Title
Beyond visual inspection: capturing neighborhood dynamics with historical Google Street View and deep learning-based semantic segmentation

Authors
Kim, Jae Hong
Ki, Donghwan
Osutei, Nene
et al.

Publication Date
2023

DOI
10.1007/s10109-023-00420-1

Beyond visual inspection: Capturing neighborhood dynamics with historical Google Street View and deep learning-based semantic segmentation

Accepted Version

(Note: Published in *Journal of Geographical Systems*
doi: 10.1007/s10109-023-00420-1
This version of the article has been accepted for publication, after peer review but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: https://link.springer.com/article/10.1007/s10109-023-00420-1 Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms)

Jae Hong Kim
Department of Urban Planning and Public Policy
University of California, Irvine
Email: jaehk6@uci.edu

Donghwan Ki
Department of City and Regional Planning
Ohio State University, Columbus, USA
Email: ki.17@osu.edu

Nene Osutei
Department of Urban Planning and Public Policy
University of California, Irvine
Email: nosutei@uci.edu

Sugie Lee
Department of Urban Planning & Engineering
Hanyang University
Email: sugielee@hanyang.ac.kr

John R. Hipp
Department of Criminology, Law and Society and Department of Sociology
University of California, Irvine
Email: hippj@uci.edu

Beyond visual inspection: Capturing neighborhood dynamics with historical Google Street View and deep learning-based semantic segmentation

Abstract: While street view imagery has accumulated over the years, its use to date has been largely limited to cross-sectional studies. This study explores ways to utilize *historical* Google Street View (GSV) images for investigation of neighborhood change. Using data for Santa Ana, California, an experiment is conducted to assess to what extent deep learning-based semantic segmentation, processing historical images much more efficiently than visual inspection, enables one to capture changes in the built environment. More specifically, semantic segmentation results are compared for (1) 248 sites with construction or demolition of buildings and (2) two sets of the same number of randomly selected control cases without such activity. It is found that the deep learning-based semantic segmentation can detect nearly 75% of the construction or demolition sites examined, while screening out over 60% of the control cases. The results suggest that it is particularly effective in detecting changes in the built environment with historical GSV images in areas with more buildings, less pavement, and larger-scale construction (or demolition) projects. False positive outcomes, however, can emerge due to the imperfection of the deep learning model and the misalignment of GSV image points over years, showing some methodological challenges to be addressed in future research.

1. Introduction

Increasing availability of street view images and advanced image analysis techniques have dramatically enhanced our ability to analyze the built environment and associated neighborhood phenomena more effectively. Over the past decade or so, researchers have started to take advantage of these promising tools in various ways. Google Street View (GSV) images, for instance, are increasingly used to precisely measure street-level greenery and other built environment characteristics that are known to have significant implications for residents' physical activities, urban vitality, neighborhood safety, and housing values, to name a few (see, e.g., Lu, 2019; Ito and Biljecki, 2021; Wang and Vermeulen, 2021; Hipp et al., 2022; Kang et al., 2021). Recent years have also seen the rise of large-scale projects in which machine learning techniques are employed for semantic segmentation and thereby more efficient use of street view imagery (Kim et al., 2021).

A relatively unexplored area is the (potential) use of *historical* street view images that could allow researchers to capture changes in the built environment over time. To date, most studies using GSV (or similar sources of street view images) have exploited the imagery in a cross-sectional fashion. The longitudinal use of historical GSV is scarce in the literature, although it can open up new avenues of research, particularly when other sources of data for what had existed or happened in the past are unavailable. The small number of existing use cases have also been limited to small-scale projects in which the historical GSV images of selected sites are manually inspected sometimes for verification of what is found from analysis of other sources of (non-imagery) information (see, e.g., Gallagher et al., 2019; Yin et al., 2019; and Section 2 for more details).

Given its great potential as a tool for detecting (micro-level) changes in the built environment over time, in this study, we examine how historical GSV can be used more efficiently for large-scale urban and environmental research. Specifically, although there are various machine learning techniques for classifying historical GSV images, we utilize one that represents the current state of the art: deep learning-based semantic segmentation. We demonstrate how this technique can capture changes in the built environment, and we devise an experiment to assess its efficacy. Using data for a city with a population of approximately 330,000 (Santa Ana, California), our experiment is designed to compare semantic segmentation results (changes in the segmentation results, more precisely) for (1) 248 construction or demolition sites and (2) two sets of the same number of randomly selected control cases without such activity in the city. In other words, this experiment enables us to assess to what extent a semantic segmentation approach can capture changes in the built environment with historical GSV images.

We anticipate that our results will shed light on the potential opportunities and challenges associated with utilizing historical street view images to capture environmental change. We not only assess the ability of the technique to capture environmental change, but also assess in which circumstances the strategy is more likely to successfully detect changes in the streetscape while minimizing false positive cases. By doing so, we aim to provide practical guidance for those who consider using historical images for longitudinal studies, which can be particularly useful in analyzing and understanding how the built environment has changed over time in areas with limited availability of high-quality spatial data including many cities in Global South (Arellana et al., 2020).

The remainder of this paper first provides a brief review of the literature. This is followed by Section 3 where we describe our study area, the design of our experiment, and the data used for the experiment. Section 4 presents our major findings from the experiment. In section 5, we discuss the implications of these findings and directions for future research.

2. Literature Review – Limited Use of Historical GSV

While it is no longer difficult to find studies actively using GSV imagery in various disciplines, as briefly mentioned above, cross-sectional studies are dominant in the literature. Although historical GSV can make it possible to investigate how a place has changed over time, longitudinal studies are extremely rare (Biljecki and Ito, 2021, p.11). Furthermore, most studies that have employed historical GSV images have focused on a small number of selected sites often for verification of findings derived from other sources of data or for measurement of place characteristics that could not be appropriately captured with conventional data sources. In other words, historical GSV images have been used as a supplementary source of information, and the great potential of this tool has yet to be realized.

Chen et al. (2016), for instance, provided an early example of using historical GSV images in this manner. The authors used historical GSV images not with semantic segmentation (or other image-processing techniques) but with two virtual auditors for their analysis of the influences of place characteristics on adolescent alcohol consumption in Taiwan. Another example was provided by Yin et al. (2019) who explored ways to better analyze the evolving spatial patterns of hotel development. They used historical GSV information to verify what they obtained from TripAdvisor or other data sources used to identify the operation years of individual hotels. Similarly, Gallagher et al. (2019) used historical GSV images in combination with other data sources to precisely identify cases of parcel amalgamation in their study area: Brisbane, Australia.

More recently, a few other studies have used historical GSV images to examine changes in neighborhood environments. These studies have focused on various elements of the neighborhood environment, such as street greenery (Li, 2021), food retailers (Cohen et al., 2020), and marijuana dispensaries (Tyndall, 2021), suggesting that historical GSV can be useful for a variety of research purposes. Some of these studies have noted that historical GSV could be a promising source of information to capture and investigate temporal changes in neighborhood environments (Cohen et al., 2020; Li, 2021), and this unconventional data source would be particularly valuable when researchers do not have reliable data for what happened or existed in the past (Cohen et al., 2020).

It should be stressed, however, that most studies using historical GSV have processed the imagery information through visual/manual inspection.[1] Cohen et al. (2020), for example, investigated changes in food environments in the Bronx, New York through manual views of historical images. A similar method of manual scanning has been employed by other researchers who have attempted to take advantage of historical GSV. Although this approach may allow one

[1] This does not mean that visual/manual inspection is still dominant in cross-sectional studies using GSV. Recent years have seen a growing number of cross-sectional studies using numerous GSV images with machine learning techniques. This mode of (large-scale) GSV use is particularly vibrant in some areas of research (Kim et al., 2021, p.3). For instance, many studies concerning street greenery have gathered and processed GSV images in an automated fashion and analyzed how street greenery can affect walking, cycling, and other forms of human behaviors (see, e.g., Li et al., 2018b; Yang et al., 2019). The public safety implications of the built environment are also increasingly examined in this fashion by criminologists and urban planners, as GSV images enable measuring micro-level built environment characteristics (see, e.g., Hipp et al., 2022; Zhanjun et al., 2022).

to detect subtle changes in neighborhood environments precisely, it would not be feasible or time-efficient to do this for a large study area. Even when a small area is of interest, it can become challenging as the number of images to be inspected can go up exponentially with additional image years and/or location points for spatio-temporal precision.

Two studies in particular motivate the present study. First, a study by Cândido et al. (2018) utilized ground-truth data to assess the validity of historical GSV images. Specifically, they identified traffic-calming features (e.g., curb extensions and speed bumps) using online audits of the Google Street View Time Machine and compared what they found with their ground-truth dataset (i.e., administrative data of traffic-calming measures). They reported that historical GSV images enabled them to achieve acceptable accuracy, suggesting that using historical GSV could offer a novel and useful way of detecting changes in neighborhood environments. This allowed them to validate the historical GSV images for capturing this type of change, but left open the question of whether such a strategy could be used for a larger-scale research site.

A second important study that motivates the present study is that of Li (2021), who conducted a large-scale investigation of changes in neighborhood environments (specifically street greenery in this study) with the use of historical GSV imagery. This study collected a large number of historical GSV images drawn from over 300,000 location points in New York City and analyzed them using deep learning and computer vision algorithms to examine how the spatial distribution of green infrastructure changed from 2008 to 2018. However, although comprehensive, the study did not provide detailed guidance on how to handle the uneven availability of historical GSV and other issues to utilize this promising data source more cautiously and effectively (see, e.g., Fry et al., 2020; Smith et al., 2021). Furthermore, in this study the image segmentation results were aggregated spatially and temporally (into census tracts at two periods: 2008-2013 and 2014-2018), leaving it unclear whether this machine learning-based approach enables one to detect changes in the streetscape at the point or street segment level.

Given that it is quite challenging to measure temporal changes in our cities with spatial precision, it is worth studying how historical street view images can be used more efficiently for a variety of research purposes. Researchers, in particular, can benefit from a method that would allow them to go beyond visual/manual inspection and identify changes in our neighborhoods in a systematic fashion for large-N projects. It is also imperative to understand what kinds of methodological challenges would arise in utilizing historical street view images. In the following sections, we present an experiment to assess to what extent deep learning-based semantic segmentation, processing historical GSV images much more efficiently than visual/manual inspection, enables one to discern locations with and without a certain type of change.

3. Data & Methods

3.1. Study area

As briefly mentioned above, we devise an experiment through which we can assess the efficacy of using historical GSV with deep learning-based semantic segmentation in detecting changes in our neighborhoods. Our experiment is carried out using data for the City of Santa Ana located at the center of Orange County, California.

Santa Ana is home to approximately 330,000 people and numerous diverse business establishments. As one of the earliest incorporated cities in Southern California (incorporated in 1886), Santa Ana contains both old and new development styles, and parts of the city possess some of the urban features common in small or mid-sized American cities. In recent years, it has been characterized by various efforts to redevelop commercial districts and residential neighborhoods, thus providing a good opportunity to evaluate the performance of various approaches to using historical GSV imagery. Furthermore, although largely urbanized, the city features a mixture of high- and low-density development patterns and various land uses, allowing one to test certain methods in various settings.

3.2. Historical GSV images in Santa Ana

GSV provides an invaluable opportunity for researchers to obtain street view images in Santa Ana and numerous other cities around the world. While acquiring static images in some formats incurs costs, panoramic (360 degree) images can be downloaded more easily without requiring payment. This is true not only for the latest imagery but also for historical street view images, presenting great potential for a variety of longitudinal research projects.

It should be noted, however, that the availability of GSV is far from uniform. Figure 1 demonstrates the unevenness by showing the temporal distribution of the GSV panoramic images available in Santa Ana. This figure was made based on the metadata of all panoramic GSV images found in the City of Santa Ana. Since Google does not provide a convenient way to retrieve historical GSV images through their official API, we employed a module obtained from GitHub (source: https://github.com/robolyst/streetview) that enabled us to access historical GSV images. This module allowed us to find all available GSV images, including current and old ones, surrounding a given location point. To check the GSV availability thoroughly, we created points with 20-meter intervals along with all road segments in Santa Ana and used these points repeatedly to identify all available GSV images in the city, given that the search boundary of the module to retrieve GSV images was greater than 20 meters (Figure 2). We removed duplicate images found through this process and ended up with a total of 570,522 panoramic images.

As shown in the figure, Google started to take/provide street view images in the city in 2007, but the number of images taken/provided was small in that year. In the case of Santa Ana, the first massive street imagery acquisition took place in June 2008. Additionally, over twenty-five thousand images were taken in March 2011, April 2015, and December 2017, whereas GSV imagery is often unavailable for months (e.g., no GSV image available from March 2009 to September 2010 for any location points in Santa Ana).

<< Insert Figure 1 about here >>

The spatial distribution of GSV images is also uneven. As reported by Kim et al. (2021) and other studies, GSV images are often unavailable for private properties (e.g., gated communities) due to difficulties in accessing them. While the geographic coverage of GSV has dramatically expanded, it is not always possible to get the same amount of street imagery information for all road segments in a city (see Figure 2). To mitigate the data availability issue, researchers need to use all street view images taken over months or even years rather than relying on the data collected on a single day. Admittedly, there is no single perfect time window or aggregation scheme, but it is important to inspect how the availability of street view images varies across time and space. In particular, when conducting longitudinal studies, consideration may need to be given to when massive street imagery acquisition took place in the study area (such as April 2015 and December 2017 in the case of Santa Ana).

<< Insert Figure 2 about here >>

3.3. Research design

Our experiment utilizes the logic of difference-in-differences analysis, comparing (1) treatment group and (2) two control groups with observations before and after the treatment. The treatment group consists of 248 sites where new construction or demolition of existing buildings took place between May 2015 and November 2017. These 248 sites (i.e., treatment group) and their exact location points were identified through a careful examination of building permit data provided by the City of Santa Ana, as detailed in Appendix A. The spatial distribution of these sites is not uniform since development projects tend to concentrate in some locations rather than being evenly distributed within a city. As shown in Figure 3, there were 39 sites of construction or demolition on/near Bristol Street, a north-south major arterial that underwent a large-scale corridor improvement project. For the control groups, we used a random sampling approach to draw two sets of the same number (248) of location points without such construction or demolition activity.[2] While it is unlikely that 248 randomly generated points will result in misleading conclusions, we nonetheless generated an additional control group (Control group #2) to increase the robustness of our experiment.

<< Insert Figure 3 about here >>

[2] This was accomplished by employing the "Create Random Points" tool in ArcGIS. In doing so, we first created a layer that meets the following three conditions and then drew random points from the layer to increase the comparability between treatment and control group points: (1) All control group locations should be located within the city boundaries, (2) Distance to the nearest road segment should be between 10 and 30 meters (since most of the 248 treatment group cases had this range of distance to the nearest road segments), and (3) Distance to the nearest construction or demolition site should be larger than 100 meters.

It is important to note that the aforementioned time period (May 2015 – November 2017) was chosen to have the largest number of GSV images possible for both before and after the period. As shown in the previous section (Figure 1), in Santa Ana, over twenty-five thousand images were taken in April 2015 and December 2017, respectively. In the experiment, to have additional images, we used the following 18-month windows:

- Pre-treatment: GSV images taken between November 2013 and April 2015 (Note: No images available for the first six months of this period, but we made it 18 months to make it consistent with the post-treatment period.)
- Post-treatment: GSV images taken between December 2017 and May 2019

For each of the 248 treatment and 496 (=2×248) control group cases, we first searched and identified the nearest GSV image taken during the pre- and post-treatment time windows. Additionally, we collected nearby GSV images from each (treatment or control) point using the following four buffers: 25, 50, 75, and 100 meters. While a larger buffer (e.g., 100 meters) allowed us to have more GSV images around each case, having these additional images does not necessarily improve our ability to detect changes in the built environment more precisely with historical GSV images. In contrast, a narrow buffer (e.g., 25 meters) can result in no image taken during the pre- and/or post-treatment time windows. In other words, there is a tradeoff between precision and data availability, and we will present and discuss our results regarding this issue in Section 4.

All nearby GSV images (for each treatment or control point) were analyzed by employing a deep learning-based semantic segmentation approach. As explained in detail below (*Section 3.4. Semantic Segmentation*), this allowed us to determine what percentage of the pixels in each image indicated buildings and many other streetscape elements (e.g., pavement, sidewalks, sky, vehicles, vegetation) and thus compare the detailed semantic segmentation results derived from treatment and control group cases. Since the treatment in this experiment is construction or demolition of buildings, our assessment focused on changes in *percent buildings*.

More specifically, for each case, we calculated the before-and-after change in *percent buildings* and used the metric in absolute term (*abs.change.pct.buildings* hereafter) for our assessment as shown below.[3] The calculation of *abs.change.pct.buildings* was conducted in two ways:

1) Using the two nearest GSV images:

$$abs.change.pct.buildings(i) = |nearest.post.pct.buildings(i) - nearest.pre.pct.buildings(i)|$$

where $nearest.post.pct.buildings(i)$ indicates the percent buildings of the nearest GSV image taken during the post-treatment period (December 2017 – May 2019) from *i*-th case point (*i* = 1, 2, …, 743, 744 because there are 248 treatment and 496 control cases), while $nearest.pre.pct.buildings(i, d)$ represents the percent buildings of the nearest GSV image taken during the pre-treatment period (November 2013 – April 2015) for the same (*i*-th) case point.

2) Using all nearby images falling within a buffer:

$$abs.change.pct.buildings(i, d) = |mean.post.pct.buildings(i, d) - mean.pre.pct.buildings(i, d)|$$

where $mean.post.pct.buildings(i)$ indicates the mean of the percent buildings values of the GSV images taken during the post-treatment period within the buffer distance *d* (25, 50, 75, or

[3] The metric was used in absolute term because our treatment in this experiment included both construction and demolition of buildings. One limitation is that this metric is not able to capture changes in building shape. If the overall building size remains the same, the metric will not yield a value that can allow one to detect the construction/demolition activity.

100 meters) from i-th case point, while $mean.pre.pct.buildings(i, d)$ represents the corresponding mean percent buildings calculated based on the GSV images taken during the pre-treatment period located within the same buffer distance (*d*) from the same (i-th) case point. A short buffer distance (e.g., 25 meters) can be more effective in detecting changes in the streetscape, but GSV images are not always available within such narrow boundaries. It is worthwhile to additionally test larger buffer distances to see how the results may differ depending on buffer sizes.

We expect the value of *abs.change.pct.buildings* to be larger for treatment group cases (involving construction or demolition of buildings) than for control group cases (without such activity), if semantic segmentation of historical GSV can precisely detect changes in the built environment. If so, one could use a threshold (or cutoff) value of *abs.change.pct.buildings* (e.g., *abs.change.pct.buildings* > 1 percentage point) to identify areas experiencing a certain form of the built environment changes by utilizing this deep learning-based semantic segmentation approach with historical GSV. However, some control group cases can have a large magnitude of *abs.change.pct.buildings* exceeding the threshold. Such a false positive outcome (or Type I error as shown in Figure 4), can occur due to the imperfection of the semantic segmentation tool and/or other reasons that could make pre-and-post-treatment GSV images inconsistent.

<< Insert Figure 4 about here >>

Although it is possible that false positive or false negative outcomes are more likely to arise in some locations, this strategy might show better performance in other parts of the city. For instance, GSV images taken on a wider street might not have the same ability to capture a single building on the street (due to the longer distance between the building and the Street View car/fleet) compared to those taken on a narrower road segment and thus could lead to false negative outcomes more frequently. To understand under what circumstances historical GSV processed through deep learning-based semantic segmentation can yield more satisfactory outcomes, we analyzed how the rates of false outcomes were associated with street segment characteristics. To accomplish this, for each road segment in Santa Ana, we calculated the mean *percent buildings*, *percent pavement*, *percent sidewalks*, *percent sky*, *percent vehicles*, *percent vegetation*, and *percent all other streetscape elements* by processing historical GSV images through semantic segmentation. This segment-level average computation used not only the GSV images located near a treatment or control case but also other street view images available on the road segment. The following section provides a detailed description of the semantic segmentation method used for this average calculation as well as our analysis of the treatment and control cases.

3.4. Semantic segmentation

To analyze the collected historical GSV images, we employed semantic segmentation, a computer vision technique to segment each image pixel as a component. While traditional image classification methods rely on RGB (red, green, blue) information or other spectral characteristics, recent years have seen the development of more advanced semantic segmentation models trained with the use of deep learning algorithms that exhibit stronger performance. These models consider the semantic meaning of objects in an image to better classify them into pre-defined categories at the pixel level by accounting for the detailed arrangement, shape, and color information embedded in the image. Some of these models, pre-trained by a deep learning algorithm with the use of a benchmark dataset containing annotated urban street images, are readily available for performing semantic segmentation.[4] Therefore, these models have been widely adopted as a method of processing street view images to quantify the neighborhood environment.

Among the several semantic segmentation models available to researchers (e.g., FCN8s, PSPNet), this study used the Deeplabv3+ model (Chen et al., 2018) which has performed well and is therefore utilized by a growing number of studies, such as Nagata et al. (2020), Wang and Vermeulen (2021), and Hipp et al. (2022). This model uses convolutional neural networks (CNN) to detect complex patterns and relationships of objects within an image (Li et al., 2020). The CNN deep learning makes it suitable for extracting features from urban street view images where RGB information alone often leads to inaccuracies (see, e.g., Li et al., 2015; Lu, 2019).

The Deeplabv3+ model was pre-trained with the Cityscapes dataset (Cordts et al., 2016), one of the most popular datasets offering both original streetscape images and annotated images of various streetscape components. In other words, the model had the ability to classify GSV image pixels into various streetscape elements. These elements included not only buildings (the main target feature in this study) but also pavement, sidewalks, sky, vehicles, vegetation, and many other objects, such as humans and traffic signs.

It is important to note that the raw panoramic GSV images that can be downloaded without payment have limitations. Specifically, a distortion issue exists in panoramic images, and this can make the outcome of semantic segmentation less desirable (Li et al., 2018a; Tsai and

---

[4] For the detection of new buildings, one could train and use an object detection model instead of employing a semantic segmentation model. However, it requires a lengthy process of data preparation, image annotation, and model training to develop such a tool tailored to capture a certain object (e.g., new buildings) with high precision. In this study, we tested the applicability of a pretrained semantic segmentation model, Deeplabv3+ (source: https://github.com/lexfridman/mit-deep-learning) that is increasingly used by researchers given its proven performance in detecting various features in urban street view images.

Chang, 2013; Yin et al., 2015). The degree of distortion is particularly severe in the upper and lower parts of panoramic images (see, e.g., Tsai and Chang, 2013; Orhan and Bastanlar, 2022). Therefore, we only used the central part of each panoramic image with the least distortion by cropping 100 pixels of the upper and lower parts of the image to handle this issue.

In addition, from each panoramic image, we derived more narrowly-focused street view images for our auxiliary analysis. This process, illustrated in Figure 5, enabled us to replicate what one could have if s/he used static GSV images. Although static images can be collected via GSV API, there is a quite strict quota on the number of freely downloadable static GSV images per month. Therefore, we processed panoramic images using the Equirec2Perspec package (source: https://github.com/fuenwang/Equirec2Perspec) that enabled us to convert an equirectangular panoramic image into the images comparable to static images. For each panoramic image, we extracted four images (front, back left, and right) with a size of 640 × 640 pixels. Since buildings tend to appear in the left- and right-side images of the street, we analyzed these two images through semantic segmentation for our auxiliary analysis (see Section 4.4. for the results of this analysis).

<< Insert Figure 5 about here >>

4. Results

4.1. Comparison of treatment and control groups

As described in the previous section, for each of the 248 treatment and 496 control group cases, we computed *abs.change.pct.buildings* by analyzing the pre- and post-treatment GSV images through semantic segmentation. First, the nearest pre- and post-treatment images were compared for this experiment. Additionally, we used all nearby pre- and post-treatment GSV images in four buffer sizes (*d* = 25, 50, 75, 100 meters), and Table 1 provides a summary of these comparisons of the treatment and control groups. It should be noted that not all 744 sites (248 treatment and 496 control cases) had nearby GSV images taken during both pre- and post-treatment periods. This issue was particularly apparent when a narrow buffer size (e.g., 25m) was used (see *n* in Table 1 for the number of treatment or control group cases with GSV images available for both pre- and post-treatment periods).

<< Insert Table 1 about here >>

Overall, the treatment group showed much larger *abs.change.pct.buildings* values than the control group, as expected. For instance, when the nearest pre- and post-treatment GSV images were analyzed, treatment group cases exhibited a mean of 3.781 for *abs.change.pct.buildings* which was significantly larger than those (0.613 and 0.642) of two control groups. Similarly, even though the magnitude of the difference varied as discussed below, all of the four buffer sizes resulted in a larger mean of *abs.change.pct.buildings* for the treatment group. Furthermore, the mean differences between the treatment and control groups were all statistically significant at $p < 0.001$, suggesting that the tested deep learning-based semantic segmentation of historical GSV images provided an ability to identify areas with real changes in the built environment.

Among the four buffer sizes tested, *d*=50m yielded the largest mean difference between the treatment and control groups. However, the t-statistic for the mean difference was larger, when a larger buffer size was used (given the larger *n*). On the contrary, *d*=25m demonstrated the poorest performance. As shown in Table 1, when this narrow buffer size was used, only 89 out of 248 sites of the treatment group had GSV images for both pre- and post-treatment periods.

Moreover, the *abs.change.pct.buildings* values derived from these 89 sites varied quite substantially, as indicated by the large standard deviation (SD=2.170) presented in the table.

As the buffer size increased, an increasing number of cases have at least one GSV image for both pre- and post-treatment periods. However, in the case of the treatment group, even when *d*=100m was used, less than 90% of the total 248 sites had GSV images for both pre- and post-treatment periods, whereas nearly all control group cases had pre- and post-treatment GSV images. This indicates that the GSV availability near the treatment group sites was more limited than that of the control groups. This happened mainly because a considerable number of treatment group cases were construction sites in new subdivisions for which GSV images during the pre-treatment period were not available due to the limited or lack of road access to those locations. The finding highlights the uneven availability of historical GSV, and this challenge deserves attention as it can limit the ability of researchers to utilize historical street imagery data platforms for their investigation of some parts of our cities (e.g., new subdivisions).

4.2. False positive and false negative rates

We next asked the likelihood of making a correct decision about the presence of new buildings when using the nearest GSV image comparison (the first strategy). The top part of Table 2 compares the distribution of *abs.change.pct.buildings* for the treatment group with that of the two control groups. As demonstrated in the table, the treatment group showed larger values of *abs.change.pct.buildings* much more frequently than the two control groups. For instance, only 25.4% of treatment cases had *abs.change.pct.buildings* less than 0.5, whereas over 60% of control cases had a value of *abs.change.pct.buildings* in this range.

<< Insert Table 2 about here >>

In the lower part of Table 2, we present the false positive and false negative rates calculated based on various thresholds to identify treatment sites with construction or demolition of buildings. That is, the identification is based simply on whether each site's *abs.change.pct.buildings* is greater than the threshold or not. These rates demonstrate the efficacy of the semantic segmentation method tested in this study in two ways: (1) to what extent does it misidentify a control case as a treatment case (false positive) and (2) to what extent does it fail to identify a treatment case (false negative).

If a researcher had used 0.5 as a cutoff to determine whether a site had treatment (i.e., involved construction or demolition of buildings) or not, s/he would correctly identify 74.6% of the total treatment cases. That is to say, in this case, the false negative rate is 25.4%. On the other hand, the cutoff of 0.5 would result in a false positive rate of 33.9% (or 39.1% in the case of Control group #2), as it could screen out 66.1% (or 60.9% in the case of Control group #2) of the control cases.

It is clear that there is a tradeoff between false positive and false negative rates. A lower cutoff (e.g., 0.5) can lead to more false positives, while it can reduce the chance of ending up with false negatives. If the goal is to keep the false positive rate low, a higher cutoff would be more desired. According to our experiment, cutoff=1.5 was found to reduce the false positive rate down to less than 10% (i.e., 5.6% in the case of Control group #1 and 9.3% in the case of Control group #2). In this case, the semantic segmentation method was able to detect 41.5% of treatment cases.

Figure 6 presents the receiver operating characteristic (ROC) curves which clearly show the tradeoff by displaying the false positive and true positive rates for various cutoff values. A method is considered good when the area under the curve (AUC) is large. The AUC values from this experiment were larger than 0.75, which can be considered fair (Lüdemann et al., 2006).

<< Insert Figure 6 about here >>

4.3. Performance variation across contexts

As noted in Section 3, the false positive or false negative outcomes may arise more frequently in some contexts than in other locations. We examined this possible variation across geographical contexts by analyzing how the values of *abs.change.pct.buildings* were associated with road segment-level characteristics, and the results of this analysis are summarized in Table 3.

<< Insert Table 3 about here >>

As seen in the table, in the case of the treatment group, *abs.change.pct.buildings* was found to be larger in areas where GSV images had more pixels of buildings, less pavement, less sidewalks, less sky, more vegetation, and less other streetscape elements (remainders). The association between *abs.change.pct.buildings* and the segment-level mean percent buildings was strong (correlation = +0.399), indicating that historical GSV yielded a large value of *abs.change.pct.buildings* and thus effectively captured changes in the built environment in locations with more buildings. In contrast, a notable negative correlation was detected with the mean percent pavement and the mean percent sky, meaning that *abs.change.pct.buildings* tended to be lower where pavement and sky accounted for a large proportion of the street view images. These two findings together suggest that false negative rates tended to be lower (due to larger *abs.change.pct.buildings* values) in environments with more buildings and less pavement+sky pixels. As a consequence, in a built environment with narrow streets and minimal setbacks new construction or demolition of old buildings can easily be reflected in GSV image pixels and thereby captured through deep learning-based semantic segmentation with fewer false negatives.

Many contemporary cities in fact have wide streets and large setbacks. In our study area, the Bristol Street mentioned in Section 3.3 presents such a situation. Among the 39 treatment sites located within 50 meters of the street, nearly half (17 out of 39, 43.4%) had *abs.change.pct.buildings* less than 0.5, suggesting that the probability of false negative results is likely to be high in this type of geographical setting. The probability was much lower (22%) for the remaining 209 treatment sites, including those in non-central locations of the city. Thus, the size and layout of the street on which the building is located, which may be a function of the time period of development, may also impact the strategy's ability to detect environmental change.

Although not presented in the table, another notable finding is that *abs.change.pct.buildings* tended to be larger in areas with multiple building permits. This implies that false negative rates can further be reduced on road segments with construction or demolition of multiple buildings. In other words, while a single (or small-scale) project might not always be detected, it is rare for semantic segmentation of historical GSV images to miss large-scale construction or demolition projects. Therefore, this deep learning-based method can be used with high confidence for detecting such projects or other types of large-scale changes in the built environment.

In the case of Control group #2, *abs.change.pct.buildings* showed a strong, positive correlation with the segment-level mean percent buildings, while this was not the case in Control group #1. This finding (combined with the similarly strong positive correlation detected from the treatment group) suggests that both false positive and false negative rates can be higher in areas with more building pixels.

Figure 7 provides historical GSV examples showing false positive (B) and false negative (C) outcomes. In the case of (B), trees on the street made it difficult to capture all the buildings in the pre-treatment period, and this resulted in a high level of *abs.change.pct.buildings* for this control case. In the case of (C), there was another building behind the new building (as shown in the pre-treatment image), and consequently *percent buildings* did not increase substantially. The small size of the new building in the post-treatment image also contributed to this false negative outcome.

<< Insert Figure 7 about here >>

4.4. Analysis with left and right-side images only

What if we used narrowly-focused (static) images instead of large panoramic GSV images? We attempted to answer this question by deriving left- and right-side street view images from each panoramic image using the Equirec2Perspec package as explained in Section 3.4. The results of this auxiliary analysis are presented in Table 4.

<< Insert Table 4 about here >>

As shown in the table, *abs.change.pct.buildings* values became significantly larger, when these narrowly-focused images were used. For instance, when the nearest pre- and post-treatment GSV images were compared, the mean of *abs.change.pct.buildings* increased from 3.781 (Table 1) to 10.098 (Table 4) for the treatment group, from 0.613 (Table 1) to 1.718 (Table 4) for Control group #1, from 0.642 (Table 1) to 2.046 (Table 4) for Control group #2. Accordingly, the mean difference (between the treatment and control groups) went up substantially. This is not surprising given that buildings are likely to account for a larger percentage of the pixels in these narrowly-focused images than they could in larger panoramic images. The t-statistics for the mean differences, however, did not increase, suggesting that the image treatment does not necessarily improve the performance of the tested deep learning-based semantic segmentation in distinguishing treatment cases from control cases. Using such images can result in inaccuracies when GSV images are not available right in front of the target (buildings in this study) or pre- and post-images were not taken from the exact same location, while larger panoramic images are less vulnerable to these situations.

5. Summary and Discussion

Recent years have witnessed the emergence of new data opportunities that can dramatically enhance our capabilities for urban and environmental research, and geocoded street view imagery is one of the most promising data sources of this kind. However, while this promising resource has been increasingly utilized by researchers in many disciplines, we still know little about how to use it more effectively and efficiently (Kim et al., 2021). This is particularly true when it comes to historical street view images and their use for longitudinal studies (Biljecki and Ito, 2021).

In this study, we devised and conducted an experiment to assess the efficacy of using historical street view images. More specifically, we assessed the feasibility of deep learning-based semantic segmentation as a method for detecting changes in the built environment with historical GSV. To do this, we analyzed the rates of false positive and false negative outcomes using data for (1) 248 treatment sites with construction or demolition of buildings and (2) two groups of randomly chosen control cases without such activity in Santa Ana, California.

Additionally, we examined under what circumstances (or geographical contexts) the method would tend to yield better results.

Our assessment presents both opportunities and challenges. On the one hand, the deep learning-based method tested in this study showed an ability to distinguish treatment cases from control cases. To be more precise, when a threshold of *abs.change.pct.buildings* > 0.5 was used, it was able to detect nearly 75% of the treatment sites with historical GSV images available for pre- and post-treatment periods, while screening out over 60% of the control cases. It performed well particularly in areas with more buildings, narrow streets (or less pavement+sky pixels), and larger-scale construction/demolition projects.

On the other hand, the method yielded a noticeably high value of *abs.change.pct.buildings* for some control cases. While the occurrence of such false positive outcomes can be mitigated by adopting a higher threshold as discussed in Section 4.2, this approach does not eliminate the possibility given that false positive outcomes can happen not only due to the imperfection of the semantic segmentation technique but also due to the uneven availability of historical GSV images and the misalignment of GSV image points over years. It is also important to note that a high threshold to reduce false positive outcomes increases false negative results, an important tradeoff that exists in devising and applying a method for detecting changes in the built environment through analysis of street view images. One could argue that false positive outcomes are less concerning given that researchers can verify the cases with positive results through an additional step of visual (eyeball) inspection, site visits, or verification with other data sources, but this may not always be feasible in large-N projects.

We also highlight that while we had building permit data for our study area—which potentially could be directly used to assess environmental change—this will not be the case for other study sites. Many cities do not provide building permit data in a geocodable format. There is an even more limited list of cities that provide historical building permit data for a long period of time. Furthermore, the data availability is often more limited in some other countries (see e.g., Arellana et al., 2020). Given all these issues, historical GSV imagery can be viewed as a valuable alternative source of information. Although our experiment focused on construction or demolition of buildings, we believe this experiment can inform those who are interested in detecting other types of changes in the streetscape using historical GSV with deep learning-based semantic segmentation tools.

Although this study has revealed some challenges and limitations to using historical street imagery, we nonetheless believe that on balance there is strong potential for using the images in research exploring neighborhood change. When combined with a machine learning technique, it appears that historical street images can feasibly capture real change in the built environment for large-N projects. Our analysis shows some limitations that researchers will want to be aware of when utilizing such images, including that the availability of street view images fluctuates over time and thus requires special attention in the research design. We therefore believe that this study provides insights that we hope will result in even more creative and broader use of historical street imagery. It is our hope that our analyses can serve as a catalyst for more experiments or other types of future research through which the deep learning-based method presented here (or its derivatives) can be refined. Such future research endeavors will greatly contribute to more efficient use of historical street view images that are indispensable for more complete understanding of the constant evolution of our cities and the complex mechanisms behind it.

References

Arellana, J., Saltarín, M., Larrañaga, A. M., Alvarez, V., & Henao, C. A. (2020). Urban walkability considering pedestrians' perceptions of the built environment: A 10-year review and a case study in a medium-sized city in Latin America. *Transport Reviews*, 40(2), 183-203.

Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning*, 215, 104217.

Cândido, R. L., Steinmetz-Wood, M., Morency, P., & Kestens, Y. (2018). Reassessing urban health interventions: Back to the future with Google Street View time machine. *American Journal of Preventive Medicine*, 55(5), 662-669.

Chen, Y. T., Cooper, H. L., Windle, M., Haardörfer, R., Crawford, N. D., Chen, W. J., & Chen, C. Y. (2016). Residential environments, alcohol advertising, and initiation and continuation of alcohol consumption among adolescents in urban Taiwan: A prospective multilevel study. *SSM-Population Health*, 2, 249-258.

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV) (pp. 801-818).

Cohen, N., Chrobok, M., & Caruso, O. (2020). Google-truthing to assess hot spots of food retail change: A repeat cross-sectional Street View of food environments in the Bronx, New York. *Health & Place*, 62, 102291.

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

Fry, D., Mooney, S. J., Rodríguez, D. A., Caiaffa, W. T., & Lovasi, G. S. (2020). Assessing Google Street View image availability in Latin American cities. *Journal of Urban Health*, 97(4), 552-560.

Gallagher, R., Liu, Y., & Sigler, T. (2019). Parcel amalgamation as a mechanism for achieving urban consolidation through densification: The fixity of property boundaries over time. *Land Use Policy*, 89, 104239.

Hipp, J. R., Lee, S., Ki, D., & Kim, J. H. (2022). Measuring the built environment with Google Street View and machine learning: Consequences for crime on street segments. *Journal of Quantitative Criminology*, 38(3), 537–565.

Ito, K., & Biljecki, F. (2021). Assessing bikeability with street view imagery and computer vision. *Transportation Research Part C: Emerging Technologies*, 132, 103371.

Kang, Y., Zhang, F., Peng, W., Gao, S., Rao, J., Duarte, F., & Ratti, C. (2021). Understanding house price appreciation using multi-source big geo-data and machine learning. *Land Use Policy*, 111, 104919.

Kim, J. H., Lee, S., Hipp, J. R., & Ki, D. (2021). Decoding urban landscapes: Google Street View and measurement sensitivity. *Computers, Environment and Urban Systems*, 88, 101626.

Li, X. (2021). Examining the spatial distribution and temporal change of the green view index in New York City using Google Street View images and deep learning. *Environment and Planning B: Urban Analytics and City Science*, 48(7), 2039-2054.

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

Li, X., Ratti, C., & Seiferling, I. (2018a). Quantifying the shade provision of street trees in urban landscape: A case study in Boston, USA, using Google Street View. *Landscape and Urban Planning*, 169, 81-91.

Li, X., Santi, P., Courtney, T. K., Verma, S. K., & Ratti, C. (2018b). Investigating the association between streetscapes and human walking activities using Google Street View and human trajectory data. *Transactions in GIS*, 22(4), 1029-1044.

Li, X., Yang, Y., Zhao, Q., Shen, T., Lin, Z., & Liu, H. (2020). Spatial pyramid based graph reasoning for semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8950-8959).

Lu, Y. (2019). Using Google Street View to investigate the association between street greenery and physical activity. *Landscape and Urban Planning*, 191, 103435.

Lüdemann, L., Grieger, W., Wurm, R., Wust, P., & Zimmer, C. (2006). Glioma assessment using quantitative blood volume maps generated by T1-weighted dynamic contrast-enhanced magnetic resonance imaging: a receiver operating characteristic study. *Acta Radiologica*, 47(3), 303-310.

Nagata, S., Nakaya, T., Hanibuchi, T., Amagasa, S., Kikuchi, H., & Inoue, S. (2020). Objective scoring of streetscape walkability related to leisure walking: Statistical modeling approach with semantic segmentation of Google Street View images. *Health & Place*, 66, 102428

Orhan, S., & Bastanlar, Y. (2022). Semantic segmentation of outdoor panoramic images. *Signal, Image and Video Processing*, 16(3), 643-650.

Smith, C. M., Kaufman, J. D., & Mooney, S. J. (2021). Google Street View image availability in the Bronx and San Diego, 2007–2020: Understanding potential biases in virtual audits of urban built environments. *Health & Place*, 72, 102701.

Tsai, V. J., & Chang, C. T. (2013). Three-dimensional positioning from Google Street View panoramas. *IET Image Processing*, 7(3), 229-239.

Tyndall, J. (2021). Getting high and low prices: Marijuana dispensaries and home values. *Real Estate Economics*, 49(4), 1093-1119.

Wang, M., & Vermeulen, F. (2021). Life between buildings from a street view image: What do big data analytics reveal about neighbourhood organisational vitality? *Urban Studies*, 58(15), 3118-3139.

Yang, Y., He, D., Gou, Z., Wang, R., Liu, Y., & Lu, Y. (2019). Association between street greenery and walking behavior in older adults in Hong Kong. *Sustainable Cities and Society*, 51, 101747.

Yin, L., Cheng, Q., Wang, Z., & Shao, Z. (2015). 'Big data' for pedestrian volume: Exploring the use of Google Street View images for pedestrian counts. *Applied Geography*, 63, 337-345.

Yin, L., Wu, L., Cole, S., & Wu, L. (2019). Synthesizing data to explore the dynamic spatial patterns of hotel development. *ISPRS International Journal of Geo-Information*, 8(10), 448.

Zhanjun, H. E., Wang, Z., Xie, Z., Wu, L., & Chen, Z. (2022). Multiscale analysis of the influence of street built environment on crime occurrence using street-view images. *Computers, Environment and Urban Systems*, 97, 101865.