Landscape and Urban Planning 183 (2019) 122–132

Contents lists available at ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

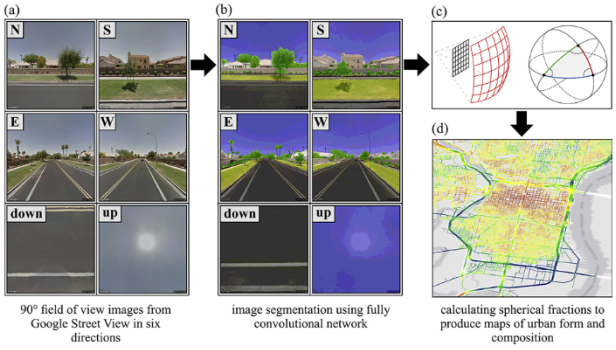
Research Paper

Urban form and composition of street canyons: A human-centric big data and deep learning approach

Ariane Middel[a,*], Jonas Lukasczyk[b], Sophie Zakrzewski[b], Michael Arnold[c], Ross Maciejewski[d]

[a] School of Arts, Media and Engineering (AME), School of Computing, Informatics, and Decision Systems Engineering (CIDSE), Arizona State University, 950 S. Forest Mall, Stauffer B, Tempe, AZ 85281, USA
[b] Department of Computer Science, Technische Universität Kaiserslautern, PO Box 3049, Kaiserslautern, Germany
[c] TerraLoupe, Germany
[d] School of Computing, Informatics and Decision Systems Engineering, Arizona State University, 699 S Mill Ave, Tempe, AZ 85281, USA

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Urban form and composition
Street canyon
Human-centric
Spherical fractions
Deep learning
Google Street View

ABSTRACT

Various research applications require detailed metrics to describe the form and composition of cities at fine scales, but the parameter computation remains a challenge due to limited data availability, quality, and processing capabilities. We developed an innovative big data approach to derive street-level morphology and urban feature composition as experienced by a pedestrian from Google Street View (GSV) imagery. We employed a scalable deep learning framework to segment 90-degree field of view GSV image cubes into six classes: sky, trees, buildings, impervious surfaces, pervious surfaces, and non-permanent objects. We increased the classification accuracy by differentiating between three view directions (lateral, down, and up) and by introducing a void class as training label. To model the urban environment as perceived by a pedestrian in a street canyon, we projected the segmented image cubes onto spheres and evaluated the fraction of each surface class on the sphere. To demonstrate the application of our approach, we analyzed the urban form and composition of Philadelphia County and three Philadelphia neighborhoods (suburb, center city, lower income neighborhood) using stacked area graphs. Our method is fully scalable to other geographic locations and constitutes an important step towards building a global morphological database to describe the form and composition of cities from a human-centric perspective.

* Corresponding author at: School of Arts, Media and Engineering, School of Computing, Informatics and Decision Systems Engineering, Arizona State University, 950 S. Forest Mall, Stauffer B, Tempe, AZ 85281, USA.
E-mail addresses: ariane.middel@asu.edu (A. Middel), j_lukasczy09@informatik.uni-kl.de (J. Lukasczyk), zakrzewski@cs.uni-kl.de (S. Zakrzewski), intelligent-design@protonmail.com (M. Arnold), rmacieje@asu.edu (R. Maciejewski).

https://doi.org/10.1016/j.landurbplan.2018.12.001
Received 22 October 2017; Received in revised form 2 December 2018; Accepted 3 December 2018
Available online 14 December 2018

1. Introduction

Urban composition and urban form—i.e. the physical characteristics of built environments including the configuration, shape, size, and density of urban features—are important parameters for analyses in urban climate, ecology, planning, and design. Research has shown that urban form and composition impact local and micro-scale climate (Coutts, Beringer, & Tapper, 2007; Middel, Häb, Brazel, Martin, & Guhathakurta, 2014; Stewart & Oke, 2012), ecosystem performance (Alberti & Marzluff, 2004; Tratalos, Fuller, Warren, Davies, & Gaston, 2007), outdoor human thermal comfort (Johansson, 2006; Middel, Lukasczyk, & Maciejewski, 2017; Middel, Selover, Hagen, & Chhetri, 2016), land surface temperature (Li, Li et al., 2016; Zhang, Murray, & Turner, 2017; Zhou, Huang, & Cadenasso, 2011), energy use (Anderson, Kanaroglou, & Miller, 1996; Ewing & Rong, 2008), travel behavior (Dieleman, Dijst, & Burghouwt, 2002; Handy, 1996), physical activity and public health (Frank & Engelke, 2001; Hankey & Marshall, 2017; Jackson, Dannenberg, & Frumkin, 2013), traffic safety (Dumbaugh & Rae, 2009), crime (Cozens, 2011), and environmental services such as water use (Shandas & Parandvash, 2010).

The composition and configuration of built-up environments is of particular interest to the urban climate modeling and urban planning communities because the vertical dimension of cities (height-to-width ratio of buildings and streets, sky view factor, etc.) plays a crucial role in model parameterizations and sustainable neighborhood design. Large scale urban morphological datasets have traditionally been subject to a tradeoff between resolution and spatial coverage. New datasets have become increasingly available – e.g., from LIDAR (Light Detection and Ranging) – but are often expensive, do not account for the rich geometric and semantic structures of cities, or do not provide city-wide, regional, or global coverage.

Due to the limited availability of extensive, yet detailed urban morphological datasets, remotely sensed images such as NAIP (National Agriculture Imagery Program), Quickbird, and Landsat have been used to describe the configuration and composition of urban areas (Kane, Connors, & Galletti, 2014; Li et al., 2011; Li, Li et al., 2016; Li, Kamarianakis, Ouyang, Turner, & Brazel, 2017; Myint et al., 2015), especially in the context of land surface architecture that seeks to address the mosaic of land units (Turner, Janetos, Verbug, & Murray, 2013; Turner, 2016). However, land cover segmentations from satellite imagery fail to capture ground surfaces that are obstructed by horizontal features such as tree canopies and do not explicitly represent the vertical dimension of urban features; therefore, satellite imagery does not explicitly account for shading effects. At the individual person scale, satellite-derived products do not meet the requirements of heterogeneous urban areas (Vanos, Middel, McKercher, Kuras, & Ruddell, 2016), and the land cover configuration and composition does not correspond to how people experience cityscapes, which is crucial for human impact assessments.

In recent years, street-level urban imagery from online products such as Google Street View (GSV), Baidu Maps, and Mapillary have increasingly become available to the public. Those images are acquired within urban street canyons and provide a human-centric view of the built environment. With the increasing availability of high-resolution imagery, big data approaches that yield detailed urban morphology have become feasible (Bechtel et al., 2017; Li, Zhang, Li, Ricard et al., 2015; Li, Zhang, Li, Kuzovkina, & Weiner, 2015; Middel, Lukasczyk, Maciejewski, Demuzere, & Roth, 2018).

We present an innovative approach to assess the configuration and composition of cities at high spatial resolution (approximately 10 m at street level) from a human-centric, within-street-canyon perspective using GSV imagery. These images are evaluated semantically through deep learning—a machine learning technique that has been successfully employed for automatic land cover classification (Xu, Zhu, Fu, Dong, & Xiao, 2017), to detect objects in photographs (Chen, Papandreou, Kokkinos, Murphy, & Yuille, 2018; Girshick, Donahue, Darrell, & Malik, 2014; Ren, He, Girshick, & Sun, 2015; Yin, Cheng, Shao, Wang, & Wu, 2017), extract buildings from remotely sensed images (Maltezos, Doulamis, Doulamis, & Ioannidis, 2017; Zhang, Zhang & Du, 2016; Zhang, Wang, Liu, Liu, & Wang, 2016), and to count building floors from photographs (Iannelli & Dell'Acqua, 2017).

We focus on urban surface type classes that are most relevant for climate and planning applications, i.e. buildings, trees and plants, impervious surfaces, pervious surfaces, sky, and non-permanent objects. To quantify the composition of surface types that pedestrians experience in the street canyon, we project segmented image cubes onto a unit sphere that surrounds each GSV image location and calculate the area of each surface type class on the sphere. Our approach is innovative because it creates a unique description of urban form and composition as experienced by a pedestrian in a street canyon and is therefore more relevant to the human experience of cities when compared to planar bird's eye views from satellite data. The methodology is fully scalable to large urban areas and, if applied city-wide, yields a high-resolution urban form and composition dataset that can inform urban design, assist in land cover and green space management, facilitate climate model parameterizations at various scales, and is a step

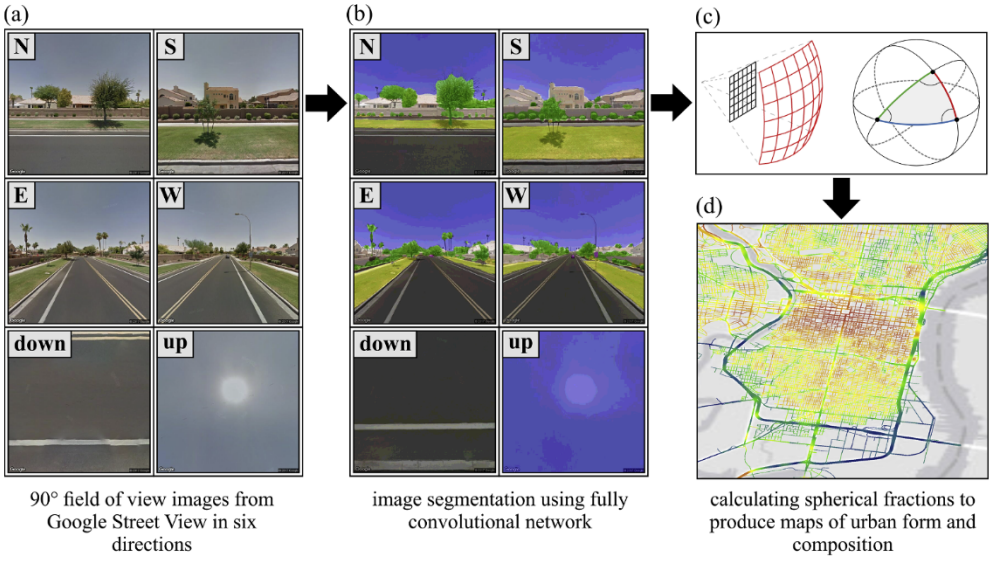


Fig. 1. Workflow for image retrieval (a), segmentation (b), spherical surface type fraction calculation (c), and mapping (d).

towards building a global morphological database of cities.

2. Materials and methods

Our approach consists of three steps: image acquisition, segmentation, and spherical fraction processing (Fig. 1). First, we retrieved all available GSV locations and corresponding imagery for a given area using the methodology presented by Middel et al. (2018). Second, we fine-tuned a pre-trained fully convolutional neural network (FCN) for three image view directions, i.e., lateral, up, and down with manually labeled GSV images from various cities around the world (Section 2.1). Finally, we derived spherical surface type fractions from the segmented images at each location using a cube-to-sphere projection and evaluated image pixel areas on the sphere through the concept of spherical excess (Section 2.2).

2.1. Image segmentation

Neural networks have gained popularity over the past few years due to the availability of parallel computing resources, large amounts of labeled data, and the availability of deep learning frameworks such as Caffe (Jia et al., 2014) and Tensorflow (Abadi et al., 2016). Such frameworks have shifted the paradigm of hand-crafting features towards end-to-end learning, where learning feature embeddings is part of the optimization process. Neural network research has shown that stacking convolutional hidden layers with non-linear activation functions can build hierarchies of abstractions. In these convolutional neural networks (CNNs), lower layers usually encode simple features, such as colors and gradients, while higher layers go from representing parts of objects and simple local features to the representation of faces and objects. A special case of CNNs is the fully convolutional network (FCN), which is built only from locally connected layers. That enables the use of variable image sizes and reduces the number of parameters and computation time. For a comprehensive comparison of various deep learning implementations over a wide range of parameter configurations and a performance analysis of CNNs we refer to Sze, Chen, Yang, and Emer (2017) and Li, Zhang, Huang, Wang, and Zheng (2016).

To segment GSV images into different urban features, we used the Caffe deep learning framework (Jia et al., 2014), which is based on an FCN. The quality of the segmentation strongly depends on the network's training and layout. It is challenging to achieve good segmentation results with a limited amount of labeled data, as small datasets are inherently biased and cannot account for the vast variability of unseen data. For the same reason, a network trained on lateral photographs of daytime scenes does not perform well if it is tasked to segment nighttime, fisheye, panorama, or upwards facing images, since such images were not represented in the training dataset (Fig. 2).

We utilized an instance of the FCN by Long, Shelhamer, and Darrell (2015) that became famous for its superior efficiency at semantic segmentation during the “Pascal Visual Object Classification” challenge in 2012. It is fully convolutional as it derives, for each individual input signal (the pixels of an image), a corresponding output signal (a pixel label) independent of pixel position. The FCN is based on the SIFT Flow dataset (Liu, Yuen, & Torralba, 2011), a collection of 2688 images with pixel labels for 33 semantic classes (Table 1). We aggregated these classes into four surface type categories (trees and plants, buildings, pervious surfaces, impervious surfaces), sky, and non-permanent objects that are not part of the urban fabric (bird, boat, bus, car, cow, person).

The original FCN was trained on lateral (landscape oriented) photographs, but we require a network that can reliably segment upwards and downwards facing views as part of the 6-directional GSV image dataset. To avoid significant segmentation errors, such as those shown in Fig. 2, we fine-tuned the FCN, i.e., initialized it with weights from the already trained network and trained on our own GSV dataset from the bottom up. To fine-tune the network, we retrieved 2,634 GSV images for locations around the world representing a wide range of urban forms, designs, and materials (Table 2).

We manually segmented the images into sky, building, impervious, pervious, tree, and non-permanent object classes for a pixel-wise prediction. To generate a ground truth dataset, a total of 257 lateral view images, 232 upwards facing view images, and 237 downwards facing view images were classified using the original FCN-8s. Segmentation errors were identified visually through a MatLab user interface and then manually corrected in a “paint-by-number” fashion. Additionally, 40 upwards facing view images and 346 downwards facing view images were available that contained only one label class and therefore did not have to be corrected. As the segmentation quality strongly depends on the size of the training dataset, yet labeling is time consuming, we introduced a void label, similar to the work done by Cordts et al. (2016): a class that is ignored in the calculation of the weight-gradient during training (Fig. 3). The void label ensures that only representative pixels of a given class are selected during the manual labeling process, while features such as fences, power lines, and small objects are ignored during training. Lastly, to increase the amount of ground truth and balance the frequency of labels, images that contained at least two different classes were mirrored vertically and added to the training dataset.

The set of labeled images for each view direction was randomly split into a training and testing dataset at a ratio of 4:1. We then trained the network using unsorted images, a stepwise refinement from 32 strides (FCN-32-s, meaning the filter convolves around the input image by shifting 32 units at a time) to 8 strides (FCN-8-s), and a separate model for each view direction (lateral, down, and up). Training and testing were performed in the deep learning framework Caffe (Jia et al., 2014) on a single NVIDIA TITAN X Pascal graphics card. Average processing

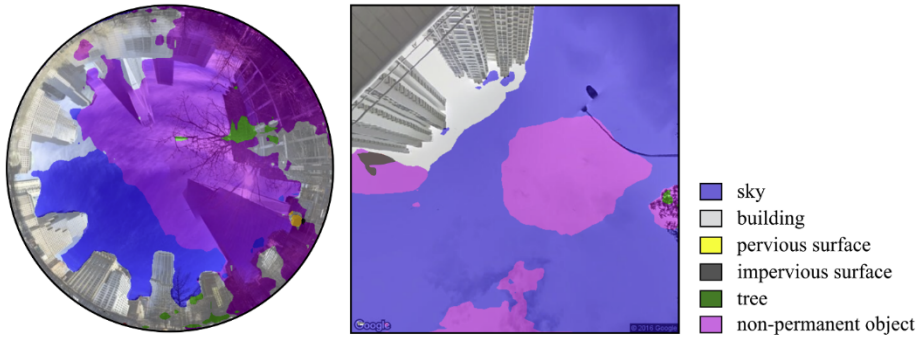


Fig. 2. Image segmentation results for a fisheye photo and upwards facing photograph, performed with a fully convolutional network trained on conventional photos and a pixel accuracy of 90.3% (Long et al., 2015).

Table 1
Mapping from 33 original SIFT flow model categories (Liu et al., 2011) to six classes that describe urban features.

(1) Sky (Blue)	(2) Trees & plants (GREEN)	(3) Buildings (LIGHT GREY)	(4) Impervious surfaces (DARK GREY)	(5) Pervious surfaces (YELLOW)	(6) Non-permanent objects (PURPLE)
moon (16)	plant (19)	awning (01)	bridge (05)	field (14)	bird (03)
sky (28)	tree (32)	balcony (02)	crosswalk (10)	grass (15)	boat (04)
sun (31)		building (06)	desert (11)	river (21)	bus (07)
		door (12)	road (22)	sea (25)	car (08)
		fence (13)	rock (23)		cow (09)
		pole (20)	sand (24)		person (18)
		sign (27)	sidewalk (26)		
		streetlight (30)	staircase (29)		
		window (33)			

Table 2
Cities around the world that are included in the Google Street View image training dataset used for fine-tuning.

North America	Europe	Elsewhere
Arlington, VA	Amsterdam, NL	Buenos Aires, AR
Boston, MA	Athens, GR	Capetown, ZA
Chicago, IL	Barcelona, ES	Dhaka, BD
Las Vegas, NV	Bonn, DE	Dubai, AE
Los Angeles, CA	Dublin, IE	Hong Kong, HK
Manhattan, NY	Gothenburg, SE	Jerusalem, IL
Philadelphia, PA	London, GB	Melbourne, AU
Phoenix, AZ	Monaco, MC	Moskau, RU
San Francisco, CA	Nizza, IT	Rio de Janeiro, BR
Seattle, WA	Oslo, NO	Seoul, KR
Tempe, AZ	Paris, FR	Singapore, SG
Vancouver, CA	Prag, CZ	Tel Aviv, IL
Washington, DC	Sofia, BG	Tokyo, JP
	Stockholm, SE	
	Zurich, CH	

time per image was 0.114 s.

2.2. Spherical surface fractions from segmented image cube

To calculate the composition of surface type fractions as experienced by a pedestrian in an urban street canyon at a given location, we projected the segmented GSV image cube onto a unit sphere and computed the area contribution for each surface type class on the sphere (Fig. 4). We determined the surface area of each pixel on the sphere – also known as the spherical excess (Zwillinger, 1995) – by partitioning each pixel into two triangles; projecting the triangle nodes onto the unit sphere; computing the arc lengths and angles between the points, and; determining the surface area of the projected triangles.

We project a corner (u, v) $\in [-1; 1]^2$ onto the unit sphere with the map:

$$p(u, v) \rightarrow \frac{1}{\sqrt{u^2 + v^2 + 1}} \begin{pmatrix} 1 \\ u \\ v \end{pmatrix}. \quad (1)$$

The latitude ϕ and longitude θ of each projected corner point x are given by

$$\phi(x) = arcsin\frac{x_2}{\|x\|} \text{ and } \theta(x) = arctan\frac{x_1}{x_0},$$

respectively. We compute the arc length d on the sphere between the two points (ϕ_0, θ_0) and (ϕ_1, θ_1) as follows:

$$d(\phi_0, \theta_0, \phi_1, \theta_1) = 2arcsin\sqrt{sin^2\left(\frac{\phi_1 - \phi_0}{2}\right) + cos(\phi_1)cos(\phi_0)sin^2\left(\frac{\theta_1 - \theta_0}{2}\right)}. \quad (2)$$

For one triangle with arc lengths A, B, and C computed by Eq. (2), we derive the angles

$$\alpha = arccos\frac{\cos(A) - \cos(B)\cos(C)}{\sin(B)\sin(C)}$$

$$\beta = arccos\frac{\cos(B) - \cos(C)\cos(A)}{\sin(C)\sin(A)}, \text{ and}$$

$$\gamma = arccos\frac{\cos(C) - \cos(A)\cos(B)}{\sin(A)\sin(B)},$$

to finally compute the spherical excess E of the projected triangle via

$$E = \alpha + \beta + \gamma - \pi. \quad (3)$$

On the unit sphere, the spherical excess corresponds to the area of the triangle projected on its surface. Hence, we can compute the area per pixel covered on the sphere's surface. Since the spherical excess per pixel is the same for all six view directions, we can pre-compute the pixel areas and store them in a lookup table. To calculate the fractions, we summarize all pixel excesses according to their labeled surface types.

3. Results

3.1. Model accuracy

We report four widely accepted metrics to assess the accuracy and performance of our fine-tuned FCN model (Table 3): (a) pixel accuracy $\sum_i n_{ii} / \sum_i t_i$, (b) mean accuracy $(1/n_{cl})(PA)$, (c) mean intersection over union (IoU) $(1/n_{cl}) \sum_i n_{ii}/\left(t_i + \sum_j n_{ji} - n_{ii}\right)$, and (d) frequency weighted intersection over union $(\sum_k t_k)^{-1} \sum_i t_i n_{ii}/\left(t_i + \sum_j n_{ji} - n_{ii}\right)$. Here, n_{ij} denotes the number of pixels of class i that were predicted to class j (out of n_{cl} classes, with the total number of pixels $t_i = \sum_j n_{ij}$ in class i). For semantic segmentation, mean IoU is a good bench-marking metric, because it controls for the total number of pixels in a class, is not dominated by the assessment of background pixels, and, as opposed to mean accuracy, penalizes for false positives. Our FCN-8s prediction achieved over 95% pixel accuracy for all fine-tuned nets, with a reasonable mean IoU for lateral views (0.841) and excellent classification results for the upwards and downwards facing views (mean IoU 0.939 and 0.984, respectively). Decreasing the stride to 4 did not further improve the quality of the classification.

The confusion matrices for the FCN-8s prediction (Fig. 5) show that the network performs best at detecting sky, roads, buildings, and trees (Fig. 6 a-h) and performs worse for detecting non-permanent objects and pervious surfaces (Fig. 6i-p). Since the non-permanent object class has the least characteristic features and can appear anywhere in the image, it is often mistaken for other classes, such as buildings or roads (Fig. 6n-p). Another problem is that some surfaces might look like other classes, e.g., dry grass is falsely classified as road (Fig. 6j-k). Accuracy metrics based on the confusion matrix (Precision, Recall, and F1) are summarized in Table 4.

Fig. 3. Manually labeled images for network fine-tuning; lateral images (a-b, e-f, i-j), downwards facing images (c, g, k), and upwards facing images (d, h, l). Segmentation classes are sky (blue), pervious surfaces (yellow), trees (green), buildings (grey), impervious surfaces (dark grey), and non-permanent objects (purple); void class (turquoise) denotes pixels that were not labeled to belong to one of the six classes and is ignored during the training process. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

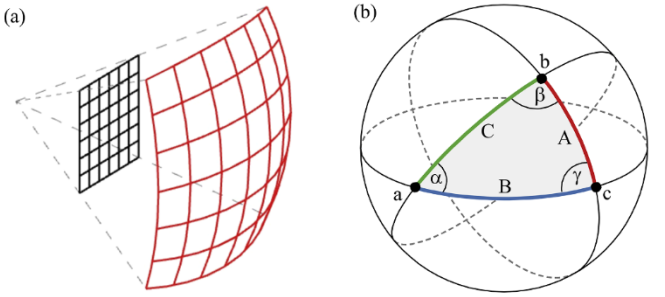


Fig. 4. (a) Projection *p (u,v)* (Eq. (1)) to map a 90-degree image (black) onto the surface of the unit sphere (red). (b) Spherical excess computation for the projected points *(a, b, c)* where we first compute their respective arc lengths *(A, B, C)* and then their spherical angles to finally compute the surface area *E* (Eq. (3)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Classification accuracy results for the fine-tuned FCN-8 s: Training loss (the error on the training set of data), accuracy, mean accuracy, mean Intersection over Union, and Frequency weighted Intersection over Union (unbalanced, similar to pixel accuracy).

	loss	accuracy	mean accuracy	mean IoU	Frequency weighted IoU
FCN-8s lateral	39,676.1	0.950	0.919	0.841	0.907
FCN-8s down	8409.5	0.989	0.971	0.939	0.979
FCN-8s up	4496.6	0.995	0.992	0.984	0.989

3.2. Urban environment composition

We computed surface fractions on a complete sphere for all GSV outdoor locations in Philadelphia County, yielding 1,149,047 data points (Fig. 7). On average, the impervious fraction is the largest fraction (0.29 ± 0.07), followed by sky (0.24 ± 0.10), trees (0.23 ± 0.17), buildings (0.13 ± 0.14), pervious surfaces (0.07 ± 0.06), and non-permanent objects (0.04 ± 0.04). The mean fractional contribution of non-permanent objects exhibits high variability and is dependent on the time of day and day of the week when the Google images were taken. The impervious surface fraction has a low standard deviation because most locations are on asphalt roads with similar coverage on the lower hemisphere. Pervious surface cover is underrepresented across the urban area because GSV images do not include complete coverage of parks, green spaces, and backyards.

As all surface fractions are evaluated on a sphere, the sky fraction generally does not exceed 0.50. Yet, in special cases when the images were acquired at elevated locations, such as a bridge or hill, the sky fraction can slightly exceed 0.50. Please note that doubling the sky fraction is not equal to the Sky View Factor (SVF), which is defined as the fraction of sky observed from a point as a proportion of the total possible sky hemisphere (Oke, 1981). In contrast to the spherical sky fraction, the SVF is evaluated on a planar surface using annular rings (Steyn et al., 1986). Here, we use the spherical sky fraction for consistency with other fraction types.

We focused on three sub areas in Philadelphia County to examine the differences in urban form and composition: Philadelphia Center City (downtown area characterized by midrise to high-rise buildings and few trees); a suburb north of Philadelphia (detached single-family homes

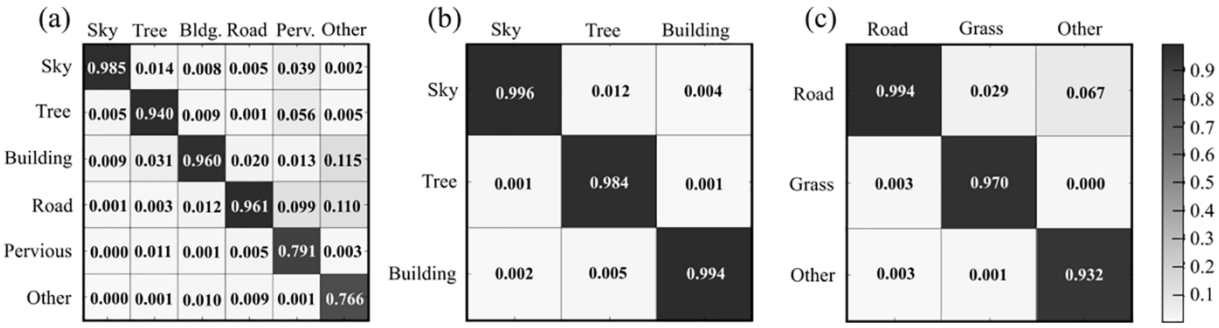


Fig. 5. Confusion matrix of ground truth (x-axis) vs. classified (y-axis) for lateral views (a), upward view (b), and downward view (c).

with sizeable yards), and; a lower-income area. These sites were selected because they exhibit significant differences in urban form (low, medium, and high density), vegetation cover (low to high), and socio-economic status. According to the American Community Survey from 2016, median income in those neighborhoods ranges from $141,302 (suburb) to $70,204 (Center City) to $52,552 (lower-income neighborhood).

We created stacked area graphs to illustrate the fractional distribution of each surface type per study area (Fig. 8) and the composition of fractions at each location per study area (Fig. 9). While the lower-income neighborhood shows a dominant building fraction for values 0.5 to 1, the total amount of locations that yield such high fractions are relatively low (Fig. 8). Buildings generally have the largest fractions except in the suburb. The lower income neighborhood has the least amount of tree canopy cover and the suburban area the highest. The distribution of impervious areas is roughly the same for all four

Fig. 6. Sample image segmentation results illustrating segmentations with an accuracy > 95% (a-h) and an accuracy < 95% (i-p).

Table 4
Classification accuracy results for the fine-tuned FCN-8s: Precision, Recall, and F1.

		Precision	Recall	F1
Lateral View	sky	0.985	0.954	0.969
	tree	0.940	0.962	0.951
	building	0.960	0.948	0.954
	road	0.961	0.964	0.963
	pervious	0.791	0.874	0.831
	non-permanent objects	0.766	0.813	0.789
	Lateral View Overall	**0.901**	**0.919**	**0.909**
Upward View	sky	0.996	0.997	0.997
	tree	0.984	0.987	0.986
	building	0.994	0.993	0.994
	Upward View Overall	**0.992**	**0.992**	**0.992**
Downward View	road	0.994	0.993	0.994
	grass	0.970	0.969	0.970
	other	0.932	0.950	0.941
	Downward View Overall	**0.966**	**0.971**	**0.968**

areas because GSV images are biased towards asphalt road ground surfaces (Fig. 8). Although non-permanent objects introduce uncertainty as they obscure the surfaces behind them, they can be used as an indicator for human activity, since they are more prominent in inner-city areas (Fig. 8 a-c) than in suburban areas (Fig. 8 d). The sky fraction does not exceed 0.5 as, for street views, the sky cannot cover more than the entire upper hemisphere. The fraction composition graph (Fig. 9) further highlights the lack of green infrastructure in the lower-income neighborhood (both trees and grass) as well as a reduced pervious surface cover in the downtown area. The suburb and Center City exhibit a similar distribution of sky fraction, which is reduced compared to the lower-income neighborhood and Philadelphia County as a whole. In the suburb, the horizon limitation is caused by trees, while the downtown area features taller buildings.

4. Discussion

Our methodology to assess urban form and composition from a human-centric perspective has several advantages over existing approaches. Compared to remote sensing techniques that analyze urban environments from a bird's eye view (Li et al., 2011; Li et al., 2017; Myint et al., 2015; Turner et al., 2013), our approach incorporates the vertical dimension of urban features and captures areas that are obstructed by horizontal features, such as surfaces under tree canopies. Heterogeneous urban areas are evaluated at the individual person scale, and the resulting composition and configuration of street-level features corresponds to how pedestrians experience cities, which is critical for human impact assessments.

GSV images are an excellent resource to derive urban parameters at street level. As opposed to previous studies that use GSV imagery (Li, Zhang, Li, Ricard et al., 2015; Shen et al., 2017), we employ a comprehensive sampling with full coverage of all available outdoor GSV locations and their entire spherical environment, yielding > 1 million points for Philadelphia County with an average spacing of ca. 10 m. However, GSV image locations are biased towards streets and do not offer complete coverage of parks and open areas at this point. To provide even more coverage, Google has started to sample parks, university campuses, and hiking trails around the world using cameras mounted on backpacks. Until full coverage of urban areas is available, GSV data could be complemented with satellite-based or drone-captured data to cover missing areas.

Several approaches exist that process segmented GSV data to derive metrics such as green cover (Li, Zhang, Li, Ricard et al., 2015; Li, Zhang, Li, Kuzovkina et al., 2015) or sky exposure (Shen et al., 2017). These GSV image-based segmentation studies use pre-trained networks such as SegNet (Badrinarayanan, Kendall, & Cipolla, 2015) and process lateral views only, thus missing tree canopies, bridges, and sometimes whole building facades. To compensate for the missing vertical information, SegNet was also used to process panorama images (Liang et al., 2017), which leads to segmentation errors due to distortions that are introduced by the cylindrical projection. These errors occur because SegNet was not trained on panorama images. Our GSV image segmentation improves existing classification strategies by fine-tuning three separate networks that distinguish between lateral, upwards facing, and downwards facing image orientations instead of using a single off-the-shelf neural network for all views that was not adequately trained to process non-lateral images. Including ground truth from cities all over the world into our fine-tuning, we achieved accuracies of $IoU = 0.841$ (lateral), $IoU = 0.939$ (down), and $IoU = 0.984$ (up) with an overall accuracy of 0.950 (lateral), 0.989 (down), and 0.995 (up). In comparison, Shen et al. (2017) reported an off-the-shelf SegNet accuracy of 0.828 (lateral), and Liang et al. (2017) reported an achieved accuracy of 0.961 for sky pixels.

Existing approaches derive urban form by counting pixels of segmented lateral images. In contrast, our method models how the environment is perceived by pedestrians at street level by evaluating the surrounding urban feature classes on a sphere. We currently employ six urban surface types, but the list of urban features could be refined to include building types or tree species, which would require considerably more ground truth and fine-tuning. A shortcoming of our segmentation is the absence of a water label. Water bodies are rare in GSV images but could be included in the segmentation if enough ground truth was labeled. We classified water bodies as pervious surfaces, which is the most appropriate class considering the thermal properties of all available categories.

Our methodology is scalable to other cities, and efforts to build a global urban morphological database are underway (Ching et al., 2018; Ching et al., under review). The approach is also transferable to other big data image repositories, such as Baidu or Open Street Map, and can be extended to high resolution drone image data or point-based LIDAR data. Efforts are currently underway to create Sky View Factor (SVF) and spherical fraction datasets from GSV imagery for 70 cities around the world (Ching et al., under review; Middel et al., 2018) that will be hosted on the World Urban Database and Access Portal Tools (WUDAPT) website along with other urban canopy parameters for multi-scale climate modeling. The SVFs and spherical fractions have already been used to evaluate local climate zones in various studies (Bechtel et al., 2019; Demuzere et al., under review; Wang et al., 2018) and to assess the accuracy of procedural 3D city model generation (Ching et al., under review). A recent study integrated the spherical fractions with traditional planar land cover fractions from remote sensing and socio-economic data to estimate daytime and nighttime land surface temperature (LST) in Phoenix, Arizona (Zhang et al., under review). Results show that the spherical fractions explain more of the variability in LST, because they capture the verticality of vegetation and building walls that results in shade.

As the spherical fractions represent urban form and land cover composition from a human-centric perspective, a natural use is to relate them to physical activity, outdoor thermal comfort, and heat stress. Kosaka et al. (2018) and Vanos et al. (2019) used the SVF from GSV imagery to model heat stress experienced by marathon runners during the Tokyo 2020 Olympics. We see various other applications and potential uses of the spherical surface fractions. For example, the non-permanent object fraction could be a proxy for the amount of human activity and anthropogenic heat in urban areas. The spherical tree fraction could be used in urban forestry studies and biodiversity assessments. Since the tree fraction represents the amount of vegetation from a vertical perspective, it could be employed to evaluate the availability of shade for heat mitigation in hot urban areas. It would further be interesting to relate the tree fraction to the number of road injuries, traffic safety, and neighborhood crime in cities.

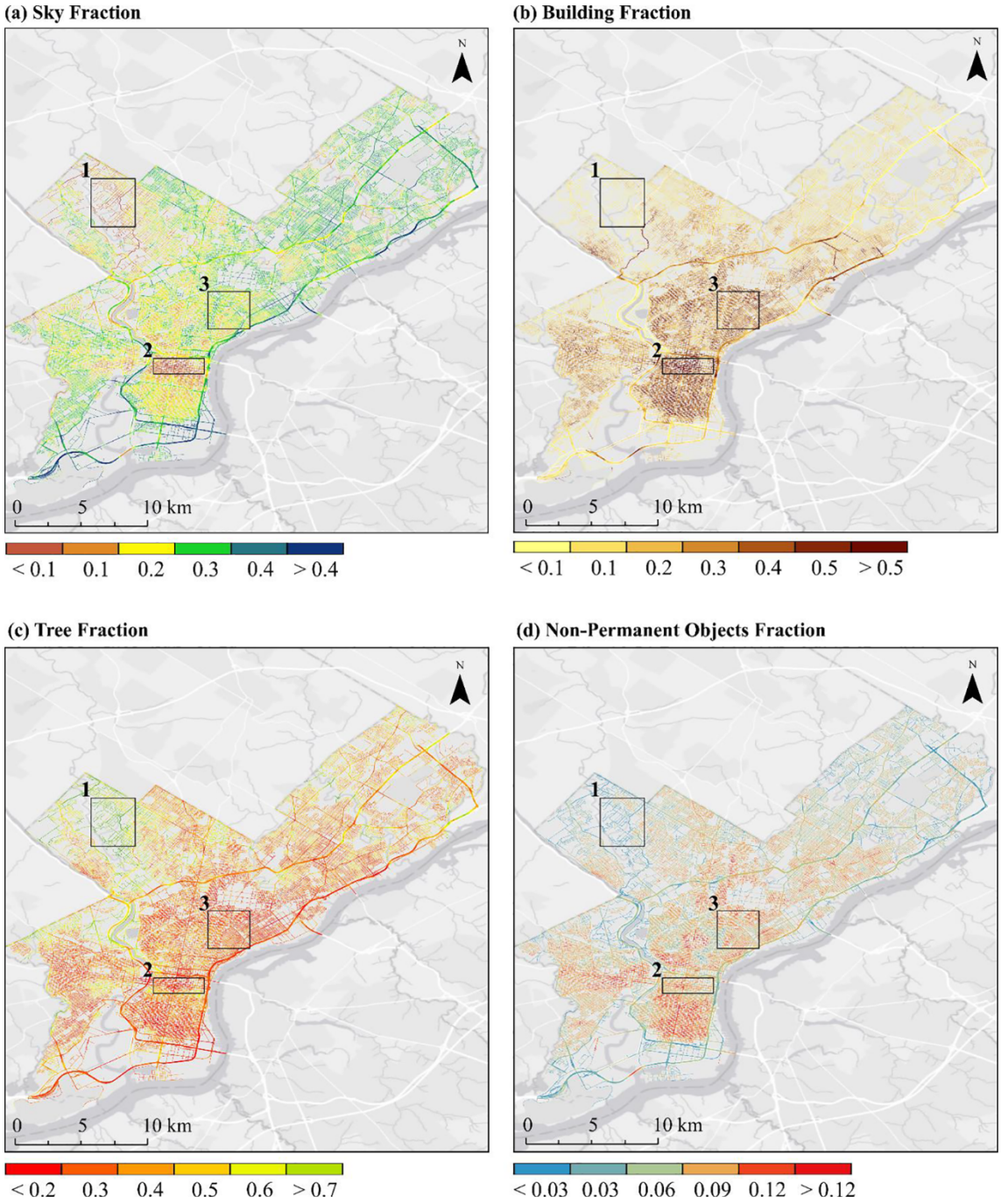


Fig. 7. Spherical sky fraction (a), building fraction (b), tree fraction (c), and non-permanent objects fraction (d) for Philadelphia County, including a suburb (1), Center City (2), and a lower-income neighborhood (3).

5. Conclusions

We presented an innovative big data approach that uses GSV imagery to assess urban form and composition of cities from a human-centric perspective. In contrast to traditional satellite-based assessments, our methodology accounts for the vertical dimension of urban features and calculates their configuration and composition as experienced by a pedestrian in urban street canyons. To calculate surface fractions of the urban environment at high spatial resolution, we fine-tuned a fully convolutional neural network for three image view directions (lateral, down, and up) and segmented GSV image cubes into classes that are relevant for urban planning and climate applications: sky, trees, buildings, impervious surfaces, pervious surfaces, and non-permanent objects. The segmented image cubes were projected onto a sphere, and the area of each feature class was evaluated using spherical excess. Our methodology is fully scalable to any geographic location where street level image cube data or point based classified LIDAR data are available. Spherical urban surface fractions at street level have high potential to inform urban design and assist in land cover and green space management. Applied over a large area, our methodology yields a novel dataset of spherical urban surface fractions for use in climate model parameterizations at various scales. The approach also establishes a universal pathway towards building a global urban morphological database. Above all, this work has the potential to transform how we describe the form and composition of cities towards a more human-centric perspective.

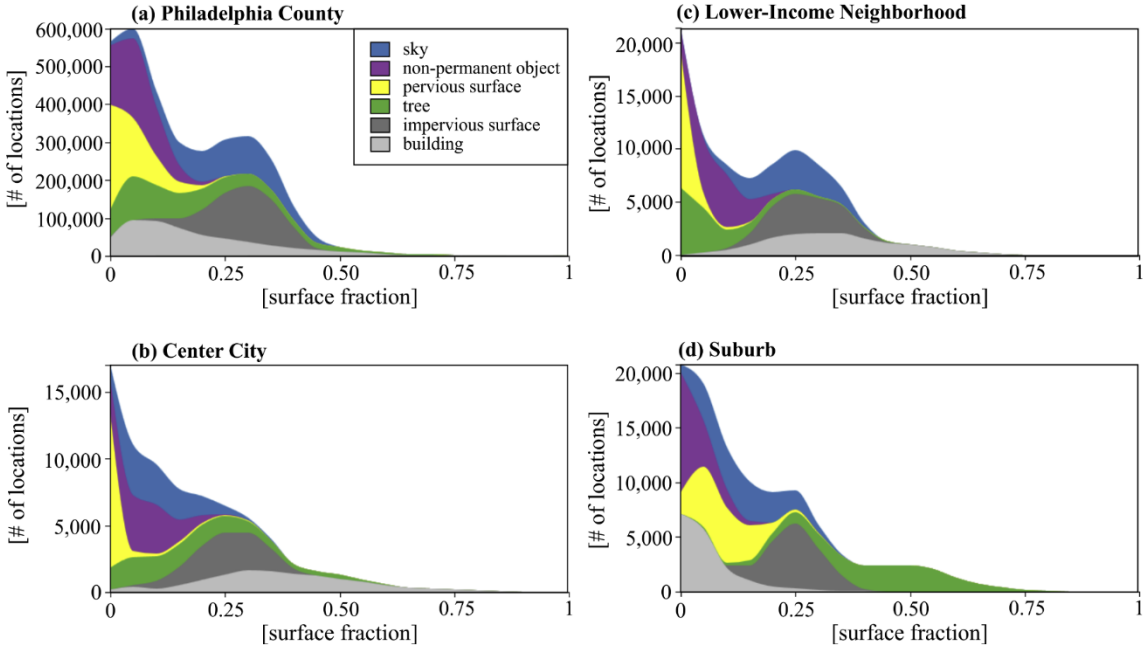


Fig. 8. Stacked area graphs that illustrate the magnitude of different surface type fractions for Philadelphia County (a), Center City (b), a lower-income neighborhood (c), and a suburb (d). The x-axis represents the fraction value; surface types are color-coded. The y-axis shows the total amount of locations per fraction value.

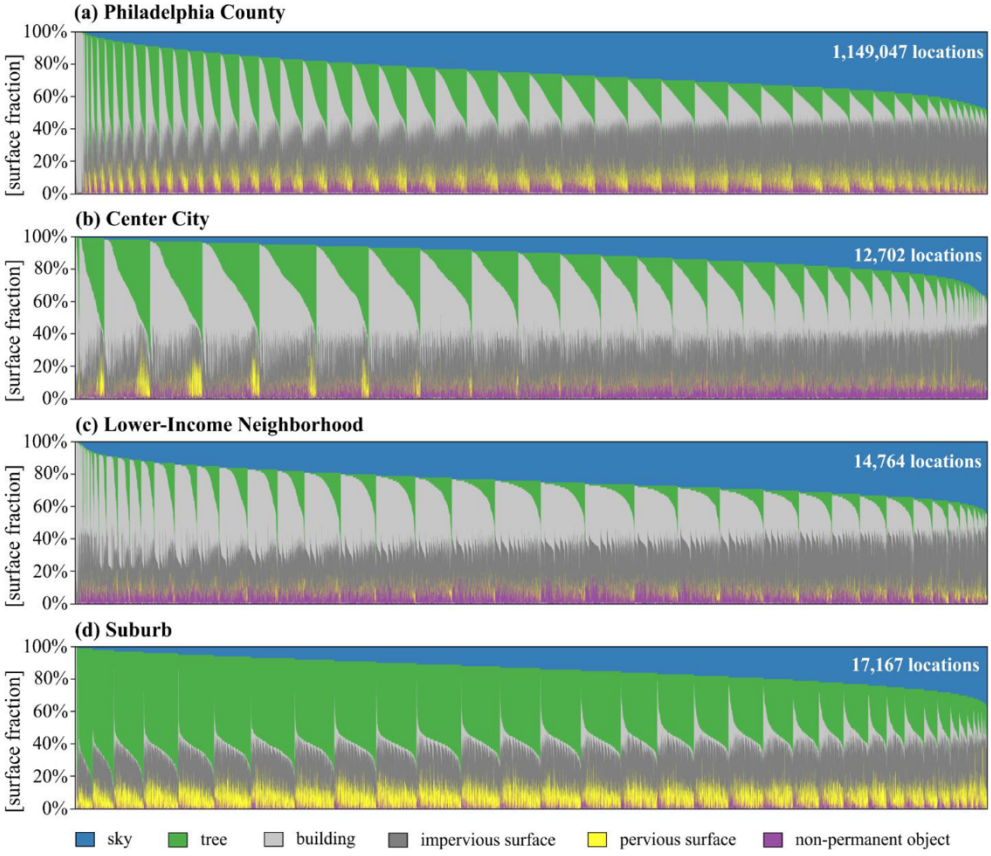


Fig. 9. Stacked area graphs illustrating the composition of surface type fractions for all locations (x-axis) in each study area (Philadelphia County, Center City, lower-income neighborhood, suburb). Locations are sorted by sky, tree, building, impervious, pervious, and non-permanent fractions from low to high.

Acknowledgements

This research was sponsored by University of Kaiserslautern, grant "Microclimate Data Collection, Analysis, and Visualization" and supported by National Science Foundation (NSF) Award Number 1635490, "A Simulation Platform to Enhance Infrastructure and Community Resilience to Extreme Heat Events." As well as NSF Award Number 1639227, "Flexible Model Compositions and Visual Representations for Planning and Policy Decisions for the Food-Energy-Water Nexus". Any opinions, findings, and conclusions or recommendations expressed in

this material are those of the authors and do not necessarily reflect the views of the sponsoring organizations.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). Tensorflow: A system for large-scale machine learning. *OSDI*, 265–283.

Alberti, M., & Marzluff, J. M. (2004). Ecological resilience in urban ecosystems: Linking urban patterns to human and ecological functions. *Urban Ecosystems, 7*(3), 241–265.

Anderson, W. P., Kanaroglou, P. S., & Miller, E. J. (1996). Urban form, energy and the environment: A review of issues, evidence and policy. *Urban Studies, 33*(1), 7–35.

Badrinarayanan, V., Kendall, A., & Cipolla, R., 2015, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, arXiv preprint arXiv: 1511. 00561.

Bechtel, B., Alexander, P. J., Beck, C., Böhner, J., Brousse, O., Ching, J., et al. (2019). Generating WUDAPT Level 0 data – Current status of production and evaluation. *Urban Climate, 27*, 24–45. https://doi.org/10.1016/j.uclim.2018.10.001.

Bechtel, B., Demuzere, M., Sismanidis, P., Fenner, D., Brousse, O., Beck, C., et al. (2017). Quality of crowdsourced data on urban morphology—the human influence experiment (HUMINEX). *Urban Science, 1*(2), 15.

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(4), 834–848.

Ching, J., Aliaga, D., Mills, G., Masson, V., See, L., & Neophytou, M., et al., under review, Pathway using WUDAPT's Digital Synthetic City tool towards generating urban canopy parameters for multi-scale urban atmospheric modeling, Urban Climate.

Ching, J., Mills, G., Bechtel, B., See, L., Feddema, J., Wang, X., et al. (2018). World Urban Database and Access Portal Tools (WUDAPT), an urban weather, climate and environmental modeling infrastructure for the Anthropocene. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/BAMS-D-16-0236.1.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., & Benenson, R., et al., 2016, The cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3213–3223).

Coutts, A. M., Beringer, J., & Tapper, N. J. (2007). Impact of increasing urban density on local climate: Spatial and temporal variations in the surface energy balance in Melbourne, Australia. *Journal of Applied Meteorology and Climatology, 46*(4), 477–493.

Cozens, P. M. (2011). Urban planning and environmental criminology: Towards a new perspective for safer cities. *Planning Practice and Research, 26*(4), 481–508.

Demuzere, M., Bechtel, B., Middel, A., Mills, G., under review, Mapping Europe into Local Climate Zones, PLOS.

Dieleman, F. M., Dijst, M., & Burghouwt, G. (2002). Urban form and travel behaviour: Micro-level household attributes and residential context. *Urban Studies, 39*(3), 507–527.

Dumbaugh, E., & Rae, R. (2009). Safe urban form: revisiting the relationship between community design and traffic safety. *Journal of the American Planning Association, 75*(3), 309–329.

Ewing, R., & Rong, F. (2008). The impact of urban form on US residential energy use. *Housing Policy Debate, 19*(1), 1–30.

Frank, L. D., & Engelke, P. O. (2001). The built environment and human activity patterns: Exploring the impacts of urban form on public health. *Journal of Planning Literature, 16*(2), 202–218.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580–587).

Handy, S. (1996). Methodologies for exploring the link between urban form and travel behavior. *Transportation Research Part D: Transport and Environment, 1*(2), 151–165.

Hankey, S., & Marshall, J. D. (2017). Urban form, air pollution, and health. *Current Environmental Health Reports, 4*(4), 491–503.

Iannelli, G. C., & Dell'Acqua, F. (2017). Extensive exposure mapping in urban areas through deep analysis of street-level pictures for floor count determination. *Urban Science, 1*(2), 16.

Jackson, R. J., Dannenberg, A. L., & Frumkin, H. (2013). Health and the built environment: 10 years after. *American Journal of Public Health, 103*(9), 1542–1544.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., & Girshick, R., 2014. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on multimedia. ACM, pp. 675–678.

Johansson, E., & Emmanuel, R. (2006). The influence of urban design on outdoor thermal comfort in the hot, humid city of Colombo, Sri Lanka. *International Journal of Biometeorology, 51*(2), 119–133.

Kane, K., Connors, J. P., & Galletti, C. S. (2014). Beyond fragmentation at the fringe: A path-dependent, high-resolution analysis of urban land cover in Phoenix, Arizona. *Applied Geography, 52*, 123–134.

Kosaka, E., Iida, A., Vanos, J., Middel, A., Yokohari, M., & Brown, R. (2018). Microclimate variation and estimated heat stress of runners in the 2020 Tokyo Olympic Marathon. *Atmosphere, 9*(5), 192.

Li, X., Kamarianakis, Y., Ouyang, Y., Turner, B. L., II, & Brazel, A. (2017). On the association between land system architecture and land surface temperatures: Evidence from a Desert Metropolis—Phoenix, Arizona, USA. *Landscape and Urban Planning, 163*, 107–120.

Li, X., Li, W., Middel, A., Harlan, S., Brazel, A., & Turner, B. (2016a). Remote sensing of the surface urban heat island and land architecture in Phoenix, Arizona: Combined effects of land composition and configuration and cadastral–demographic–economic factors. *Remote Sensing of Environment, 174*, 233–243.

Li, J., Song, C., Cao, L., Zhu, F., Meng, X., & Wu, J. (2011). Impacts of landscape structure on surface urban heat islands: A case study of Shanghai, China. *Remote Sensing of Environment, 115*(12), 3249–3263.

Li, X., Zhang, G., Huang, H. H., Wang, Z., & Zheng, W. (2016b). Performance analysis of GPU-based convolutional neural networks. *45th international conference on parallel processing (ICPP)* (pp. 67–76). IEEE. https://doi.org/10.1109/ICPP.2016.15.

Li, X., Zhang, C., Li, W., Kuzovkina, Y. A., & Weiner, D. (2015b). Who lives in greener neighborhoods? The distribution of street greenery and its association with residents' socioeconomic conditions in Hartford, Connecticut, USA. *Urban Forestry & Urban Greening, 14*(4), 751–759.

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2015a). Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry & Urban Greening, 14*(3), 675–685.

Liang, J., Gong, J., Sun, J., Zhou, J., Li, W., Li, Y., et al. (2017). Automatic sky view factor estimation from street view photographs—A big data approach. *Remote Sensing, 9*(5), 411.

Liu, C., Yuen, J., & Torralba, A. (2011). Nonparametric scene parsing via label transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33*(12), 2368–2382.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431–3440).

Maltezos, E., Doulamis, N., Doulamis, A., & Ioannidis, C. (2017). Deep convolutional neural networks for building extraction from orthoimages and dense image matching point clouds. *Journal of Applied Remote Sensing, 11*(4), 042620.

Middel, A., Häb, K., Brazel, A. J., Martin, C., & Guhathakurta, S. (2014). Impact of urban form and design on microclimate in Phoenix, AZ. *Landscape and Urban Planning, 122*, 16–28.

Middel, A., Lukasczyk, J., & Maciejewski, R. (2017). Sky View factors from synthetic fisheye photos for thermal comfort routing—A case study in Phoenix, Arizona. *Urban Planning, 2*(1), 19–30.

Middel, A., Lukasczyk, J., Maciejewski, R., Demuzere, M., & Roth, M. (2018). Sky View Factor footprints for urban climate modeling. *Urban Climate, 25*, 120–134.

Middel, A., Selover, N., Hagen, B., & Chhetri, N. (2016). Impact of shade on outdoor thermal comfort—A seasonal field study in Tempe, Arizona. *International Journal of Biometeorology, 60*(12), 1849–1861.

Myint, S. W., Zheng, B., Talen, E., Fan, C., Kaplan, S., Middel, A., et al. (2015). Does the spatial arrangement of urban landscape matter? Examples of urban warming and cooling in Phoenix and Las Vegas. *Ecosystem Health and Sustainability, 1*(4), 1–15.

Oke, T. R. (1981). Canyon geometry and the nocturnal urban heat island: Comparison of scale model and field observations. *Journal of Climatology, 1*(3), 237–254.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 91–99.

Shandas, V., & Parandvash, G. H. (2010). Integrating urban form and demographics in water-demand management: An empirical case study of Portland, Oregon. *Environment and Planning B: Planning and Design, 37*(1), 112–128.

Shen, Q., Zeng, W., Ye, Y., Arisona, S. M., Schubiger, S., Burkhard, R., & Qu, H. (2017). StreetVizor: Visual exploration of human-scale urban forms based on street views. *IEEE Transactions on Visualization and Computer Graphics*.

Stewart, I. D., & Oke, T. R. (2012). Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society, 93*(12), 1879–1900.

Steyn, D. G., Hay, J., Watson, I. D., & Johnson, G. T. (1986). The determination of sky view-factors in urban environments using video imagery. *Journal of Atmospheric and Oceanic Technology, 3*(4), 759–764.

Sze, V., Chen, Y. H., Yang, T. J., & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE, 105*(12), 2295–2329. https://doi.org/10.1109/JPROC.2017.2761740.

Tratalos, J., Fuller, R. A., Warren, P. H., Davies, R. G., & Gaston, K. J. (2007). Urban form, biodiversity potential and ecosystem services. *Landscape and Urban Planning, 83*(4), 308–317.

Turner, B., II (2016). Land system architecture for urban sustainability: New directions for land system science illustrated by application to the urban heat island problem. *Journal of Land Use Science, 11*(6), 689–697.

Turner, B., Janetos, A. C., Verbug, P. H., & Murray, A. T. (2013). *Land system architecture: Using land systems to adapt and mitigate global environmental change.* Richland, WA (US): Pacific Northwest National Laboratory (PNNL).

Vanos, J., Kosaka, E., Iida, A., Yokohari, M., Middel, A., & Brown, B. (2019). Planning for spectator thermal comfort and health in the face of extreme heat: The Tokyo 2020 Olympic marathons. *Sci. Total Environ.*. https://doi.org/10.1016/j.scitotenv.2018.11.447.

Vanos, J. K., Middel, A., McKercher, G. R., Kuras, E. R., & Ruddell, B. L. (2016). Hot playgrounds and children's health: A multiscale analysis of surface temperatures in Arizona, USA. *Landscape and Urban Planning, 146*, 29–42.

Wang, C., Middel, A., Myint, S. W., Kaplan, S., Brazel, A. J., & Lukasczyk, J. (2018). Assessing local climate zones in arid cities: The case of Phoenix, Arizona and Las Vegas, Nevada, ISPRS. *Journal of Photogrammetry and Remote Sensing, 141*, 59–71.

Xu, G., Zhu, X., Fu, D., Dong, J., & Xiao, X. (2017). Automatic land cover classification of geo-tagged field photos by deep learning. *Environmental Modelling & Software, 91*, 127–134.

Yin, L., Cheng, Q., Shao, Z., Wang, Z., & Wu, L. (2017). 'Big Data': Pedestrian volume using google street view images. *Seeing cities through big data* (pp. 461–469). Cham: Springer.

Zhang, Q., Wang, Y., Liu, Q., Liu, X., & Wang, W. (2016b). CNN based suburban building detection using monocular high resolution Google Earth images. In Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International (pp. 661–664). IEEE.

Zhang, Y., Middel, A., & Turner, B. L., under review, Evaluating the Effects of Vertical

Urban Forms on Neighborhood Land Surface Temperature Using Google Street View Images, Landscape Ecology.
Zhang, Y., Murray, A. T., & Turner, B. (2017). Optimizing green space locations to reduce daytime and nighttime urban heat island effects in Phoenix, Arizona. *Landscape and Urban Planning, 165*, 162–171.
Zhang, L., Zhang, L., & Du, B. (2016a). Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine, 4*(2), 22–40.
Zhou, W., Huang, G., & Cadenasso, M. L. (2011). Does spatial configuration matter? Understanding the effects of land cover pattern on land surface temperature in urban landscapes. *Landscape and Urban Planning, 102*(1), 54–63.
Zwillinger, D. (Ed.). (1995). *CRC Standard Mathematical Tables and Formulae* (pp. 469). Boca Raton, FL: CRC Press.