Granularity at Scale: Estimating Neighborhood Socioeconomic Indicators From High-Resolution Orthographic Imagery and Hybrid Learning

Ethan Brewer, Giovani Valdrighi, Parikshit Solunke, Joao Rulff, *Graduate Student Member, IEEE*, Yurii Piadyk, Zhonghui Lv, Jorge Poco, *Member, IEEE*, and Claudio Silva, *Fellow, IEEE*

***Abstract*—Many areas of the world are without basic information on the socioeconomic well-being of the residing population due to limitations in existing data collection methods. Overhead images obtained remotely, such as from satellite or aircraft, can help serve as windows into the state of life on the ground and help "fill in the gaps" where community information is sparse, with estimates at smaller geographic scales requiring higher resolution sensors. Concurrent with improved sensor resolutions, recent advancements in machine learning and computer vision have made it possible to quickly extract features from and detect patterns in image data, in the process correlating these features with other information. In this work, we explore how well two approaches—a supervised convolutional neural network and semisupervised clustering based on bag-of-visual-words—estimate population density, median household income, and educational attainment of individual neighborhoods from publicly available high-resolution imagery of cities throughout the United States. Results and analyses indicate that features extracted from the imagery can accurately estimate the density (R^2 up to 0.81) of neighborhoods, with the supervised approach able to explain about half the variation in a population's income and education. In addition to the presented approaches serving as a basis for further geographic generalization, the novel semisupervised approach provides a foundation for future work seeking to estimate fine-scale information from aerial imagery without the need for label data.**

***Index Terms*—Aerial imagery, computer vision, deep learning, remote sensing, sustainable development.**

Manuscript received 13 September 2023; revised 11 December 2023; accepted 13 February 2024. Date of publication 21 February 2024; date of current version 6 March 2024. This work was supported in part by the National Council for Scientific and Technological Development, CNPq under Grant #311144/2022-5, in part by the Carlos Chagas Filho Foundation for Research Support of Rio de Janeiro State, FAPERJ under Grant #E-26/201.424/2021, in part by the Sáo Paulo Research Foundation, FAPESP, under Grant #2021/07012-0, and in part by the School of Applied Mathematics at Fundação Getulio Vargas. *(Corresponding author: Ethan Brewer.)*

Ethan Brewer is with the Spectral Sciences, Inc., Burlington MA 01803-3304 USA (e-mail: ebrewer@spectral.com).

Giovani Valdrighi and Jorge Poco are with the Fundação Getulio Vargas, Rio de Janeiro 20000-000, Brazil (e-mail: giovani.valdrighi@fgv.br; jorge.poco@fgv.br).

Parikshit Solunke, Joao Rulff, Yurii Piadyk, and Claudio Silva are with the New York University, New York NY 10012 USA (e-mail: parikshit.s@nyu.edu; jlrulff@nyu.edu; ypiadyk@nyu.edu; csilva@nyu.edu).

Zhonghui Lv is with the William and Mary, Williamsburg VA 23185 USA (e-mail: zlv@wm.edu).

Digital Object Identifier 10.1109/JSTARS.2024.3368018

I. Introduction

CENSUSES and other surveys administered to collect socioeconomic data are expensive and time-consuming [1]. For this reason, there is often an undesirably long gap between surveys in developing countries, hindering the appropriate formulation of public policies. The ability to measure socioeconomic metrics is essential for evaluating progress toward targets (such as the United Nations' Sustainable Development Goals [2]), promoting accountability, enabling evidence-based decision-making, and providing a basis for informed actions and interventions to improve human well-being [3], [4]. Measurements help determine areas and populations that require the most attention and resources [5]. By quantifying development indicators such as urbanization, education levels, healthcare access, and income, policymakers and development practitioners can identify the most vulnerable and disadvantaged groups and design targeted interventions to address their specific needs and optimize resource allocation [5], [6], [7].

Such metrics are traditionally measured through national accounts data, household surveys, and administrative records, such as tax filings [1]. Even in regions where such data are collected, they are expensive [8]. Furthermore, at the neighborhood level, socioeconomic conditions can undergo drastic changes in very short periods of time [9]. Hence, it is imperative to explore faster and more cost-effective techniques for estimating vital socioeconomic outcomes between neighborhoods. Since 1990s, analysis of nighttime light intensity from remote sensing technologies, such as from sensors onboard satellites and aircraft, has effectively contributed to approximating development and socioeconomic metrics at large scales [10], [11]. Beginning around 2016, analysis of remotely sensed data (often high-resolution[1] daytime imagery) with machine learning methods, particularly neural networks, has rapidly grown in popularity and enabled finer scale approximations. These techniques have broadly illustrated that analyzing aerial imagery with machine learning is an effective strategy to remotely monitor the natural and built environment and to estimate and track development and socioeconomic statistics [12]. Technical challenges arise in

[1] We define high resolution to be each pixel dimension representing a geographic length of ≤ 3 m.

this line of research as aerial images can cover vast areas and, at higher resolutions, be too large to input directly into computer vision networks. This complexity is further compounded by the often irregular shapes of neighborhoods, and the aggregation of survey statistics at different levels across various geographical areas.

Our objective: In this article, we investigate the potential of machine learning models trained on high-resolution aerial imagery to estimate the following metrics at the U.S. Census block group level (approximately the size of a neighborhood).

1) Population density.
2) Median household income (MHI).
3) Educational attainment (% of the population with at least a bachelor's degree).

This article is a feasibility study in how well aerial photography and machine learning can detect where and how people live between neighborhoods. UN sustainable development goals that may benefit specifically from the measurement of neighborhood density, income, and education conditions include goals 1 (poverty reduction), 4 (well-being improvement), and 10 (within-country inequality reduction) [2].

We carry out our investigation by training and testing models on 94 of the 100 largest U.S. cities by gross domestic product (GDP). By automatically extracting spatial features in urban settings, variations in city infrastructure, such as roads, parks, and buildings, can be quantified and related to the census variables. Two methodologies are employed—one driven by a supervised convolutional neural network (CNN) and the other by a semisupervised framework utilizing bag-of-visual-words (BoVW) to generate simplified but interpretable representations of census blocks.

The most notable **contributions** of this article are as follows.

1) Demonstration of the ability of contemporary aerial imagery to resolve features related to socioeconomic variables at the scale of a neighborhood.
2) Finding that supervised learning and semisupervised clustering of image patches can respectively explain 81% and 61% of the variation in neighborhood population density.

Paper structure: The rest of this article is organized as follows. Section II provides an overview of the existing literature on estimating poverty, population, and other socioeconomic indicators from aerial imagery with machine learning. In Section III, we detail how the image and annotation data are acquired, fused, and processed. We describe our methods in Section IV, including supervised and semisupervised learning approaches. Next, we present results in Section V, and analyze and discuss the limitations of our work in Section VI. Finally, Section VII concludes this article.

II. Related Work

Throughout much of the world, there is a lack, or a complete absence, of data on the social and economic well-being of people due to conflict, natural disasters, pandemics, and the effort, expense, and time periods between surveys [1], [13]. In recent years, remotely sensed images in combination with machine learning have helped fill in these critical information gaps. Use of such techniques has extended into the estimation of population [14], wealth [15], poverty [13], conflict [16], migration [17], education [18], land use [19], and infrastructure [20], [21], among other applications [22], [23]. In this section, we focus on studies related to poverty, population density, education, and related metrics. Income and wealth are correlated with self-reported happiness and well-being [24], [25]. Postsecondary educational attainment is associated with higher levels of income [26], [27], satisfaction with life [28], and lifelong well-being [29]. Educational attainment for women of reproductive age is linked to reduced child and maternal mortality, lower fertility, and improved reproductive health [30]. Existing studies show a mixed correlation between population density and quality of life [31], [32], [33]. Findings within the city of Oslo, Norway suggest that, compared to residents of lower density neighborhoods, residents in higher density neighborhoods have higher levels of personal relationship satisfaction and perceived physical health, similar levels of leisure satisfaction, but lower levels of emotional response to neighborhood and higher levels of anxiety driven primarily by noise and safety concerns [33].

Poverty: Gaining momentum in the mid-2010s, several studies have focused on estimating poverty. In [34], a fully CNN was trained to predict nighttime light intensity from daytime imagery, simultaneously learning features that are useful for poverty prediction at 1 km resolution in Uganda. The model identified different terrains and manmade structures, including roads, buildings, and farmlands, without supervision beyond nighttime lights. Their results approached the predictive performance of survey data collected in the field. Jean et al. [13] showed that a CNN trained on Google Maps daytime imagery and existing survey data can identify image features that can explain 37%–75% of the variation in local-level economic outcomes, such as wealth and consumption across countries in Africa. In [35], household wealth in Bangladesh was estimated at 10 km resolution with random forest regression from multiple sources such as nighttime lights, daytime imagery, and land cover maps. Yeh et al. [15] showed multispectral 30-m Landsat imagery can help estimate African village wealth in countries where the model was not trained with errors comparable to existing ground data. Other related work to identify poverty via remote sensing and machine learning include [36], which validated a slum severity index using gray level co-occurrence matrix features extracted from high-resolution satellite images of Mexico, [37] which used high-resolution imagery and geospatial covariates to characterize degrees of intraurban deprivation in Nairobi, Kenya ($R^2 = 0.65$), and [38] in which deprivation in Liverpool, U.K. was measured by extracting features from Google Earth images ($R^2 = 0.54$). Additional related poverty studies include [39], [40], [41], and [42].

Population: Commonly used techniques for small-area population estimation typically redistribute population "top-down" from higher to lower administrative units using areal weighting interpolation or dasymetric mapping techniques [43]. Open-source population products that use this approach include Landscan, Meta's high resolution settlement layer, gridded population of the world, WorldPop, and global human settlement layer. Existing studies have focused on redistributing population counts using a random forest-based weighting scheme in Cambodia, Vietnam, and Kenya [44], redistributing population density in Peru using satellite imagery-based covariates employing regression and tree-based methods [45], and downscaling population counts using one billion mobile phone call records from Portugal and France [46]. Most population density studies do not validate the accuracy of their estimates against a census [43]. Other studies have used coarse nighttime lights for large administrative areas [47], 3-D city models [48], or focused on a subset of the population (such as children under five years of age) [49]. Moving further, [43] estimated local population density for in-between census years in Bangladesh by combining household surveys with geospatial data, including an assortment of satellite imagery-based indicators. The data were analyzed with Poisson regression models, with out-of-sample results approximating the density of subdistricts (larger than a village) with an R^2 of up to 0.83.

Additional metrics: Other closely related work includes [50] in which a Siamese-like CNN, integrating ridge regression and Gaussian process regression, was developed for the estimation of income for districts and zip codes in New York City. Their model makes use of a pairwise comparison of location-based house price information, daytime satellite images, street views, and spatial location information, achieving an R^2 of 0.72 at the census tract level. Castro and Álvarez [1] used daytime and nighttime satellite imagery and transfer learning to estimate average income, GDP per capita, and a water index at the city level in two Brazilian states, explaining up to 64% of the variation in the target variables. In [51], the authors estimate American Community Survey (ACS) socioeconomic variables, such as income, race, education, and voting patterns in 200 U.S. cities at the zip code and precinct level solely through 50 million images of street scenes from Google Street View and computer vision detection of the make, model, and year of all motor vehicles present in the images. Finally, Graetz et al. [30] explored educational inequalities across Africa by estimating years of schooling across a 5 × 5 km grid based on geocoded survey data, generating estimates of average educational attainment by age and sex.

Bag-of-visual-words: Widely used in natural language processing, bag-of-words is a numerical representation of text by counting individual words [52]. Despite being a simple formulation, this methodology has shown positive results in diverse language tasks. Inspired by it, computer vision studies have proposed an adaption called bag-of-features or BoVW that have shown positive results in natural scene classification [53], [54]. By creating a set of low-level visual "features" that describe the images, the frequency of the visual features in each image can be used for predictive tasks. The method has also shown positive results [55], [56], [57] in the domain of remote sensing imagery. Despite this existing work, these techniques have not been tested to estimate high-resolution census variables.

Most existing remote sensing studies analyze variables for entire nations, states, or cities, potentially obscuring neighborhood-level prosperity and inequality patterns. Our study pushes beyond the limitations of existing work by investigating population density, income, and education at an unprecedentedly precise scale across a country, using free, publicly available data.

III. Data

In this section, we detail how the data for the 94 cities are collected and processed.

Imagery: For the United States, orthographic imagery is retrieved from the National Agriculture Imagery Program (NAIP) [58], administered by the U.S. Department of Agriculture. For a given point in the U.S., RGBIR aerial imagery is acquired approximately every 2–3 years at a resolution of 60-cm ground sample distance during the agricultural growing season, or "leaf on" conditions. The images are orthorectified, which combines the image characteristics of an aerial photograph with the georeferenced qualities of a map. We utilize the most recent NAIP tiles (2019–2021) for 94 prominent cities across the United States, including the ten largest (by GDP). Our selection process involved filtering cities of the 100 largest metropolitan statistical areas (MSAs) based on 2021 GDP, followed by identifying the largest city by area within each MSA. This selection is used to study a diverse set of cities while considering the most significant ones.

Annotation data: Census data for the United States are acquired through the ACS [59]. Every year, the U.S. Census Bureau contacts approximately 3.5 million households (1 in 40 total households) across the country to participate in the ACS, with a 2021 response rate of 85.3%. The survey includes various demographic, social, economic, and housing data on residents, such as age, race, occupation, income, disability status, housing type (e.g., single-family, multiunit), languages spoken, and highest degree earned. The resulting data products are aggregated at various levels from country, to state, to county, to census tract, to block group. In this study, data are examined at the finest possible level, block group, to extract the maximum benefit from the resolution of the imagery. Block groups contain an average of approximately 1500 residents and may henceforth be referred to as neighborhoods.

The following five-year[2] ACS variables for neighborhoods are downloaded via API for all counties containing the cities (since the ACS aggregates by county but not city) from the year their associated imagery was captured.

1) Total population, P_t.
2) Population >25 years old, P_{25}.
3) MHI.
4) Four education variables: Population >25 years old whose highest degree completed is: i) bachelor's, P_b, ii) master's, P_m, iii) professional, P_p, iv) doctoral, P_d

An educational attainment metric, E, representing the percent of the population with at least a bachelor's degree is calculated

[2] Five-year estimates aggregate data from the preceding 60-month period. For example, a five-year estimate from the 2021 ACS aggregates data from Jan 1, 2017 to Dec 31, 2021 [60].

TABLE I
NEIGHBORHOOD SPECIFICATIONS

Specification	**Value**
# images	43,497
Image width range	103–17 101pixels
Median width	1 353 pixels
σ in width	1 124 pixels
Image height range	137–22 174 pixels
Median height	1 350 pixels
σ in height	1 141 pixels
Density range	2–47 791 $\frac{\text{ppl}}{\text{km}^2}$
Median density	634 $\frac{\text{ppl}}{\text{km}^2}$
Mean density	1 457 $\frac{\text{ppl}}{\text{km}^2}$
σ in density	2 519 $\frac{\text{ppl}}{\text{km}^2}$
MHI range	2499–250 001 USD
Median MHI	63 232 USD
Mean MHI	72 997 USD
σ in MHI	42 595 USD
Educational attainment range	0.0–100
Median educational attainment	31.3
Mean educational attainment	36.2
σ in educational attainment	24.7

with

$$E = \frac{P_b + P_m + P_p + P_d}{P_{25}} \times 100. \tag{1}$$

The population density metric, D, representing people per square kilometer is calculated with

$$D = 10^6 \frac{P_t}{A} \tag{2}$$

where A is the geographic area of a neighborhood in square meters found by

$$A = \frac{\text{\# nonzero pixels in image}}{\text{Resolution of image}}. \tag{3}$$

See Table I for all specifications of the neighborhoods analyzed.

Image-label pairing: To generate geographic boundaries for the imagery, shapefiles of the neighborhoods are downloaded from the U.S. Census Bureau's TIGER archive [61] (see Fig. 1). This geographic information (polygons of neighborhoods) is then merged with their corresponding ACS variables. In the process, neighborhoods containing census errors or zero population are dropped.

Fig. 2(a) shows all the cities examined (in orange outlines) overlaid with the ACS MHI by neighborhood, as an illustration. The cities of New York and Chicago are enlarged in Fig. 2(b) and (c) to provide a more detailed view.

Next, the imagery is cropped by neighborhood based on the bounding boxes of the neighborhood polygons. This results in a total of 43 497 images (see Table I for other specifications on the neighborhood crops).

Crop processing: The crops are processed for CNN input for the supervised method in two ways, "patching" and "resizing" (an example is visualized in Fig. 3).

Patching: With this technique, neighborhoods are split into 512 × 512 patches, as in Fig. 3(a). If either of the original dimensions of an image is not a multiple of 512, it is padded by zeros before being split. Only patches composed of >50% nonzero pixels are kept. This results in a total of 339 413 patches. Patching allows an image of a neighborhood to retain its resolution and shape, but results in the CNN treating each patch as a separate image, thus breaking apart the spatial relationship within a neighborhood.

TABLE II
DATASET SIZES

Dataset	**Method**		
	Sup. patching	Sup. resizing	Semi sup.
Train (# images)	237 589	30 447	1 364 790
Validation (# images)	50 911	6524	294 483
Test (# images)	50 913	6526	293 322
Total (# images)	339 413	43 497	1 952 595

Resizing. With this technique, neighborhoods are resized (through bilinear interpolation) to the median size of a neighborhood, i.e., a width of 1353 pixels and a height of 1350 pixels. Resizing allows a neighborhood to be read as a single image by the CNN, but results in upsampling/downsampling for crops that have a dimension(s) less/greater than the median, and shape distortion for crops with a $\frac{\text{width}}{\text{height}}$ ratio different from $\frac{1353}{1350}$.

Semisupervised. In the semisupervised methodology, the imagery is cropped into a square grid, each cell measuring 112 × 112 pixels. Therefore, each neighborhood is composed of a mosaic of patches. This high granularity serves the purpose of separating distinct urban structures within each patch. Due to the NAIP tiles covering areas not examined in this study, only the subset of obtained patches that have an intersection to any neighborhood are used. In addition, considering that some neighborhoods are very large (i.e., they have low population density), further filtering is implemented to limit the maximum number of patches to 50 per neighborhood to reduce computational costs. The patches are semirandomly selected at a probability proportional to the percentage of their area contained in the neighborhood. As a result of these filtering measures, around two million 112 × 112 patches are generated (see Table II). Fig. 4 depicts two example neighborhoods and their division of patches.

For both supervised and semisupervised approaches, the resulting data are separated into training, validation, and testing sets for model input in 70%–15%–15% splits. Dataset sizes are shown in Table II.

IV. METHODS

We now present the formulations of each methodology and their details.

A. Supervised

The overall supervised methodology for both the image patching and resizing techniques is executed in the following manner.

1) A ResNet50-based [62] architecture (shown in Fig. 5) is trained in separate instances on each of the three target variables (density, MHI, education).
2) The trained networks are evaluated on the independent test set not included in the training process.

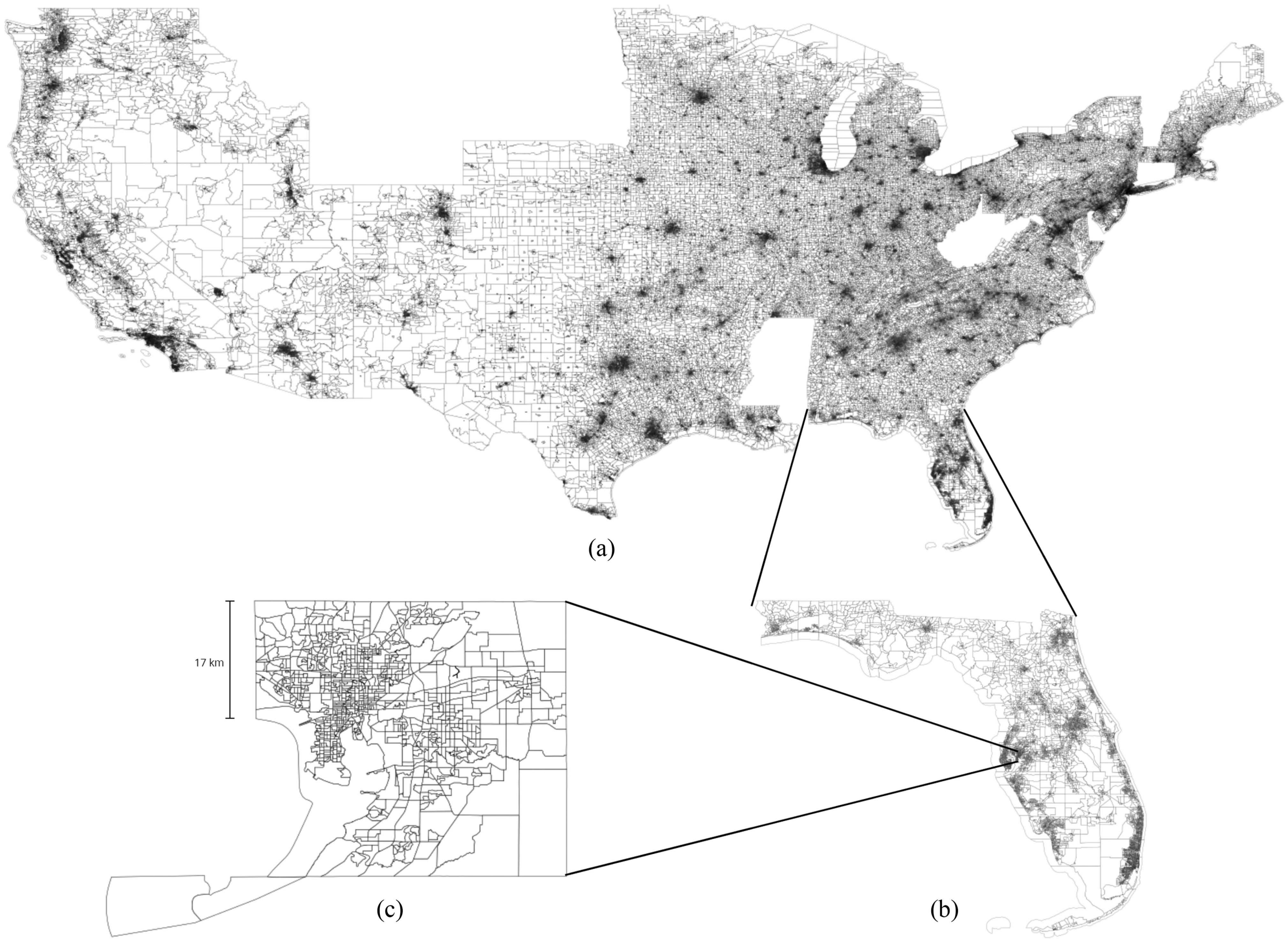


Fig. 1. (a) Illustration of the neighborhoods in states containing cities analyzed in this study. (b) Blow up of neighborhoods in the state of Florida. (c) Blow up of neighborhoods in the county of Hillsborough, Florida, which contains the city of Tampa.

The ResNet50-based architecture of Fig. 5 has its base model pre-trained on ImageNet.[3] In this study, seven fully connected layers are added after the base model to gradually scale down the feature space to a single output estimation.

For model training on each metric, updating all weights in all layers for patching and updating only the fully connected layers for resizing produces optimal results. A batch size of 16, AdamW optimization, and L1 loss [i.e., mean absolute error, (MAE)] are utilized throughout. For each training epoch, the models are trained on the training set and then accessed with the validation set. Models are trained until there is no improvement in validation accuracy after five epochs. Model weight configurations with the lowest validation loss are evaluated on the test set.

B. Semisupervised

As discussed in Section. II, we use ideas from BoVW to produce a rich set of features from the high-resolution imagery that are interpretable and correspond to distinct urban infrastructures. These features are later used to fit a supervised model with the target variables. The methodology (see Fig. 6) can be separated into two steps: clustering of patches and calculation of the cluster distribution of each neighborhood.

Clustering of patches: While cities in the U.S. exhibit variations in culture and environment, there are shared characteristics in their urban infrastructure that can be organized into clusters. However, due to the high dimensionality of images and the distances defined between pixel colors, clustering algorithms can present better results when applied to a learned representation of the images. Diverse methodologies have already used the representational power of deep neural networks to improve clustering results [63]. A simple and effective technique is to run k-means in the latent representation learned from an autoencoder. Deep embedding clustering (DEC) [64] is a more sophisticated technique that trains an autoencoder in two stages: after the first stage optimizes for reconstruction, the embeddings are clustered by k-means and the resulting centroids are added as parameters jointly with the encoder. The loss penalizes the distance between the embeddings and their respective centroids in the second stage.

[3] A test was conducted with weights pre-trained on classification and segmentation tasks with Sentinel imagery, but it did not perform better than ImageNet weights.

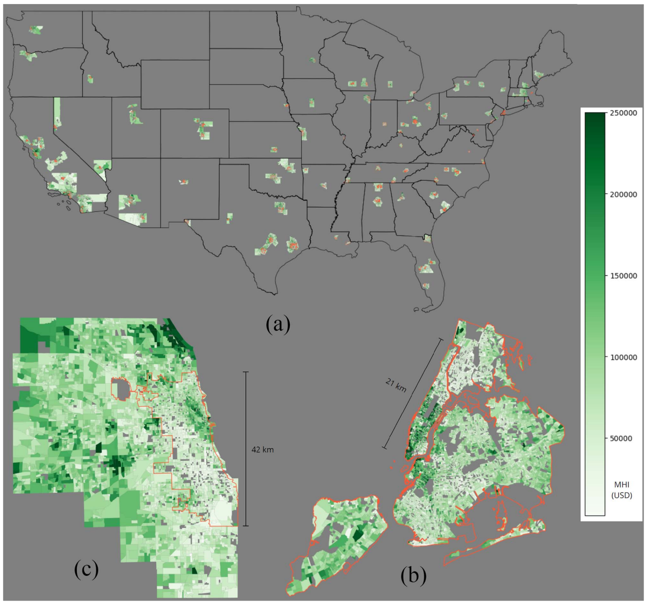


Fig. 2. (a) Illustration of MHI (in 2021 USD) of counties containing the 94 cities examined. City boundaries are in orange. (b) Expanded view of New York City in which its boroughs are coterminous with counties. (c) Expanded view of Chicago in which its city limits are within Cook and DuPage counties (mostly Cook).

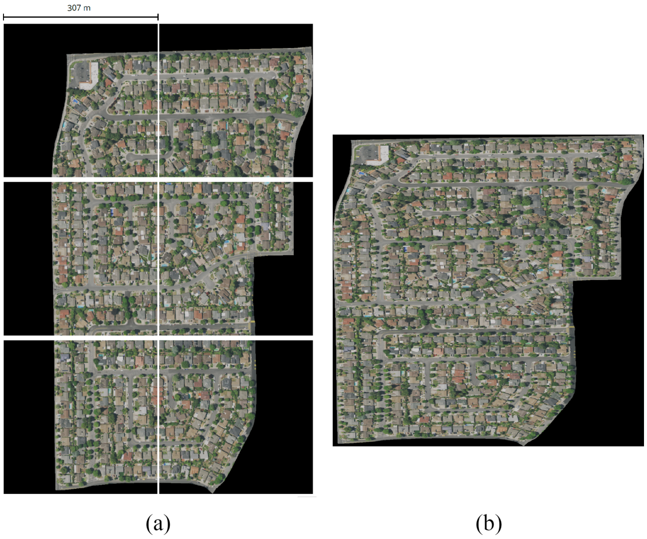


Fig. 3. Processing of a typical neighborhood (this one is in San Jose, CA) for the two processing methods for supervised learning. (a) Patching: The image is split into six 512 × 512 patches. (b) Resizing: The image is resized to 1353 × 1350 pixels (the median width and height of a neighborhood).

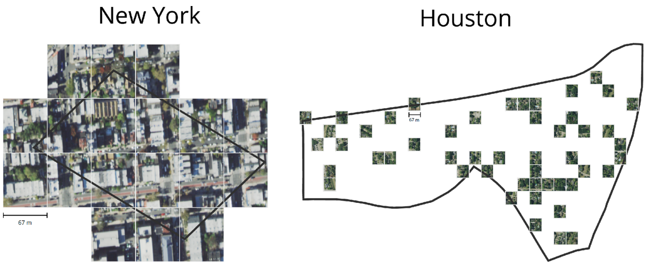


Fig. 4. For the semisupervised approach, examples of 112 × 112 patches for neighborhoods in different cities. Patch boundaries are denoted with white borders, and census block groups (i.e., neighborhoods) with black borders. In the New York neighborhood, all patches overlapping with the neighborhood are used. For the larger Houston neighborhood, only 50 samples are selected.

TABLE III
SUPERVISED RESULTS

Metric	Method			
	Patching		Resizing	
	MAE	R^2	MAE	R^2
Density $\left(\frac{\text{ppl}}{\text{km}^2}\right)$	**142**	0.73	461	**0.81**
MHI (USD)	23,816	0.41	**21,579**	**0.48**
Education (%)	**12.9**	0.50	13.2	**0.51**

In our work, we evaluate using both k-means in a regular autoencoder and k-means in an autoencoder trained with DEC. For both techniques, a ResNet50 pre-trained on ImageNet is used as a feature extractor from patches, generating a 2048-D representation for each. The autoencoder is defined as a feed-forward network with the architecture depicted in Fig. 6, i.e., four layers in the encoder and decoder. The choice of the latent space dimension (d_Z) hyperparameter is crucial to balance the expressive power of the autoencoder and the effectiveness of the k-means clustering. A higher dimension allows for lower reconstruction loss, but it may hinder the clustering process since k-means relies on the Euclidean distance between embeddings. Therefore, the latent dimension is tested using $\{32, 64\}$. A second necessary parameter is the number of clusters, k, which is also selected between $\{50, 100, 200\}$ through experimentation.

Cluster distribution: In this step, we use the patch clusters to build two different sets of features for the neighborhoods. The first set of features is the frequency of each of the k clusters among the patches. The second set of features calculates the distance in the latent space of the patches to each of the k centroids defined in the latent space. Both methods attempt to describe each neighborhood as a composition of clusters, i.e., a composition of distinguished urban infrastructures. The features then can be used in a regression model. We choose to evaluate with random forest.

As mentioned, the methodology has hyperparameters that need to be selected: the dimension of latent space d_Z, the number of clusters k, the type of the set of features, and random forest hyperparameters. The autoencoder and random forest training are made using only the training dataset, with hyperparameters selected based on performance on the validation set. We present separate results comparing k-means and DEC.

V. RESULTS

A. Supervised

Table III displays supervised results on the test data, in terms of MAE and R^2, for patching and resizing image processing techniques. In Table III, bold fonts highlight the most accurate MAE and R^2 results. Models performed best at measuring density, with models trained on resized neighborhoods able to

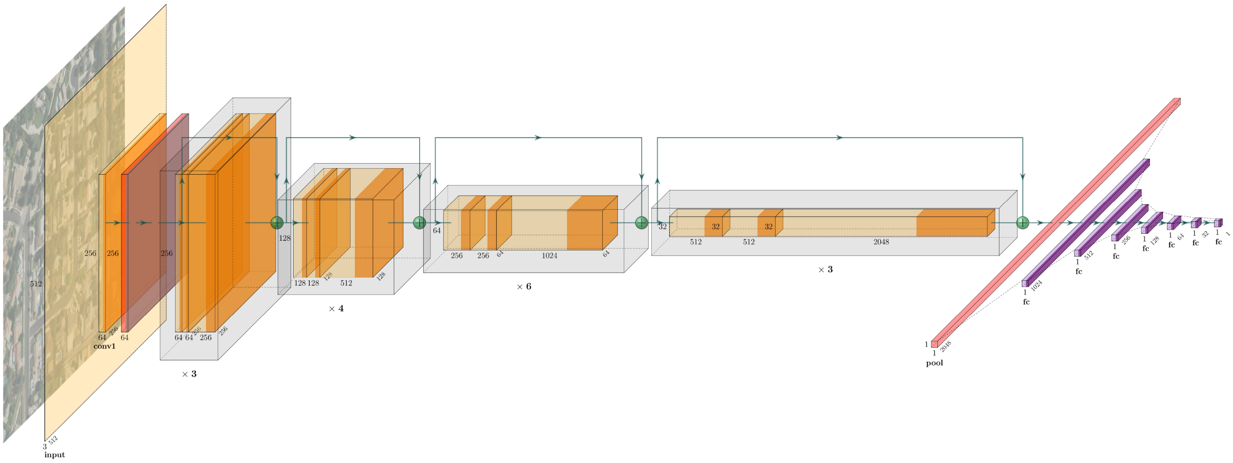


Fig. 5. Visual representation of the ResNet50-based architecture used in the supervised approach. 30% dropout layers are embedded after the first four fully connected layers.

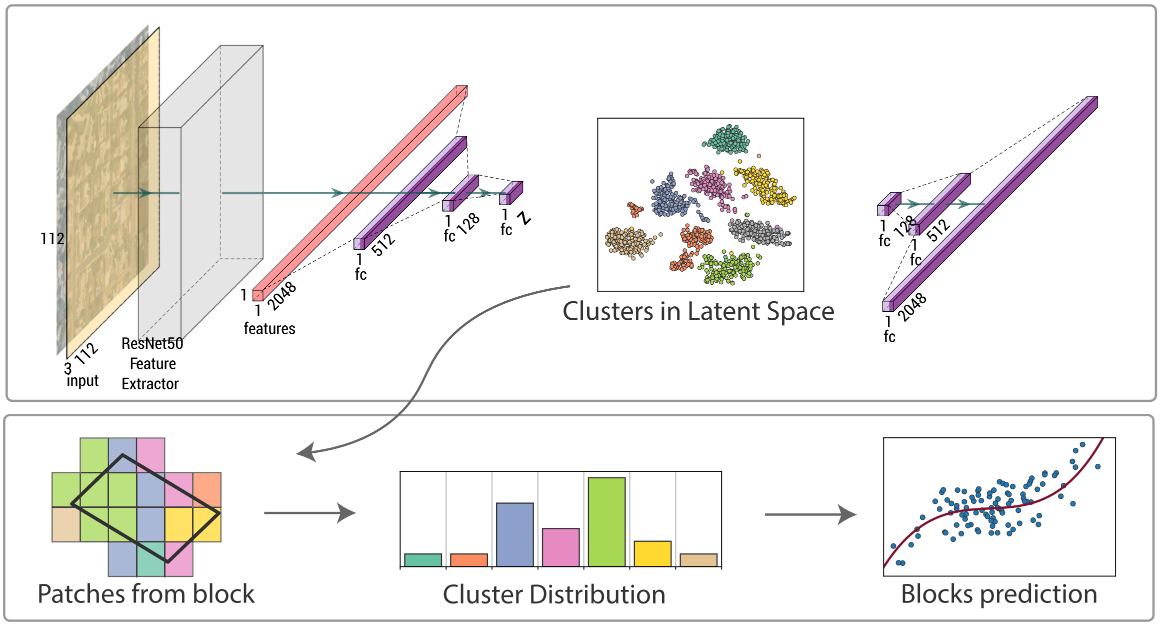


Fig. 6. Overall steps of the semisupervised methodology. First, an unsupervised clustering algorithm is used to cluster small patches of neighborhoods from aerial imagery. The clustering uses ResNet50 as a feature extractor and an autoencoder. The second step is supervised regression on the target variables using the distribution of clusters composed of neighborhood patches.

explain 81% of the variation in density across the study area. These models were able to estimate density to within 461 $\frac{\text{ppl}}{\text{km}^2}$, on average (for reference, the density variable has a ground truth standard deviation of 2519 $\frac{\text{ppl}}{\text{km}^2}$). With resizing, models trained to measure MHI and educational attainment are able to explain about half the variance in the ground truth (R^2 of 0.48 and 0.51, respectively). Models trained on patches performed as well as resized images for estimating education level, however, they were about 7%–8% worse (in terms of R^2) at estimating density and MHI than their resize-based counterparts. In addition, it is worth noting that while the R^2 score is similar for estimating education level, the MAE is lower using patches compared to resizing.

TABLE IV
SEMISUPERVISED RESULTS

Metric	Method			
	k-means		DEC	
	MAE	R^2	MAE	R^2
Density $\left(\frac{\text{ppl}}{\text{km}^2}\right)$	**3,583**	**0.61**	3,743	0.55
MHI (USD)	**31,366**	**0.03**	31,534	0.02
Education (%)	**20.3**	**0.04**	20.4	0.03

B. Semisupervised

Table IV presents MAE and R^2 results obtained on the test data with the best-set parameters from Table V. Results show

TABLE V
SEMISUPERVISED OPTIMAL HYPERPARAMETERS

Metric	**Method**			
	k-means		DEC	
	k	d_Z	k	d_Z
Density	50	64	100	64
MHI	100	64	20	64
Education	100	64	100	32

that k-means in the latent space performs better than DEC. Only small values of k were evaluated in [64] and using a larger k ($>$100) could result in cluster collapse during training (since not all clusters have samples linked to them). The designed features of BoVW are able to explain some degree of variation in density ($R^2 = 0.61$), however they not well-suited for estimating income and education, as we discuss later in Section VI. Also in Section VI, we discuss the capabilities of this method and apply explainability techniques.

VI. DISCUSSION

A. Supervised

As shown in Table III, resizing the neighborhoods generally produces more accurate results than splitting them into patches. The MAEs in measuring population density and education are better through patching, but their R^2s are less than resizing indicating those models do not generalize and explain the variation in density and education as well. A possible explanation for the performance gap in the processing techniques is that resizing the neighborhoods, as opposed to splitting them up, retains more of the spatial and geographic relationships throughout the image amenable to inferring the target metrics. This makes more sense when taken to the logical extreme—a model trained and tested on individual pixels will perform no better than random.

CNN interpretation: A frequent criticism of deep learning models is the difficulty of interpreting the relative importance of features in prediction. To help explore what factors contribute to a density model's estimates, we apply SHapley Additive exPlanations (SHAP) saliency map visualizations [65] on patches from two neighborhoods of contrasting density (see Fig. 7). In Fig 7, the first column shows patches in high-density (top) and low-density (bottom) neighborhoods. The second column displays the SHAP values at the pixel level. SHAP values represent the relative importance of features within the selected images. In this example, red pixels represent those that contribute to a higher density estimate while blue pixels contribute to a lower estimate. Fig. 7 is one of many examples showing manmade structures contributing to a higher density estimate. In the examples, the outlines of smaller dwellings, in particular, are relatively important features. In comparison, in the low-density area, the model demonstrates a tendency to assign a relatively lower density value to larger single-family homes and other features within this more rural area. It is important to note these features would be much less visible from lower resolution imagery, such as Sentinel-2.

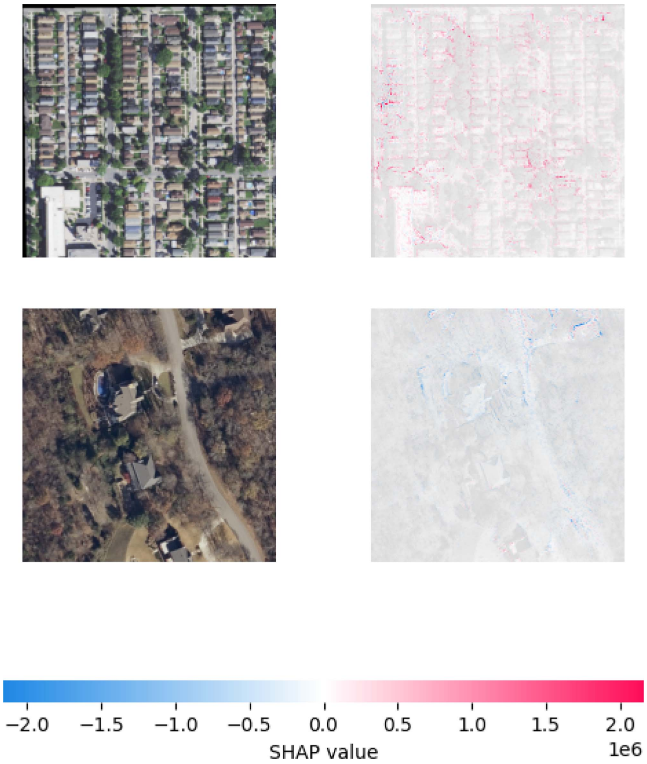


Fig. 7. SHAP of two patches from the supervised approach: (top row) within a high-density area of 1814 $\frac{\text{ppl}}{\text{km}^2}$, and (bottom row) within a low-density area of 11 $\frac{\text{ppl}}{\text{km}^2}$.

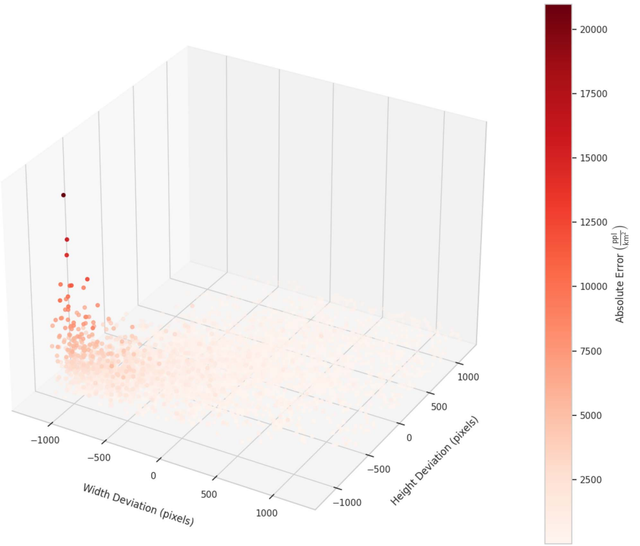


Fig. 8. 3-D plot of the absolute error as a function of width and height deviation from the resized values (when estimating density using the supervised resizing approach).

Effect of resizing: To better understand the effect of resizing on model estimation, for each image in the density test set, absolute error is plotted against the degree to which the image is resized (see Fig. 8). In Fig. 8, width (and height) deviation is the difference between an original image's width (and height) and the median value the images are resized to, i.e., 1353 (and

1350 pixels). Interestingly, the result is exponential decay in error as the original images become larger. A possible explanation for this is that larger neighborhoods contain more information, both in the amount of geographic context and in the number of pixels (i.e., in the bilinear interpolation resizing process, a neighborhood image smaller than the median must upsample while an image greater than the median must downsample). This phenomenon also occurs with MHI and education metrics, though to a lesser degree.

B. Semisupervised

Hyperparameters: The semisupervised method utilizes two important hyperparameters: the dimension of the latent space, d_Z, and the number of clusters, k. As previously mentioned, different values for the parameters are evaluated, and the ones that provided the best results on the validation data are selected (and displayed in Table V). Focusing on the results through k-means, it can be seen that the optimal number of clusters is $k = 100$ for both MHI and educational attainment, in contrast with the density variable that obtained optimal results with $k = 50$. This is intuitive because more detailed clusters ("visual words") are necessary to predict the small variations in MHI and education attainment. The latent dimension with the best results is $d_Z = 64$ (the larger of the two tested), and this result indicates that smaller dimensions do not represent the information in the images as accurately.

Clustering: By studying the output clusters, it can be seen that clustering primarily results in groups of undeveloped (natural) geographic areas and groups of urban infrastructure without substantial differentiation within the two groups (though some clusters exhibit congregation of certain feature subtypes, such as roads and bodies of water). The method clusters most prominently on natural environment versus built, i.e., the degree of urbanization and development, which are closely correlated with population density but not income or education (at least not within city boundaries in the United States). As an illustration, Fig. 9 shows random samples from two clusters, with cluster 7 containing patches from urban areas but without a particular infrastructure type. Despite its shortcomings, the semisupervised approach can take advantage of a large unlabeled dataset to create a corpus of BoVW features, which may be an interesting aspect to exploit in future work.

Interpretation: The proposed semisupervised methodology presents two learning steps that make use of complex models. First, a deep neural network is employed to cluster, and then a random forest model is used. A t-SNE projection [66] is used to visualize the latent representation learned by the autoencoder and to comprehend and validate the neighboring relations learned by the network. Similar to the analysis of the supervised method, we use SHAP to study how the random forest regression model interprets cluster features to generate estimations. To exemplify, we select two neighborhoods in New York, one with high density and one with low density. In the low-density neighborhood, the most important features are from cluster 3, and when analyzed in further detail, it is possible to identify that it is a cluster of water patches (shown in Fig. 9). In the high-density neighborhood, the most important features are related to clusters 31, 39, and 7 (despite the neighborhood having no patches in cluster 39). Similarly, by inspecting the patches of these clusters, it can be seen that they are densely built areas with attributes, such as high-rise apartments (see Fig. 9).

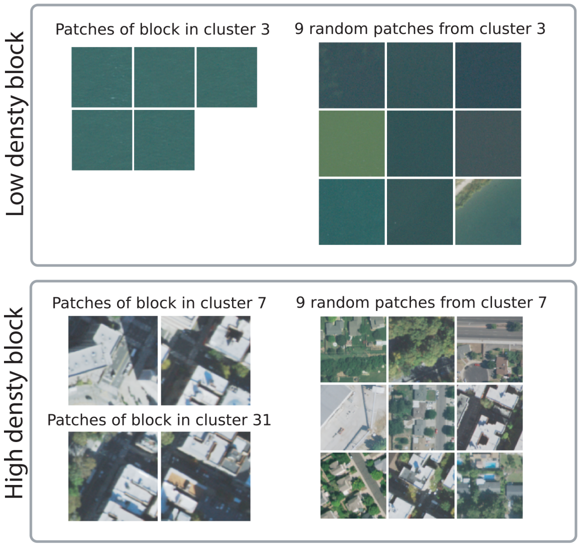

Low densty block
Patches of block in cluster 3
9 random patches from cluster 3
High densty block
Patches of block in cluster 7
Patches of block in cluster 31
9 random patches from cluster 7

Fig. 9. Analysis of the most important features of two neighborhoods in New York using SHAP. The most important cluster for the low-density neighborhood is related to water, and for the high-density neighborhood, two clusters corresponding to densely built areas.

C. Limitations

For supervised learning, only one resize dimension and patch size are studied, so conclusions can only be drawn based on the arrangements presented. That is, a greater patch size, for example, may lead to better or worse performance than the one chosen (512 × 512). Our experiments show that despite the semisupervised approach exhibiting interesting results; it is not able to surpass the performance obtained in supervised learning, particularly when measuring income and education. Clustering is unable to separate nuanced variations in urban infrastructure, which could have been helpful in estimating certain socioeconomic variables, such as income and education. Conversely, the supervised models are trained directly on the predicted variables, directly associating image features to metric values. Overall, income and education results from both methods highlight a general limitation of machine learning to extract useful features from aerial imagery correlated with neighborhood-level census variables. Also, it remains a question to what extent the results gathered here in the United States generalize to different social and ecological environments, with or without fine-tuning the models. Finally, it should be noted that many factors contribute to human well-being and not all of them can be quantified, including from aerial or remotely sensed data [67]. To most accurately understand well-being and development, a holistic approach that considers a complex mesh of personal freedoms, institutional capacity and stability, mental and physical health, cultural values, etc. is required [68]. Nevertheless, the techniques employed in this study can serve as a foundation for further refinement, allowing for more precise estimations of socioeconomic variables at a finer granularity than existing literature. This progress can pave the way for future endeavors aimed at estimating the well-being of neighborhoods.

VII. Conclusion

Census data require resources and coordination to collect, so are therefore produced relatively infrequently in developing countries. Such data are also usually disseminated with a lag, making it difficult to rapidly assess changes in living standards, especially at local levels. In this work, we explored how well CNNs trained on census data and a semisupervised clustering approach can estimate census variables in urban neighborhoods throughout the United States. Results show promise in accurately approximating certain metrics (i.e., density), while uncovering limitations for others (i.e., income, education).

Our findings raise several questions for further research including generalizability, whether changes in aerial imagery could be used to forecast changes in neighborhood metrics over time, and how the latest aerial data could be fused with survey-based measures (including "now-casting" [42]). Different methods may also be explored at this scale such as the use of semantic models to explicitly extract features such as roads (and their quality), number of buildings (and their type), amount of vegetation, etc. (possibly compiled into an "urban well-being index"), for a regressor downstream. As more and more high-resolution aerial imagery products become available ([69], [70]) including from cheaply produced unmanned aerial vehicles deployed outside the United States ([71], [72]), the techniques introduced here provide a foundational benchmark for researchers and reveal potentially fruitful avenues for future work.

Code Availability

Our code is available at.[4]

References

[1] D. A. Castro and M. A. Álvarez, "Predicting socioeconomic indicators using transfer learning on imagery data: An application in Brazil," *GeoJournal*, vol. 88, no. 1, pp. 1081–1102, 2023.

[2] United Nations Department of Economic and Social Affairs, "The sustainable development goals: Report," [Online]. Available: https://sdgs.un.org/goals

[3] S. R. Psaki et al., "Measuring socioeconomic status in multicountry studies: Results from the eight-country MAL-ED study," *Popul. Health Metrics*, vol. 12, no. 8, pp. 1–11, 2014.

[4] J. Wilson, P. Tyedmers, and R. Pelot, "Contrasting and comparing sustainable development indicator metrics," *Ecological Indicators*, vol. 7, no. 2, pp. 299–314, 2007.

[5] M. Sapena, L. A. Ruiz, and H. Taubenböck, "Analyzing links between spatio-temporal metrics of built-up areas and socio-economic indicators on a semi-global scale," *ISPRS Int. J. Geo- Inf.*, vol. 9, no. 7, pp. 1–22, 2020.

[6] R. Stephen, L. Ross, and N. Kalathil, "Innovative metrics for economic development: Final report," *Center for Innovation Strategy and Policy*, 2017. [Online]. Available: https://eda.gov/files/performance/Innovative-Metrics-ED-Report.pdf

[7] L. Kelly, D. Nogueira-Budny, and J. Chelsky, "Conflicting results: Measuring outcomes in situations of conflict," World Bank blogs, 2020.

[8] B. Edmonston, "The case for modernizing the U.S. census," *Society*, vol. 39, no. 1, pp. 42–53, Nov. 2001.

[9] A. S. Schnake-Mahl, J. L. Jahn, S. V. Subramanian, M. C. Waters, and M. Arcaya, "Gentrification, neighborhood change, and population health: A systematic review," *J. Urban Health*, vol. 97, no. 1, pp. 1–25, Jan. 2020.

[10] C. D. Elvidge, K. E. Baugh, E. A. Kihn, H. W. Kroehl, E. R. Davis, and C. W. Davis, "Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption," *Int. J. Remote Sens.*, vol. 18, no. 6, pp. 1373–1379, 1997.

[11] C. D. Elvidge et al., "A global poverty map derived from satellite data," *Comput. Geosciences*, vol. 35, no. 8, pp. 1652–1660, 2009.

[12] M. Burke, A. Driscoll, D. B. Lobell, and S. Ermon, "Using satellite imagery to understand and promote sustainable development," *Science*, vol. 371, no. 6535, 2021, Art. no. eabe8628.

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

[14] W. Hu et al., "Mapping missing population in rural India: A deep learning approach with satellite imagery," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc.*, 2019, pp. 353–359.

[15] C. Yeh et al., "Using publicly available satellite imagery and deep learning to understand economic well-being in africa," *Nature Commun.*, vol. 11, pp. 1–11, 2020.

[16] S. Goodman, A. BenYishay, and D. Runfola, "A convolutional neural network approach to predict non-permissive environments from moderate-resolution imagery," *Trans. GIS*, vol. 25, no. 2, pp. 674–691, 2021.

[17] D. Runfola, H. Baier, L. Mills, M. Naughton-Rockwell, and A. Stefanidis, "Deep learning fusion of satellite and social information to estimate human migratory flows," *Trans. GIS*, vol. 26, no. 6, pp. 2495–2518, 2022.

[18] D. Runfola, A. Stefanidis, and H. Baier, "Using satellite data and deep learning to estimate educational outcomes in data-sparse environments," *Remote Sens. Lett.*, vol. 13, no. 1, pp. 87–97, 2022.

[19] A. Stoian, V. Poulain, J. Inglada, V. Poughon, and D. Derksen, "Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems," *Remote Sens.*, vol. 11, no. 17, pp. 1–26, 2019.

[20] E. Brewer, J. Lin, P. Kemper, J. Hennin, and D. Runfola, "Predicting road quality using high resolution satellite imagery: A transfer learning approach," *Plos One*, vol. 16, no. 7, pp. 1–18, 2021.

[21] A. Van Etten, D. Hogan, J. M. Manso, J. Shermeyer, N. Weir, and R. Lewis, "The multi-temporal urban development spacenet dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 6394–6403.

[22] P. Helber, B. Bischke, A. Dengel, and D. Borth, "EurosAT: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2217–2226, Jul. 2019.

[23] Z. Lv, K. Nunez, E. Brewer, and D. Runfola, "PyShore: A deep learning toolkit for shoreline structure mapping with high-resolution orthographic imagery and convolutional neural networks," *Comput. Geosciences*, vol. 171, 2023, Art. no. 105296.

[24] M. N. Asadullah and N. Chaudhury, "Subjective well-being and relative poverty in rural Bangladesh," *J. Econ. Psychol.*, vol. 33, no. 5, pp. 940–950, 2012.

[25] J. F. Helliwell, R. Layard, J. D. Sachs, J.-E. De Neve, L. B. Aknin, and S. Wang, "World happiness report, Technical report," *Sustain. Develop. Solutions Netw.*, 2022. [Online]. Available: https://happiness-report.s3.amazonaws.com/2022/WHR22.pdf

[26] National Center for Education Statistics, "Annual earnings by educational attainment," Condition of Education. U.S. Department of Education, Institute of Education Sciences, 2023. [Online]. Available: https://nces.ed.gov/programs/coe/indicator/cba

[27] M. Zhan and S. Pandey, "Economic well-being of single mothers: Work first or postsecondary education," *J. Soc. Soc. Welfare*, vol. 31, pp. 87–112, 2004.

[28] J. Jongbloed, "Higher education for happiness? investigating the impact of education on the hedonic and eudaimonic well-being of europeans," *Eur. Educ. Res. J.*, vol. 17, no. 5, pp. 733–754, 2018.

[29] M. Zhan and S. Pandey, "Postsecondary education and the well-being of women in retirement," *Social Work Res.*, vol. 26, no. 3, pp. 171–184, 2002.

[30] N. Graetz et al., "Mapping local variation in educational attainment across africa," *Nature*, vol. 555, no. 7694, pp. 48–53, Mar. 2018.

[31] M. Greenberg and D. Schneider, "Population density: What does it really mean in geographical health studies?," *Health Place*, vol. 81, 2023, Art. no. 103001.

[4] [Online]. Available: https://github.com/VIDA-NYU/GDPFinder

[32] V. Cramer, S. Torgersen, and E. Kringlen, “Quality of life in a city: The effect of population density,” *Social Indicators Res.*, vol. 69, no. 1, pp. 103–116, 2004.

[33] K. Mouratidis, “Compact city, urban sprawl, and subjective well-being,” *Cities*, vol. 92, pp. 261–272, 2019.

[34] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon, “Transfer learning from deep features for remote sensing and poverty mapping,” in *Proc. Thirtieth AAAI Conf. Artif. Intell.*, 2016, pp. 3929–3935.

[35] X. Zhao et al., “Estimation of poverty using random forest regression with multi-source data: A case study in Bangladesh,” *Remote Sens.*, vol. 11, no. 4, pp. 1–18, 2019.

[36] D. Roy, D. Bernal, and M. Lees, “An exploratory factor analysis model for slum severity index in Mexico city,” *Urban Stud.*, vol. 57, no. 4, pp. 789–805, 2020.

[37] E. Luo, M. Kuffer, and J. Wang, “Urban poverty maps - from characterising deprivation using geo-spatial data to capturing deprivation from space,” *Sustain. Cities Soc.*, vol. 84, 2022, Art. no. 104033.

[38] D. Arribas-Bel, J. E. Patino, and J. C. Duque, “Remote sensing-based measurement of living environment deprivation: Improving classical approaches with machine learning,” *Plos One*, vol. 12, no. 5, pp. 1–25, 2017.

[39] A. Sandborn and R. N. Engstrom, “Determining the relationship between census data and spatial features derived from high-resolution imagery in Accra, Ghana,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 5, pp. 1970–1977, May 2016.

[40] S. Pandey, T. Agarwal, and N. C. Krishnan, “Multi-task deep learning for predicting poverty from satellite images,” in *Proc. AAAI Conf.*, 2018, pp. 7793–7798.

[41] G. Li, Z. Cai, Y. Qian, and F. Chen, “Identifying urban poverty using high-resolution satellite imagery and machine learning approaches: Implications for housing inequality,” *Land*, vol. 10, no. 6, pp. 1–16, 2021.

[42] R. Engstrom, J. Hersh, and D. Newhouse, “Poverty from space: Using high resolution satellite imagery for estimating economic well-being,” *World Bank Econ. Rev.*, vol. 36, no. 2, pp. 382–412, 2022.

[43] R. Engstrom, D. Newhouse, and V. Soundararajan, “Estimating small-area population density in Sri Lanka using surveys and geo-spatial data,” *Plos One*, vol. 15, no. 8, pp. 1–20, 2020.

[44] F. R. Stevens, A. E. Gaughan, C. Linard, and A. J. Tatem, “Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data,” *Plos One*, vol. 10, no. 2, pp. 1–22, 2015.

[45] W. Anderson, S. Guikema, B. Zaitchik, and W. Pan, “Methods for estimating population density in data-limited areas: Evaluating regression and tree-based models in Peru,” *Plos One*, vol. 9, no. 7, pp. 1–15, 2014.

[46] P. Deville et al., “Dynamic population mapping using mobile phone data,” *Proc. Nat. Acad. Sci.*, vol. 111, no. 45, pp. 15888–15893, 2014.

[47] P. Sutton, D. Roberts, C. Elvidge, and K. Baugh, “Census from heaven: An estimate of the global human population using night-time satellite imagery,” *Int. J. Remote Sens.*, vol. 22, no. 16, pp. 3061–3076, 2001.

[48] F. Biljecki, K. A. Ohori, H. Ledoux, R. Peters, and J. Stoter, “Population estimation using a 3 D city model: A multi-scale country-wide study in The Netherlands,” *Plos One*, vol. 11, no. 6, pp. 1–22, 2016.

[49] V. A. Alegana et al., “Fine resolution mapping of population age-structures for health and development applications,” *J. Roy. Soc. Interface*, vol. 12, no. 105, 2015, Art. no. 20150073.

[50] R. Bai, J. C. K. Lam, and O. K. Victor Li, “Siamese-like convolutional neural network for fine-grained income estimation of developed economies,” *IEEE Access*, vol. 8, pp. 162533–162547, 2020.

[51] T. Gebru et al., “Using deep learning and Google street view to estimate the demographic makeup of neighborhoods across the United States,” *Proc. Nat. Acad. Sci.*, vol. 114, no. 50, pp. 13108–13113, 2017.

[52] T. Joachims, “Text categorization with support vector machines: Learning with many relevant features,” in *Proc. Eur. Conf. Mach. Learn.*, 1998, pp. 137–142.

[53] A. Coates, A. Ng, and H. Lee, “An analysis of single-layer networks in unsupervised feature learning,” in*Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 215–223.

[54] A. Bosch, A. Zisserman, and X. Muñoz, “Scene classification via pLSA,” in *Proc. Comput. Vis.–ECCV*, 2006, pp. 517–530.

[55] S. Bouteldja and A. Kourgli, “High resolution satellite image indexing and retrieval using SURF features and bag of visual words,” in *Proc. 9th Int. Conf. Mach. Vis.*, vol. 10341, 2017, Art. no. 1034120.

[56] E. Kalinicheva, J. Sublime, and M. Trocan, “Unsupervised satellite image time series clustering using object-based approaches and 3 D convolutional autoencoder,” *Remote Sens.*, vol. 12, no. 11, pp. 1–20, 2020.

[57] A. Barbara Metzler et al., “Phenotyping urban built and natural environments with high-resolution satellite images and unsupervised deep learning,” *Sci. Total Environ.*, vol. 893, Oct. 2023, Art. no. 164794.

[58] United States Department of Agriculture, “National Agriculture Imagery Program (NAIP).,” [Online]. Available: https://catalog.data.gov/dataset/national-agriculture-imagery-program-naip

[59] U. S. Census Bureau, “American community survey.,” [Online]. Available: https://www.census.gov/programs-surveys/acs

[60] U. S. Census Bureau, “When to use 1-year or 5-year estimates.” [Online]. Available: https://www.census.gov/programs-surveys/acs/guidance/estimates.html

[61] U. S. Census Bureau, “Tiger/line shapefiles.,” [Online]. Available: https://www.census.gov/cgi-bin/geo/shapefiles/index.php

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

[63] S. Zhou et al., “A comprehensive survey on deep clustering: Taxonomy, challenges, and future directions, 2022, *arXiv:2206.07579*.

[64] J. Xie, R. Girshick, and A. Farhadi, “Unsupervised deep embedding for clustering analysis,” in *Proc. 33rd Int. Conf. Int. Conf. Mach. Learn.*, 2016, vol. 48, pp. 478–487.

[65] M. Scott Lundberg and Su-In Lee, “A unified approach to interpreting model predictions,” in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 4768–4777.

[66] Laurens van der Maaten and G. Hinton, “Visualizing data using t-SNE,” *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, 2008.

[67] J. Chelsky, “Collateral damage: Pitfalls in quantitative measures of success,” World Bank Blogs, Feb. 12, 2020. [Online]. Available: https://ieg.worldbankgroup.org/blog/collateral-damage-pitfalls-quantitative-measures-success

[68] A. Sen, *Development as Freedom*. New York City, NY, USA: Alfred A. Knopf, 1999.

[69] Eric Hand, “Startup liftoff,” *Science*, vol. 348, no. 6231, pp. 172–177, 2015.

[70] Maxar Technologies, “High-resolution satellite imagery.” [Online]. Available: https://www.maxar.com/products/satellite-imagery

[71] S. Nezami and E. Khoramshahi, Olli nevalainen, ilkka pölönen, and eija honkavaara, “Tree species classification of drone hyperspectral and RGB imagery with deep learning convolutional neural networks,” *Remote Sens.*, vol. 12, no. 7, pp. 1–20, 2020.

[72] A. Retallack, G. Finlayson, B. Ostendorf, and M. Lewis, “Using deep learning to detect an indicator arid shrub in ultra-high-resolution UAV imagery,” *Ecological Indicators*, vol. 145, 2022, Art. no. 109698.

Ethan Brewer received the Ph.D. degree in computational geography from William and Mary, Williamsburg, VA, USA, in 2022.

He is a Senior Scientist with Spectral Sciences, Inc., Boston, MA, USA. Prior to SSI, he was a Research Assistant Professor of computer science and engineering with New York University, New York, NY, USA. His research interests include harnessing geospatial data, computer vision, and machine learning (geoAI) to better understand patterns in development, population, and the environment. His research has been featured in IEEE, *ACM*, *PLoS One*, and *Sensors*, among others.

Giovani Valdrighi received the B.Sc. degree in applied mathematics in 2021 from the School of Applied Mathematics, Fundação Getulio Vargas, Rio de Janeiro, Brazil, where he is currently working toward the M.Sc. degree in mathematical modeling. He is also working toward the Ph.D. degree in computer science with the University of Campinas, Campinas, Brazil.

His research interests include visual analytics and machine learning for spatiotemporal data.

Parikshit Solunke received the B.E. degree in computer engineering from the University of Pune, Pune, India, in 2018, and the M.S. degree in computer science from the University of Illinois at Chicago, Chicago, IL, USA, in 2021. He is currently working toward the Ph.D. degree in computer science with the NYU Tandon School of Engineering—VIDA Center, New York University, Brooklyn, NY, USA.

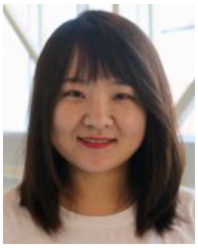
His research interests include the intersection of data science and human computer interaction.

Joao Rulff (Graduate Student Member, IEEE) received the B.S. degree in computer science from Fluminense Federal University, Niterói, Brazil, in 2017. He is currently working toward the Ph.D. degree in computer science with the Tandon School of Engineering—VIDA Center, New York University, Brooklyn, NY, USA.

He has authored or coauthored papers on top-tier venues, such as IEEE VIS and EuroVis. His research interests include visualization, visual analytics, human–computer interaction, and urban computing.

Yurii Piadyk received the M.S. degree in high energy physics from the Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, in 2016, and the Ph.D. degree in computer science from New York University, New York, NY, USA, in 2023.

He was a Research Associate with Visualization and Data Analytics Lab, NYU Tandon, Brooklyn, NY, USA. His research interest include imaging for urban applications, sports analytics, and 3-D scanning.



Zhonghui Lv received two M.S. degrees in environmental science and policy and geographic information science in development and environment, respectively, from Clark University, Worcester, MA, USA, in 2012 and 2014, respectively. She is currently working toward the Ph.D. degree in computational geography with William and Mary, Williamsburg, VA, USA.

She is a Senior Geospatial Researcher with the Virginia Institute of Marine Science, Gloucester Point, VA, USA. Her research focuses on how machine learning and satellite imagery can help promote data-driven management of coastal resources, such as coastal vegetation and shoreline infrastructure.

Jorge Poco (Member, IEEE) received the B.E. degree in system engineering from the National University of San Agustín, Arequipa, Peru, in 2008, the M.Sc. degree in computer science from the University of São Paulo, São Paulo, Brazil, in 2010, and the Ph.D. degree in computer science from New York University, New York, NY, USA, in 2015.

He is an Associate Professor with the School of Applied Mathematics at Fundação Getulio Vargas, Rio de Janeiro, Brazil. His research interests include data visualization, visual analytics, machine learning, and data science.

Dr. Poco has served on several program committees, including IEEE SciVis, IEEE InfoVis, VAST, and EuroVis.

Claudio Silva (Fellow, IEEE) received the B.S. degree in mathematics from the Federal University of Ceará, Fortaleza, Brazil, in 1990, and the M.S. and Ph.D. degrees in computer science from the State University of New York at Stony Brook, Stony Brook, NY, USA, in 1996.

He is the Institute Professor of computer science and engineering and data science with New York University, New York, NY, USA. He is also affiliated with the Center for Urban Science and Progress, Brooklyn, NY, USA, (which he helped co-found in 2012) and the Courant Institute of Mathematical Sciences, New York University. His research interests include visualization, visual analytics, machine learning, reproducibility and provenance, geometric computing, urban computing, computer graphics, and computer vision.

Dr. Silva was a recipient of the IEEE Visualization Technical Achievement Award, and 20 best paper awards, including the IEEE VIS 2023 Test of Time Award for his work on urban visualization.