IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 25, 2020, accepted August 18, 2020, date of publication August 26, 2020, date of current version September 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3019239

Siamese-Like Convolutional Neural Network for Fine-Grained Income Estimation of Developed Economies

RUIQIAO BAI[1], JACQUELINE C. K. LAM[1,2,3], (Member, IEEE), AND VICTOR O. K. LI[1], (Life Fellow, IEEE)

[1]Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong
[2]Energy Policy Research Group, Judge Business School, University of Cambridge, Cambridge CB2 1AG, U.K.
[3]Department of Computer Science and Technology, University of Cambridge, Cambridge CB2 1AG, U.K.

Corresponding author: Jacqueline C. K. Lam (jcklam@eee.hku.hk)

This work was supported in part by the Theme-Based Research Scheme of the Research Grants Council of Hong Kong under Grant T41-709/17-N.

ABSTRACT Estimating the per-capita income and the household income at a fine-grained geographical scale is critical but challenging, even across the developed economies. In this article, a novel Siamese-like Convolutional Neural Network, integrating Ridge Regression and Gaussian Process Regression, has been developed for fine-grained estimation of income across different parts of New York City. Our model (the *GP-Mixed-Siamese-like-Double-Ridge model*) makes good use of the pairwise comparison of location-based house price information, daytime satellite image, street view and spatial location information as the inputs. Taking the per-capita income and the median household income in New York City as the ground truths, our model outperforms ($R^2 = 0.72$-0.86 for five-fold validation) other state-of-the-art income estimation models and achieves good performance in cross-district and cross-scale validation. We also find that models which partially share our model architecture, including the *Spatial-Information-GP* and the *Mixed-Siamese-like model,* perform well under certain spatial granularity and data availability. Since such models rely on less data input types and simpler architectures, they can be used to save resources on data collection and model training. Hence, using our model for fine-grained income estimation does not mean excluding these models that share similar architectures. Our fine-grained income estimation model can allow the per-capita and the household income data generated in fine-grained resolution to couple with other types of data, such as the air pollution or the epidemic data, of the same scale, to ensure that any location-specific socio-economic-related study and evidence-based decision-making at the fine-grained resolution can be conducted. Future research will focus on extending our model for fine-grained income estimation in developing metropolises, and for developing other socio-economic indicators.

INDEX TERMS Daytime satellite image, developed metropolis, fine-grained resolution, *GP-mixed-Siamese-like-double-ridge model*, house price, household income, per-capita income, Siamese-like convolutional neural network, street view.

I. INTRODUCTION

Measuring income[1] distribution at a high spatial resolution is critical but challenging, even for developed economies [1]–[3]. Accurate income data are mainly obtained from field surveys, which can be highly capital intensive [2]. Over the past few decades, attempts have been made to overcome data scarcity and to estimate fine-grained income distribution across developing or non-urban areas [4]–[7]. Few studies have attempted to make good use of proxy data and deep learning models for high accuracy, fine-grained income estimation in developed and urban contexts. Such studies should advance our understanding of income distribution and variation at the fine-grained geographical level, so far as the developed and urban contexts are concerned [2], [8].

The associate editor coordinating the review of this manuscript and approving it for publication was Jinjia Zhou.

[1]According to the definition of American Community Survey, "Total income" refers to the sum of incomes reported separately for wage or salary income; net self-employment income; interest, dividends, or net rental or royalty income, or income from estates and trusts; Social Security or Railroad Retirement Income; Supplemental Security Income (SSI); public assistance or welfare payments; retirement, survivor, or disability pensions; and all other incomes [3].

Income is an important indicator critical for socio-economic studies in the developed world. First, income can largely reflect citizens' accessibility to a number of goods and services in most developed economies [9]. Second, income is closely related to ones' living standards in developed economies. Given better welfare allocation (e.g. retirement plans, free health care, unemployment compensation), citizens of developed economies have less incentives to save their income to mitigate future financial risks due to illness and unemployment etc., and more incentives to spend their income over the short-term to maintain their standards of living [10]–[13]. In the United States, the savings rate has substantially fallen to below 3% during the late 2000s [14], [15]. Third, collecting income data is a relatively easy task across the developed economies when field survey resources/services are freely provided/supported by the government and other NGOs [9].

In this article, our fine-grained income estimation study estimates income at the district-level of a city. Estimating income at such a level is beneficial for our understanding of the relationship between income and other socio-economic variables, such as air pollution exposure or COVID-19 pandemic. Such fine-grained analysis can allow policymakers to provide recommendations on any socio-economic related environmental/public health challenges that are location-specific [16]. However, the validity of the analysis will ultimately be dependent upon the accuracy of income estimation at the fine-grained resolution.

Collecting accurate fine-grained income data/conducting accurate income estimation is crucial for developed economies. First, as compared to developing economies characterized mostly by low-income distribution, developed economies are facing a higher risk of intra-city income inequality. Specifically, some citizens of developed economies may earn extremely high levels of incomes, whilst other citizens who lack the needed capabilities may be forced to accept extremely low levels of incomes [17]. Second, developed economies usually are associated with a higher level of democracy, and a higher social awareness and demand for data transparency [18]. Publishing fine-grained income data can meet the public demand and can facilitate better understanding of such issues as socio-economic-related environmental exposure inequality or COVID-19 infection imbalance.

In any developed economy, such as the United States, the income data obtained via large-scale surveys are not immediately updated; data collection is highly expensive [19], [20]. In fact, the United States spends more than USD250 million per year on discharging the American Community Survey (ACS), a door-to-door survey that collects statistics such as per-capita income and household income [21]. Due to high manpower, smaller geographical units (areas having <65,000 residents) are investigated less frequently, and income data surveyed are not published until one or two years later [19]. Delays in data-reporting may impede timely policy decisions and weaken the effectiveness of public resource allocation [22].

To reduce the manpower needed for fine-grained income surveys and to speed up fine-grained income data collection, researchers have used house price as a proxy for income. Previous studies have identified a positive correlation between house price and income [23]–[29], whilst house price data are easily accessible and downloadable online in the developed world. However, estimation models that depend on house price as the input and income as the output have yielded a low estimation accuracy. A study that estimates yearly household income with a kernel regression model, using as inputs the household-level house price information of six cities across the United States, has achieved very low estimation performance [30]. The Spearman rank correlation between house price and income at the household-level has achieved a correlation coefficient as low as 0.38 to 0.52 [30]. In another study, a polynomial model is used to estimate the household income in London, also taking house price as an explanatory variable, but no validation accuracy has been provided [31]. Furthermore, in most developed metropolises, house price data distribution is uneven. Some parts of the city may have more house price data points than others. Due to data skewness, income estimation using house price as the input may be inaccurate. More obstacles have to be overcome when alternative advanced machine learning techniques that use house price as the input are being considered.

In addition, as house price to income ratio can vary greatly across different times and spaces [32], [33], other auxiliary factors may need to be properly taken into account. To better capture the interactions between house price and other variables, scholars have suggested that an advanced machine learning method may be useful for overcoming the compounding effect/multi-collinearity of input variables, which are commonly found in traditional statistical models [34].

Apart from the house price-based income estimation model, other resource-efficient methods for fine-grained district-level income estimation in the developed economies have been identified. In Table 1, we classify these models into four different categories. The first category is based on the visual appearance of the district, which is normally captured by the night-time/daytime satellite image or street view [2], [35]–[37]. The assumption being that buildings, roads, vegetations and nightlight intensities can vary from place to place, when the income level of these places vary [2], [35]–[37]. Some researchers have claimed that combining the visual data with the spatial data of individual districts can contribute to higher income estimation accuracy [1]. The second category focuses on transportation [19], [38]–[40]. Some studies have extracted features of human mobility or car attributes of a small area to represent an average income level of that area [19], [32], [38]–[41]. The third category is based on the quality of local food restaurants/stores, as an area with low-end restaurants or food

TABLE 1. Literature review on fine-grained income estimation across the developed world.

Classification	Paper	Indicator Estimated	Key Input	Country (district/city)	Geographical Level	Machine Learning/ Statistical Technique	Major Conclusion
Visual Appearance	(Glaeser *et al.*, 2015) [35]	Median family income	Street view	United States (NYC, Boston)	Block	Geometric layout algorithm; ν-support vector regression	For NYC, the model obtains validation R^2 = 0.77. When being transferred to estimate income in Boston, R^2 = 0.71.
	(Mellander *et al.*, 2015) [36]	Wage income density	Nightlight	Sweden	Square	Geographically weighted regression	R^2 = 0.49, no validation dataset.
	(Suel *et al.*, 2018) [1]	Mean household income	Street view, latitude and longitude of district centroids	United Kingdom (Greater London)	Lower-super output area	VGG16-based CNN; Gaussian processes model	Validation r = 0.93
	(Suel *et al.*, 2018) [37]	Mean household income decile	Street view	United Kingdom (Greater London)	Lower-super output area	VGG16-based CNN;	When transforming the 10 deciles to ten integers (1-10) in a regression task, validation r = 0.86
	(Piaggesi *et al.*, 2019) [2]	Mean household income	Daytime satellite image, nightlight	Chile (Santiago Metropolitan Area); United States (Los Angeles, Philadelphia, Boston, Chicago, Houston)	Chile: zona, comuna; United States: ZIP code, tract	VGG-F/ResNet50-based CNN; ridge regression	ResNet50 always outperforms the VGG-F model. Hence the following results from ResNet50 models: Santiago Metropolitan Area: R^2 = 0.477 for zona, R^2 = 0.643 for comuna. No such result in United States cities. Train the model using ZIP code-level income, and validate using tract-level income: Santiago Metropolitan Area: R^2 = 0.411; 5 cities in United States: R^2 = 0.447-0.631
Transportation	(Smith *et al.*, 2013) [40]	Income score (defined in Index of Multiple Deprivation)	Transportation card record	United Kingdom (London)	Lower-super output area	Linear Regression, support vector machine	Validation R^2 = 0.39 (only accounted for 10% of the census areas in London where rail stations are present).
	(Gebru *et al.*, 2015) [39]/ (Gebru *et al.*, 2017) [38]	Median household income	Street view, car attribute dataset	United States (200 large cities)	ZIP code	Deformable part model, CNN	R^2 = 0.49, no validation dataset.
	(Gebru *et al.*, 2017) [19]	Median household income	Street view	United States (Tampa)	ZIP code	Deformable part model, CNN, ridge regression	ZIP code-level in Tampa (a city in Florida) R^2 = 0.76, no validation dataset.
Food & Shopping	(Block *et al.*, 2014) [43]	Median household income	Fast-food restaurants distribution	United States (New Orleans)	Tract	Bivariate analysis	The number of restaurants per square mile was negatively correlated with median household income (r = -0.275, p < 0.001).
	(Glaeser *et al.*, 2017) [42]	Median household income	Yelp recommended review	United States (NYC)	ZIP code	Random forest algorithm	Yelp has more estimation power in denser (validation R^2 = 0.194), wealthier (validation R^2 = 0.256), and more educated (validation R^2 = 0.234) ZIP code areas.
Online Social Network Platform	(Hristova et al., 2016) [44]	Income score (defined in Index of Multiple Deprivation)	Twitter record, Foursquare record	United Kingdom (London)	Borough	Spearman rank correlation	r = 0-0.5, p-values < 0.05, no validation dataset

stores is assumed to have lower-income residents [42], [43]. The fourth category is based on data collected from the online social network platforms [44]. It assumes that people having more complex social networks may earn higher incomes due to better accessibilities to higher paid jobs [45]–[49], or better entrepreneurship opportunities [50], [51]. However, such fine-grained income estimation models are yet to address the followings:

First, these studies are highly dependent on non-public data, and that some of these data are not easily obtainable and may invite privacy concerns. For instance, transportation card records [40] have been used to extract features of social network structures and human mobility patterns but can be hardly accessible without making prior agreements with the relevant organizations, such as the transport authorities. The models that rely on social media records, such as Twitter [44], might expose the personal information of Twitter users and raise privacy concerns.

Second, some of these estimation studies have been based on indicators which have low correlations with the district-level incomes. For example, indicators derived from the distributions of fast-food restaurants [43] and business/ restaurant reviews, or profiles from Yelp [42], have relatively low correlations with district-level incomes. Besides, the nightlight intensity has been widely used to estimate Gross Domestic Product (GDP) in developing economies [52]–[57], but its application on fine-grained income estimation tends to achieve low accuracy across

developed economies [36], [58]–[61]. As both rich and poor parts of cities in the developed world have been equipped with sufficient lighting facilities, nightlight intensity can hardly be used to differentiate the poor areas from the rich areas in the developed cities [36], [58]. To cite an example, in New York City (NYC), the nightlight intensity is high across all districts; hence, the intensity variation appears to be too low to signify any change in income at the fine-grained resolution [59]. Some researchers have developed a nightlight-based transfer learning methodology relevant for estimating assets [62] and consumer expenditures in African countries [4], but such studies may not be directly applicable to cities in the developed economies [2].

Third, regarding district-level income estimation, previous studies have yet combined daytime satellite image with street view as a model input. Some studies have shown that combining daytime satellite image with street view can contribute to good model performance in house price estimation [63], which may be extended to income estimation. Further, previous visual-based income estimation methods have yet exploited features extracted from both aerial and ground-level street view [1], [2], [35], [37].

II. NOVELTY

Given such background, we propose the adoption of a transfer learning methodology for fine-grained per-capita income and median household income estimation in developed economies, which outperforms state-of-the-art models and achieves a higher estimation accuracy at a district-level of a city. Specifically, our proposed method combines four data categories, including house price, daytime satellite image, street view, and spatial information (latitude and longitude of district centroid) as data inputs. Based on pair-wise comparison results of house price information, we develop a novel Siamese-like Convolutional Neural Network (CNN) to enhance the effectiveness of image feature extraction. The model does not require one to input all house price information from all parts of a city, which may solve the problem of house price data sparsity due to information skewness. Our model presents high generalizability.

The rest of the paper is organized as follows. Section III details the methodology of our Siamese-like CNN model. Section IV reveals and discusses our income estimation model results. We further compare the performance of our model with our selected state-of-the-art income models. Finally, Section V concludes our study and puts forward suggestions for future research.

III. METHODOLOGY

Our overall methodology consists of four parts (see Fig. 1). In Part 1, we develop a Siamese-like CNN to extract house price-related features from the daytime satellite images and the street views collected from NYC. In Part 2, our image features are averaged at the district-level and taken as the inputs to the Ridge Regression model for district-based income estimation. Since house price is positively correlated with

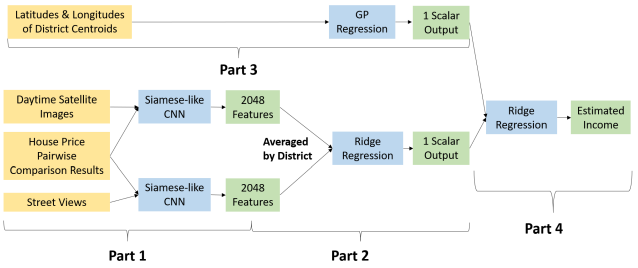

FIGURE 1. Methodological framework.

income, it is expected that our house price-related features would be correlated with the income values (ground truths) obtained from NYC. Given the richer information derivable from the daytime satellite images and the street views, they are expected to outperform house price information in income estimation, given their better spatial representation. In Part 3, we take the latitude and the longitude of a district centroid as the inputs to a Gaussian Processes (GP) model to extract a scalar value from the spatial information for income estimation. In Part 4, we concatenate the scalar outputs generated by Part 2 and Part 3 and feed them into another Ridge Regression model for final income estimation. We take NYC in the United States, a metropolis with a highly developed economy as our case study. We use a Siamese-like CNN to estimate the district-based income levels of NYC in 2018.

A. LABELLED DATA

The ground truths of the district incomes in 2018 in NYC are obtained from the 2014-2018 American Community Survey (a 5-year estimate), a national-level door-to-door field survey [21]. Two types of district-level average income data in NYC are used as labels: per-capita income and median household income. The average income data at two different geographical levels are used to test the model performance at different granularities: the tract-level (2067 tracts), and the ZIP code-level (211 ZIP codes). Specifically, the data is gathered from Census Reporter[2] [64].

B. INPUT DATA

Four types of inputs are used in this research: the house price, the daytime satellite image, the street view, and the spatial location information of each district (see Fig. 2). The house price information in 2018 is obtained from NYC Department of Finance, which is the official and a highly comprehensive information source [65]. Each piece of house price data corresponds to one real estate transaction, and the exact location of the building is used. 21,144 items of house price information (sales price divided by gross square feet) are extracted after data cleaning (excluding sale price 0, gross square feet 0, or location not found). The latitude and the longitude of each building is located by the official map searching tool [66].

[2]Table B19301 for per-capita income, and B19013 for median household income.

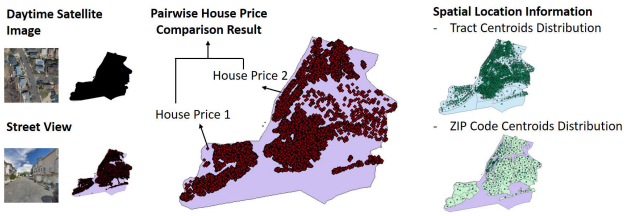

FIGURE 2. Four types of input data to the Siamese-like CNN.

Daytime satellite images, captured in 2018, are gathered from the NYC government [67]. All satellite images are gathered at zoom-level 18; successive images are taken approximately every 0.001 degree with no overlappings (a total of 89889 images, at 256 × 256 pixels per image). The spatial resolution of the daytime satellite image is approximately 4×10^{-7} degrees per pixel. Although many previous studies have obtained the daytime satellite image from Google Static Maps API, we do not collect our images from these data sources as the API does not provide the exact year of the image captured. Our street views are directly obtained from Google Street View Static API which provides the captured year of the street views. One image is taken with a change in view every 0.001 degree if it exists (within a default 50 meter searching radius, 640 × 640 pixels per image, 54246 images). Data cleaning is conducted to filter any street views that are invalid/dark/interior/blurred/duplicated/ obstructed by an object [63], [68]. Only the street views taken between 2017 to 2019 are used. It is assumed that the physical appearance of NYC did not change significantly from 2017 to 2019. The spatial location information refers to the latitude and the longitude of any district centroid in NYC, which is obtained from the district boundary shapefile via Census Reporter [64].

C. TRANSFER LEARNING

Transfer learning is a machine learning technique that learns certain knowledge during one process of problem-solving, then transfers such knowledge to another area of problem-solving [69]. In this study, we adopt the method of transfer learning and extract image features to compare house prices in NYC, then apply the knowledge learnt from house price to income estimation. The overall framework of transfer learning consists of four steps, as detailed below (see Fig. 1).

The first step of our study aims at extracting the house price-related features from the daytime images and the street views. Before training, each piece of house price data is matched with the nearest daytime satellite image and street view. To extract image features, an intuitive approach is to establish a Regression model between the image and the house price information (normalized by the maximum value). Following this method, a model based on CNN is constructed. As shown in Fig. 3, the image is inputted to a Resnet-50 (a 50-layer residual CNN) and a tensor consisting of 2048 features is extracted [70]; the predicted house price is generated after a dense layer (with activation function *tanh*). The mean square error is used as the loss function. However, there are

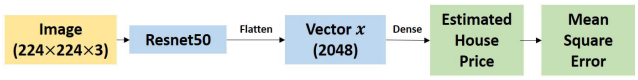

FIGURE 3. Architecture of non-Siamese-like CNN.

limitations with regard to this method during the training process. In particular, the loss function can hardly converge without excluding the outliers of the house prices. Experimental results show that the features extracted are not highly correlated with the actual income values.

To improve the feature extraction performance, we design a novel Siamese-like CNN to more effectively extract house price-related features for fine-grained income estimation. The traditional Siamese CNN has been a few-shot learning technique, designed originally for image classification [71]. As shown in Fig. 4, the inputs to Siamese CNN normally cover a pair of images; each image can be treated by one CNN, and the outputs of the two CNNs can be concatenated. After some fully connected layers with the Rectified Linear Unit (ReLU) activation function, the model produces a scalar value indicating the similarity between two images [71]. A unique characteristic of a Siamese-like CNN is that the two CNN models always share the same architecture and weight. Researchers have designed a Siamese-like CNN to predict the human judgment of pairwise image comparisons [72]. In this study, we develop a novel Siamese-like CNN for extracting house price-related image features for fine-grained income estimation.

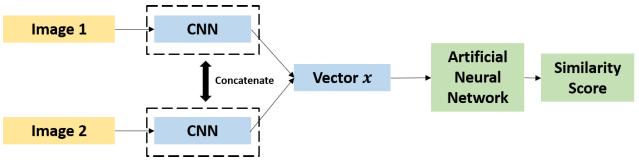

FIGURE 4. Architecture of Siamese CNN.

Fig. 5 shows the architecture of our newly designed Siamese-like CNN. It is a classification model that predicts the comparison results of the house prices at two locations. Specifically, the input consists of a pair of daytime satellite images, or a pair of street views. Each house price-related image is treated by a Resnet-50, and two Resnet-50s always generate the same weights. The Resnet-50 can extract one set of 2048 features from each image (see x_1 and x_2 in Fig. 5), with the two image feature sets being subtracted element-wise and fed into a dense layer to generate a 3-element vector for representing the image captured in the location with a

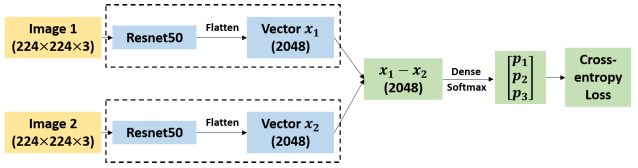

FIGURE 5. Architecture of Siamese-like CNN.

higher house price. In Fig. 5, p_1, p_2, p_3 represent the three elements and each of them ranges from 0 to 1 to represent the likelihood of each of the three possible results; Image 1 is higher in house price, Image 1 and Image 2 are equivalent in house price, or Image 2 is higher in house price. The label used in our study is transformed to a one-hot vector with 3 values ([1,0,0] indicates that Image 1 is higher in values, [0,1,0] indicates that Image 1 and 2 are equivalent in values, and [0,0,1] indicates that Image 2 is higher in values).

There are reasons why features extracted by Siamese-like CNN can outperform non-Siamese-like CNN for fine-grained income estimation. Siamese-like CNN is a classification model and converges more easily as compared to Regression. Non-Siamese-like CNN requires that image features extracted estimate the exact house price, whereas Siamese-like CNN relaxes this requirement and allows the image features to be less strongly correlated with the exact house price. Such difference leads to the next question: Are features that are good for house price estimation also good for income estimation? In reality, studies have shown that given the same house price, difference in the income level can still be significant [73], [74]. Hence, the expectation for a one-to-one correspondence between house price and income is unrealistic (which is the underlying assumption taken by non-Siamese-like CNN), whereas it is much more likely that any districts having a higher house price would have a higher income level (which is the underlying assumption of the pairwise comparison adopted by Siamese-like CNN). Hence, though the features extracted by Siamese-like CNN are less strongly correlated with the exact house price values, they can better capture any factors that simultaneously influence both house price and income (instead of factors that only influence house price), and eventually achieve a higher correlation with the actual income. The better performance of Siamese-like CNN (see Section IV) also confirms our intuition that by relaxing some irrelevant and redundant restrictions on feature extractions, the classification model can obtain features more relevant to the income values of the local contexts. Our model consists of four steps:

Step 1: We train our Siamese-like CNN. The compared results of 100,000 house price-related image pairs are randomly generated based on daytime satellite images and street views separately. The cross-entropy loss is used as the loss function for classification and the Resnet-50 is initiated by the weights pre-trained on ImageNet [75]. The batch size is 32, the training epoch is 10, *ReLu* is used as the activation function, except for the *softmax* layer that is used for calculating the cross-entropy loss. The optimizer is Stochastic Gradient Descent (SGD) (momentum = 0.9; initial learning rate = 0.001, reduced by a factor of 10 when the loss value increases, minimum learning rate = 0.0000001) and L2 regularization (0.01) is applied.

Step 2: Second, a Ridge Regression model is trained for dimension reduction via supervised learning, to reduce overfitting. Ridge Regression can be taken as Linear Regression with L2 regularization (without penalizing the intercept term) [76], which has been verified as an effective model in tackling a large number of image features extracted for income estimation [2]. The district income gathered from a field survey can be used as the label of the Regression model, and the model can generate a scalar output. Our inputs contain two sets of data: the first input is obtained from the daytime satellite images, and the second input is obtained from the street views. As each district (tract/ZIP code) contains multiple daytime satellite images, 2048 features are calculated for each district, by averaging the features of all images within the same district. The features of street views are calculated in the same way.

Step 3: Third, a GP model is used to extract a scalar value from the spatial information for income estimation. The GP model is a non-linear model built upon a Bayesian approach which specifies a Gaussian prior over the parameters [77]. Suel *et al.* have pointed out that adding the spatial data by the GP model can further enhance the income estimation accuracy of the district [1]. The inputs to the GP model cover both the latitudes and the longitudes of the district centroids. The labels of the model are the district income collected by field surveys. The GPy package is used to fit GP with the Matern-3/2 kernel [78], based on [1]. Following the default settings in the package sample codes, training has been repeated twice and the mean output scalar value is further used for income estimation.

Step 4: Finally, a new Ridge Regression model, *Image-Spatial-Info-Ridge-Regression model*, is used to combine the image features with the spatial features, and to estimate the final income via supervised learning. The scalar outputs generated from Step 2 and Step 3 are concatenated for each district and taken as the inputs. The district income data collected from the field surveys are taken as the ground truths. Ridge Regression is used here due to its ability to avoid overfitting and achieve good estimation performance.

Three types of cross-validation are used to evaluate model performance, including R^2, the Root Mean Square Error (RMSE), and the Mean Absolute Error (MAE) [2], [4], [79], [80]. First, a five-fold validation is used to evaluate the model's overall income estimation accuracy [4]. Here, our five-fold validation, which masks the income data representing one-fifth of the districts in each fold, is limited to Steps 2 to 4, since Siamese-like CNN conducted in Step 1 does not require any income data input. To ensure comparability, the same set of five-fold district division is used across all experiments. Second, a cross-district validation is utilized to test our model's spatial generalizability. Here, in each fold, all four parts of our model are trained on the data obtained from only one-fifth of the districts and validated using the data from the rest of the districts. Third, a cross-scale validation is conducted by applying the ZIP code-level (coarse-scale) model to tract-level (fine-scale) income estimation [2]. The hyperparameter of the Ridge Regression model in each fold is determined by a grid-searching procedure, that aims at maximizing the R^2 of another five-fold validation conducted on the training dataset [2], [4].

TABLE 2. Comparison of The Mixed Siamese GP Model with Models of Alternative Data Inputs and Architectures (Five-fold Validation R^2, RMSE, MAE).

Model	Input Data	Relevance	Model Architecture	Tract-level						ZIP code-level					
				Per-capita income			Median Household Income			Per-capita Income					
				R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
GP Mixed Siamese-like Double Ridge	house price, daytime satellite image, street view, latitude and longitude of district centroid	districts where street views are captured	Siamese-like CNN (Resnet50), ridge regression, GP model	0.85	10905	6689	0.72	17665	12556	0.86	11525	7472	0.72	19245	12674
Mixed Siamese-like GP	house price, daytime satellite image, street view, latitude and longitude of district centroid	districts where street views are captured	Siamese-like CNN (Resnet50), ridge regression, GP model	0.80	12317	7296	0.68	18830	13129	0.83	12530	8312	0.71	19461	12723
Spatial Information GP	latitude and longitude of district centroid	all districts	GP model	0.84	11050	6600	0.68	18827	12907	0.87	12138	8197	0.74	19730	13196
Mixed Siamese-like Random Forest	house price, daytime satellite image, street view, latitude and longitude of district centroid	districts where street views are captured	Siamese-like CNN (Resnet50), Random Forest	0.79	12796	8139	0.64	20093	14553	0.81	13799	8922	0.69	20929	13922
Mixed Spatial Siamese-like	house price, daytime satellite image, street view, latitude and longitude of district centroid	districts where street views are captured	Siamese-like CNN (Resnet50), ridge regression	0.78	13055	8583	0.64	19869	14566	0.86	11819	8064	0.71	19873	13555
House Price[a]	house price	districts where house prices are available	ridge regression	0.46	20186	13976	0.27	28134	22537	0.62	20110	13472	0.36	29812	20917
Mixed Siamese-like	house price, daytime satellite image, street view	districts where street views are captured	Siamese-like CNN (Resnet50), ridge regression	0.78	13094	8447	0.65	19690	14230	0.84	12357	8158	0.70	20274	13215
Mixed Non-Siamese-like	house price, daytime satellite image, street view	districts where street views are captured	Non-Siamese-like CNN (Resnet50), ridge regression	0.65	16502	10981	0.37	26139	20533	0.80	14918	10753	0.49	26751	20104
Satellite Siamese-like	house price, daytime satellite image	all districts	Siamese-like CNN (Resnet50), ridge regression	0.77	13338	8584	0.64	19761	14285	0.85	12635	8698	0.71	20153	14133
Satellite Non-Siamese-like[a]	house price, daytime satellite image	all districts	Non-Siamese-like CNN (Resnet50), ridge regression	0.65	16563	11015	0.36	26393	20822	0.80	15168	10898	0.49	26848	20178
Street View Siamese-like	house price, street view	districts where street views are captured	Siamese-like CNN (Resnet50), ridge regression	0.70	15093	9672	0.56	21983	16052	0.78	13776	8943	0.63	22181	14414
Street View Non-Siamese-like[a]	house price, street view	districts where street views are captured	Non-Siamese-like CNN (Resnet50), ridge regression	0.56	27561	17358	0.31	27472	21445	0.76	22400	16606	0.42	30834	22584

[a]House price values of the largest and the smallest 2.5 percentiles, are taken as outliers and removed.

IV. RESULT AND DISCUSSION

A. FIVE-FOLD VALIDATION

1) COMPARISON OF SIAMESE-LIKE CNNS VS NON-SIAMESE-LIKE CNNS OF DIFFERENT DATA INPUTS AND ARCHITECTURES

In Table 2, the models of different data inputs and architectures are compared. It shows that our proposed *GP-Mixed-Siamese-like-Double-Ridge model* achieves outstanding performance ($R^2 = 0.72 - 0.86$), as compared to other models, which only use part of the available data.

Specifically, we compare our Siamese-like CNN models with different input image datasets. The *Mixed-Siamese-like model* is based on imagery features from both the daytime satellite image and the street views, the *Satellite-Siamese-like model* and *Street-view-Siamese-like model* are merely based on one type of image corresponding to their name. It is observed that the *Mixed-Siamese-like model* always attains the highest R^2 value at both the tract-level and the ZIP code-level. Besides, the *Mixed-Siamese-like model* achieves a higher R^2 on the per-capita income-level as compared to the median household income-level. In addition, it should be mentioned the *Satellite-Siamese-like model* has a wider applicability as compared to the *Mixed-Siamese-like model*, as street views may not be available across all small districts, while satellite images are. This would not present a major challenge for NYC, as the street views are available across most of the districts.

We also compare with the models that are only based on the house price or the spatial location information. The *House Price model* only estimates district income based on the local average house price. The *Satellite-Siamese-like model*, the *Street-view-Siamese-like model* and the *House Price model* all generate the estimated income value by Ridge Regression in the final step, but the former two perform much better than the *House Price model*. This indicates that the house price-related image features extracted by Siamese-like CNN can outperform the house price for income estimation. It is worth noting that, when GP is used, the *Spatial-Information-GP model* [1], which only takes the latitudes and the longitudes of district centroids as the inputs, achieves a high five-fold validation accuracy, especially for the per-capita income estimation. Intuitively, there is hardly any direct causal relationship between the latitudes, the longitudes and income. Hence, the efficacy of the spatial information for income estimation can be attributable to the spatial autocorrelation of the income distributions. Specifically, our

TABLE 3. Comparison of state-of-the-art models.

Classification	Paper	Model Architecture	Data Input	Data Availability	Privacy Protection	Transferability Between the Developed and the Developing Countries	Expected Estimation Accuracy
Visual Appearance	This paper	Siamese-like CNN (Resnet50), Ridge Regression, GP	Daytime satellite image, street view, house price, latitude and longitude of district centroid	High	High	Medium	High
	(Piaggesi *et al.*, 2019) [2]	Resnet50-based CNN, Ridge Regression	Daytime satellite image, nightlight	High	High	Low	Medium
	(Suel *et al.*, 2018) [1]	VGG16-based CNN, GP Residual Regression	Street view, latitude and longitude of district centroid	High	High	Medium	High (lower than this paper, see TABLE V)
Transportation	(Gebru *et al.*, 2017) [38]	Deformable Part Model, CNN	Street view, car attribute dataset	Medium	High	Medium	High
Food & Shopping	(Glaeser *et al.*, 2017) [42]	Random Forest	Yelp recommended review	High	Medium	Medium	Low
Online Social Network Platform	(Hristova et al., 2016) [44]	Spearman Rank Correlation	Twitter record, Foursquare record	Low	Low	Medium	Low

results have shown that for the people living in nearby districts, their per-capita income might be more similar.

The performance of models with and without Siamese-like CNN architecture is compared. We compare the performance of the *Mixed-Siamese-like model* vs. the *Mixed-non-Siamese-like model*; the *Satellite-Siamese-like model* vs. the *Satellite-non-Siamese-like model*; the *Street-view-Siamese-like model* vs. the *Street-view-non-Siamese-like model*. Given the same input data, Siamese-like CNN always outperforms non-Siamese-like CNN that extracts the house price-related features via Regression (see Section III).

We combine Ridge Regression and GP as a multi-step regressor in our study (see Steps 2 to 4). With the prior knowledge that individual image feature (among 4096 image features) is less important than individual spatial feature (i.e. the latitude or the longitude) for income estimation, we believe a multi-step approach is desirable. If all features are fed into a regressor indiscriminately, we will not be able to make full use of this prior knowledge. Specifically, we test the *Mixed-Siamese-like-Random-Forest model* by taking all 4098 features (i.e. 4096 image features and 2 spatial features) as the inputs to a Random Forest model in a single step [81]. The model tends to overfit. In addition, we compare our *Mixed-Siamese-like-Double-Ridge* model with the *Mixed-Siamese-like-GP model*, which excludes Step 3 and Step 4 from our model, and takes the scalar output of Step 2, the latitude and the longitude as the inputs to a GP model for income estimation. The results show that our proposed model outperforms the *Mixed-Siamese-like-GP model*, indicating that Step 3 and Step 4 are capable of reducing overfitting.

We test the *Mixed-Spatial-Siamese-like model*, which takes the spatial information as an input to a Siamese-like CNN. In this model, the latitude and the longitude of each image are combined with 2048 image features by a dense layer in each branch of a Siamese-like CNN, and the dense layer produces 2050 outputs. Other architectures of this Siamese-like CNN are the same as the one described in Section III. All features are then taken as the inputs to a Ridge Regression model for income estimation. It performs less well than the *Spatial-Information-GP model*. One reason being that the Siamese-like CNN aims at extracting the house price-related features. Taking the latitude and the longitude as the inputs to the model, and by transforming the spatial information to house price-related features, it increases the difficulty for the model to directly comprehend the spatial autocorrelation of the income distributions.

2) COMPARISON OF SIAMESE-LIKE CNNS WITH STATE-OF-THE-ART MODELS

To compare our Siamese-like CNNs with state-of-the-art models, we select five district-level income estimation Regression models by the following criteria. For each of the four methods shown in Table 1, at least one model is selected, and the model shall deploy state-of-the-art methodology and achieve outstanding performance when compared to other studies of the same class. The architectures of the selected models are summarized in Tables 3 and 4.

The comparison analysis shows that our model outperforms the five state-of-the-art Regression models. Specifically, as shown in Table 3, model performance is compared across four dimensions, including, data availability, privacy protection, transferability between the developed and the developing countries, and the overall expected estimation accuracy. Details of our proposed model's major advantage over the other models are outlined in Table 4. In addition, given the available data, we obtain the performance of three state-of-the-art models on income estimation in NYC, and show that their validation R^2 values are lower than our proposed model (see Table 5). *Benchmark Model 1* utilizes

TABLE 4. Advantage of Mixed Siamese-like CNN relative to state-of-the-art models.

Classification	Paper	Model Architecture	Data Input	Advantage of Our Mixed Siamese-like CNN Relative to State-of-the-art Models
Visual Appearance	(Piaggesi *et al.*, 2019) [2]	VGG-F/Resnet50-based CNN, Ridge Regression	Daytime satellite image, nightlight	Our model does not take the nightlight as a proxy for the economic strength of a district, as nightlight intensity tends to be identical or similar across different districts of a developed metropolis. By so doing our model accuracy can be improved.
	(Suel *et al.*, 2018) [1]	VGG16-based CNN, GP Residual Regression	Street view, latitude and longitude of district centroid	Our Siamese-like CNN can outperform those extracted from the non-Siamese-like CNN, by extracting imagery features and feeding them into our model, which increases the model's correlation strength with income.
Transportation	(Gebru *et al.*, 2017) [38]	Deformable Part Model, CNN	Street view, car attribute dataset	Our model is not human capital intensive, and we do not take car attributes as a proxy for income, as cars identified from a district may not necessarily belong to the residents of the district. They can belong to someone who work there or someone who travel to the district for other purposes. Our model that depends on street view and house price information as model inputs may provide a more accurate prediction.
Food & Shopping	(Glaeser *et al.*, 2017) [42]	Random Forest	Yelp recommended review	Compared to the information obtained from Yelp's recommended review, the inputs of our model, including both house price and street view, have a higher correlation with district-level income, which gives higher income estimation accuracy.
Online Social Network Platform	(Hristova et al., 2016) [44]	Spearman Rank Correlation	Twitter record, Foursquare record	Our model, which does not rely on any input from the online social network platforms, reduces possible privacy concerns when data could have been collected from the online social networks.

TABLE 5. Comparison of state-of-the-art models (five-fold validation R^2, RMSE, MAE).

Model	Data Input	Relevance	Model Architecture	Tract-level						ZIP code-level					
				Per-capita Income			Median Household Income			Per-capita Income			Median Household Income		
				R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
GP Mixed Siamese-like Double Ridge	house price, daytime satellite image, street view, latitude and longitude of district centroid	districts where street views are captured	Siamese-like CNN (Resnet50), ridge regression, GP model	0.85	10905	6689	0.72	17665	12556	0.86	11525	7472	0.72	19245	12674
Benchmark Model 1 (Suel *et al.*, 2018) [1]	street view, latitude and longitude of district centroid	districts where street views are captured	VGG16-based CNN, GP residual Regression	0.84	11078	6607	0.67	18852	12914	0.84	12111	8185	0.70	19681	13176
Benchmark Model 2 (Glaeser *et al.*, 2015) [34]	street view	districts where street views are captured	Geometric layout algorithm; ν-support vector regression	0.46	20539	13102	0.30	27769	20169	0.66	19843	14388	0.55	25126	18313
Benchmark Model 3 (Piaggesi *et al.*, 2019) [2]	daytime satellite image, nightlight	all districts	Resnet50-based CNN, ridge regression	0.23	24519	15831	0.19	29815	21938	0.17	30209	23826	0.17	33832	25422

CNN for street view feature extraction and uses the spatial location information and GP for performing a residual regression task to boost the income estimation accuracy [1]. The model takes the four street views of each location as the inputs (views of the north, east, south, and west), Whereas in our model, we only take one street view of each location (views facing the searching center). Hence, we keep the street view locations constant between the two models to ensure model comparability.[3] *Benchmark Model 2* extracts 7480 pre-defined features from each street view (i.e. GIST [82], texton, and color histogram features [83]) and deploys a Support Vector Regression model for income estimation [35]. The same street views deployed in our model are taken as the input. The implementation is slightly different from the original work due to the training speed limitation. The original work first trains an image-level Support Vector Regression model and then averages the income estimation of each image at the district-level. Since our model feeds in a large number of images, our training speed becomes extremely slow. Hence, we first average the image features at the district-level, then train a district-level Support Vector Regression model to generate district-level income estimation. *Benchmark Model 3* is an income estimation model that utilizes the daytime satellite image and different CNN techniques [2]. To ensure comparability, the same daytime satellite image dataset used by our model is used as the input dataset for *Benchmark Model 3*. Specifically, *Benchmark Model 3* is based on a transfer learning technique, which uses CNN to first extract imagery features that are useful in estimating nightlight intensity, then applies the features for income estimation by Ridge Regression [2], [4]. It is seen in Table 5 that the *GP-Mixed-Siamese-like-Double-Ridge model* significantly outperforms the benchmark models.

[3] 7 locations are excluded as some locations do not have pictures covering four directions and some are lost when the Google database has been updated.

TABLE 6. Cross-district validation.

Model	Tract-level						ZIP code-level					
	Per-capita Income			Median Household Income			Per-capita Income			Median Household Income		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
GP Mixed Siamese-like Double Ridge	0.76	13987	8694	0.63	20253	14325	0.61	21024	13781	0.47	27935	18848
Mixed Siamese-like GP	0.69	15731	9776	0.56	22192	15762	0.71	19073	12590	0.46	29160	20109
Spatial Information GP	0.76	13640	8082	0.59	21269	14733	0.59	22374	14813	0.37	31504	21150
Mixed Siamese-like	0.69	15954	10026	0.54	22754	16514	0.77	17927	11983	0.49	28776	19367
Satellite Siamese-like	0.67	16215	10185	0.52	23091	16671	0.74	21332	15553	0.47	29522	20678
Street View Siamese-like	0.58	18633	11617	0.43	25355	18478	0.70	23374	15957	0.39	31449	21389

TABLE 7. Cross-scale validation.

	Per-capita Income			Median Household Income		
	R^2	RMSE	MAE	R^2	RMSE	MAE
GP Mixed Siamese-like Double Ridge	0.78	13091	8739	0.59	22588	16913
Mixed Siamese-like GP	0.75	13792	9189	0.58	22819	16914
Spatial Information GP	0.77	13466	8127	0.59	21514	15182
Mixed Siamese-like	0.75	14212	9640	0.57	22709	16818
Satellite Siamese-like	0.70	15390	10353	0.49	25151	18331
Street View Siamese-like	0.69	15546	10285	0.53	23082	17062

B. CROSS-DISTRICT VALIDATION

The cross-district validation results are presented in Table 6. The cross-district validation is conducted by randomly separating the small districts (tract/ZIP code) into five sets each set containing the same number of districts. Each time, the model is trained on input data (the daytime satellite image, the street view, the house price and the spatial location information) from one set of districts and is evaluated based on a validation dataset composed of all other districts. The average R^2, RMSE and MAE of five models trained on the five sets of district data are calculated and presented in Table 6. Here we find models sharing partially the architecture of our model, including the *Spatial-Information-GP model* and the *Mixed-Siamese-like model*, may outperform other models under different circumstances. Specifically, the *Spatial-Information-GP model* has achieved very good tract-level cross-district validation performance, when such validation is based on the per-capita income. However, it performs less well in ZIP code-level cross-district validation. As the optimal performance of the *Spatial-Information-GP model* depends on the availability of a relatively large training dataset, one-fifth of the ZIP code-level regions is too small a training dataset to maintain the model performance. Besides, the *Mixed-Siamese-like model*, which relies on satellite images, street views and house prices, has achieved an outstanding performance on ZIP code-level cross-district validation. Hence, when data of different spatial granularities are available, different models may be preferred. We would discuss this further in Section IV Part E. Further, our results also imply that when performing income estimation, instead of conducting a labor-intensive door-to-door survey across all districts, researchers can instead develop a CNN that estimates the income levels of a subset of all districts, say one-fifth, then use the trained CNN to estimate the income levels across the remaining four-fifths of the districts of the city.

C. CROSS-SCALE VALIDATION

Table 7 presents R^2, RMSE and MAE when the model trained on a less fine-grained spatial scale (ZIP code-level) is applied to income estimation at the more fine-grained scale (tract-level). With the same statistical significance, the less fine-grained scale model requires a smaller number of households to be interviewed, hence more resource friendly. The comparison results imply that the combined use of satellite images and street views can enhance the cross-scale validation accuracy. The *Spatial-Information-GP model* has achieved a good performance on cross-scale validation based on both per-capita and median household income.

D. ESTIMATED INCOME DISTRIBUTION

Fig. 6 presents the estimated income distribution, generated by the *GP-Mixed-Siamese-like-Double-Ridge model* following a five-fold validation procedure. We notice that our model generates a negative estimated median household income for a tract in the center of NYC. We investigate this problem by first checking the intermediate model outputs generated by the images and the spatial information. We find the scalar output generated by image features in Step 2 of our model is negative for this tract, and this leads to the final negative estimation. Hence, we further check the images in this tract. The street views in this tract look normal, whereas the satellite images in this tract contain some large shadows caused by high buildings. This indicates that dark building shadows might bring extra noises and reduce the model's estimation accuracy. We verify this idea by checking the estimated median household income generated by the *Satellite-Siamese-like model* and the *Street-view-Siamese-like model*. The *Satellite-Siamese-like model* also generates a negative estimated median household income for this tract, and the *Street-view-Siamese-like model* could generate a normal positive income estimation. Those results also support

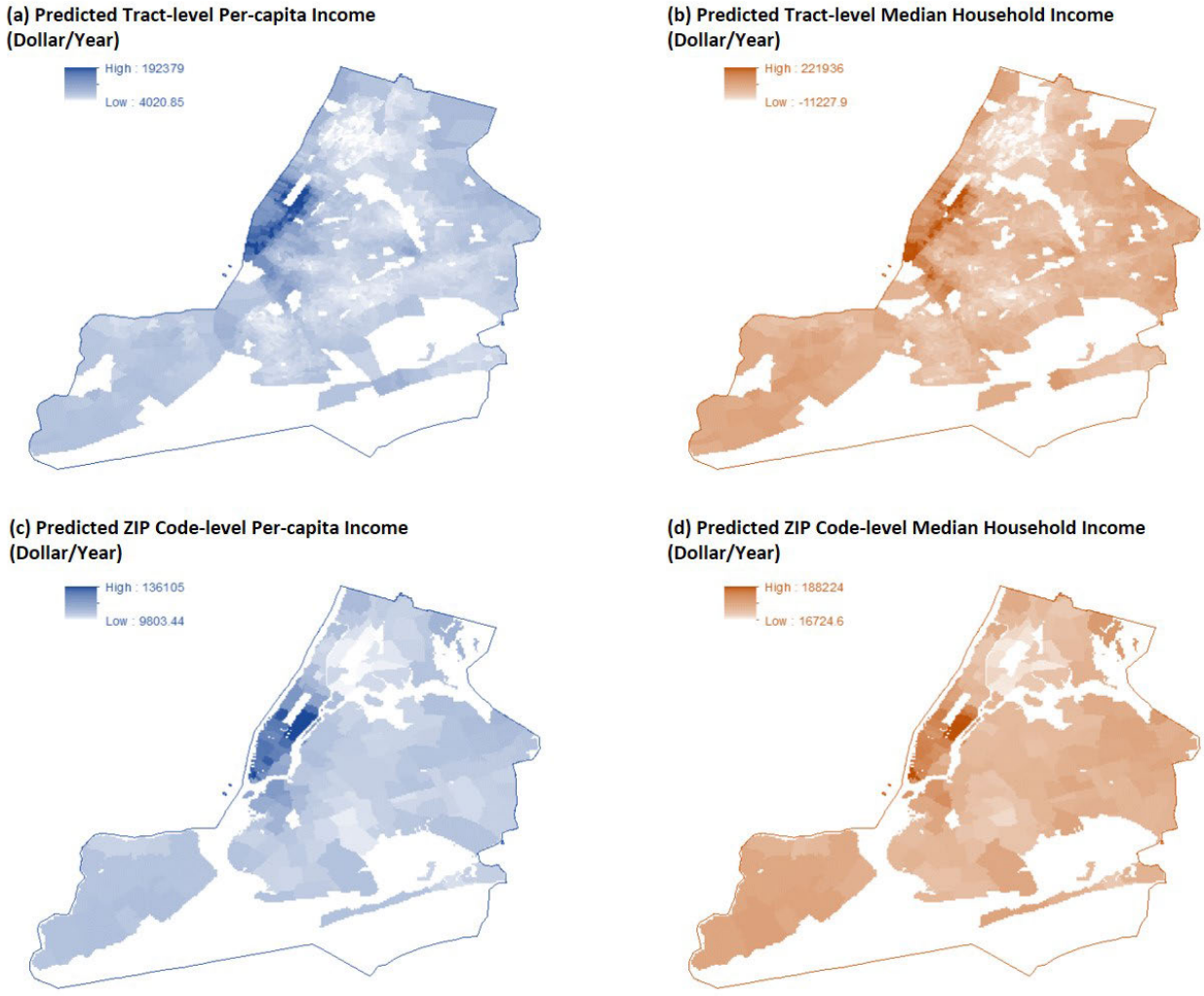

FIGURE 6. Predicted income distribution of NYC by Siamese-like CNN.

that abnormal satellite images would lead to low income estimation accuracy. Hence, we would suggest scholars filter out the daytime satellite images with large shadows before using our model, and we also encourage scholars to further investigate more specific criteria used for filtering abnormal daytime satellite images. Besides, any district that does not have ground truth data is not included in the Figure. Since some of these areas are covered/surrounded by water, whereas our model is trained on the ground truths collected from the land areas, the estimated income values in these districts are considered not credible and discarded.

E. DISCUSSION

Our study brings forth an important question, can we input additional data types as proxies to income, or integrate more powerful techniques to our model to improve our income estimation accuracy? First, as illustrated in Section I, other types of data that have been used in existing literature present various challenges; transportation card records are not open to the public [40], social media records can induce privacy concern [44], nightlights have a low correlation with incomes in developed countries [36], [58]–[61], and restaurant/business information exhibits a low estimation power in previous studies [42], [43]. Second, more powerful techniques, or additional data input types do not guarantee a higher estimation accuracy. For instance, studies from the Stanford University have shown that a powerful Generative Adversarial Network (GAN) and multiple types of data inputs may perform more poorly than a simple CNN, also perform less better than GAN with fewer input types, and have suggested that the failure may be attributable to the model's propensity of overfitting, as the model may be fitted to noise [80]. Third, an important performance evaluation criterion is whether the model maintains a good balance between simplicity and accuracy [84]. Our model achieves a relatively high R^2 of 0.72-0.86 (a five-fold validation), indicating that integrating more powerful models or providing extra inputs to this study may increase model complexity, but hardly improve estimation accuracy. However, income

TABLE 8. Computational burden of the income estimation models.

Computational Burden	Model
High	GP Mixed Siamese-like Double Ridge Mixed Siamese-like GP Mixed Siamese-like Random Forest Mixed Spatial Siamese-like Mixed Siamese-like Satellite Siamese-like Street View Siamese-like Benchmark Model 1 Benchmark Model 2
Medium	Mixed Non-Siamese-like Satellite Non-Siamese-like Street View Non-Siamese-like Benchmark Model 3
Low	Spatial Information GP House Price

estimation models tend to overfit when combining image features with spatial information. Hence, estimation models that better fit the spatial information with other regional features without overfitting, such as Graph Neural Network (GNN), are recommended for the future study [85].

Besides, the computational burdens of different models are compared and presented in Table 8. We divide the computational burden of our assessed models into three levels, 'high', 'medium' and 'low'. Models of a 'high' computational burden are models that concatenate a number of sub-models and spend long time on training and making inference (>6 hours). Models of 'low' computational burdens are models of a single sub-model and the results are generated within a short time (<1 hour). Other models falling between the two extremes are categorized as 'medium'. In our baseline models, many have been rated as ''high'' in computational costs. Unlike some real-time model training and inference operations, regional income estimation usually does not require high computational speed. Hence, such shortcoming might not affect income estimation operation seriously, even though efforts to reduce the computational costs can be further pursued in future studies.

Although our *GP-Mixed-Siamese-like-Double-Ridge model* has achieved a good estimation performance, several models sharing partially our proposed architecture can achieve comparable or even better performance under different circumstances (i.e. different spatial granularities and data availability). Hence, when data of different spatial granularities are available, different models may be preferred (see Table 9). Specifically, the high-performance models under high income data availability are selected according to the five-fold validation results (where field survey-based income data from 80% of districts are available). The high-performance models suitable to be used under low income data availability are selected according to the cross-district validation results (where 20% of the ground truth income data are available).

V. CONCLUSION AND FUTURE RESEARCH

We propose a novel methodology, Mixed Siamese-like CNN, which integrates Ridge Regression and GP for the fine-grained, district-level per-capita income and median household income estimation for NYC in the United States. Our new model (the *GP-Mixed-Siamese-like-Double-Ridge model*) makes good use of a rich array of data types, including the house price, the daytime satellite image, the street view and the spatial location information. Our model outperforms other state-of-the-art income estimation Regression models (R^2 = 0.72-0.86 under a five-fold validation). A good performance has been achieved with regard to cross-district and cross-scale validation, which can be used to replace field surveys to reduce manpower and financial resources. We also identify models that share partially the architecture of our model, including the *Spatial-Information-GP model* and the *Mixed-Siamese-like model*. Each of them can perform better than other baselines under certain spatial granularity and data availability. Since each of those models relies on less data input types and simpler architectures, utilizing them can save resources spent on data collection and model training. We recommend that these model architectures can be flexibly utilized under different circumstances to optimize the estimation performance.

TABLE 9. Suggested models under different conditions.

Availability of Ground Truth Income	Tract-level		ZIP code-level	
	Per-capita Income	Median Household Income	Per-capita Income	Median Household Income
High	GP Mixed Siamese-like Double Ridge; Spatial Information GP	GP Mixed Siamese-like Double Ridge	GP Mixed Siamese-like Double Ridge; Spatial Information GP	GP Mixed Siamese-like Double Ridge; Spatial Information GP
Low	GP Mixed Siamese-like Double Ridge; Spatial Information GP	GP Mixed Siamese-like Double Ridge	Mixed Siamese-like	Mixed Siamese-like

Even though our income estimation model is applicable to the developed economies, it can be modified and extended to developing countries where no historical fine-grained income data is readily available. First, instead of estimating the exact income, due to the lack of accurate ground truths, one can transform this model into an income classification model based on unsupervised learning. Second, other types of fine-grained data inputs can be utilized in the developing metropolises. Third, field surveys can be conducted in some districts of the targeted developing metropolis to verify model accuracy. To apply our model to estimate the fine-grained income values in other cities, it is necessary to perform verification. First, when any cities to be examined are lying within the same country, one may wish to check whether the geographical features, the population densities, the house prices and socio-economic status are showing similar characteristics. If the differences are significant, the model parameters trained in one city may not be directly transferrable. Even if the city characteristics are similar, field surveys are best conducted beforehand across some parts of the new city to be examined. Second, for cities located in a different country, it is desirable to examine if the general socio-economic features of the two countries are similar first.

In future, we plan to study how to transfer our proposed Siamese-like CNN to other unsupervised learning-based models. Our proposed Siamese-like CNN enjoys the following advantages. First, training a Siamese-like CNN does not need income data collected from field surveys, rendering it suitable for unsupervised learning. Second, house price is highly correlated with other socio-economic indicators, such as wealth [86]. Hence, the house price-related features extracted from Siamese-like CNN can facilitate the estimation of composite indicators that represent multi-dimensional concepts [87]. In reality, comprehensive house price information like that available in NYC may not be readily available in all cities. Hence, our model can still work well even when only a fraction, for instance, when one-fifth of house price information is available (see the cross-district validation result). Further, our Siamese-like CNN can also be used for estimating other socio-economic indicators. Depending on the nature and the type of the socio-economic values to be estimated, instead of using the house price for image feature extraction, multi-modal big data can be used for transfer learning. The features extracted from relevant dimensions can then serve as the inputs for unsupervised learning model (e.g. Principal Component Analysis [88]).

ACKNOWLEDGMENT

The authors would like to thank two anonymous reviewers for their detailed recommendations; Google for offering Street View Static API; National Oceanic and Atmospheric Administration (NOAA) for providing the nightlight intensity; ACS for collecting income data and Census Reporter for providing income data and spatial information; NYC Department of Finance for providing house price data; NYC government for providing the daytime satellite images of NYC, and the official searching tool, NYCityMap. Thanks also go to Google for providing Street View Static API, and NOAA for providing the nightlight intensity data for Benchmark Model 3.

REFERENCES

[1] E. Suel, M. Boulleau, M. Ezzati, and S. Flaxman, "Combining street imagery and spatial information for measuring socioeconomic status," in *Proc. Workshop Modeling Decis.-Making Spatiotemporal Domain, 32nd NIPS*, Montreal, QC, Canada, 2018, pp. 1–5.

[2] S. Piaggesi, L. Gauvin, M. Tizzoni, C. Cattuto, N. Adler, S. Verhulst, A. Young, R. Price, L. Ferres, and A. Panisson, "Predicting city poverty using satellite imagery," in *Proc. IEEE Conf. CVPR Workshops*, Jun. 2019, pp. 90–96.

[3] USCB, "Income in the past 12 months," in *American Community Survey and Puerto Rico Community Survey 2018 Subject Definitions*. Suitland, MD, USA: U.S. Census Bur, 2018, p. 81.

[4] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, vol. 353, no. 6301, pp. 790–794, Aug. 2016.

[5] J. E. Steele, P. R. Sundsøy, C. Pezzulo, V. A. Alegana, T. J. Bird, J. Blumenstock, J. Bjelland, K. Engø-Monsen, Y.-A. de Montjoye, A. M. Iqbal, K. N. Hadiuzzaman, X. Lu, E. Wetter, A. J. Tatem, and L. Bengtsson, "Mapping poverty using mobile phone and satellite data," *J. Roy. Soc. Interface*, vol. 14, no. 127, Feb. 2017, Art. no. 20160690.

[6] T. Gutierrez, G. Krings, and V. D. Blondel, "Evaluating socio-economic state of a country analyzing airtime credit and mobile phone datasets," 2013, *arXiv:1309.4496*. [Online]. Available: http://arxiv.org/abs/1309.4496

[7] S.-P. Kam, M. Hossain, M. L. Bose, and L. S. Villano, "Spatial patterns of rural poverty and their relationship with welfare-influencing factors in bangladesh," *Food Policy*, vol. 30, nos. 5–6, pp. 551–567, Oct. 2005.

[8] S. Chakravorty, "Urban inequality revisited: The determinants of income distribution in U.S. Metropolitan areas," *Urban Affairs Rev.*, vol. 31, no. 6, pp. 759–777, Jul. 1996.

[9] T.-K. Pfoertner, H.-J. Andress, and C. Janssen, "Income or living standard and health in Germany: Different ways of measurement of relative poverty with regard to self-rated health," *Int. J. Public Health*, vol. 56, no. 4, pp. 373–384, 2011.

[10] I. Gough, "Welfare regimes in East Asia and Europe," in *Proc. Annu. World Bank Conf. Develop. Econ. Eur.*, Washington, DC, USA, 2000, pp. 1–39.

[11] J. Gibson and G. Scobie, "A cohort analysis of household income, consumption and saving," *New Zealand Econ. Papers*, vol. 35, no. 2, pp. 196–216, Dec. 2001.

[12] O. P. Attanasio, *A Cohort Analysis of Saving Behavior by US Households*. Cambridge, MA, USA: National Bureau of Economic Research, 1993.

[13] J. K. Gibson and G. M. Scobie, "Household saving behaviour in New Zealand: A cohort analysis," New Zealand Treasury Working Paper 01/18, 2001.

[14] Y. Chen, M. Mazzocco, and B. Személy, "Explaining the decline of the US saving rate: The role of health expenditure," *Int. Econ. Rev.*, vol. 60, no. 4, pp. 1823–1859, Nov. 2019.

[15] BEA. (Dec. 15, 2019). *Personal Saving Rate*. [Online]. Available: https://www.bea.gov/data/income-saving/personal-saving-rate

[16] R. Bai, J. C. K. Lam, and V. O. K. Li, "A review on health cost accounting of air pollution in China," *Environ. Int.*, vol. 120, pp. 279–294, Nov. 2018.

[17] A. S. Tsui, G. Enderle, and K. Jiang, "Income inequality in the united states: Reflections on the role of corporations," *Acad. Manage. Rev.*, vol. 43, no. 1, pp. 156–168, Jan. 2018.

[18] S. J. Piotrowski and G. G. Van Ryzin, "Citizen attitudes toward transparency in local government," *Amer. Rev. Public Admin.*, vol. 37, no. 3, pp. 306–323, Sep. 2007.

[19] T. Gebru, J. Krause, Y. Wang, D. Chen, J. Deng, E. L. Aiden, and L. Fei-Fei, "Using deep learning and Google street view to estimate the demographic makeup of neighborhoods across the united states," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 50, pp. 13108–13113, Dec. 2017.

[20] T. Smith, M. Noble, S. Noble, G. Wright, D. McLennan, and E. Plunkett, *The English Indices of Deprivation 2015 Research Report*. London, U.K.: Department for Communities and Local Government, 2015.

[21] ACS. *American Community Survey*. Accessed: Feb. 11, 2020. [Online]. Available: https://americancommunitysurveye.com/

[22] L. Dong, S. Chen, Y. Cheng, Z. Wu, C. Li, and H. Wu, "Measuring economic activity in China with mobile big data," *EPJ Data Sci.*, vol. 6, no. 1, p. 29, Dec. 2017.
[23] J. Y. Campbell and J. F. Cocco, "How do house prices affect consumption? Evidence from micro data," *J. Monetary Econ.*, vol. 54, no. 3, pp. 591–621, Apr. 2007.
[24] M. M. Iacoviello, "Housing wealth and consumption," in *FRB International Finance Discussion Paper*. Washington, DC, USA: Board of Governors of the Federal Reserve System, 2011.
[25] J. Steegmans and W. Hassink, "Financial position and house price determination: An empirical study of income and wealth effects," *J. Housing Econ.*, vol. 36, pp. 8–24, Jun. 2017.
[26] J. Gallin, "The long-run relationship between house prices and income: Evidence from local housing markets," *Real Estate Econ.*, vol. 34, no. 3, pp. 417–438, Aug. 2006.
[27] M. Harter-Dreiman, "Drawing inferences about housing supply elasticity from house price responses to income shocks," *J. Urban Econ.*, vol. 55, no. 2, pp. 316–337, Mar. 2004.
[28] Y. Xu, A. Belyi, I. Bojic, and C. Ratti, "Human mobility and socioeconomic status: Analysis of Singapore and Boston," *Comput., Environ. Urban Syst.*, vol. 72, pp. 51–67, Nov. 2018.
[29] S. Ding, H. Huang, T. Zhao, and X. Fu, "Estimating socioeconomic status via temporal-spatial mobility analysis—A case study of smart card data," in *Proc. 28th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul. 2019, pp. 1–9.
[30] N. Määttänen and M. Terviö, "Income distribution and housing prices: An assignment model approach," *J. Econ. Theory*, vol. 151, pp. 381–410, May 2014.
[31] G. Piggott. (Dec. 15, 2015). *GLA Household Income Estimates*. [Online]. Available: https://data.london.gov.uk/blog/gla-household-income-estimates/
[32] J. Tan and Y. Zhao, "An empirical study of the housing-price-to-income ratio: Evidences from Beijing and China," *China Land Sci.*, vol. 26, no. 9, pp. 66–70, 2012.
[33] Y. Hu and L. Oxley, "Bubbles in US regional house prices: Evidence from house price–income ratios at the state level," *Appl. Econ.*, vol. 50, no. 29, pp. 3196–3229, Jun. 2018.
[34] W. Heo, J. M. Lee, N. Park, and J. E. Grable, "Using artificial neural network techniques to improve the description and prediction of household financial ratios," *J. Behav. Experim. Finance*, vol. 25, Mar. 2020, Art. no. 100273.
[35] E. L. Glaeser, S. D. Kominers, M. Luca, and N. Naik, "Big data and big cities: The promises and limitations of improved measures of urban life," *Econ. Inquiry*, vol. 56, no. 1, pp. 114–137, Jan. 2018.
[36] C. Mellander, J. Lobo, K. Stolarick, and Z. Matheson, "Night-time light data: A good proxy measure for economic activity?" *PLoS ONE*, vol. 10, no. 10, Oct. 2015, Art. no. e0139779.
[37] E. Suel, J. W. Polak, J. E. Bennett, and M. Ezzati, "Measuring social, environmental and health inequalities using deep learning and street imagery," *Sci. Rep.*, vol. 9, no. 1, p. 6229, Dec. 2019.
[38] T. Gebru, J. Krause, Y. Wang, D. Chen, J. Deng, and L. Fei-Fei, "Fine-grained car detection for visual census estimation," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4502–4508.
[39] T. Gebru, J. Krause, Y. Wang, D. Chen, J. Deng, and L. Fei-Fei, "Visual census: Using cars to study people and society," in *Proc. CVPR Bigvision*, 2015, pp. 1–2.
[40] C. Smith, D. Quercia, and L. Capra, "Finger on the pulse: Identifying deprivation using transit flow analysis," in *Proc. Conf. Comput. Supported Cooperat. Work CSCW*, 2013, pp. 683–692.
[41] N. Lathia, D. Quercia, and J. Crowcroft, "The hidden image of the city: Sensing community well-being from urban mobility," in *Proc. Int. Conf. Pervasive Comput.*, 2012, pp. 91–98.
[42] E. L. Glaeser, H. Kim, and M. Luca, *Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity*. Cambridge, MA, USA: National Bureau of Economic Research, 2017.
[43] J. Block, R. Scribner, and K. Desalvo, "Fast food, race/ethnicity, and incomeA geographic analysis," *Amer. J. Preventive Med.*, vol. 27, no. 3, pp. 211–217, Oct. 2004.
[44] D. Hristova, M. J. Williams, M. Musolesi, P. Panzarasa, and C. Mascolo, "Measuring urban social diversity using interconnected geo-social networks," in *Proc. 25th Int. Conf. World Wide Web WWW*, 2016, pp. 21–30.
[45] D. J. Brass, "Men's and Women's networks: A study of interaction patterns and influence in an organization," *Acad. Manage. J.*, vol. 28, no. 2, pp. 327–343, Jun. 1985.
[46] R. S. Burt, *Structural Holes: The Social Structure of Competition*. Cambridge, MA, USA: Harvard Univ. Press, 2009.
[47] P. V. Marsden and J. S. Hurlbert, "Social resources and mobility outcomes: A replication and extension," *Social Forces*, vol. 66, no. 4, pp. 1038–1059, Jun. 1988.
[48] H. D. Flap and N. D. De Graaf, "Social capital and attained occupational status," *Netherlands J. Sociology*, vol. 22, no. 2, pp. 145–161, 1986.
[49] S. E. Seibert, M. L. Kraimer, and R. C. Liden, "A social capital theory of career success," *Acad. Manage. J.*, vol. 44, no. 2, pp. 219–237, Apr. 2001.
[50] D. L. Sexton, *The Art and Science of Entrepreneurship*. Pensacola, FL, USA: Ballinger Pub Co, 1986.
[51] P. Dubini and H. Aldrich, "Personal and extended networks are central to the entrepreneurial process," *J. Bus. Venturing*, vol. 6, no. 5, pp. 305–313, Sep. 1991.
[52] C. D. Elvidge, P. C. Sutton, T. Ghosh, B. T. Tuttle, K. E. Baugh, B. Bhaduri, and E. Bright, "A global poverty map derived from satellite data," *Comput. Geosci.*, vol. 35, no. 8, pp. 1652–1660, Aug. 2009.
[53] J. V. Henderson, A. Storeygard, and D. N. Weil, "Measuring economic growth from outer space," *Amer. Econ. Rev.*, vol. 102, no. 2, pp. 994–1028, Apr. 2012.
[54] J. Wu, C. Wang, X. He, X. Wang, and N. Li, "Spatiotemporal changes in both asset value and GDP associated with seismic exposure in China in the context of rapid economic growth from 1990 to 2010," *Environ. Res. Lett.*, vol. 12, no. 3, Mar. 2017, Art. no. 034002.
[55] S. Basihos, "Nightlights as a development indicator: The estimation of gross provincial product (GPP) in Turkey," in *Proc. SSRN*, May 2016, Art. no. 2885518.
[56] X. Wang, M. Rafa, J. Moyer, J. Li, J. Scheer, and P. Sutton, "Estimation and mapping of sub-national GDP in uganda using NPP-VIIRS imagery," *Remote Sens.*, vol. 11, no. 2, p. 163, Jan. 2019.
[57] T. Bundervoet, L. Maiyo, and A. Sanghi, *Bright Lights, Big Cities: Measuring National and Subnational Economic Growth in Africa From Outer Space, With an Application to Kenya and Rwanda*. Washington, DC, USA: World Bank, 2015.
[58] A. Venerandi, G. Quattrone, L. Capra, D. Quercia, and D. Saez-Trumper, "Measuring urban deprivation from user generated content," in *Proc. 18th ACM Conf. Comput. Supported Cooperat. Work Social Comput. CSCW*, 2015, pp. 254–264.
[59] NOAA. *DMSP-OLS Nighttime Lights Time Series*. Accessed: Jan. 15, 2019. [Online]. Available: http://www.ngdc.noaa.gov/eog/dmsp/downloadV4composites.html
[60] J. Graesser, A. Cheriyadat, R. R. Vatsavai, V. Chandola, J. Long, and E.Bright, "Image based characterization of formal and informal neighborhoods in an urban landscape," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1164–1176, Aug. 2012.
[61] R. Engstrom, J. Hersh, and D. Newhouse, *Poverty in HD: What Does High Resolution Satellite Imagery Reveal About Economic Welfare*. Washington, DC, USA: World Bank, 2016.
[62] DHS. (Mar. 25, 2020). *Wealth Index Construction*. [Online]. Available: https://www.dhsprogram.com/topics/wealth-index/Wealth-Index-Construction.cfm
[63] S. Law, B. Paige, and C. Russell, "Take a look around: Using street view and satellite images to estimate house prices," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 5, pp. 1–19, Nov. 2019.
[64] CR. *Income*. Accessed: Feb. 1, 2020. [Online]. Available: https://censusreporter.org/topics/income/
[65] NYCDF. *Annualized Sales Update*. Accessed: Feb. 11, 2020. [Online]. Available: https://www1.nyc.gov/site/_nance/taxes/property-annualized-sales-update.page
[66] NYCGOV. *NYCityMap*. Accessed: Feb. 12, 2020. [Online]. Available: http://maps.nyc.gov/doitt/nycitymap/
[67] NYCGOV. *NYC Then and Now*. Accessed: Nov. 21, 2019. [Online]. Available: https://maps.nyc.gov/then&now/
[68] S. Law, Y. Shen, and C. Seresinhe, "An application of convolutional neural network in street image classification: The case study of london," in *Proc. 1st Workshop Artif. Intell. Deep Learn. Geographic Knowl. Discovery GeoAI*, 2017, pp. 5–9.
[69] S. Jialin Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[70] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[71] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4353–4361.
[72] A. Dubey, N. Naik, D. Parikh, R. Raskar, and C. A. Hidalgo, "Deep learning the city: Quantifying urban perception at a global scale," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 196–212.
[73] W. Si and D. Zhu, "Discussion on the income-house price ratio index," *Construct. Economy*, vol. 11, no. 6, pp. 33–35, Nov. 2007.
[74] M.-C. Chen, I.-C. Tsai, and C.-O. Chang, "House prices and household income: Do they move apart? Evidence from taiwan," *Habitat Int.*, vol. 31, no. 2, pp. 243–256, Jun. 2007.
[75] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[76] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, Feb. 1970.
[77] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*, vol. 3. Cambridge, MA, USA: MIT Press, 2006.
[78] GPy. (2012). *GPy: A Gaussian Process Framework in Python*. [Online]. Available: http://github.com/SheffieldML/GPy
[79] A. Head, M. Manguin, N. Tran, and J. E. Blumenstock, "Can human development be measured with satellite imagery?" in *Proc. 9th Int. Conf. Inf. Commun. Technol. Develop.*, Nov. 2017, pp. 1–11.
[80] A. Perez, S. Ganguli, S. Ermon, G. Azzari, M. Burke, and D. Lobell, "Semi-supervised multitask learning on multispectral satellite images using wasserstein generative adversarial networks (GANs) for predicting poverty," 2019, *arXiv:1902.11110*. [Online]. Available: http://arxiv.org/abs/1902.11110
[81] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[82] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.
[83] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, "SUN database: large-scale scene recognition from abbey to zoo," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3485–3492.
[84] B. Y. Lee and S. M. Bartsch, "How to determine if a model is right for neglected tropical disease decision making," *PLOS Neglected Tropical Diseases*, vol. 11, no. 4, Apr. 2017, Art. no. e0005457.
[85] F. Scarselli, M. Gori, A. Chung Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2009.
[86] J. C. K. Lam, Y. Han, R. Bai, V. O. K. Li, J. Leong, and K. J. Maji, "Household wealth proxies for socio-economic inequality policy studies in China," *Data Policy*, vol. 2, no. 1, pp. 1–21, Jan. 2020.
[87] OECD, "The OECD-JRC handbook on practices for developing composite indicators," presented at the OECD Committee Statist., Paris, France, Jun. 2004.
[88] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.

RUIQIAO BAI was born in Beijing, China. In 2017, she received B.Eng. in Environmental Engineering (Global Environment Program) from Tsinghua University in Beijing, China. She is currently pursuing a Ph.D. in Electrical and Electronic Engineering at the University of Hong Kong. Her research interests include big data, machine learning, AI, interdisciplinary socio-economic and environmental studies.

JACQUELINE C. K. LAM (Member, IEEE) is a Visiting Scholar at the Department of Computer Science and Technology, University of Cambridge, also an Associate Professor at the Department of Electrical and Electronic Engineering, the University of Hong Kong. She was visiting MIT CEEPR as a Visiting Fellow in 2019. Since 2018 Jacqueline has co-established HKU-AI WiSe with Prof. Victor OK Li, Chair of Information Engineering at the University of Hong Kong. She is the Co-Director of HKU-Cambridge Clean Energy and Environment Research Platform, and of HKU-Cambridge AI to Advance Well-being and Society Research Platform. She was the Hughes Hall Visiting Fellow, before she takes up the Visiting Senior Research Fellow and Associate Researcher in Energy Policy Research Group, Judge Business School in Cambridge. Her research focuses on the interface between AI, society and environment, and their policy applications for the US, the UK and China. Her recent research uses big data and machine learning techniques to study personalized air pollution monitoring and health management. Her work is published in IEEE, *Environment International, Applied Energy, Environmental Science and Policy*, and *Energy Policy*. Jacqueline has received four times the research grants awarded by the Research Grants Council, HKSAR Government, during 2011-2020. The funded amount was USD 7.9 M in PI/Co-PI capacity. In 2019, she co-co-organized the first HKU-Cambridge AI for Social Good Symposium, with Prof. Jonathan Crowcroft, FRS (CST Dept., Cambridge) and Prof. Victor OK Li (Clare Hall Life Fellow, Cambridge). Jacqueline also serves as the co-editor of a special issue on AI for environmental decision-making, published by the SCI journal, *Environmental Science and Technology*. Her interdisciplinary research study, in joint collaboration with Yang Han and Victor OK Li on $PM_{2.5}$ and environmental inequality in Hong Kong, was published in Environmental Science and Policy, and widely covered by more than 30 local and overseas newspapers and TVs.

VICTOR O. K. LI (Life Fellow, IEEE) received SB, SM, EE and ScD degrees in Electrical Engineering and Computer Science from MIT. Prof. Li is Chair of Information Engineering and Cheng Yu-Tung Professor in Sustainable Development at the Department of Electrical & Electronic Engineering (EEE) at the University of Hong Kong. He is the Director of the HKU-Cambridge Clean Energy and Environment Research Platform, and of the HKU-Cambridge AI to Advance Wellbeing and Society Research Platform, which are interdisciplinary collaborations with Cambridge University. He was Visiting Professor in the Department of Computer Science and Technology at the University of Cambridge from April to August 2019. He was the Head of EEE, Assoc. Dean (Research) of Engineering and Managing Director of Versitech Ltd. He serves on the board of Sunevision Holdings Ltd., listed on the Hong Kong Stock Exchange and co-founded Fano Labs Ltd., an artificial intelligence (AI) company with his PhD student. Previously, he was Professor of Electrical Engineering at the University of Southern California (USC), Los Angeles, California, USA, and Director of the USC Communication Sciences Institute. His research interests include big data, AI, optimization techniques, and interdisciplinary clean energy and environment studies. In Jan 2018, he was awarded a USD 6.3M RGC Theme-based Research Project to develop deep learning techniques for personalized and smart air pollution monitoring and health management. Sought by government, industry, and academic organizations, he has lectured and consulted extensively internationally. He has received numerous awards, including the PRC Ministry of Education Changjiang Chair Professorship at Tsinghua University, the UK Royal Academy of Engineering Senior Visiting Fellowship in Communications, the Croucher Foundation Senior Research Fellowship, and the Order of the Bronze Bauhinia Star, Government of the HKSAR. He is a Fellow of the Hong Kong Academy of Engineering Sciences, the IEEE, the IAE, and the HKIE.