642

Han, G., Zhou, T., Sun, Y., & Zhu, S. (2022). The relationship between night-time light and socioeconomic factors in China and India. *Plos One*, *17*(1), e0262503. https://doi.org/10.1371/journal.pone.0262503

Han, J., Chen, W.-Q., Zhang, L., & Liu, G. (2018). Uncovering the spatiotemporal dynamics of urban infrastructure development: A high spatial resolution material stock and flow analysis. *Environmental Science & Technology*, *52*(21), 12122–12132. https://doi.org/10.1021/acs.est.8b03111

Hattori, R., Horie, S., Hsu, F.-C., Elvidge, C. D., & Matsuno, Y. (2014). Estimation of in-use steel stock for civil engineering and building using nighttime light images. *Resources, Conservation and Recycling*, *83*, 1–5. https://doi.org/10.1016/j.resconrec.2013.11.007

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(8), 832–844. https://doi.org/10.1109/34.709601

Hu, M., Van Der Voet, E., & Huppes, G. (2010). Dynamic material flow analysis for strategic construction and demolition waste management in Beijing. *Journal of Industrial Ecology*, *14*(3), 440–456. https://doi.org/10.1111/j.1530-9290.2010.00245.x

Huang, T., Shi, F., Tanikawa, H., Fei, J., & Han, J. (2013). Materials demand and environmental impact of buildings construction and demolition in China based on dynamic material flow analysis. *Resources, Conservation and Recycling*, *72*, 91–101. https://doi.org/10.1016/j.resconrec.2012.12.013

Huang, Z., Qi, H., Kang, C., Su, Y., & Liu, Y. (2020). An ensemble learning approach for urban land use mapping based onremote sensing imagery and social sensing data. *Remote Sensing*, *12*(19), 3254. https://doi.org/10.3390/rs12193254

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790–794. https://doi.org/10.1126/science.aaf7894

Kleemann, F., Lederer, J., Rechberger, H., & Fellner, J. (2017). GIS-based analysis of Vienna's material stock in buildings. *Journal of Industrial Ecology*, *21*(2), 368–380. https://doi.org/10.1111/jiec.12446

Krausmann, F., Wiedenhofer, D., Lauk, C., Haas, W., Tanikawa, H., Fishman, T., Miatto, A., Schandl, H., & Haberl, H. (2017). Global socioeconomic material stocks rise 23-fold over the 20th century and require half of annual resource use. *Proceedings of the National Academy of Sciences*, *114*(8), 1880–1885. https://doi.org/10.1073/pnas.1613773114

Lanau, M., Liu, G., Kral, U., Wiedenhofer, D., Keijzer, E., Yu, C., & Ehlert, C. (2019). Taking stock of built environment stock studies: Progress and prospects. *Environmental Science & Technology*, *53*(15), 8499–8515. https://doi.org/10.1021/acs.est.8b06652

Langevin, J., Harris, C. B., & Reyna, J. L. (2019). Assessing the potential to reduce U.S. building CO_2 emissions 80% by 2050. *Joule*, *3*(10), 2403–2424. https://doi.org/10.1016/j.joule.2019.07.013

Levin, N., Kyba, C. C. M., Zhang, Q., Sánchez de Miguel, A., Román, M. O., Li, X., Portnov, B. A., Molthan, A. L., Jechow, A., Miller, S. D., Wang, Z., Shrestha, R. M., & Elvidge, C. D. (2020). Remote sensing of night lights: A review and an outlook for the future. *Remote Sensing of Environment*, *237*, 111443. https://doi.org/10.1016/j.rse.2019.111443

Li, X., Zhou, Y., Yu, S., Jia, G., Li, H., & Li, W. (2019). Urban heat island impacts on building energy consumption: A review of approaches and findings. *Energy*, *174*, 407–419. https://doi.org/10.1016/j.energy.2019.02.183

Liang, H., Tanikawa, H., Matsuno, Y., & Dong, L. (2014). Modeling in-use steel stock in China's buildings and civil engineering infrastructure using time-series of DMSP/OLS nighttime lights. *Remote Sensing*, *6*(6), 4780–4800. https://doi.org/10.3390/rs6064780

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4768–4777. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Lv, Z., Jia, Y., Zhang, Q., & Chen, Y. (2017). An adaptive multifeature sparsity-based model for semiautomatic road extraction from high-resolution satellite images in urban areas. *IEEE Geoscience and Remote Sensing Letters*, *14*(8), 1238–1242.

Mao, R., Bao, Y., Huang, Z., Liu, Q., & Liu, G. (2020). High-Resolution mapping of the urban built environment stocks in Beijing. *Environmental Science & Technology*, *54*(9), 5345–5355. https://doi.org/10.1021/acs.est.9b07229

Marcellus-Zamora, K. A., Gallagher, P. M., Spatari, S., & Tanikawa, H. (2016). Estimating materials stocked by land-use type in historic urban buildings using spatio-temporal analytical tools. *Journal of Industrial Ecology*, *20*(5), 1025–1037. https://doi.org/10.1111/jiec.12327

Müller, D., Wang, T., & Duval, B. (2010). Patterns of iron use in societal evolution. *Environmental Science & Technology*, *45*(1), 182–188.

Müller, D. B. (2006). Stock dynamics for forecasting material flows—Case study for housing in The Netherlands. *Ecological Economics*, *59*(1), 142–156. https://doi.org/10.1016/j.ecolecon.2005.09.025

Müller, E., Hilty, L. M., Widmer, R., Schluep, M., & Faulstich, M. (2014). Modeling metal stocks and flows: A review of dynamic material flow analysis methods. *Environmental Science & Technology*, *48*(4), 2102–2113. https://doi.org/10.1021/es403506a

Oezdemir, O., Krause, K., & Hafner, A. (2017). Creating a resource Cadaster—A case study of a district in the Rhine-Ruhr metropolitan area. *Buildings*, *7*(2), 45. https://doi.org/10.3390/buildings7020045

Peled, Y., & Fishman, T. (2021). Estimation and mapping of the material stocks of buildings of Europe: A novel nighttime lights-based approach. *Resources, Conservation and Recycling*, *169*, 105509. https://doi.org/10.1016/j.resconrec.2021.105509

Runting, R. K., Phinn, S., Xie, Z., Venter, O., & Watson, J. E. (2020). Opportunities for big data in conservation and sustainability. *Nature Communications*, *11*(1), 1–4. https://doi.org/10.1038/s41467-020-15870-0

Sartori, I., Bergsdal, H., Müller, D. B., & Brattebø, H. (2008). Towards modelling of construction, renovation and demolition activities: Norway's dwelling stock, 1900–2100. *Building Research & Information*, *36*(5), 412–425.

Schandl, H., Marcos-Martinez, R., Baynes, T., Yu, Z., Miatto, A., & Tanikawa, H. (2020). A spatiotemporal urban metabolism model for the Canberra suburb of Braddon in Australia. *Journal of Cleaner Production*, *265*, 121770. https://doi.org/10.1016/j.jclepro.2020.121770

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *2017 IEEE International Conference on Computer Vision (ICCV)*, 618–626. https://doi.org/10.1109/ICCV.2017.74

Shao, Z., Tang, P., Wang, Z., Saleem, N., Yam, S., & Sommai, C. (2020). BRRNet: A fully convolutional neural network for automatic building extraction from high-resolution remote sensing images. *Remote Sensing*, *12*(6), 1050. https://doi.org/10.3390/rs12061050

Stathakis, D., & Baltas, P. (2018). Seasonal population estimates based on night-time lights. *Computers, Environment and Urban Systems*, *68*, 133–141. https://doi.org/10.1016/j.compenvurbsys.2017.12.001

Sun, M., Han, C., Nie, Q., Xu, J., Zhang, F., & Zhao, Q. (2022). Understanding building energy efficiency with administrative and emerging urban big data by deep learning in Glasgow. *Energy and Buildings*, *273*, 112331. https://doi.org/10.1016/j.enbuild.2022.112331

Takahashi, K. I., Terakado, R., Nakamura, J., Adachi, Y., Elvidge, C. D., & Matsuno, Y. (2010). In-use stock analysis using satellite nighttime light observation data. *Resources, Conservation and Recycling*, *55*(2), 196–200. https://doi.org/10.1016/j.resconrec.2010.09.008

Tang, Z., Chuang, K. V., DeCarli, C., Jin, L.-W., Beckett, L., Keiser, M. J., & Dugger, B. N. (2019). Interpretable classification of Alzheimer's disease pathologies with a convolutional neural network pipeline. *Nature Communications*, *10*(1), 2173. https://doi.org/10.1038/s41467-019-10212-1

Tanikawa, H., & Hashimoto, S. (2009). Urban stock over time: Spatial material stock analysis using 4d-GIS. *Building Research & Information*, *37*(5–6), 483–502. https://doi.org/10.1080/09613210903169394

Vilaysouk, X., Islam, K., Miatto, A., Schandl, H., Murakami, S., & Hashimoto, S. (2021). Estimating the total in-use stock of Laos using dynamic material flow analysis and nighttime light. *Resources, Conservation and Recycling*, *170*, 105608.

Wang, J., Hu, X., Meng, Q., Zhang, L., Wang, C., Liu, X., & Zhao, M. (2021). Developing a method to extract building 3D information from GF-7 data. *Remote Sensing*, *13*(22), 4532. https://doi.org/10.3390/rs13224532

Wang, M., Zhang, X., Niu, X., Wang, F., & Zhang, X. (2019). Scene classification of high-resolution remotely sensed image based on ResNet. *Journal of Geovisualization and Spatial Analysis*, *3*(2), 16. https://doi.org/10.1007/s41651-019-0039-9

Xiao, H., Ma, Z., Mi, Z., Kelsey, J., Zheng, J., Yin, W., & Yan, M. (2018). Spatio-temporal simulation of energy consumption in China's provinces based on satellite night-time light data. *Applied Energy*, *231*, 1070–1078. https://doi.org/10.1016/j.apenergy.2018.09.200

Xing, X., Huang, Z., Cheng, X., Zhu, D., Kang, C., Zhang, F., & Liu, Y. (2020). Mapping human activity volumes through remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 5652–5668. https://doi.org/10.1109/JSTARS.2020.3023730

Xu, L., & Chen, Q. (2019). Remote-sensing image usability assessment based on ResNet by combining edge and texture maps. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(6), 1825–1834. https://doi.org/10.1109/JSTARS.2019.2914715

Ye, R., Huang, Z., Li, L., & Shan, X. (2022). GeoUNet: A novel AI model for high-resolution mapping of ecological footprint. *International Journal of Applied Earth Observation and Geoinformation*, *112*, 102803. https://doi.org/10.1016/j.jag.2022.102803

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., & Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications*, *11*(1), 1–11.

Yin, G., Huang, Z., Bao, Y., Wang, H., Li, L., Ma, X., & Zhang, Y. (2022). ConvGCN-RF: A hybrid learning model for commuting flow prediction considering geographical semantics and neighborhood effects. *GeoInformatica*, https://doi.org/10.1007/s10707-022-00467-0

Yu, B., Deng, S., Liu, G., Yang, C., Chen, Z., Hill, C. J., & Wu, J. (2018). Nighttime light images reveal spatial-temporal dynamics of global anthropogenic resources accumulation above ground. *Environmental Science & Technology*, *52*(20), 11520–11527. https://doi.org/10.1021/acs.est.8b02838

Yu, B., Lian, T., Huang, Y., Yao, S., Ye, X., Chen, Z., Yang, C., & Wu, J. (2019). Integration of nighttime light remote sensing images and taxi GPS tracking data for population surface enhancement. *International Journal of Geographical Information Science*, *33*(4), 687–706. https://doi.org/10.1080/13658816.2018.1555642

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., & Zhang, L. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, *241*, 111716. https://doi.org/10.1016/j.rse.2020.111716

Yuan, X., Jia, L., Menenti, M., Zhou, J., & Chen, Q. (2019). Filtering the NPP-VIIRS nighttime light data for improved detection of settlements in Africa. *Remote Sensing*, *11*(24), 3002. https://doi.org/10.3390/rs11243002

Zheng, Q., Weng, Q., & Wang, K. (2021). Characterizing urban land changes of 30 global megacities using nighttime light time series stacks. *ISPRS Journal of Photogrammetry and Remote Sensing*, *173*, 10–23. https://doi.org/10.1016/j.isprsjprs.2021.01.002

Zhu, H., Ma, M., Ma, W., Jiao, L., Hong, S., Shen, J., & Hou, B. (2021). A spatial-channel progressive fusion ResNet for remote sensing classification. *Information Fusion*, *70*, 72–87. https://doi.org/10.1016/j.inffus.2020.12.008

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bao, Y., Huang, Z., Wang, H., Yin, G., Zhou, X., & Gao, Y. (2023). High-resolution quantification of building stock using multi-source remote sensing imagery and deep learning. *Journal of Industrial Ecology*, *27*, 350–361. https://doi.org/10.1111/jiec.13356