**DATA DICTIONARY**

We utilized Google’s Street View Image API to collect image data from Chicago, Illinois; Charleston, West Virginia; and Salt Lake City, Utah between December 2016 and February 2017. Details on data collection and image processing are included below. This dataset provides zip code and census tract summarized data for the following neighborhood characteristics: 1) street greenness/landscaping (street trees and street landscaping comprised at least 30% of the image; yes/no), 2) building type (single-family detached house vs. other), and 3) presence of crosswalks (yes/no).

**Variables:**

Green30: proportion of images in which street greenness/street landscaping comprised at least 30% of the image  
Crosswalk: proportion of images containing a crosswalk

Notsfh: proportion of images containing a building type other than single family detached house

**METHODS**

**Street View Image Collection**

We built algorithms to download and process image data for Chicago, Charleston, and Salt Lake City between December 2016-February 2017. Using Google’s Street View Image API, we downloaded image data on all road intersections and a random sample of street segments approximately 50 meters apart. At each search location, we obtained street view images to the west, east, north and south in order to more fully capture features of the environment. Image resolution was 640x640 pixels. In total 227,000 images were downloaded for Chicago, 150,000 images for Salt Lake City, and 53,000 images for Charleston.

**Image data processing**

Convolutional Neural Networks (ConvNets)21 achieve state-of-the-art accuracy for several computer vision tasks including but not limited to object recognition, object detection, and scene labeling. To create a training dataset for our computer vision models, we manually annotated 14,000 images. Each image had three binary labels for the following neighborhood characteristics: 1) street greenness/landscaping (street trees and street landscaping comprised about 30% of the image; yes/no), 2) building type (single-family detached house vs. other), and 3) presence of crosswalks (yes/no). A cut-point of approximately 30% was utilized to assist with inter-rater reliability in manual annotations of street greenness. Moreover, we found that most images had some street greenery and aimed to create a neighborhood indicator to distinguish between ample vs. sparse street landscaping. Images were manually labeled by four of the authors. Inter-rater agreement was above 85% for all neighborhood indicators (86-96% for crosswalks; 85-90% for building type; 85-95% for green30). We randomly divided each dataset into a training set and a test set. The training set contained 80% of total labeled images and the remaining 20% was used as a test set to evaluate the model’s accuracy. We used a deep convolutional network (Visual Geometry Group (VGG-16 model) that is commonly used for object recognition. Three separate networks were trained, one for each indicator of interest. We obtained 84.59%, 85.40% and 93.03% accuracy for the “commercial buildings/apartments,” “green-30%” and “crosswalk” recognition tasks, respectively.

We next experimented with semi-supervised learning. In addition to the labeled dataset, we used an unlabeled set containing 226,879 images captured from Google Street View of different neighborhoods in Chicago. The following steps were used to process the unlabeled images. We resized each unlabeled image to 256 x 256 and then replicate the image 4 times. Next, we cropped each replication at a random location into 224 x 224. Except the first replication, we also matched the histogram of each replicated image to the next 3 unlabeled images. This approach increased the accuracy to 86.07%, 86.55% and 93.37% for the “commercial buildings/apartments”, “green-30%” and “crosswalk” recognition tasks, respectively. (Note: We were not able to leverage unlabeled images in the deep convolutional network approach because unlabeled images can only be used with semi-supervised learning approaches). Image values of neighborhood features (e.g., crosswalks) were then aggregated to produce small area summaries. We utilized zip code and census tract boundaries as a working definition of neighborhoods.