METHOD

Mapping built environments from UAV imagery: a tutorial on mixed methods of deep learning and GIS

Xin Hong[1]* , Scott Sheridan[2] and Dong Li[3]

Abstract

Evidence has suggested that built environments are significantly associated with residents' health and the conditions of built environments vary between neighborhoods. Recently, there have been remarkable technological advancements in using deep learning to detect built environments on fine spatial scale remotely sensed images. However, integrating the extracted built environment information by deep learning with geographic information systems (GIS) is still rare in existing literature. This method paper presents how we harnessed deep leaning techniques to extract built environments and then further utilized the extracted information as input data for analysis and visualization in a GIS environment. Informative guidelines on data collection with an unmanned aerial vehicle (UAV), greenspace extraction using a deep learning model (specifically U-Net for image segmentation), and mapping spatial distributions of greenspace and sidewalks in a GIS environment are offered. The novelty of this paper lies in the integration of deep learning into the GIS decision-making system to identify the spatial distribution of built environments at the neighborhood scale.

Keywords: Deep learning, GIS, UAV mapping, Greenspace, Sidewalks, Image segmentation

1 Introduction

Accumulating evidence has suggested that greenspace and sidewalks are vital built environment attributes associated with physical activities (Gunn et al., 2014; Lu 2018). Detecting sidewalk infrastructure using fine spatial scale images (e.g., street view images) has become a critical part of health studies to measure neighborhood built environments (Janssen & Rosu, 2012; Rundle et al., 2011; Vargo et al., 2012). In addition to the presence of sidewalks, microclimate conditions along sidewalks may also be a major determinant of physical activities. Kim et al. (2018), Mayer et al. (2009), and Park et al. (2017) suggested that sidewalks with more trees and wider grass verges may help to reduce air temperature, and hence

lead to increased use by pedestrians and bicyclists. As much as studies have shown the associations between built environments and health behaviors, disparities of built environments and health outcomes between neighborhoods with different socioeconomic levels are also sufficiently documented. For instance, non-White and low-income residents are more likely living in a neighborhood facilitated with lower sidewalk continuity and worse or absent pedestrian amenities (Kolosna & Spurlock, 2019; Lowe, 2016; Thornton et al., 2016). In addition, they tend to have lower outcomes of physical and mental health (Sallis et al., 2018; Shen, 2022). The goal of this study was to quantify the built environment attributes of greenspace and sidewalks at the neighborhood scale and explore their spatial distribution across neighborhoods with different economic levels. It is expected that this study can be a new framework for structuring deep learning techniques inside a GIS framework and improve the analytical performance in health studies.

*Correspondence: xhong21@ku.edu.tr
[1] Research Center for Anatolian Civilizations, Koç University, Istiklal Cd. No:181, Merkez Han 34433 Beyoğlu, Istanbul, Turkey
Full list of author information is available at the end of the article

Hong *et al. Computational Urban Science*   (2022) 2:12

Page 2 of 15

There are active geographic information system (GIS) methods proposed to analyze built environments at the local level (Campos-Sánchez et al., 2020; Moghadam et al., 2018; Thornton et al., 2011). However, data related to built environments (i.e., street connectivity, land use, building information) in these studies were acquired by traditional means, either field survey or secondary data sources. The analysis could be restrained by either the limited attributes and volume or the availability of the data. With the advancement of computational powers in recent years, Convolutional Neural Network (CNN), a deep learning technique, has been introduced to the field of remote sensing and proven to be a powerful technique for classifying high-volume and fine-spatial resolution images in a fast and automated fashion (Zhang et al., 2016; Zhu et al., 2017). Some examples are building detection, and land use and land cover classification (Konstantinidis et al., 2020; Zhang et al., 2019). There also has been research documenting the use of deep learning to classify built environment attributes. For instance, Ning et al. (2021) extracted sidewalks for four counties in the US from aerial and street view images using the deep learning model YOLACT. Lu et al. (2018) analyzed urban greenness for the whole city of Hong Kong on street view using deep learning techniques. However, the majority of existing research related to image classification and deep learning focuses on the technical performance, e.g., the efficiency of the deep learning model on classifying big data (Konstantinidis et al., 2020, Zhang et al., 2019). The potential utility of structuring the geographical variables extracted using deep learning on remotely sensed data into a GIS framework is not fully explored. Little research has integrated deep learning into image classification as part of the analytical process in GIS in the past few years. Xiong et al. (2018) mapped mineral prospectivity through big data analytics in a GIS environment supported by deep learning. Zhao et al. (2020) developed a hybrid model to measure land-use characteristics that combines GIS feature extraction techniques and Deep Neural Network (DNN), and the proposed model achieved promising and resilient results. Another example can be Wu and Biljecki (2021)' s research on sustainable roofscape detection in cities. They integrated deep learning with geospatial techniques to develop a model that can successfully map the spatial distribution and area of solar and green roofs and automatically detect sustainable roofs of over one million buildings across 17 cities.

The recent studies mentioned above highlighted the promising potentials of the effective and precise scene classification of deep learning in contributing to the creation of new types of geographical variables and the decision- making process. But the integration of deep learning and GIS is still rare in existing literature. Such current deficiency can be explained by the fact that the emergence of deep learning in classifying large datasets is rather recent. Along with the advancement of computing power, the deep convolutional neural networks trained to efficiently classify large volumes of high-resolution images (1.2 million) was first published in 2012 by Krizhevsky et al. (2012). Since then, this groundbreaking work has opened doors for numerous research domains in the applications of deep learning (Ayyoubzadeh et al., 2020; Cao et al., 2018; Hamidinekoo et al., 2018). Mohan and Giridhar's (2022) review on the integration of deep learning and GIS pointed out that the existing GIS software is tailored for specific types of decision making, and they may perform poorly or fail for processing large and complicated datasets with many features. The maturing and promoting of deep learning techniques per se in handling large remotely sensed data naturally arose prior to further integrating the large and complicated geographic variables processed by deep learning inside a GIS decision-making system. The process of exploring deep learning techniques in the applications of GIS utilizing remote sensing data is bound to take time, and it is currently at the growing stage in the remote sensing and GIS communities.

As mentioned earlier, identifying spatial disparities of built environments at the local scale, such as the neighborhood level, is a necessity in health studies, while research on this topic supported by the geographical attributes extracted on large and fine spatial resolution images in a GIS environment is still not explored yet. Based on the review paper by Mohan and Giridhar (2022) analyzing 55 papers related to GIS and deep learning, only 26 papers published recently (during 2016 and 2021) explored integrating deep learning into various fields of GIS, and none of them applied such integration for understanding spatial disparities of the built environment at a fine spatial scale. This study took up this opportunity to marry deep learning in image processing and GIS for studying built environments at the neighborhood scale This method paper aims to provide detailed guidelines on image collection from an unmanned aerial vehicle (UAV), greenspace extraction using a deep learning model, and how to visualize the spatial patterns of sidewalks and greenspace, as proxy for build environments, in a GIS environment. In the meantime, two research objectives were achieved: a) to test the potentials of deep learning in mapping built environments from fine spatial scale UAV images; and b) to identify spatial disparities of built environments at the neighborhood scale.

Fig. 1 Workflow chart
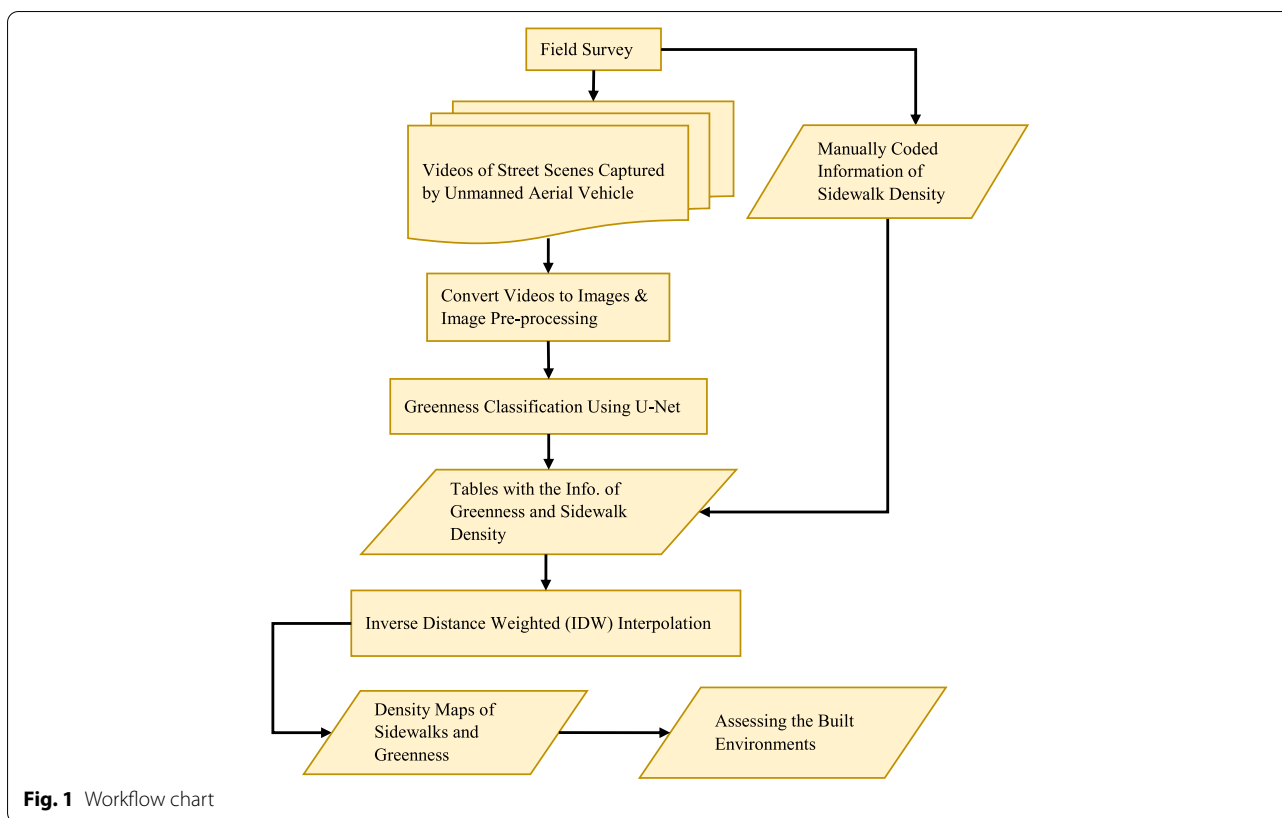
2 Methods

The practice of the study started with data collection using a UAV, followed by preprocessing the datasets and applying the pre-trained deep learning architecture U-Net for greenness detection. The next step was using the greenness information resulting from U-Net and the in-situ sidewalk density information to map and assess the sidewalk environments using GIS. The workflow chart is presented in Fig. 1.

2.1 Study area and data collected from UAV

The *sidewalk-homogenous neighborhood*, as an operational spatial unit, was proposed for this study. It is bounded by the level of estate values and intersections of Local (residential) and Collector (tertiary) roadways.[1] That is—two types of neighborhoods are delineated in this research: one has higher monetary estate value, and another one has lower estate value, and both are bounded by the intersections between Local and Collector roadways. This concept respects the economic component of a neighborhood through using the estate values. It also respects the social-behavioral component through

utilizing the intersection between Local Roads and Collectors. Because automobile traffic adjacent to sidewalks can also greatly influence residents' usage of sidewalks, and Local Roads have substantially slower traffic, thus have more favorable conditions to pedestrians, than Collectors. Residents are more likely to limit their sidewalk usage zone within their local roadways which have less traffic volume rather than crossing a busy state route (Tefft, 2011).

Four neighborhoods in Northeast Ohio based on the concept of *sidewalk-homogenous neighborhood* were selected as the study sites. Data of the boundaries between the Local and Collector roadways was acquired from Ohio Department of Transportation (2020). Electronic searches for house price were performed on multiple real estate brokerage websites, including Redfin, Trulia, and Zillow, with final search completed on June 18, 2018. The neighborhoods with corresponding estate values in the year of 2018 are listed as follows: River Bend ($\sim$ \$300 K and $\sim$\$ 900 K), Summit Mall ($\sim$\$220 K to $\sim$\$500 K), Sandy Lake ($\sim$\$170 K to $\sim$\$300 K), and North Cherry ($\sim$ 70 K to $\sim$ 150 K). Their corresponding areas are 0.61 km^2 (River Bend), 0.8 km^2 (North Cherry), 4.48 km^2 (Sandy Lake), and 1.13 km^2 (Summit Mall). The neighborhoods' limits and their locations in Northeast Ohio are shown in Fig. 2.

---

[1] For information about definitions of Local Roads and Collectors, refer to Ohio Department of Transportation (2020).

Hong *et al. Computational Urban Science* (2022) 2:12
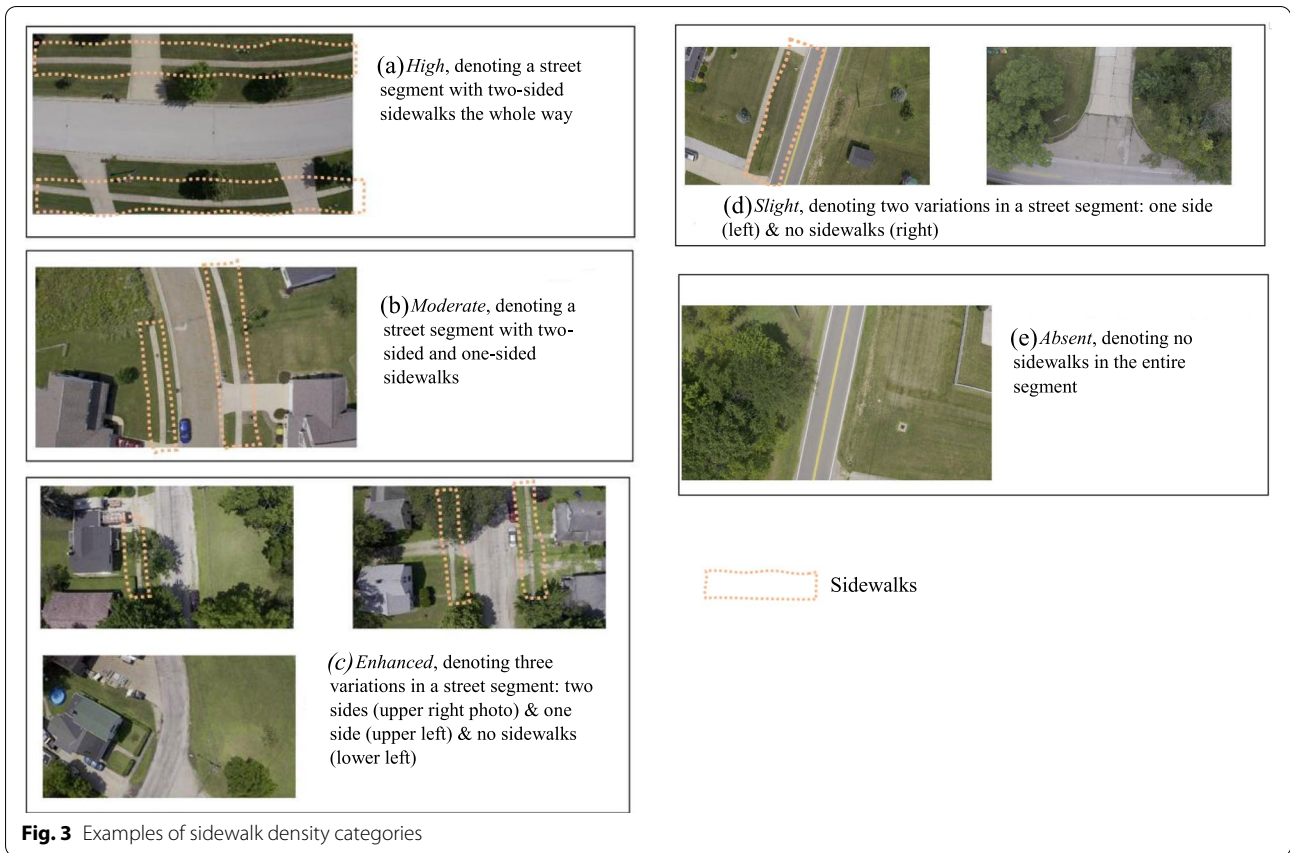
Page 4 of 15

Fig. 2 Study area
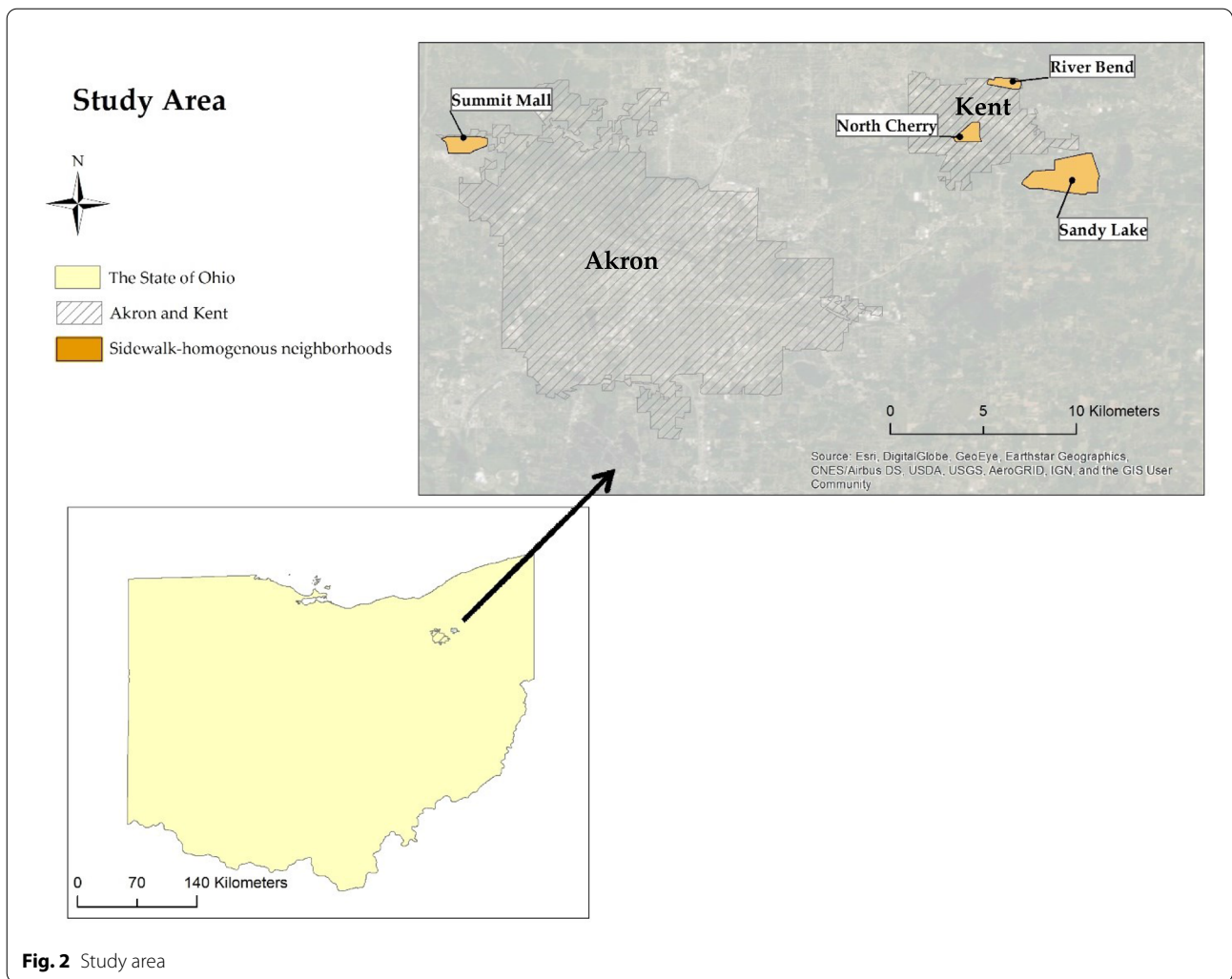
The model DJI® Mavic 2 Pro with an ultra-high spatial resolution sensor was deployed for data collection. Flight missions were carried out in summer from July to August 2019, which is the season for vegetation growth with most of the sunny and clear sky days, and the acquisition time was between 10 a.m. and 2 p.m. *Waypoints* navigation mode was set to autonomously fly above every street within the study sites with speed of 10.1 kph and altitude at 40 m above ground level (AGL). During flight missions, the gimbal was tilted to the angle of − 90 degrees (nadir facing), and therefore scenes were captured directly straight down. The scenes were recorded as videos with ultra-high definition (4 k) 3840×2160 resolution. After field survey, the videos were converted to image frames at a frame rate of 1/20 fps (one frame every 20 s) The frame rate of 1/20 fps equals 56- m ground distance between images. The corresponding GPS files of each image frame was also reformatted to. CSV files. In total, there are 570 images, and the spatial resolution is 0.58 cm.

While videotaping the street scenes, sidewalk density of each street segment was also simultaneously coded into five categories: a) *high*, denoting two-sided sidewalks the whole way; (b) *moderate*, denoting a street segment with two-sided and one-sided sidewalks; (c) *enhanced*, denoting three variations in a street segment: two sides, one side, and no sidewalks; (d) *slight*, denoting a segment with one side and no sidewalks; and (e) *absent*, denoting no sidewalks in the entire segment. Figure 3 illustrates examples of sidewalk categories.

2.2 Greenspace pixel segmentation using deep learning
U-net, published by Ronneberger et al. (2015), was initially developed for segmenting biomedical images in medical diagnostics. It was built upon Long et al. (2015) 's Fully Convolutional Networks (FCNs) and extended to work with a small volume of training images and produces more precise segmentations. U-net takes the name of its architectural shape. It appears similar to the

Hong *et al. Computational Urban Science* (2022) 2:12

Page 5 of 15

Fig. 3 Examples of sidewalk density categories

letter *U*, as shown in Fig. 4. Its architecture consists of an encoder network and a decoder network. The encoder is the first half in the architecture (on the left). It applies convolutional layers and followed by a maxpool downsampling layer to encode the input image into feature representations at multiple levels. The network does not have a fully-connected layer in the end as most Convolutional Neural Networks (CNNs) do in image classification. Instead, it continues with a decoder as the second half, which aims to semantically project the feature representations learnt by the encoder back onto the pixel space. Intuitively, the U-net model takes an image of any size in and produces a segmentation map out (Kizrak, 2019).[2]

For this study, U-net was adopted to segment the images and extract greenspace pixels. One of the main advantages of U-net is the application of data augmentation for increasing the size of the training dataset, as the training sample are limited in this study. It performs image deformations (specifically, random elastic deformations) to teach the network to simulate realistic deformations based on the input samples and generate more training samples. For image data, data augmentation is a technique to expand the size of a sample dataset by creating modified versions of existing sample images (Brownlee, 2019). The elastic deformation is performed by generating a random displacement field μ (x, y), and that each location (x, y) specifies a unit displacement vector. The displacement vector is defined as $Rw = Ro + \alpha\mu$, where Rw and Ro describe the data space locations of the pixels in the original and deformed images, respectively. The α gives the strength of the displacement to pixels. *The parameter* σ controlled the smoothness of the displacement, which is the standard deviation of the Gaussian distribution of the random values in the displacement field μ (x, y) (Wong et al., 2016). For the data augmentation process in U-net, the random displacement field is μ *(3,3)*, and the displacements are sampled from a Gaussian distribution with standard deviation $\sigma = 10$ pixels.
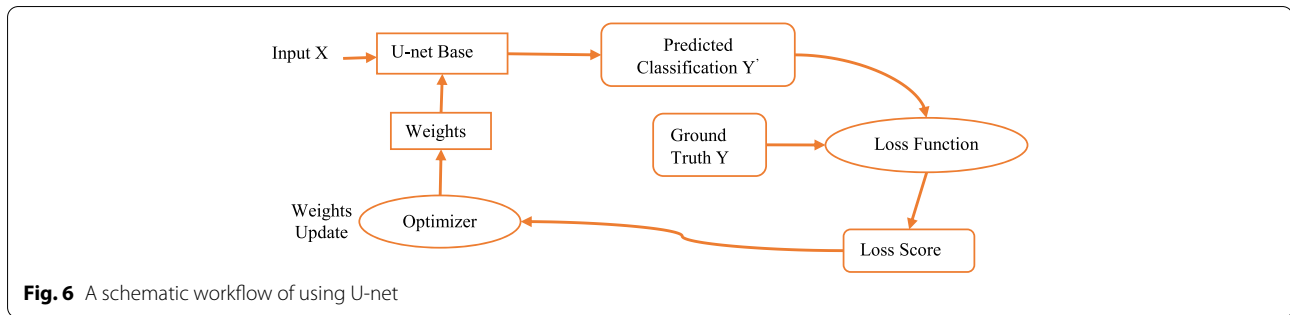
Another unique character in U-net is preserving the high-level abstraction during the contracting and expansive paths. A traditional encoder-decoder CNN reduces the size of the input information in a linear behavior on the encoder half, which results in only low dimensional

---

[2] More details of the U-net architecture can be referred to Ronneberger et al. (2015).

Fig. 4 U-net model (Ronneberger et al., 2015)

Fig. 5 Illustration of a traditional encoder-decoder or autoencoder model (Weng, 2018)
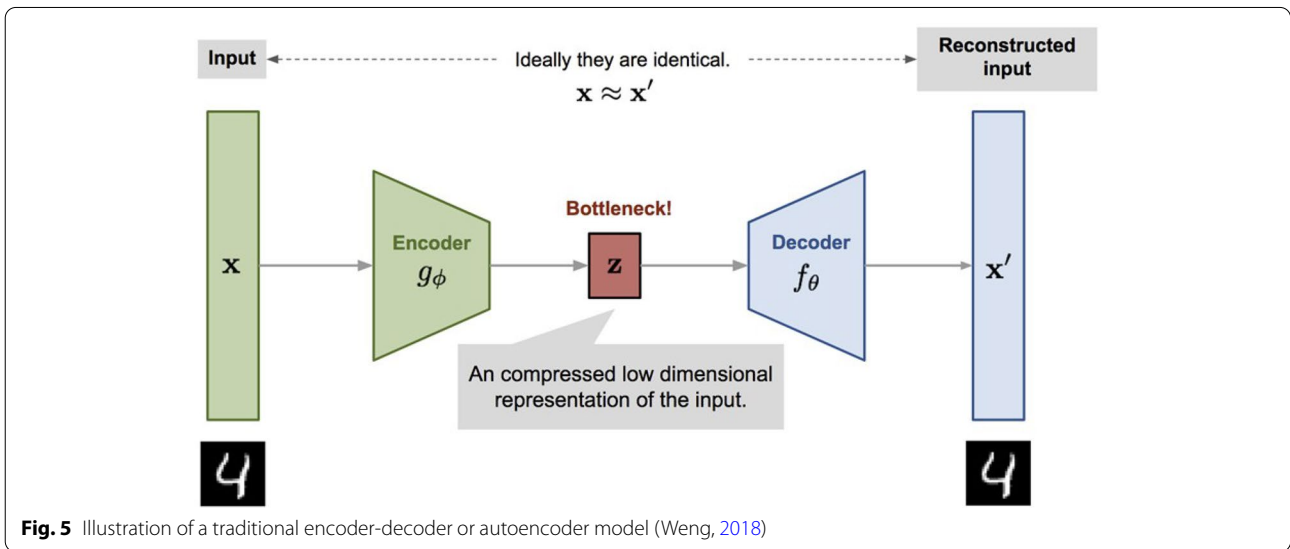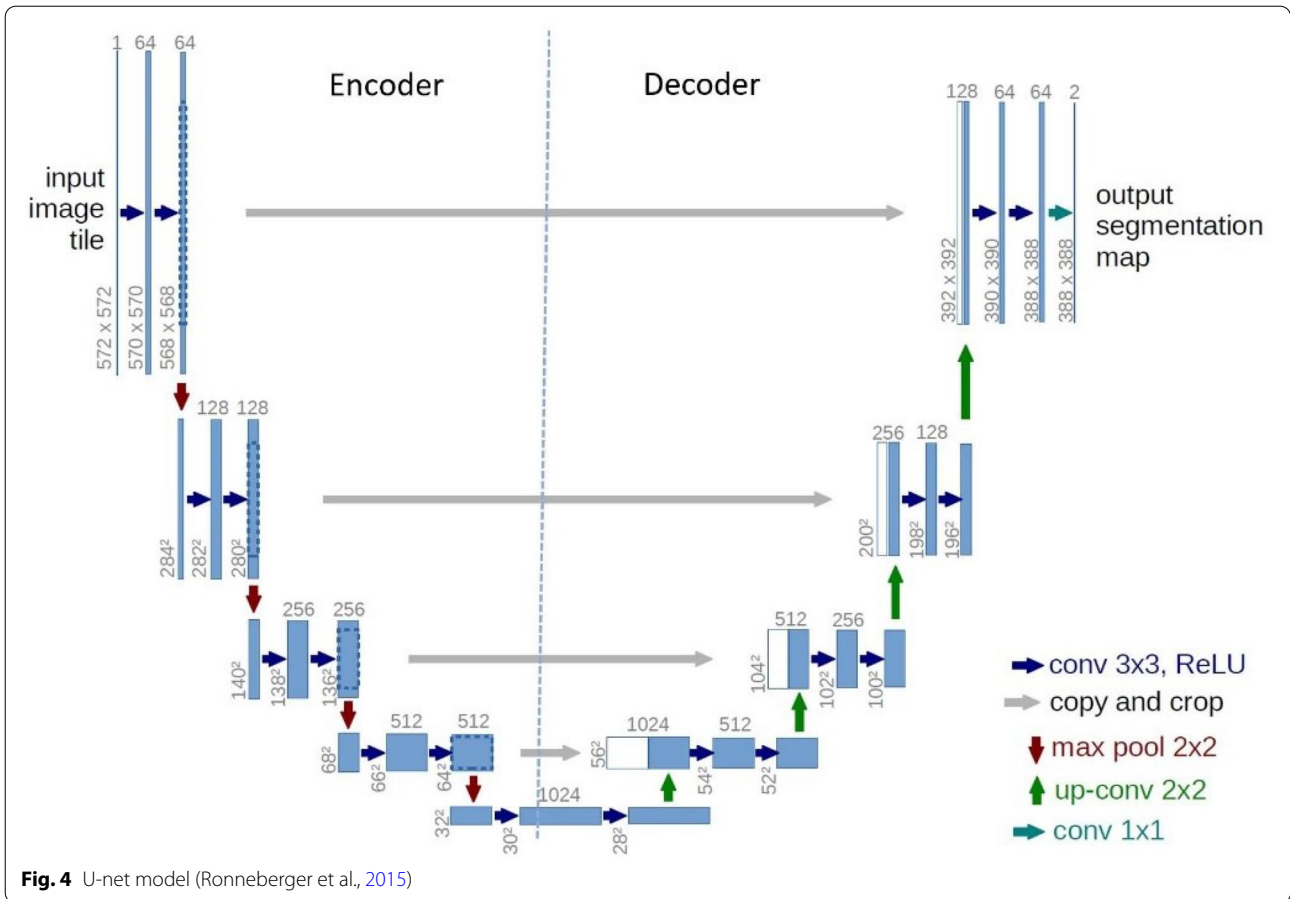
representation of the input being transmitted to the decoder side. This bottleneck issue in the transmission from encoder to decoder significantly reduces the classification accuracy for the output images (Fig. 5). This is where U-net is unique. U-net operates upsampling on the decoder side (i.e. in the second half) and can overcome this bottleneck issue by implementing skip connections to copy the uncompressed activations from

Hong *et al. Computational Urban Science*   (2022) 2:12

Page 7 of 15

Fig. 6 A schematic workflow of using U-net

downsampling blocks to the corresponding upsampling blocks by concatenation (Fig. 5), and such feature extractor, which preserves high dimensional representation, provides more accurate segmentation maps (Ronneberger et al., 2015).

The backbone of the U-net model was used to segment greenness pixels from the images. Figure 6 demonstrates the schematic workflow of the training U-net. As the first step, the training dataset was fed to the pre-trained U-net base and passed through the whole network. Followed by the forward pass, the predicted greenness pixels in the images were assessed by comparing with the ground truth data using the loss function, and a loss score was generated. Afterward, a backward pass was performed to determine how weights change can minimize the loss score by using an optimizer. Once the ideal weights were computed, the weights in the network were updated. The process that the dataset passes forward and backward through the entire network once counts as one epoch, and the model was trained for 40 epochs using a combination of BCE-Dice loss function and Adam optimizer[3].

The annotated images contain two categories: greenspace and non-greenspace pixels. They were split into three groups: training (58 images), validation (6 images), and test sets (5 images). It may appear that the number of images in the test set is relatively small, but the large size (3840×2160 pixels) of the image and the simplicity of the classification task (only needs to classify greenness and non-greenness pixels) suggest that a small test set is sufficient (Neurohive, 2018). The test set was used as the ground truth data to compare with the prediction results. Although the street scenes of all the four neighborhoods do not vary much in terms of green cover, the test images were carefully curated to include all possible types of greenness. One image each was selected for River Bend, Sandy Lake, and North Cherry, and two images for Summit Mall. During the testing, the standard binary

classification metrics were computed, and their formulas are as follows (Tharwat, 2020):

$$Overall\ accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$User\ accuracy\ (precision) = \frac{TP}{TP + FP} \quad (2)$$

$$Producer\ accuracy\ (recall) = \frac{TP}{TP + FN} \quad (3)$$

$$F1\ score = \frac{2PR}{P + R} \quad (4)$$

where:

- TP (true positives) is the number of correctly identified greenness pixels.
- TN (true negatives) is the number of correctly identified non-greenness pixels.
- FP (false positives) is the number of non-greenness pixels in the ground reference but is mislabeled as greenness pixels.
- FN (false negatives) is the number of missed greenness pixels.
- F1 score is the harmonic mean of precision and recall.
- P is the precision.
- R is the recall.

The abovementioned TP, TN, FP, FN, and F1 Score are the conventional measures in assessing image segmentation performance.

2.3 Mapping greenness and sidewalk densities

The greenness ratio of each street, as the derivatives from the greenness segmentation outputs, and the manually coded sidewalk density information recorded during field survey were further analyzed in a GIS environment. Estimated greenness and sidewalk densities at street level

---

[3] Documentation and tutorials about using U-net for image segmentation is available in GitHub (https://github.com/qubvel/segmentation_models).

Hong *et al. Computational Urban Science*   (2022) 2:12

Page 8 of 15

Table 1 Accuracy of the trained model on the test set (%)

Trained model	Precision	Recall	Overall Accuracy	F1 Score
	96	98	97	97

were produced using Inverse Distance Weighted (IDW) interpolation. In IDW, In the process of IDW interpolation, the unmeasured value is predicted as the weighted distance average from nearby locations by the following function:

$$z_0 = \frac{\sum_{i=1}^{s} z_i \frac{1}{d_i^k}}{\sum_{i=1}^{s} \frac{1}{d_i^k}} \qquad (5)$$

Where z_0 is the predicted value at location 0, z_i is the value of known point i, d_i is the distance between known point i and predicted point 0, s is the number of know points used in the prediction, ad k is the user selected power (Chang, 2018, pp. 334–5). In addition, correlation coefficients between the greenness and sidewalk densities were computed to understand their statistical relationship.

3 Results

3.1 Evaluation metrics of greenspace segmentation

The trained U-net model achieved Precision of 96%, Recall of 98%, Overall Accuracy of 97%, and F1 Score of 97% on the test set (Table 1). The Precision value suggests that among the total predicted greenness pixels (total predicted positive), 96% of them were actual greenness pixels, but 4% of them were non-greenness. The Recall value indicates that 98% of the actual greenness pixels (total actual positive) were predicted accurately while 2% of the actual greenness pixels were predicted as non-greenness. The Overall Accuracy, as a ratio of correctly predicted values to the total observations, indicates that 97% of the total pixels were accurately predicted. However, we may see from the formulas (in Section 2.2) that Overall Accuracy only focuses on the True Negatives but might neglect the False Negatives and False Positives. Therefore, F1 Score, as the weighted average of Precision and Recall, complements the oversights of Overall Accuracy. As suggested by the F1 Score, which the weighted harmonic mean between Precision and Recall is as high as Overall Accuracy, the classification performance is high (Tharwat, 2020). Overall, the evaluation values suggest that the model achieved high classification accuracy. For the purpose of visually identifying the accuracy of the classification performance, the graphs of the classified
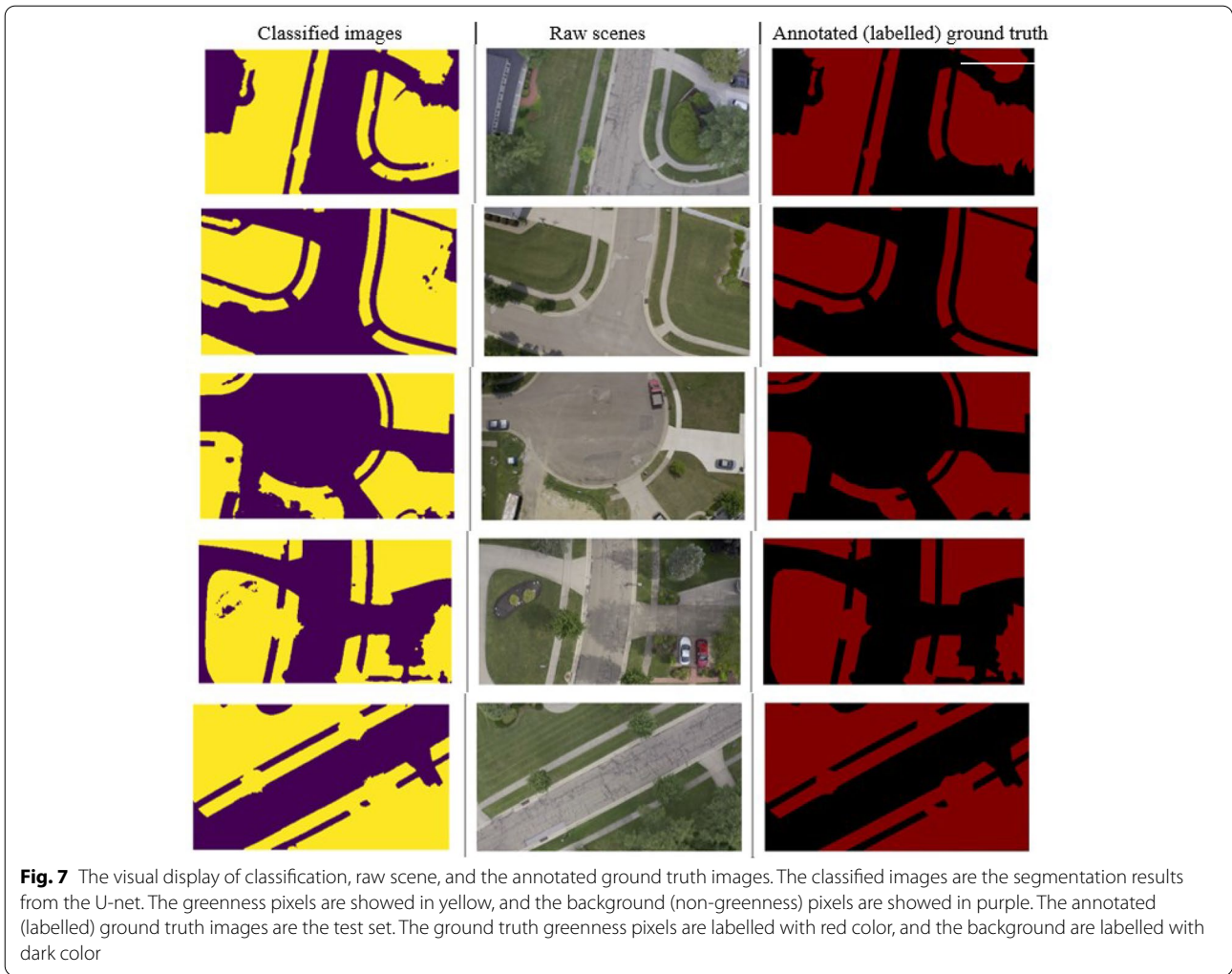
images generated by the trained U-net, the raw drone photos, and the annotated ground truth images (the test set) are displayed in Fig. 7. We can see that most of the greenness pixels were segmented accurately in the classification images as compared with the annotated ground truth images and the raw scenes.

3.2 The spatial patterns of built environments

Sidewalk built environments, represented by greenness and sidewalk densities, were assessed at street level of neighborhoods with different economic levels. The approximate house price, the proportion of sidewalk density categories (*high, moderate, enhanced, slight, and absen*t), and the corresponding greenness in sidewalk density categories are given in Table 2. In general, neighborhoods in higher economic levels are facilitated with a greater presence of sidewalks and greenness. For example, River Bend as the richest neighborhood, has sidewalks in all street segments, and 73% of the segments are accompanied with two-sided sidewalks. About 59% of the two-sided sidewalks are facilitated with greenspace. Similarly, second to River Bend in terms of economic level, Summit Mall also has nearly 72% of the streets in high sidewalk density, and about 63% of such sidewalks are surrounded by greenness. However, as a lower economic neighborhood, North Cherry is lower in sidewalk presence. Most of the North Cherry streets are identified as Enhanced, Slight and Absent in sidewalk density, although they are sufficiently surrounded by green. The wide disparities in sidewalk density between neighborhoods is also presented in Fig. 8, where we can see that most streets in River Bend and Summit Mall are in higher sidewalk density, while North Cherry and Sandy Lake are in much lower sidewalk density.

We may have the general conception that sidewalks are more likely to be facilitated with greenspace. The spatial combination of sidewalks and greenspace can be seen in the economically favored River Bend and Summit Mall, while not in the economically disadvantaged Sandy Lake and North Cherry (Fig. 8). As we can see, the east of North Cherry is mostly facilitated with High density sidewalks (two-sided sidewalks) but with low greenness (Fig. 9 and Fig. 10). The same pattern can be found on the west part of Sandy Lake, where it is facilitated with no sidewalks but dense greenness. Furthermore, the statistical correlation between sidewalk and greenness in each neighborhood was produced to explore the spatial arrangements of built environments. Except Summit Mall, in where has negligible correlation, all the other three neighborhoods have either low or moderate inverse correlation between sidewalk and greenness densities (Table 3).

Fig. 7 The visual display of classification, raw scene, and the annotated ground truth images. The classified images are the segmentation results from the U-net. The greenness pixels are showed in yellow, and the background (non-greenness) pixels are showed in purple. The annotated (labelled) ground truth images are the test set. The ground truth greenness pixels are labelled with red color, and the background are labelled with dark color

Table 2 Presence of sidewalks and average greenness in sidewalk categories (%)

Neighborhood	Approximate house price (thousand US $)	High	Greenness in high	Moderate	Greenness in moderate	Enhanced	Greenness in enhanced
North Cherry	70–150	26.17	46.69	0.00	0.00	60.40	58.23
Sandy Lake	170–300	31.38	49.55	17.02	53.99	13.83	49.34
Summit Mall	220–500	71.70	62.65	7.55	55.51	15.09	56.89
River Bend	300–900	72.97	59.10	14.86	56.33	12.16	72.56
Neighborhood	**Slight**	**Greenness in slight**	**Absent**	**Greenness in absent**	**Total greenness**		
North Cherry	11.41	66.13	2.01	79.36	56.54		
Sandy Lake	5.85	72.11	31.91	67.08	57.19		
Summit Mall	0.00	0.00	5.66	61.61	61.18		
River Bend	0.00	0.00	0.00	0.00	60.33		

(a) high, denoting two-sided sidewalks the whole way; (b) denoting a street segment with two-sided and one-sided sidewalks; c) enhanced, denoting three variations in a street segment: two sides, one side, and no sidewalks; (d) slight, denoting a segment with one side and no sidewalks; and (e) absent, denoting no sidewalks in the entire segment

Hong *et al. Computational Urban Science* (2022) 2:12

Page 10 of 15

Presence of Sidewalks in Neighborhoods

Fig. 8 Percentage of sidewalks. Description of the sidewalk classes: **a** high, denoting two-sided sidewalks the whole way; **b** denoting a street segment with two-sided and one-sided sidewalks; **c** enhanced, denoting three variations in a street segment: two sides, one side, and no sidewalks; **d** slight, denoting a segment with one side and no sidewalks; and **e** absent, denoting no sidewalks in the entire segment
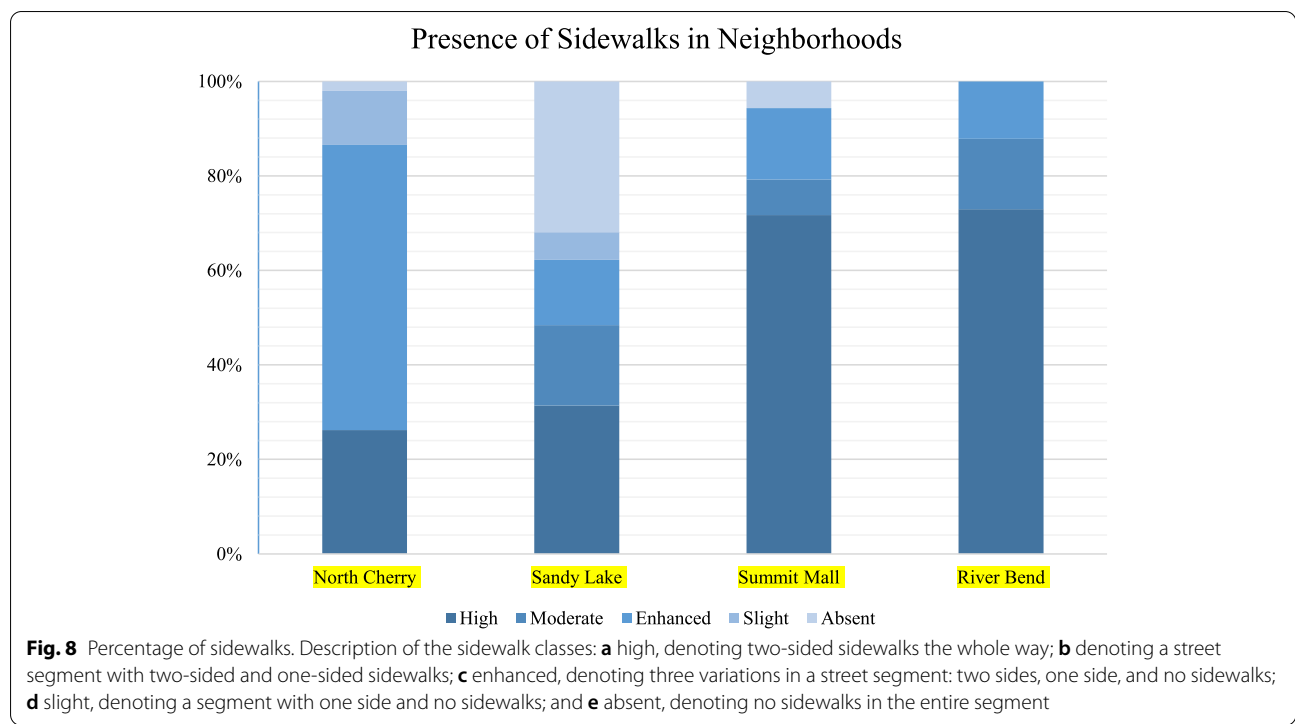
4 Discussion

By recognizing that accurately describing and mapping health-related built environments at a fine spatial scale is critical in understanding the relations between place and health (Duncan et al., 2018; Gullón & Lovasi, 2018), this study makes several key contributions. The contributions of this study lie in its combined use of deep learning and first-hand UAV images in a GIS environment to facilitate detailed spatial explorations of health-related built environments at neighborhood scale.

The experimental results demonstrated the capability of a U-net based model, trained with a completely different dataset, in classifying greenspace on high volume drone images. Along with other proven U-net applications in remote sensing, this study may highlight a new capability of U-net on classifying greenspace on UAV images. U-net was originally designed by Ronneberger et al. (2015) for classifying biomedical images, and its applications have been extended to solve some challenging segmentation tasks on datasets (e.g., Lidar point cloud data and very fine spatial resolution satellite images) that is beyond the capacity of traditional remote sensing algorithms (Belloni et al., 2020; Wagner et al., 2019). Recently, Jiang et al. (2020) experimented crack segmentation on drone-captured airport runway pavement images using U-net, and the trained U-net model achieved high performance with limited train images. Another recent publication of using U-net to segment high-resolution drone images is

Ahmed et al. (2020)'s building extraction project from noisy labeled images. They compared the performance of U-net model with the widely used Geographic Information System (GIS) software, e.g., ArcGIS, with the same noisy labeled images for building segmentation, and their empirical results revealed that U-net produced more accurate and less noisy classification results (Ahmed et al., 2020). The high accuracy of U-net on classifying green cover on UAV images in this study aligns with existing literature on the efficiency of the U-net model. Although this study addresses the transferability and efficiency of deep neural networks, U-net in particular, it is worth noticing that only two classes (greenness and non-greenness pixels) were detected, and the features and inner spatial arrangements in the images were relatively homogenous across the study sites. This study did not confirm the reliability of the U-net model or other CNN models if the classification tasks were complex, such as classifying scenes into more categories or when the texture of greenness is complicated. If further studies were to be performed for testing the reliability of CNNs in classifying complex scenes, Wagner et al. (2019)'s study of using U-net to segment multiple forest types with very high-resolution images could be a reference.

The presence of sidewalks was manually coded with field survey data rather than segmented with deep learning methods. Manual coding may be feasible for a small study area; however, the process can be expensive and

Hong *et al. Computational Urban Science*   (2022) 2:12

Page 11 of 15
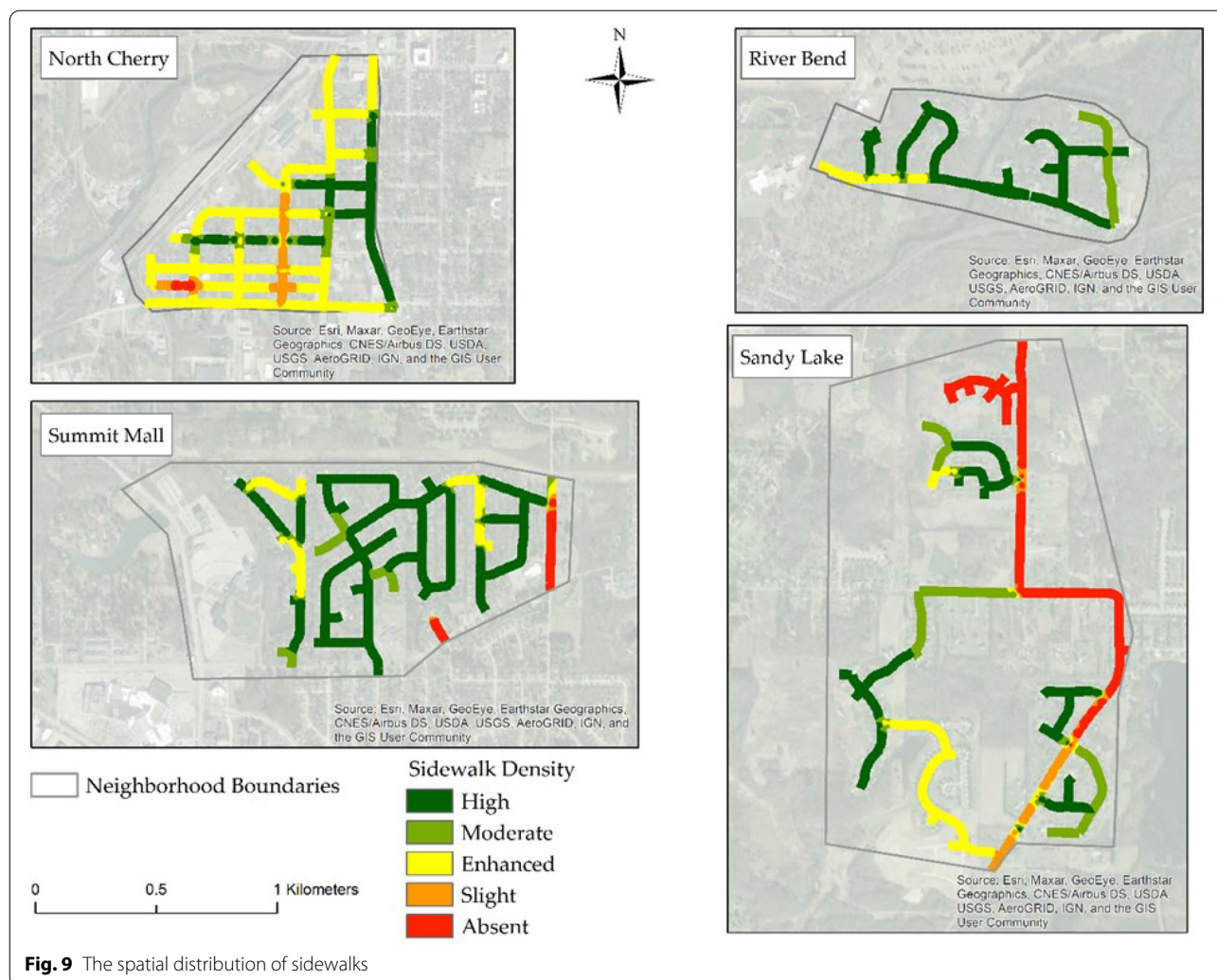
Fig. 9 The spatial distribution of sidewalks

labor intensive for large areas regardless of its high accuracy. There were two main reasons for not being able to use deep learning in sidewalk classification in this study: (a) the similarity of road and sidewalk pavements, and (b) canopy occlusion in drone scenes. As most sidewalks are adjacent to roads and have similar pavement texture as roads, it would be challenging to distinguish them using U-net. Secondly, drone images captured from the bird's view perspective may fail to show sidewalk segments in shadows or occluded by canopies or buildings. Before this research was finished, deep learning for image classification had already been more advanced. Solutions to these two issues are presented in a recently published article authored by Ning et al. (2021). They used neural networks model YOLACT (Bolya et al., 2019), which can recognize the boundaries of sidewalks more accurately, for extracting sidewalks from aerial images. They also used Google Street View images as a supplementary data source to fill the gaps from canopy occlusion. For a study

area that is too big and manual coding is impossible, using YOLACT and Google Street View images as Ning et al. (2021) or similar ways to automatically classify sidewalks can be the next steps of improvement.

This empirical study identified the disparities of built environments between neighborhoods and within the low-income neighborhood at the street level using the combination of deep learning and UAV images. It addressed the environmental injustice issues to some extent, but the methodology and the interpretations of results were limited by the delineations of *built environments* and *neighborhoods* in this study. Built environments were represented by the presence of sidewalks and the density of green cover of streets in the study. As defined by Centers for Disease Control and Prevention (2016), built environment "includes all of the physical parts of where we live and work (e.g., homes, buildings, streets, open spaces, and infrastructure)" (p. 1). This study may not capture the full perspectives of built

Hong *et al. Computational Urban Science*   (2022) 2:12

Page 12 of 15
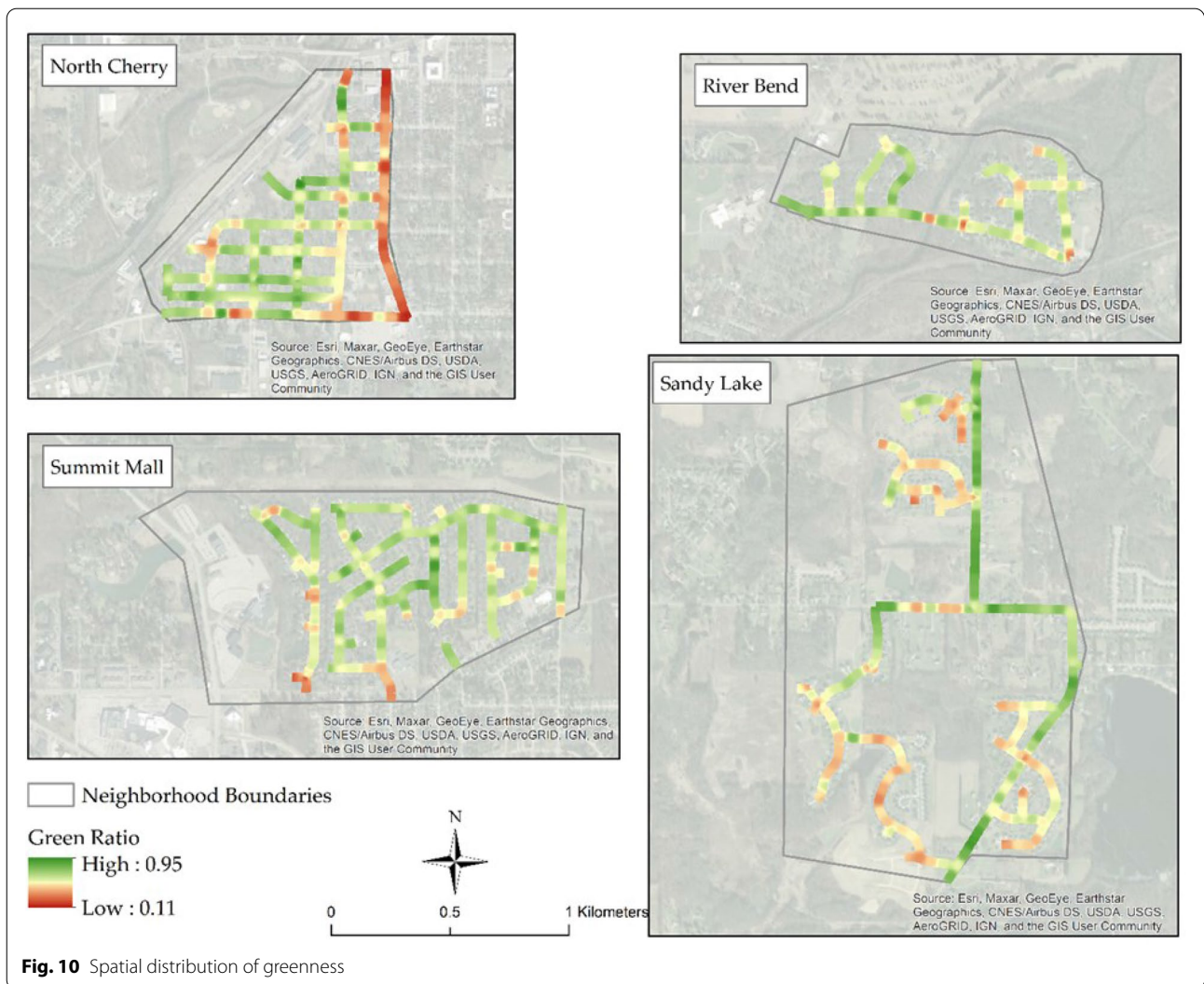
Fig. 10 Spatial distribution of greenness

Table 3 Correlation between sidewalk and greenness density

Neighborhood	Correlation coefficient, r	Interpretation
North Cherry	−0.40	Low inverse correlation
Sandy Lake	−0.54	Moderate inverse correlation
Summit Mall	0.09	Negligible correlation
River Bend	−0.34	Low inverse correlation

environment as sidewalks and green cover are merely two (among many other) physical parts of our living environments. Other physical parts, such as sidewalk conditions and connectivity may also be the important components of built environment (Carr et al., 2010). Besides, the amenities of sidewalk and green cover in a neighborhood may not be the determining factors for a higher level of physical activity. People's physical activity is found to be promoted when they live in close proximity to a park and the safety is ensured (Cohen et al., 2006). In addition, only four neighborhoods were selected for the analysis, the results regarding to the built environment disparities found across the neighborhoods may not be representative or statistically significant. Expanding the sample size could be included in future studies.

5 Conclusion
In this method paper, we briefly introduced the significance of greenspace and sidewalks in health and the inequity of built environment between neighborhoods. The main part lied on the informative guidelines of: a) the design of study sites and data collection using UAV, b) greenspace extraction using a deep learning model U-net,

Hong *et al. Computational Urban Science*   (2022) 2:12

Page 13 of 15

and c) mapping sidewalks and greenness densities using GIS methods. There were three major findings from this study. Firstly, the deep learning model U-net was proven efficient in classifying greenspace from high volume and fine spatial scale remote sensing images. Secondly, economically favored neighborhoods appear to be facilitated with better sidewalk environments than disadvantaged neighborhoods. The third main finding indicated that there is no strong correlation between sidewalk and greenness densities across the selected neighborhoods.

The key contributions of study are the combined use of deep learning on classifying UAV images and GIS to facilitate detailed spatial explorations of health-related built environment across neighborhoods of different economic levels. By the time of the study, there was no dataset or any pre-trained model available for us to train the model. All the data/images used in this work must be handcrafted. Thus, another significance of this work is to help generate the data automatedly for similar studies, and this work can also serve as a baseline for future studies.

This study can be also extended in several aspects. The first-hand drone images collected from this study can be expanded with street view images to extract sidewalk information using deep learning, such as sidewalk conditions and junctions with crosswalks. This may be helpful to characterize built environments and identify the unequal distributions of resources with more representations and greater spatial details across neighborhoods for identifying what and where the inequalities are, and thus assist urban planning practices to be better in ensuring access and physical activity opportunities for all residents. In addition, with the integration of drone data of this study with street view images, the state-of-the-art performance of CNNs could be further expanded to classify more features in built environment and detect street activities. As long as there are sufficient training datasets (annotated images) to exemplify the variety of streetscape, it is possible to train CNN models to detect desired features, such as public facilities (waste bins and light poles) and pedestrian or traffic volumes.

Since *sidewalk-homogenous neighborhoods* is a newly proposed operational neighborhood unit and was employed in delineating a limited number of neighborhoods in this study, the feasibility and applicability of this measurement of neighborhood may be tested in more study sites and/or regions that have very dissimilar landscapes or physical environments than the ones in this study. For example, the desert regions in Southwest United States and water canals of Venice, Italy where the facility of sidewalks and greenspace may require a different perspective with different variables (e.g., temperature effects or water levels; Brown et al., 2013; Gorrini &

Bertini, 2018). Only with the integration of multidisciplinary knowledge and local situations can we accurately and comprehensively explore the dynamics and mechanisms between neighborhood built environment and health.

Author details
[1]Research Center for Anatolian Civilizations, Koç University, Istiklal Cd. No:181, Merkez Han 34433 Beyoğlu, Istanbul, Turkey. [2]The Department of Geography, Kent State University, Kent, OH 44242, USA. [3]The Department of Computer Science, Kent State University, Kent, OH 44242, USA.

References
Ahmed, N., Mahbub, R. B., & Rahman, R. M. (2020). Learning to extract buildings from ultra-high-resolution drone images and noisy labels. *International Journal of Remote Sensing, 41*(21), 8216–8237. https://doi.org/10.1080/01431161.2020.1763496

Ayyoubzadeh, S. M., Ayyoubzadeh, S. M., Zahedi, H., Ahmadi, M., & Kalhori, S. R. N. (2020). Predicting COVID-19 incidence through analysis of Google trends data in Iran: Data mining and deep learning pilot study. *JMIR Public Health and Surveillance, 6*(2), e18828. https://doi.org/10.2196/18828

Belloni, V., Sjölander, A., Ravanelli, R., Crespi, M., & Nascetti, A. (2020). Tack project: Tunnel and bridge automatic crack monitoring using deep learning and photogrammetry. *International archives of the photogrammetry, remote sensing and spatial information sciences, XLIII-B4-2020*, 741–745. https://doi.org/10.5194/isprs-archives-XLIII-B4-2020-741-2020.

Bolya, D., Zhou, C., Xiao, F., & Lee, Y. J. (2019) YOLACT: Real-time instance segmentation. Computer Vison Foundation. *Proceedings of the IEEE/CVF international conference on computer vision (ICCV)*. 9157-9166.

Hong *et al. Computational Urban Science*   (2022) 2:12

Page 14 of 15

Brown, H. E., Comrie, A. C., Drechsler, D. M., Barker, C. M., Basu, R., Brown, T., et al. (2013). Human health. In G. Garfin, A. Jardine, R. Merideth, M. Black, & S. LeRoy (Eds.), *Assessment of climate change in the Southwest United States: A report prepared for the national climate assessment* (pp. 312–339). Island Press https://www.swcarr.arizona.edu/chapter/15

Brownlee, J. (2019). How to Configure Image Data Augmentation in Keras. Machine Learning Mastery. Retrieved on Janurary 2, 2022. https://machinelearningmastery.com/how-to-configure-image-data-augmentation-when-training-deep-learning-neural-networks/

Campos-Sánchez, F. S., Abarca-Álvarez, F.J., Molina-García, J., & Chillón, P. (2020). A GIS-based method for analysing the association between school-built environment and home-school route measures with active commuting to school in urban children and adolescents. *International journal of environmental research and public health 17*(7). Article 2295. https://doi.org/10.3390/ijerph17072295.

Cao, R., Zhu, J., Tu, W., Li, Q., Cao, J., Liu, B., Zhang, Q., & Qiu, G. (2018). Integrating aerial and street view images for urban land use classification. *Remote Sensing, 10*(10), 1553–1586. https://doi.org/10.3390/rs10101553

Carr, L. J., Dunsiger, S. I., & Marcus, B. H. (2010). Walk score™ as a global estimate of neighborhood walkability. *American Journal of Preventive Medicine, 39*(5), 460–463. https://doi.org/10.1016/j.amepre.2010.07.007

Centers for Disease Control and Prevention. (2016). *Impact of the built environment on health.* https://www.cdc.gov/nceh/publications/factsheets/impactofthebuiltenvironmentonhealth.pdf

Chang, K. (2018). Introduction to geographic information systems. McGraw-Hill Education.

Cohen, D. A., Sehgal, A., Williamson, S., Sturm, R., McKenzie, T. L., Lara, R., & Lurie, N. (2006). *Park use and physical activity in a sample of public parks in the City of Los Angeles* (publication no. TR-357-HLTH). RAND Corporation, https://www.rand.org/pubs/technical_reports/TR357.html

Duncan, D. T., Goedel, W. C., & Chunara, R. (2018). Quantitative methods for measuring neighborhood characteristics in neighborhood health research. In D. T. Duncan & I. Kawachi (Eds.), *Neighborhoods and health*. Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/oso/9780190843496.001.0001/oso-9780190843496-chapter-3

Gorrini, A., & Bertini, V. (2018). Walkability assessment and tourism cities: The case of Venice. *International Journal of Tourism Cities, 4*(3), 355–368. https://doi.org/10.1108/IJTC-11-2017-0072

Gullón, P., & Lovasi, G. S. (2018). Designing healthier built environments. In D. T. Duncan & I. Kawachi (Eds.), *Neighborhoods and health*. Oxford University Press.

Gunn, L. D., Lee, Y., Geelhoed, E., Shiell, A., & Giles-Corti, B. (2014). The cost-effectiveness of installing sidewalks to increase levels of transport-walking and health. *Preventive Medicine, 67*, 322–329. https://doi.org/10.1016/j.ypmed.2014.07.041

Hamidinekoo, A., Denton, E., Rampun, A., Honnor, K., & Zwiggelaar, R. (2018). Deep learning in mammography and breast histology, an overview and future trends. *Medical Image Analysis, 47*, 45–67. https://doi.org/10.1016/j.media.2018.03.006

Janssen, I., & Rosu, A. (2012). Measuring sidewalk distances using Google earth. *BMC Medical Research Methodology, 12*, 39–48.

Jiang, L., Xie, Y., & Ren, T. (2020). *A deep neural networks approach for pixel-level runway pavement crack segmentation using drone-captured images.* arXiv. http://arxiv.org/abs/2001.03257

Kim, Y.J., Lee, C., & Kim, J.H. (2018). Sidewalk landscape structure and thermal conditions for child and adult pedestrians. *International Journal of Environmental Research and Public Health, 15*(1), 148. https://doi.org/10.3390/ijerph15010148

Kızrak, A. (2019). Deep Learning for Image Segmentation: U-Net Architecture. Medium. Retrieved on Janurary 2, 2022. https://heartbeat.fritz.ai/deep-learning-for-image-segmentation-u-net-architecture-ff17f6e4c1cf

Kolosna, C., & Spurlock, D. (2019). Uniting geospatial assessment of neighborhood urban tree canopy with plan and ordinance evaluation for environmental justice. *Urban Forestry & Urban Greening, 40*, 215–223. https://doi.org/10.1016/j.ufug.2018.11.010

Konstantinidis, D., Argyriou, V., Stathaki, T., & Grammalidis, N. (2020). A modular CNN-based building detector for remote sensing images. *Computer networks, 168*, article 107034. https://doi.org/10.1016/j.comnet.2019.107034.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84–90. https://doi.org/10.1145/3065386

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 39 (4), 640–651. https://ieeexplore.ieee.org/document/7478072

Lowe, K. (2016). Environmental justice and pedestrianism: Sidewalk continuity, race, and poverty in New Orleans, Louisiana. *Transportation Research Record, 2598*(1), 119–123. https://doi.org/10.3141/2598-14

Lu, Y. (2018). The association of urban greenness and walking behavior: Using google street view and deep learning techniques to estimate residents' exposure to urban greenness. *International Journal of Environmental Research and Public Health, 15*(8), 1576. https://doi.org/10.3390/ijerph15081576

Lu, Y., Sarkar, C., & Xiao, Y. (2018). The effect of street-level greenery on walking behavior: Evidence from Hong Kong. *Social Science & Medicine, 208*, 41–49. https://doi.org/10.1016/j.socscimed.2018.05.022

Mayer, H., Kuppe, S., Holst, J., Imbery, F., & Matzarakis, A. (2009). Human thermal comfort below the canopy of street trees on a typical central European summer day. In *5th Japanese-German meeting on urban climatology* (pp.211–219), Albert-Ludwigs-University of Freiburg, Germany.

Moghadam, S. T., Toniolo, J., Mutani, G., & Lombardi, P. (2018). A GIS-statistical approach for assessing built environment energy use at urban scale. *Sustainable Cities and Society, 37*, 70–84. https://doi.org/10.1016/j.scs.2017.10.002

Mohan, S., & Giridhar, M. V. S. S. (2022). A brief review of recent developments in the integration of deep learning with GIS. *Geomatics and Environmental Engineering, 16*(2), 21–38. https://doi.org/10.7494/geom.2022.16.2.21

Neurohive. (2018). *U-Net: Image segmentation network*. Retrieved June 18, 2018, from https://neurohive.io/en/popular-networks/u-net/

Ning, H., Ye, X., Chen, Z., Liu, T., & Cao, T. (2021). Sidewalk extraction using aerial and street view images. *Environment and Planning B; Urban Analytics and City Science, 49*(1), 7–22. https://doi.org/10.1177/2399808321995817

Ohio Department of Transportation. (2020, December 1). *Ohio roadway functional class.* Retrieved June 15, 2021, from https://www.transportation.ohio.gov/wps/portal/gov/odot/working/funding/resources/ohio-roadway-functional-class

Park, J., Kim, J. H., Lee, D. K., Park, C. Y., & Jeong, S. G. (2017). The influence of small green space type and structure at the street level on urban heat island mitigation. *Urban Forestry & Urban Greening, 21*, 203–212. https://doi.org/10.1016/j.ufug.2016.12.005

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. Wells, & A. Frangi (Eds.), *Medical image computing and computer-assisted intervention – MICCAI 2015* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28

Rundle, A. G., Bader, M. D. M., Richards, C. A., Neckerman, K. M., & Teitler, J. O. (2011). Using Google street view to audit neighborhood environments. *American Journal of Preventive Medicine, 40*(1), 94–100. https://doi.org/10.1016/j.amepre.2010.09.034

Sallis, J. F., Conway, T. L., Cain, K. L., Carlson, J. A., Frank, L. D., Kerr, J., Glanz, K., Chapman, J. E., & Saelens, B. E. (2018). Neighborhood built environment and socioeconomic status in relation to physical activity, sedentary behavior, and weight status of adolescents. *Preventive Medicine, 110*, 47–54. https://doi.org/10.1016/j.ypmed.2018.02.009

Shen, Y. (2022). Race/ethnicity, built environment in neighborhood, and children's mental health in the US. *International Journal of Environmental Health Research, 32*(2), 277–291. https://doi.org/10.1080/09603123.2020.1753663

Zhang, W., Tang, P., & Zhao, L. (2019). Remote sensing image scene classification using CNN-*CapsNet. Remote Sensing, 11*(5), 494. https://doi.org/10.3390/rs11050494

Tefft, B.C. (2011). Impact speed and a pedestrian's risk of severe injury or death [technical report]. AAA Foundation for Traffic Safety. https://aaafoundation.org/impact-speed-pedestrians-risk-severe-injury-death/

Tharwat, A. (2020). Classification assessment methods. *Applied Computing and Informatics, 17*(1), 168–192. https://doi.org/10.1016/j.aci.2018.08.003

Thornton, C. M., Conway, T. L., Cain, K. L., Gavand, K. A., Saelens, B. E., Frank, L. D., Geremia, C. M., Glanz, K., King, A. C., & Sallis, J. F. (2016). Disparities in pedestrian streetscape environments by income and race/ethnicity. *SSM - Population Health, 2*, 206–216. https://doi.org/10.1016/j.ssmph.2016.03.004

Thornton, L. E., Pearce, J. R., & Kavanagh, A. M. (2011). Using Geographic Information Systems (GIS) to assess the role of the built environment in influencing obesity: a glossary. *International Journal of Behavioral Nutrition and Physical Activities, 8*(71), 1–9. https://doi.org/10.1186/1479-5868-8-71

Vargo, J., Stone, B., & Glanz, K. (2012). Google walkability: A new tool for local planning and public health research? *Journal of Physical Activity & Health, 9*(5), 689–697. https://doi.org/10.1123/jpah.9.5.689

Wagner, F. H., Sanchez, A., Tarabalka, Y., Lotte, R. G., Ferreira, M. P., Aidar, M. P. M., Gloor, E., Phillips, O. L., & Aragão, L. E. O. C. (2019). Using the U-net convolutional network to map forest types and disturbance in the Atlantic rainforest with very high resolution images. *Remote Sensing in Ecology and Conservation, 5*(4), 360–375. https://doi.org/10.1002/rse2.111

Weng, L. (2018). From Autoencoder to Beta-VAE. Lil'Log. Retrived August 12, 2018. https://lilianweng.github.io/2018/08/12/from-autoencoder-to-beta-vae.html

Wong, S. C., Gatt, A., Stamatescu, V., & McDonnell, M. D. (2016). Understanding data augmentation for classification: when to warp?. 2016 International Conference on Digital Image Computing: Techniques and Spplications (DICTA) (pp. 1-6). *IEEE.* https://doi.org/10.1109/DICTA.2016.7797091

Wu, A. N., & Biljecki, F. (2021). Roofpedia: Automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability. *Landscape And Urban Planning, 214*, Article 104167. https://doi.org/10.1016/j.landurbplan.2021.104167

Xiong, Y., Zuo, R., & Carranza, E. J. M. (2018). Mapping mineral prospectivity through big data analytics and a deep learning algorithm. *Ore Geology Reviews, 102*, 811–817. https://doi.org/10.1016/j.oregeorev.2018.10.006

Zhang, L., Zhang, L., & Du, B. (2016). 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. https://doi.org/10.1109/MGRS.2016.254079

Zhao J., Fan W., & Zhai X. (2019) Identification of land-use characteristics using bicycle sharing data: A deep learning approach. *Journal of Transport Geography, 82*, 2020, Article 102562. https://doi.org/10.1016/j.jtrangeo.2019.102562

Zhu, X. X., Tuia, D., Mou, L. C., Xia, G. S., Zhang, L. P., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine, 5*(4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307

Publisher's Note