Article

Training Computers to See the Built Environment related to Physical Activity: Detection of Micro-scale Walkability Features using Computer Vision

Marc A. Adams[1]*+, Christine B. Phillips[2]+, Akshar Patel[3] and Ariane Middel[4]

[1] College of Health Solutions, Arizona State University, Phoenix, Arizona; marc.adams@asu.edu
[2] Department of Psychology, Clemson University, Clemson, South Carolina; cbphill@clemson.edu
[3] College of Health Solutions, Arizona State University, Phoenix, Arizona;   aksharpatel@asu.edu
[4] School of Arts, Media and Engineering, Herberger Institute for Design and the Arts, Arizona State University, Phoenix, Arizona; Ariane.Middel@asu.edu
* Correspondence: marc.adams@asu.edu; Tel.: +011-619-220-0048
+ These authors contributed equally to this work: Marc A. Adams and Christine B. Phillips

Abstract: The study purpose was to train and validate a deep-learning approach to detect micro-scale streetscape features related to pedestrian physical activity. This work innovates by combining computer vision techniques with Google Street View (GSV) images to overcome impediments to conducting audits (e.g., time, safety, and expert labor cost). The EfficientNETB5 architecture was used to build deep-learning models for eight micro-scale features guided by the Microscale Audit of Pedestrian Streetscapes-Mini tool: sidewalks, sidewalk buffers, curb cuts, zebra and line crosswalks, walk signals, bike symbols, and streetlights. We used a train--correct loop, whereby images were trained on a training dataset, evaluated using a separate validation dataset, and trained further until acceptable performance metrics were achieved. Further, we used trained models to audit participant (N=512) neighborhoods in the WalkIT Arizona trial. Correlations were explored between micro-scale features and GIS-measured- and participant reported-neighborhood macro-scale walkability. Classifier precision, recall, and overall accuracy were all over >84%. Total micro-scale was associated with overall macro-scale walkability (r=0.30,p<.001). Positive associations were found between model-detected and self-reported sidewalks (r=0.41,p<.001) and sidewalk buffers (r=0.26,p<.001). Computer vision model results suggest an alternative to trained human raters, allowing for audits of hundreds or thousands of neighborhoods for population surveillance or hypothesis testing.

Keywords: Computer vision, Google Street View, Built Environment, Walkability, Micro-scale, Deep learning

1. Introduction

The health and well-being benefits of physical activity, and its environmental and economic co-benefits are well-established [1]. Ecological models posit, and evidence consistently shows, that approaches for promoting physical activity must address multiple levels of influence, including built environments. Features of the built environment can influence physical activity behaviors directly through accessibility, pedestrian safety, comfort, and the affective experiences of walking and active travel [1–7]. In the context of behavioral interventions, evidence suggests the built environment interacts with intervention components to impact physical activity adoption and maintenance [3,8–11]. A supportive built environment can facilitate walking and active travel [6,11], while an unsupportive environment may be a barrier to physical activity engagement [6,11], warranting behavioral intervention to overcome [8,9].

Features supportive of walking and active travel can be measured at the macro-scale using geographic information systems (GIS) or at the micro-scale (street-level) with in-person streetscape audits, and either may be closer to measuring what exists than perceptions [12]. Features measured on a macro-scale, such as land-use mix, intersection density, park, transit, or residential density, are generally difficult to modify but are easily assessed using publicly available data sources. Micro-scale built environment features such as pedestrian amenities that increase the safety and comfort of active travel can explain further variance in physical activity, even after adjusting for macro-scale walkability [4,13].

Micro-scale features are more cost-effectively modified than macro-scale features but are measured much less often. This observation is likely due to the limited feasibility of conducting neighborhood streetscape audits. Traditional in-person micro-scale audits are expensive, require extensive travel and audit time, and expose auditors to crime, traffic, and weather safety concerns. As a result, relationships between micro-scale neighborhood environments, physical activity, and the shared health, environmental and economic co-benefits of active living [1] are not broadly researched. Substantial population differences may be linked to relatively inexpensive and easily modifiable street-level features, but evidence to show inequities across neighborhoods is lacking.

Virtual micro-scale audits by human raters using online mapping tools such as Google Street View (GSV) to scroll down and audit a streetscape are reliable alternatives to traditional in-person audits [14–22]. Virtual audits effectively eliminate travel, weather, and safety challenges of standard in-person audits. However, virtual audits conducted by human content experts remain time-intensive [23–25], limited to small areas or short routes [4,21], require extensive auditor training and retraining and are susceptible to auditor fatigue [26]. Thus, the scalability of virtual audits remains dependent on the amount of available trained human labor. The challenge of scaling in-person or virtual audits to assess hundreds or thousands of neighborhoods continues to be a primary obstacle to surveillance (e.g.., changes over time) or hypothesis testing.

Combining computer-enabled deep learning and computer vision techniques is an emerging approach for increasing the scalability of collecting street-level environmental data. Deep learning is a subset of artificial intelligence that uses algorithms (i.e., neural networks) to learn to recognize and interpret patterns in data. The algorithms used in deep learning are self-adaptive, meaning that the networks get smarter when given more training data or training time. Computer vision is a broad term describing how computers see and understand digital visual data. When deep learning is combined with computer vision, neural network models can be trained to recognize built environment patterns in GSV images for classification tasks [27] such as land-use (e.g., building classification [28], scene classification (e.g., perceived streetscape safety [29]), and object detection (e.g., detecting and classifying automobiles [30] or cataloging trees [31]).

Among classification tasks, image classification using neural networks has been the most commonly used approach for remotely detecting specific features present in GSV images. For example, Hara and colleagues combined crowdsourcing with deep learning and computer vision to detect curb ramps in GSV imagery to assess sidewalk accessibility for disabled individuals [32]. Since then, others have worked to automate the detection of curb cuts [33], crosswalks [34], and other built environment objects visible in GSV images [35–37]. For example, Koo et al. selected eight of 150 categories available from the existing Pyramid Scene Parsing Network (PSPN)

model to represent mesoscale streetscapes including building, house, sidewalk, tree, road, grass, car, and plant [38]. They calculated three indices from these categories: building-to-street-ratio, greenness, and sidewalk-to-street proportion, and found that building-to-street-ratio and greenness were associated with reported walking trips. These advances have led to the promise of developing an automated or semi-automated approach for conducting pedestrian streetscape audits. However, the development of automated micro-scale tools remains an open problem. Existing research relies heavily on deep-learning models developed for broad classifications, which were not developed or validated for features related to physical activity. Further, custom trained models for physical-activity related behaviors focused on a small number of micro-scale features (e.g., curb ramps), or trained and/or validated models in one geographic region (e.g., Atlanta), or did not examine how model-detected features align with perceptions of features (or feature indices) by individuals living in evaluated neighborhoods, or some combinations of these issues. Developing a reliable and validated automated tool for detecting micro-scale features that reduces reliance on human labor is crucial to addressing current issues of scalability. Increasing scalability will enable greater numbers of studies investigating the influence of micro-scale built environment features, as well as including larger and more diverse samples of participants and neighborhoods to allow generalizability.

The current study explores whether eight micro-scale features selected mainly from an existing validated tool, the Microscale Audit of Pedestrian Streetscapes (MAPS)-Mini [13], could be reliably trained and validated using computer vision and deep learning techniques from a sample of GSV images across 5 cities. We further inferred the presence or absence of these micro-scale features in GSV images within buffers around 512 homes for participants enrolled in the baseline phase of the WalkIT Arizona physical activity trial. Additionally, we examined the correlations between our automated micro-scale audit against macro-scale walkability and participant perceptions of these neighborhood features and related subscales.

2. Materials and Methods

Study design and participant recruitment. The current study used data collected from participants enrolled in the WalkIT Arizona trial as described by Adams et al. [10]. Briefly, participant enrollment was balanced across four neighborhood types in Maricopa County, AZ, based on census block group socioeconomic status (SES) and GIS-measured macro-level walkability. For participant sampling, we computed block group SES and walkability using available Census median income data and public regional datasets for net residential density, land uses, intersection density, and public transit density. Following Frank et al [6], block groups in the 1st through 5th deciles of SES were categorized as "lower SES" and the 7th through 10th deciles were categorized as "higher SES." The 6th decile was omitted to minimize mis-categorization for participants on the boundaries. Similarly, block groups were ranked and categorized into "lower walkable" (1st through 4th deciles) and "higher walkable" (7th through 10th deciles) with the 5th and 6th walkability deciles excluded to minimize the likelihood of mis-categorization. Finally, block groups were classified according to their combined walkability and SES yielding four neighborhood strata: "higher SES/higher walkable," "lower SES/higher walkable" "higher SES/lower walkable," and "lower SES/lower walkable." Study marketing materials targeted eligible block groups from these strata.

Enrolled participants (N=512) met the following inclusion criteria: (1) lived in one of the four neighborhood strata in Maricopa County, (2) 18-60 years of age, (3) generally healthy, and (4) insufficiently active. The number of participants for each

neighborhood type ranged from 108 in the 'lower walkable/lower SES' to 136 in the 'higher walkable/higher income' and lower walkable/higher income' neighborhoods. The mean age was 45.5 (SD = 9.1) years, with the majority of the sample reporting being female (64.3%), white (84%), non-Hispanic or Latino (81.2%), and married or living with a partner (67.5%). The sample reported a median household income of $60,000-79,999, median educational attainment of college graduate, and a median distance to work of 10.1 miles (16,316 meters), and a median time at current residence of 52 months. See Adams et al., for full inclusion/ exclusion criteria and sample characteristics [10].

Perceived Walkability Attributes. Participants evaluated their perceived neighborhood attributes using the Neighborhood Environment Walkability Scale (NEWS) [39], completed at the baseline appointment. Seven NEWS subscales were computed using scoring guidelines published at https://drjimsallis.org/Documents/Measures_documents/NEWS_scoring.pdf. These included residential density, proximity to non-residential land uses, street connectivity, presence of walking and cycling facilities, aesthetics, traffic safety, and crime safety. Higher scores on each of the subscales and index score indicate higher perceived walkability. The NEWS subscales have demonstrated good to excellent test-retest reliability and the ability to discriminate between high and low walkable neighborhoods [40,41].

GIS-Measured Macro-scale Walkability. In addition to block group walkability described above for recruitment and enrollment purposes, we also calculated individual-level walkability components and the overall index around enrolled participant homes. Participants' home residential addresses were geocoded using ArcGIS 10.5 (ESRI, Redlands, CA) with US Census Tigerline address feature. Geocoded addresses were used to create a 500-m 'individual-level' buffer throughout the street network and to geoprocess spatial datasets and create 'individual-level' GIS variables for following components: net residential density (i.e., number of housing units divided by residential parcel land area), land use mix (i.e., diversity of several land uses with normalized scores ranging from 0 for single use to 1 indicating an even distribution across residential, retail, recreational, office, civic, food, and entertainment parcel land uses), intersection density (i.e., # of 3-leg or more intersections), and public transit access. A composite walkability index was calculated with the following formula: *Walkability Index* = [(z-score for net residential density) + (z-score for land use mix) + (z-score for intersection density) + (z-score for transit access)]. Figure 1 provides a visual example of these variables.

Figure 1. Example of individual-level buffers and macro-scale walkability components and index values for a single participant's neighborhood.
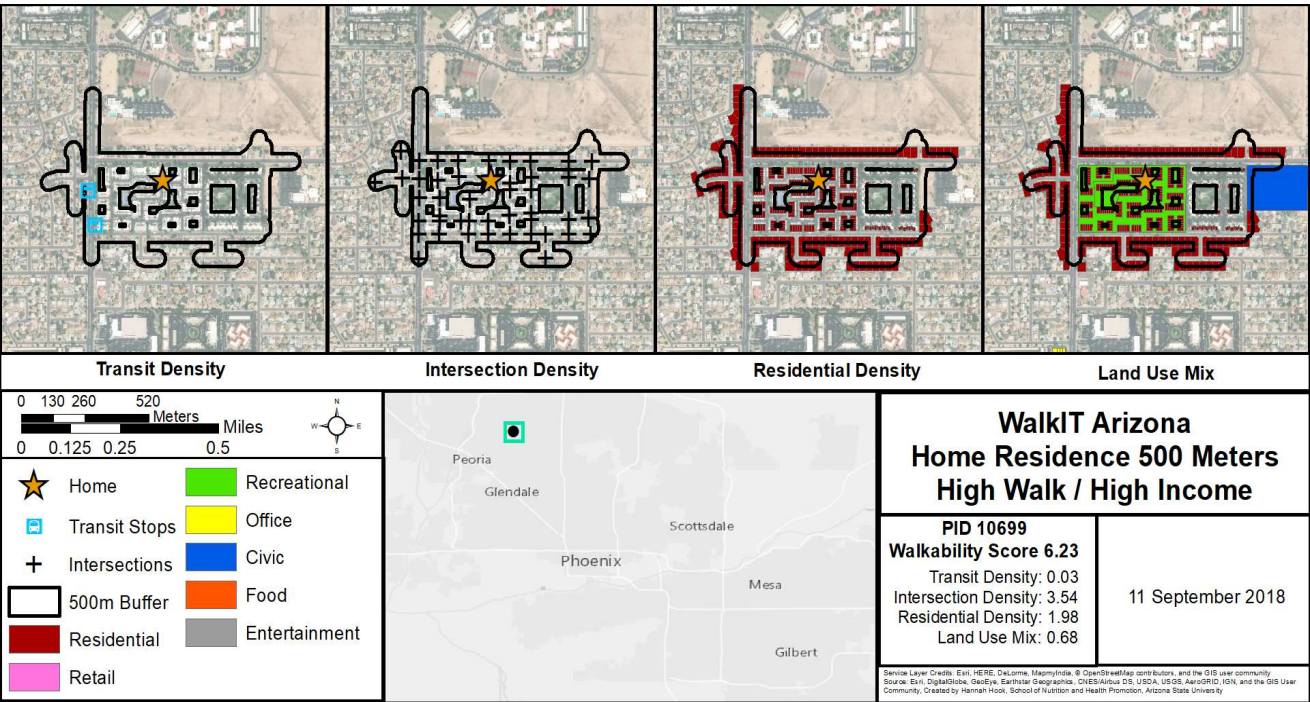
Micro-scale Features and Training and Validation Datasets. To curate a high-quality dataset of labeled features for the micro-scale features of interest for training and validation, we relied on existing GSV images of street intersections from Phoenix, AZ (133,235 images), Washington, DC (20,784 images), San Diego, CA (8,000 images), Seattle, WA (8,000 images), and Baltimore, MD (8,578 images). In addition, we relied on non-intersection images in Phoenix (1,331,994 images). Because images did not necessarily contain a feature of interest, we relied on a larger set of images than used for any single feature. The advantage of dividing the images into these intersection and non-intersection categories was to allow us to use only images that were necessary for a specific image classifier. For example, to train a zebra crosswalk classifier, we only needed intersection images and for sidewalks and sidewalk buffers we required both intersection and non-intersection images. The classes: (1) sidewalk (2) sidewalk buffer (3) curb cut (4) zebra crosswalk (5) line crosswalk (6) walk signal (7) bike symbol and (8) streetlight, were used to create their respective image classifiers.

Creating Image Classifiers: To study associations between model-detected micro-scale street features and GIS-measured and perceived neighborhood walkability, we wanted the system to determine the presence or absence of each of the eight street features at every audit point within participants' neighborhood network buffers. To accomplish this, we created a separate image classifier for each street feature based on EfficientNetB5 neural architecture [42]. For each input image, the classifier output the probability of street feature presence (i.e., crosswalk, curb cut, etc.) using the visual features it identified in the image. The classifier training and evaluation process consisted of the steps as shown in Figure 2 and further described below.

Creating Initial Maricopa Datasets: The first step in creating our classifiers was to create a set of images labeled with the appropriate classification (i.e., presence or absence of street feature) to train the classifier and a separate set of labeled images to validate classifier performance after each step of the training process. We used our knowledge of Phoenix neighborhoods to select initial GSV images from existing image datasets to be

used in the training and validation datasets. To label the datasets, we used the open annotation tool, wkentaro/labelme, available at https://zenodo.org/record/5711226#.YhzeS-jMJaQ [43].

Figure 2. Classifier training and evaluation process steps.
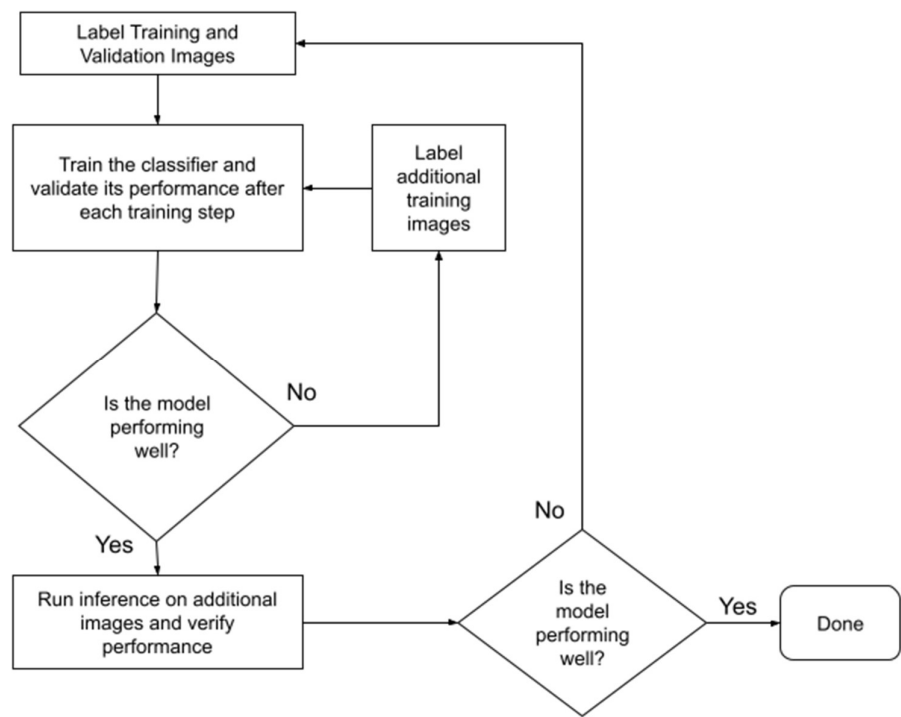
Training & Validation Loop: Using the initial training and validation datasets in place for each street feature, we trained a classifier to recognize the presence or absence of a street feature in an image. Each step in the classifier training process was one pass through the entire training dataset. After each step, the trained classifier was saved for potential future use. The training continued until the classifier started overfitting or the performance metrics did not improve after each step. To prevent overfitting, we tracked training and validation error values. We determined that the classifier had overfit when the training error continued to decrease but the validation error began to increase. When training was complete, we selected the classifier with the best performance metrics from all the saved classifiers while making sure that the classifier had not overfit [44].

If the performance metrics of the trained classifier were not satisfactory, we examined the validation dataset to understand on which images was the classifier showed false positive and false negative results. We then found additional images from the Phoenix dataset that were similar to the ones where the classifier failed, labeled them and added them to the training dataset. With the new training dataset, we restarted the training process.

If the trained classifier performed well, we ran inference on additional Phoenix images and used that to understand how the classifier was performing on those additional images. If the classifier could identify the street features in the new images well, we considered that classifier to be trained. If not, we tried to understand the images where the classifier failed, labeled the images, and added those and other similar labeled images to the training and validation datasets and restarted the training process.

In cases where images from Phoenix datasets weren't enough to improve the performance of a classifier, we used images from other cities (i.e., San Diego, Washington DC, Seattle, and Baltimore) and added those images to the training datasets.

Considerations in Training the Classifiers: *Single vs Multiple Classifiers*. We selected an approach using a separate classifier for each feature instead of a single classifier that would simultaneously detect all features because it allowed us to iterate and improve on each feature classifier effectively and efficiently. Additionally, a single classifier approach can be problematic due to the discrepancy in prevalence across features. Training a single classifier to improve on a specific feature that is less prevalent in images, such as zebra crosswalk, can lead to overfitting [44] on detecting a feature that is highly prevalent in the dataset such as curb cut. Thus, a single model approach may require settling on a poorer-performing model overall to balance these issues across features. *Selecting the classifier architecture*. Based on the results of how different neural network architectures performed on the ImageNet challenge [45] as well as the availability of pre-trained weights, we decided to base our classifier on the EfficientNetB5 architecture. *Selecting a Deep Learning Framework*. Frameworks such as TensorFlow, Keras, PyTorch, Caffe, and Fast.ai are popular in the deep learning field to create neural networks that solve a variety of computer vision problems. We evaluated the different frameworks for the purpose of creating image classifiers and selected Fast.ai due to its ease of use, inbuilt data augmentation capabilities and the simplicity of accomplishing transfer learning. *Transfer Learning*. Training a classifier as deep as EfficientNetB5 is a time-consuming process. To reduce classifier training time and quickly iterate to improve model performance, we used transfer learning. To achieve transfer learning, we used weights from a classifier pre-trained on the ImageNet dataset as the initial weights for training our classifiers [46]. The pre-trained classifiers can identify patterns for ~1000 different classes, which facilitated training our image classifiers. *Data Augmentation Techniques*. When selecting the data augmentation techniques, an important consideration was ensuring that augmenting did not result in losing information that was critical for the classifier to infer a feature. For example, if a sidewalk was only a small section of the image at one of the edges, zoom, crop, warp, cutout, and rotation augmentations could completely remove the sidewalk from the image, resulting in the classifier learning incomplete information. Thus, we applied only three data augmentation techniques: (1) horizontal flip, (2) brightness, and (3) contrast adjustments.

Quantifying Neighborhood Micro-scale Features. The trained models were used to infer (i.e., detect) eight street features in 765,869 previously unexamined photos available in participant buffers in Phoenix, AZ. The probability threshold to classify "presence" of a detected feature in images was set to ≥0.50 and used for each cardinal direction associated with every coordinate (i.e., if the model probability of specific feature presence in an image was ≥0.50, we classified the feature as present). To capture the presence of sidewalks on a block regardless of the side of the street, we averaged the four model-detected probabilities from the four directional images. This approach was used for sidewalks only to ensure we did not miss sidewalks only on one side of the block or visible in only one of four images. For sidewalk buffers, we estimated the presence of sidewalk buffers for coordinates with a model-detected sidewalk only. To summarize each participant's neighborhood, we summed the count of coordinates with positive instances of each feature within the neighborhood buffer and divided by the count of coordinates within the buffer to obtain an average count of neighborhood coordinates with positive instances of each feature. Because the denominator (number of coordinates) could vary by intersection vs. non-intersection feature, the averages of each feature were Z-scored to create a ranking relative to the sample mean. A "total micro-scale feature score" for each participant's home neighborhood was created by summing individual Z-scores for each of the micro-scale features detected within the neighborhood buffer.

Analytic Plan. The performance of the image classifier was assessed using the validation dataset for Phoenix, AZ. Because the task was classifying images by street feature presence or absence, we calculated precision, recall, negative predictive value, specificity, and accuracy for each feature. Precision was the probability that following a positive model-detected observation, the image truly had the feature present (i.e., true positives / true positives + false positives). Recall was the probability that a model-detected observation was truly present in an image (i.e., true positives / true positives + false negatives). Negative predictive value was the probability that following a negative model-detected observation, the image truly did not have a feature present (i.e., true negatives / true negatives + false negatives). Specificity was the proportion of images classified as not having a feature among all images that truly do not have the feature present (i.e., true negatives / true negatives + false positives). Accuracy was the probability of a correct observation (i.e., true positives + true negatives / all observations).

Additionally, we examined the extent to which model-detected neighborhood micro-scale features corresponded with GIS-measured macro-level walkability and self-reported neighborhood walkability attributes (i.e., convergent validity) by conducting Spearman rank correlations between: (1) Model-detected micro-scale neighborhood features (Z-scored individual features and total micro-scale) and GIS-measured neighborhood walkability (Z-scored individual components and overall walkability), (2) Model-detected micro-scale neighborhood features (Z-scored individual and total micro-scale) and participants' NEWS items (e.g., sidewalks, curb cuts) and subscales (e.g., walking and cycling facilities).

3. Results

3.1 Image Classifier Performance

Eight image classifiers were trained to identify their respective street features (i.e., sidewalk, sidewalk buffer, curb cut, zebra crosswalk, line crosswalk, walk signal, bike symbol, and streetlight). Table 1 provides the number of images used for training and validation for each of the eight classifiers.

Table 1. Summary of the number of images used for training and validation datasets.

Street Feature	Image Counts									
	Present				*Absent*				*Total*	
	All Training	*All Validation*	*Phoenix only Training*	*Phoenix only Validation*	*All Training*	*All Validation*	*Phoenix only Training*	*Phoenix only Validation*	*Training*	*Validation*
Sidewalk	8868	2851	5177	1745	3702	1254	2298	429	12570	4105
Sidewalk Buffer	3530	629	1519	347	6066	1773	4461	1567	9596	2402
Curb Cuts	5947	599	2406	268	6059	767	2459	599	12006	1366
Zebra Crosswalk	1687	2456	412	100	5604	6121	2971	879	7291	8577
Line Crosswalk	1762	1053	1693	758	4057	2462	3798	2257	5819	3515
Walk Signal	3126	509	1951	216	4722	1221	2747	1014	7848	1730

Bike Symbol	1127	152	853	132	9306	2138	6908	2078	10433	2290
Streetlight	1380	288	808	170	1213	273	761	171	2593	561

The performance metrics for each of the image classifiers using the validation dataset is displayed in Table 2. Generally, accuracy was high and ranged from 99.59% for zebra crosswalks to 90.03% for streetlights. The precision values (when the model indicated the presence of a feature, how likely was the model to be correct compared to human raters) ranged from 100% for zebra crosswalks to 86.73% for sidewalk buffers. Negative predive values (i.e., when the model indicated the absence of a feature, how likely was the model to be correct compared to human raters) ranged from 99.66% for bike symbols to 89.93% for sidewalks. See Appendix A for a table of validation performance using pooled data from all 5 cities.

Table 2. Validation performance of image classifiers for Phoenix, AZ.

Street Feature	Performance				
	Precision	*Recall*	*Negative Pre-dictive Value*	*Specificity*	*Accuracy*
Sidewalk	97.93%	97.48%	89.93%	91.61%	96.32%
Sidewalk Buffer	86.73%	84.73%	96.63%	97.13%	94.88%
Curb Cut	95.38%	92.54%	96.71%	98.00%	96.31%
Zebra Cross-walk	100%	96.00%	99.55%	100%	99.59%
Line Cross-walk	95.97%	94.20%	98.06%	98.67%	97.55%
Walk Signals	96.77%	97.22%	99.41%	99.31%	98.94%
Bike Symbols	93.28%	94.70%	99.66%	99.57%	99.28%
Streetlight	88.64%	91.76%	91.52%	88.30%	90.03%

3.2 Model Inference Results

The prevalence of model-detected features across the 512 participant neighborhoods in Phoenix, Z was highest for sidewalks (89.8%), followed by streetlights (31.5%), curb cuts (26.2%) and sidewalk buffers (15.9%), line crosswalks (4.9%), walk signals (3.7%), bike symbols (0.5%), and zebra crosswalks (0.3%).

3.2.1. Associations between Model-Detected Micro-scale Feature and GIS-Measured Macro-level Walkability

Spearman correlations between model-detected micro-scale features and GIS-measured walkability attributes are presented in Table 3. The macro-scale walkability index correlated with nine micro-scale features, while the four individual macro-scale components correlated with seven micro-scale features. There was a general pattern of significant weak-to-moderate positive associations between GIS-measured macro-scale walkability and micro-scale features (rs - 0.11 to 0.52, ps < .05). There were two exceptions to this pattern: (1) model-detected curb cuts had weak but significant negative relationships with intersection density, transit density, and overall macro-level walkability, and (2) GIS-measured intersection density had weak but significant negative associations with model-detected sidewalk buffers, curb cuts, line crosswalks, walk signals and the total micro-scale feature score. Overall, the magnitude of associations with model-detected micro-scale features was greatest for transit density, land use mix diversity and overall GIS-measured walkability. Model-detected micro-scale features generally showing the greatest magnitude of associations with GIS-measured walkability were crosswalks, walk signals, bike symbols, streetlights, and total micro-scale feature scores (rs = .19-.52, p < .05).

3.2.2. Associations between Model-Detected Micro-scale Feature and Perceived Neighborhood Walkability.

Spearman correlations between model-detected micro-scale features and NEWS subscales walkability attributes are presented in Table 3. Subscales for perceived residential density, land use mix diversity, presence of walking and cycling facilities, and perceived aesthetics were positively associated with one or more model-detected micro-scale features (rs = .11-.31). Significant negative associations were found between model-detected curb cuts and perceived residential density (r = -.19, p = .000) and between model-detected sidewalks and perceived aesthetics (r = -.24, p = .000). Perceived street connectivity, pedestrian safety and crime safety were not related to any model-detected micro-scale feature or to the total micro-scale feature score. Among model-detected micro-scale features with corresponding individual NEWS items, there were significant positive associations between model-detected and perceived sidewalks (r = .41, p = .000), model-detected sidewalk buffers and perceived grass/dirt sidewalk buffers (r = .26, p = .000), and model-detected and perceived crosswalks and walk signals (r =.15, p < .01).

Table 3. Model-Detected Micro-scale Feature Correlations with GIS-Measured Macro-level Walkability and Perceived NEWS Scales

Model-Detected Micro-scale Feature	GIS-Measured Macro-scale Neighborhood Walkability					Perceived Neighborhood Walkability						
	Residential Density	Land Use Mix Diversity	Intersection Density	Transit Density	Overall Walkability Index	Residential Density	Land Use Mix Diversity	Street Connectivity	Walking and Cycling Facilities	Aesthetics	Pedestrian Safety	Crime Safety
Sidewalks	.12**	.05	.18**	-.06	.02	-.06	-.02	-.03	.11*	-.24***	.01	-.02
Sidewalk Buffers	.18***	.30***	-.14**	.01	.17***	.07+	-.01	.05	.17***	.19**	-.08+	.01

Curb Cuts	.04	.16*	-.16***	-.20***	-.11*	-.19***	-.06	.06	.17***	-.03	.08+	.04
Zebra Cross-walks	.16***	-.07	.04	.37***	.02	.15**	.04	-.01	-.04	-.06	-.04	-.07
Line Cross-walks	.06	.42***	-.14**	.13**	.39***	.28***	.24***	.01	.02	.03	-.01	-.02
All Cross-walks	.07+	.39***	-.12**	.38**	.38***	.30***	.23***	.00	.01	.01	-.01	-.03
Walk Signals	.09*	.37***	-.10*	.52***	.46***	.31***	.23**	.02	-.00	.07	-.07+	-.07
Bike Symbols	.17**	.22***	.06	.20***	.28***	.25***	.15**	-.01	.02	-.03	-.03	-.05
Streetlights	.23***	.38***	.00	.12**	.35***	.17***	.07	-.00	.14**	-.03	-.06	-.07
Total Micro-scale	.19***	.38***	-.12*	.11*	.30***	.13**	.07+	.02	.21***	.04	-.02	-.02

Notes: Spearman rank correlation coefficients. $^+p < .10$, $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$. Model detected features assessed by z-scoring the average count of positive feature instances for coordinates within a 500 meter street network buffer around participants' homes. Perceived neighborhood features assessed by the Neighborhood Environment Walkability Scale (NEWS). All crosswalks = sum of zebra and line crosswalks. Total micro-scale score = sum of z-score averages for bike symbols, all crosswalks, curb cuts, walk signals, sidewalks, sidewalk buffers, and streetlights within each participant's 500 meter neighborhood buffer.

4. Discussion

This paper demonstrates that use of computer vision to detect intersection and street segment features conceptually related to pedestrian physical activity (i.e., zebra and line crosswalks, curb ramps, walk signals, sidewalks, sidewalk buffers, bike symbols, streetlights) is feasible with high correspondence to human raters of the same GSV images. Individual micro-scale features and a summary index of micro-scale features correlated with both GIS-measured macro-scale walkability and with human participants' reports of their neighborhood environment. These expected correlations offer a degree of validly to the computer vision models of micro-scale features. The development of machine-learning models for detecting micro-scale features opens the possibility of conducting research across broad regions and new research questions for micro-scale research.

Computer model-detected micro-scale features correlated with both the individual components of macro-scale walkability and the macro-scale walkability index around participant neighborhoods in Phoenix. These correlations were expected, as higher levels of macro-scale walkability are often complemented by improvements to more affordable micro-scale improvements (e.g., curb cuts, sidewalks) that further enhance the streetscape. Previous studies have shown that individual micro-scale features and macro-scale walkability indices are weak-to-moderately correlated, with both contributing unique measures of the built environment for walking [4]. While most component correlations between macro- and micro-scale features were positive, the macro-scale component of intersection density was negatively correlated with all micro-scale features except sidewalks. Higher intersection densities are typically observed in more dense urban settings with shorter, more connected street blocks. One could expect urban

settings with shorter blocks to have a greater prevalence of sidewalks and related safety features such as crosswalks and curb cuts, but surprisingly the WalkIT AZ participants' perceptions of street connectivity and pedestrian safety did not correlate with any model-detected features.

Model-detected micro-scale features correlated with four perceived subscales of the built environment, specifically residential density, land use mix, walking and cycling facilities, and aesthetics. WalkIT participants evaluated their neighborhoods using the previously validated NEWS, which has been validated against GIS-measured macro-scale features and used in dozens of studies as a predictor of pedestrian walking for transportation with weak to moderate correlations (i.e., $r = <0.40$) [40, 41]. The NEWS does not offer an index to summarize its seven subscales, but our index of model-detected micro-scale features did correlate with perceived residential density and walking and cycling facilities subscales ($r = 0.13$ and 0.21, respectively). The strongest micro-scale relationships occurred for the residential density subscale, which had 8 significant correlations with model-detected walk signals and combined crosswalks (zebra and line) correlating the strongest ($r = 0.30 - 0.31$). This suggests that higher levels of model-detected safety features for pedestrians correlated with higher levels of independent perceptions of residential density, which aligns with expectations that these features would be more prevalent in areas with more people. The present results were consistent with previous studies showing weak-to-modest agreement between objective and subjective assessments of neighborhood walkability overall, with lower concordance among those with less physical activity and higher BMI [12,47]. Because the current study included only insufficiently active individuals with a median BMI of 33.0, results may not generalize to other populations. However, based on previous research, we would expect higher correspondence between model-detected features and perceived neighborhood walkability in a more physically active sample.

Methodological Considerations. Two major strengths should be noted. First, in previous studies, human raters audited only parts of participant neighborhoods - usually limited to a quarter mile route or sample of blocks in a neighborhood – by in person or virtual observation. The current computer vision approach was used to audit all blocks and crossings for the entire neighborhood for all participants, limited only by the number of GSV photos available and the timeframe of the study. Second WalkIT participants were purposefully recruited in similar numbers from neighborhoods high and low in walkability and high and low in neighborhood socioeconomic status. Therefore, our analyses between micro-scale and macro-scale and participant perceptions reflect the full spectrum of walkability and income environments present in the Phoenix, AZ region.

Several considerations also should be noted in the development of computer vision models. First, we explored existing annotated image datasets such as Mapillary [48] and CAMVID [49] and found labeling did not capture the features of interest, incorrect classifications, and bucketing of categories that precluded utilizing existing datasets and developing good computer vision models of the features of interest. Second, while a standardized number of images for each feature for training and validation datasets would have been conceptually clearer to report, we found that additional training or validation images were needed for certain features (e.g., bike symbols, zebra crosswalks) because of the low prevalence of such features in the Phoenix, AZ region. We also collected a greater number of training samples to ensure that we captured inherent variability in feature design and photos of features that exist in the real world in the training dataset. For example, the variability in in the design of crosswalks varies even by small geographic regions to include different patterns and colors (e.g., crosswalk with LGBT rainbow flag colors). We also considered image artifacts (e.g., shadows that appear similar to zebra crossings, distance from GSV camera to streetlight) that resulted in small or

fuzzy training samples. Related to this concern is that some photos captured elements that confused the models (e.g., actual bike vs. painted bike symbol) and necessitated additional training samples. Although results presented in the current analyses used a Phoenix, AZ validation dataset, we expect the generalizability of our trained classifiers is enhanced with the inclusion of images from four additional geographically diverse cities (i.e. Seattle, San Diego, Washington, D.C., and Baltimore), in classifier training (see Appendix A table 1).

5. Conclusions

The current results demonstrated that computer vision models were valid for conducting neighborhood audits of pedestrian streetscape features, and that model results correlate with both objective and self-reported macro-scale neighborhood walkability. Future research will examine the relationship between model-detected micro-scale features and physical activity and chronic disease outcomes. The computer vision approach to auditing neighborhoods promises to accelerate the pace of micro-scale research and open new lines of micro-scale research for urban planning and public health. Results suggest automated virtual streetscape audits may provide a scalable alternative to human audits, enabling advancements in the field currently constrained by time and cost. Reducing reliance on trained auditors will enable scaling-up audits to assess hundreds or thousands of neighborhoods or even entire cities for surveillance, hypotheses testing, environmental disparities, or change detection research related to pedestrian streetscapes. For example, given sufficient resources, such models could be applied at scale to map all sidewalks in the U.S., evaluate whether motivational physical activity interventions perform better in neighborhoods with more vs. fewer sidewalks, determine whether the prevalence of sidewalks differ by neighborhoods that vary by race/ethnicity or income, or even evaluate change in sidewalks before and after a new development or passage of a complete street policy or transportation infrastructure tax.

Appendix A

Appendix Table 1. Validation Performance of Image Classifiers in Pooled Dataset including Phoenix AZ, San Diego CA, Washington D.C., Seattle WA, and Baltimore MD.

Street Feature	Performance				
	Precision	*Recall*	*Negative Predictive Value*	*Specificity*	*Accuracy*
Sidewalk	97.25%	96.81%	92.82%	93.78%	95.88%
Sidewalk Buffer	87.10%	85.85%	95.01%	95.49%	92.96%
Curb Cut	83.21%	65.86%	52.32%	73.81%	68.54%
Zebra Crosswalk	97.33%	84.97%	93.61%	98.95%	94.62%
Line Crosswalk	89.20%	75.59%	71.20%	86.83%	80.20%
Walk Signals	86.00%	73.38%	68.80%	83.09%	77.40%
Bike Symbols	95.00%	95.00%	98.33%	98.33%	97.50%
Streetlight	84.30%	86.44%	83.84%	81.37%	84.09%

References

1. Sallis, J. F.; Spoon, C.; Cavill, N.; Engelberg, J. K.; Gebel, K.; Parker, M.; Thornton, C. M.; Lou, D.; Wilson, A. L.; Cutter, C. L.; Ding, D. Co-benefits of designing communities for active living: An exploration of literature. *Int. J. Behav. Nutr. Phys. Act.* **2015**, *12*, 30, doi:10.1186/s12966-015-0188-2.

2. Alfonzo, M.; Boarnet, M. G.; Day, K.; Mcmillan, T.; Anderson, C. L. The relationship of neighbourhood built environment features and adult parents' walking. *Journal of Urban Design* **2008**, *13*, 29–51, doi:10.1080/13574800701803456.

3. Sallis, J. F.; Cerin, E.; Kerr, J.; Adams, M. A.; Sugiyama, T.; Christiansen, L. B.; Schipperijn, J.; Davey, R.; Salvo, D.; Frank, L. D.; De Bourdeaudhuij, I.; Owen, N. Built Environment, Physical Activity, and Obesity: Findings from the International Physical Activity and Environment Network (IPEN) Adult Study. *Annu. Rev. Public Health* **2020**, *41*, 119–139, doi:10.1146/annurev-publhealth-040218-043657.

4. Cain, K. L.; Millstein, R. A.; Sallis, J. F.; Conway, T. L.; Gavand, K. A.; Frank, L. D.; Saelens, B. E.; Geremia, C. M.; Chapman, J.; Adams, M. A.; Glanz, K.; King, A. C. Contribution of streetscape audits to explanation of physical activity in four age groups based on the Microscale Audit of Pedestrian Streetscapes (MAPS). *Soc. Sci. Med.* **2014**, *116*, 82–92, doi:10.1016/j.socscimed.2014.06.042.

5.   Casagrande, S. S.; Whitt-Glover, M. C.; Lancaster, K. J.; Odoms-Young, A. M.; Gary, T. L. Built environment and health be-haviors among African Americans: a systematic review. *Am. J. Prev. Med.* 2**009**, 3*6,* 174–181, doi:10.1016/j.ame-pre.2008.09.037.

6.   Frank, L. D.; Sallis, J. F.; Saelens, B. E.; Leary, L.; Cain, K.; Conway, T. L.; Hess, P. M. The development of a walkability in-dex: application to the Neighborhood Quality of Life Study. *Br. J. Sports Med.* **2010**, *44*, 924–933, doi:10.1136/bjsm.2009.058701.

7.   Bornioli, A.; Parkhurst, G.; Morgan, P. L. Affective experiences of built environments and the promotion of urban walking. *Transportation Research Part A: Policy and Practice* **2019**, *123*, 200–215, doi:10.1016/j.tra.2018.12.006.

8.   Kerr, J.; Norman, G. J.; Adams, M. A.; Ryan, S.; Frank, L.; Sallis, J. F.; Calfas, K. J.; Patrick, K. Do neighborhood environments moderate the effect of physical activity lifestyle interventions in adults? *Health Place* **2010**, *16*, 903–908, doi:10.1016/j.health-place.2010.05.002.

9.   Adams, M. A.; Hurley, J. C.; Todd, M.; Bhuiyan, N.; Jarrett, C. L.; Tucker, W. J.; Hollingshead, K. E.; Angadi, S. S. Adaptive goal setting and financial incentives: a 2 × 2 factorial randomized controlled trial to increase adults' physical activity. *BMC Public Health* **2017**, *17*, 286, doi:10.1186/s12889-017-4197-8.

10.   Adams, M. A.; Hurley, J. C.; Phillips, C. B.; Todd, M.; Angadi, S. S.; Berardi, V.; Hovell, M. F.; Hooker, S. Rationale, design, and baseline characteristics of WalkIT Arizona: A factorial randomized trial testing adaptive goals and financial reinforce-ment to increase walking across higher and lower walkable neighborhoods. *Contemp. Clin. Trials* **2019**, *81*, 87–101, doi:10.1016/j.cct.2019.05.001.

11.   Mayne, D. J.; Morgan, G. G.; Jalaludin, B. B.; Bauman, A. E. The contribution of area-level walkability to geographic varia-tion in physical activity: a spatial analysis of 95,837 participants from the 45 and Up Study living in Sydney, Australia. *Popul. Health Metr.* **2017**, *15*, 38, doi:10.1186/s12963-017-0149-x.

12.   Gebel, K.; Bauman, A. E.; Sugiyama, T.; Owen, N. Mismatch between perceived and objectively assessed neighborhood walkability attributes: prospective relationships with walking and weight gain. H*ealth Place* 2**011,** 1*7,* 519–524, doi:10.1016/j.healthplace.2010.12.008.

13.   Sallis, J. F.; Cain, K. L.; Conway, T. L.; Gavand, K. A.; Millstein, R. A.; Geremia, C. M.; Frank, L. D.; Saelens, B. E.; Glanz, K.; King, A. C. Is your neighborhood designed to support physical activity? A brief streetscape audit tool. *Prev Chronic Dis* **2015**, *12*, E141, doi:10.5888/pcd12.150098.

14.   Phillips, C. B.; Engelberg, J. K.; Geremia, C. M.; Zhu, W.; Kurka, J. M.; Cain, K. L.; Sallis, J. F.; Conway, T. L.; Adams, M. A. Online versus in-person comparison of Microscale Audit of Pedestrian Streetscapes (MAPS) assessments: reliability of alter-nate methods. *Int J Health Geogr* **2017**, *16*, 27, doi:10.1186/s12942-017-0101-0.

15.   Ben-Joseph, E.; Lee, J. S.; Cromley, E. K.; Laden, F.; Troped, P. J. Virtual and actual: relative accuracy of on-site and web-based instruments in auditing the environment for physical activity. *Health Place* **2013**, *19*, 138–150, doi:10.1016/j.health-place.2012.11.001.

16.   Kurka, J. M.; Adams, M. A.; Geremia, C.; Zhu, W.; Cain, K. L.; Conway, T. L.; Sallis, J. F. Comparison of field and online ob-servations for measuring land uses using the Microscale Audit of Pedestrian Streetscapes (MAPS). *J. Transp. Health* **2016**, *3*, 278–286, doi:10.1016/j.jth.2016.05.001.

17.   Zhu, W.; Sun, Y.; Kurka, J.; Geremia, C.; Engelberg, J. K.; Cain, K.; Conway, T.; Sallis, J. F.; Hooker, S. P.; Adams, M. A. Relia-bility between online raters with varying familiarities of a region: Microscale Audit of Pedestrian Streetscapes (MAPS). *Landsc. Urban Plan.* **2017**, *167*, 240–248, doi:10.1016/j.landurbplan.2017.06.014.

18.   Clarke, P.; Ailshire, J.; Melendez, R.; Bader, M.; Morenoff, J. Using Google Earth to conduct a neighborhood audit: reliability of a virtual audit instrument. *Health Place* **2010**, *16*, 1224–1229, doi:10.1016/j.healthplace.2010.08.007.

19.   Rundle, A. G.; Bader, M. D. M.; Richards, C. A.; Neckerman, K. M.; Teitler, J. O. Using Google Street View to audit neighbor-hood environments. *Am. J. Prev. Med.* **2011**, *40*, 94–100, doi:10.1016/j.amepre.2010.09.034.

20.   Silva, V.; Grande, A. J.; Rech, C. R.; Peccin, M. S. Geoprocessing via google maps for assessing obesogenic built environ-ments related to physical activity and chronic noncommunicable diseases: validity and reliability. *J. Healthc. Eng.* **2015**, *6*, 41–54, doi:10.1260/2040-2295.6.1.41.

21.   Vanwolleghem, G.; Ghekiere, A.; Cardon, G.; De Bourdeaudhuij, I.; D'Haese, S.; Geremia, C. M.; Lenoir, M.; Sallis, J. F.; Verhoeven, H.; Van Dyck, D. Using an audit tool (MAPS Global) to assess the characteristics of the physical environment related to walking for transport in youth: reliability of Belgian data. *Int J Health Geogr* **2016**, *15*, 41, doi:10.1186/s12942-016-0069-1.

22.   Vanwolleghem, G.; Van Dyck, D.; Ducheyne, F.; De Bourdeaudhuij, I.; Cardon, G. Assessing the environmental characteris-tics of cycling routes to school: a study on the reliability and validity of a Google Street View-based audit. *Int J Health Geogr* **2014**, *13*, 19, doi:10.1186/1476-072X-13-19.

23.   Wilson, J. S.; Kelly, C. M.; Schootman, M.; Baker, E. A.; Banerjee, A.; Clennin, M.; Miller, D. K. Assessing the built environ-ment using omnidirectional imagery. *Am. J. Prev. Med.* **2012**, *42*, 193–199, doi:10.1016/j.amepre.2011.09.029.

24.   Badland, H. M.; Opit, S.; Witten, K.; Kearns, R. A.; Mavoa, S. Can virtual streetscape audits reliably replace physical streetscape audits? *J Urban Health* **2010**, *87*, 1007–1016, doi:10.1007/s11524-010-9505-x.

25.   Yi, L.; Wu, T.; Luo, W.; Zhou, W.; Wu, J. A non-invasive, rapid method to genotype late-onset Alzheimer's disease-related apolipoprotein E gene polymorphisms. *Neural Regen. Res.* **2014**, *9*, 69–75, doi:10.4103/1673-5374.125332.

26.   Bader, M. D. M.; Mooney, S. J.; Lee, Y. J.; Sheehan, D.; Neckerman, K. M.; Rundle, A. G.; Teitler, J. O. Development and de-ployment of the Computer Assisted Neighborhood Visual Assessment System (CANVAS) to measure health-related neigh-borhood conditions. *Health Place* **2015**, *31*, 163–172, doi:10.1016/j.healthplace.2014.10.012.

27.   Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B. A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing* **2019**, *152*, 166–177, doi:10.1016/j.isprsjprs.2019.04.015.

28.   Kang, J.; Körner, M.; Wang, Y.; Taubenböck, H.; Zhu, X. X. Building instance classification using street view images. *ISPRS Journal of Photogrammetry and Remote Sensing* **2018**, *145*, 44–59, doi:10.1016/j.isprsjprs.2018.02.006.

29.   Naik, N.; Philipoom, J.; Raskar, R.; Hidalgo, C. Streetscore -- Predicting the Perceived Safety of One Million Streetscapes. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*; IEEE, 2014; pp. 793–799.

30.   Gebru, T.; Krause, J.; Wang, Y.; Chen, D.; Deng, J.; Aiden, E. L.; Fei-Fei, L. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proc. Natl. Acad. Sci. US*A **2017**, *114*, 13108–13113, doi:10.1073/pnas.1700035114.

31.   Branson, S.; Wegner, J. D.; Hall, D.; Lang, N.; Schindler, K.; Perona, P. From Google Maps to a fine-grained catalog of street trees. *ISPRS Journal of Photogrammetry and Remote Sensing* **2018**, *135*, 13–30, doi:10.1016/j.isprsjprs.2017.11.008.

32.   Hara, K.; Sun, J.; Moore, R.; Jacobs, D.; Froehlich, J. Tohme: Detecting curb ramps in google street view using crowdsourc-ing, computer vision, and machine learning. In *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST' ' '14*; ACM Press: New York, New York, USA, 2014; pp. 189–204.

33.   Abbott, A.; Deshowitz, A.; Murray, D.; Larson, E. C. WalkNet: A Deep Learning Approach to Improving Sidewalk Quality and Accessibility. *SMU Data Science Review* **2018**, *1*, Article 7.

34.   Berriel, R. F.; Rossi, F. S.; de Souza, A. F.; Oliveira-Santos, T. Automatic large-scale data acquisition via crowdsourcing for crosswalk classification: A deep learning approach. *Comput Graph* **2017**, *68*, 32–42, doi:10.1016/j.cag.2017.08.004.

35.   Krylov, V.; Kenny, E.; Dahyot, R. Automatic Discovery and Geotagging of Objects from Street View Imagery. *Remote Sens (Basel)* **2018**, *10*, 661, doi:10.3390/rs10050661.

36.   Balali, V.; Ashouri Rad, A.; Golparvar-Fard, M. Detection, classification, and mapping of U.S. traffic signs using google street view images for roadway inventory management. *Vis. in Eng.* **2015**, *3*, 15, doi:10.1186/s40327-015-0027-1.

37.   Hebbalaguppe, R.; Garg, G.; Hassan, E.; Ghosh, H.; Verma, A. Telecom inventory management via object recognition and localisation on google street view images. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*; IEEE, 2017; pp. 725–733.

38.   Koo, B. W.; Guhathakurta, S.; Botchwey, N. How are Neighborhood and Street-Level Walkability Factors Associated with Walking Behaviors? A Big Data Approach Using Street View Images. *Environ Behav* **2021**, 001391652110146, doi:10.1177/00139165211014609.

39.   Saelens, B. E.; Sallis, J. F.; Black, J. B.; Chen, D. Neighborhood-based differences in physical activity: an environment scale evaluation. *Am. J. Public Health* **2003**, *93*, 1552–1558, doi:10.2105/ajph.93.9.1552.

40.   Cerin, E.; Conway, T. L.; Saelens, B. E.; Frank, L. D.; Sallis, J. F. Cross-validation of the factorial structure of the Neighborhood Environment Walkability Scale (NEWS) and its abbreviated form (NEWS-A). *Int. J. Behav. Nutr. Phys. Act.* **2009**, *6*, 32, doi:10.1186/1479-5868-6-32.

41.   Adams, M. A.; Ryan, S.; Kerr, J.; Sallis, J. F.; Patrick, K.; Frank, L. D.; Norman, G. J. Validation of the Neighborhood Environment Walkability Scale (NEWS) items using geographic information systems. *J. Phys. Act. Health* **2009**, *6 Suppl 1*, S113–23, doi:10.1123/jpah.6.s1.s113.

42.   Tan, M.; Le, Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *CoRR* **2019**, *abs/1905.11946*.

43.   Wada, K.; Mpitid; Buijs, M.; Zhang Ch., N.; なるみ; Kubovčík, B. M.; Myczko, A.; Latentix; Zhu, L.; Yamaguchi, N.; Fujii, S.; Iamgd67; IlyaOvodov; Patel, A.; Clauss, C.; Kuroiwa, E.; Iyengar, R.; Shilin, S.; Malygina, T.; Kawaharazuka, K.; Engelberts, J.; J, A.; AlexMa; Song, C.; Rose, D.; Livingstone, D.; Toft, H. wkentaro/labelme: v4.6.0. *Zenodo* **2021**, doi:10.5281/zenodo.5711226.

44.   Ying, X. An overview of overfitting and its solutions. *J. Phys.: Conf. Ser.* **2019**, *1168*, 022022, doi:10.1088/1742-6596/1168/2/022022.

45.   Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; Berg, A. C.; Fei-Fei, L. ImageNet large scale visual recognition challenge. *Int J Comput Vis* **2015**, *115*, 211–252, doi:10.1007/s11263-015-0816-y.

46.   Huh, M.; Agrawal, P.; Efros, A. A. What makes ImageNet good for transfer learning? **2016**.

47.   Gebel, K.; Bauman, A.; Owen, N. Correlates of non-concordance between perceived and objective measures of walkability. *Ann Behav Med* **2009**, *37*, 228–238, doi:10.1007/s12160-009-9098-3.

48.   Neuhold, G.; Ollmann, T.; Bulo, S. R.; Kontschieder, P. The mapillary vistas dataset for semantic understanding of street scenes. In *2017 IEEE International Conference on Computer Vision (ICCV)*; IEEE, 2017; pp. 5000–5009.

49.   Brostow, G. J.; Fauqueur, J.; Cipolla, R. Semantic object classes in video: A high-definition ground truth database. *Pattern Recognit Lett* **2009**, *30*, 88–97, doi:10.1016/j.patrec.2008.04.005.