Subjectively Measured Streetscape Qualities for Shanghai with Large-Scale Application of Computer Vision and Machine Learning

Waishan Qiu[1]([⊠]), Wenjing Li[2], Xun Liu[3], and Xiaokai Huang[4]

[1] Department of City and Regional Planning, Cornell University, Ithaca, NY 14850, USA
[2] Center for Spatial Information Science, The University of Tokyo, Tokyo 113-8654, Japan
[3] School of Architecture, University of Virginia, Charlottesville, VA 22904, USA
[4] Graduate School of Design, Harvard University, Cambridge, MA 02138, USA

Abstract. Recently, many new studies emerged to apply computer vision (CV) to street view imagery (SVI) dataset to objectively extract the view indices of various streetscape features such as trees to proxy urban scene qualities. However, human perceptions (e.g., imageability) have a subtle relationship to visual elements which cannot be fully captured using view indices. Conversely, subjective measures using survey and interview data explain more human behaviors. However, the effectiveness of integrating subjective measures with SVI dataset has been less discussed. To address this, we integrated crowdsourcing, CV, and machine learning (ML) to subjectively measure four important perceptions suggested by classical urban design theory. We first collected experts' rating on sample SVIs regarding the four qualities which became the training labels. CV segmentation was applied to SVI samples extracting streetscape view indices as the explanatory variables. We then trained ML models and achieved high accuracy in predicting the scores. We found a strong correlation between predicted complexity score and the density of urban amenities and services Point of Interests (POI), which validates the effectiveness of subjective measures. In addition, to test the generalizability of the proposed framework as well as to inform urban renewal strategies, we compared the measured qualities in Pudong to other five renowned urban cores worldwide. Rather than predicting perceptual scores directly from generic image features using convolution neural network, our approach follows what urban design theory suggested and confirms various streetscape features affecting multi-dimensional human perceptions. Therefore, its result provides more interpretable and actionable implications for policymakers and city planners.

Keyword: Subjective measure · Human perception · Street view image · Computer vision · Global comparison

1 Introduction

Urban design qualities such as the enclosure directly affect a person's appreciation of a place (Ewing and Handy 2009). Recently, with prevalence of Street View Imagery

(SVI) data in environment auditing (Yin and Wang 2016), computer vision (CV) have been widely applied to extract streetscape features, making the large-scale urban scene understanding possible (Yin et al. 2015). However, studies are limited to the objective measures. Only the view index of individual features such as tree and building are analysed, while the viewers' overall perceptions are ignored. Human perceptions have subtle relationships which cannot be fully represented by individual view indices nor a simple combination of them (Ewing and Handy 2009; Lin and Moudon 2010).

Conversely, the "subjective measure" which refers to evaluative scores collected from surveys questions can capture more subtle relationships (Lin and Moudon 2010). It is more user centered (Naik et al. 2014), although the definitions of perceptual qualities are inconsistent across studies (Ewing and Handy 2009). However, few studies have addressed subjective measures' effectiveness in capturing more subtle perceptions using SVI data.

To bridge the gap, we took Shanghai as an example and applies CV and ML to subjectively measure four perceptual qualities, namely the **enclosure, human scale, complexity, and imageability**. These perceptions have been identified important in affecting pedestrians' behaviors, residences' mode choices, and home buyers' willingness to pay (Ma et al. 2021). Our work enriches subjectively-measured urban perception studies. It is also the first cross-study for global cities. Urban renewal implications are derived for policymakers based on the global comparison. Furthermore, we contribute to future studies by proposing a framework integrating AI applications with classical urban measurement frameworks.

2 Literature Review

2.1 Objective and Subjective Measures

Street environment significantly affects people's appreciation of a place, as well as residence's physical activities, mode choice, and willingness to pay (Ewing and Cervero 2010). Street qualities have been mostly measured using objective quantities such as building height, street width, and number of trees (Cervero and Kockelman 1997). However, physical features alone cannot represent people's overall perception which have more subtle relationships (Ewing and Handy 2009).

Conversely, subjective measures often derive from interview and surveys. They explain people's behavior more completely, as behavior is mediated by the "cognitive map" of the environment (Lynch 1960). Conventional approaches relied on interview or telephone survey to collect people's overall perception have problems (Ewing and Handy 2009). First, the consistency and reliability of the operation can be questioned due to individual differences. Secondly, measurements based on surveys is time-consuming and expensive. The low throughput method limits subjective measure's application to larger geographic contexts (Naik et al. 2014). Third, the results are difficult to interpret, hence providing less instructive implications to policymakers (Lin and Moudon 2010).

Nevertheless, the subjective and objective measures could have been integrated. Ewing and Handy (2009) reviewed 51 subjective perceptual qualities from a pile of urban design literature. They statistically correlated subjective scores rated by experts for watching street view video clips, to the objectively-quantified elements like people

and tree from field survey. They successfully operationalized to objectively measure five seemingly subjective perceptions.

2.2 Computer Vision and Machine Learning in Street Measures

Recently, new studies emerged to take the advantages of open-source big data and AI algorithms. First, SVI data covers a handful of cities and spreads to new cities rapidly since 2007, which can be used to measure street-level human eye views inaccessible from the bird view (Li et al. 2015). A few recent studies measure the built environment using SVIs. For example, Rundle et al. (2011) used Google SVI to manually audit neighbor environment. Later, with the advance of AI such as CV and ML, automatically extracting features from images became possible. Yin and Wang (2016) applied ML to measure visual enclosure from SVI. Their results showed that the ML algorithms performed well to recognize and calculate the sky areas, allowing the measurement to be done reproducibly. Other researches have measured pedestrian, trees, sky, building, façade etc., respectively (Chen et al. 2020; Li et al. 2015; Ma et al. 2021). However, as discussed formerly, these objective view indices cannot represent viewers' overall feelings with the street scenes (Qiu et al. 2021).

Besides open-source SVI data, integrating crowdsourcing with AI has become viable to uncover large-scale public perceptions (Naik et al. 2014). Online data collection allows greater number of participants to evaluate perceived qualities from images, largely increasing the accessibility of urban perception data (Naik et al. 2014; Salesses et al. 2013). Naik et al. (2014) collected perceived safety online by asking participants to rank pairwise street photos. These preferences were converted to ranked safety scores and became the training data to train ML models to predict perceived safety score for 21 cities worldwide. The method was also applied to investigate the correlation between urban appearance and neighborhood income as well as housing prices (Glaeser et al. 2018).

Despite the effectiveness of subjective measures in incorporating more subtle human perceptions, most studies using SVI data are limited to objectively extracted visual elements. Little has been done to construct global maps of the subjectively-measured perceptions for the many perceptual qualities identified by classical urban design studies, such as imageability and complexity (Ewing and Handy 2009).

Therefore, our work sets to enrich the subjective measures of urban perceptions. It contributes to analytical frameworks by extending classical urban design framework with AI and big data (Fig. 1a). While Ewing and Handy (2009) relied on human labor to manually count physical features from video clips, we applied CV to extract the pixel ratios or counts of each important feature. While Naik et al. (2014) only mapped perceived safety score, we measured four important qualities identified by literatures in urban design and validate the scores with objective POI data. Furthermore, it is the first cross-study for several global cities with application of CV and ML which sheds light on urban renewal implications for global studies.
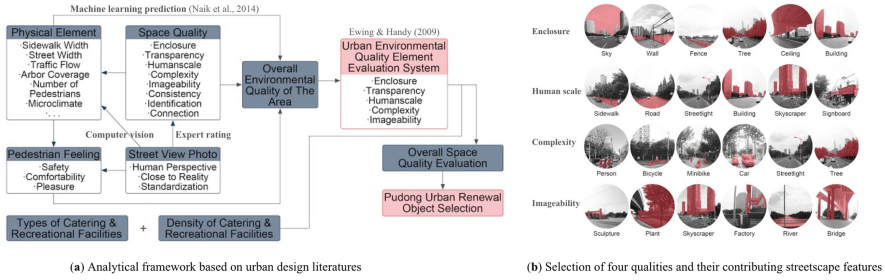
(a) Analytical framework based on urban design literatures

(b) Selection of four qualities and their contributing streetscape features

Fig. 1. Analytical framework (a) based on literatures in urban design qualities; and (b) the selection of the four perceptual qualities and their contributing features

3   Data and Methods

3.1   Study Area and Data Preparation

Pudong District in Shanghai is the financial center of China. Since the housing reform in 1998, Pudong has become one of the most expensive and vibrant housing markets in China (Chen et al. 2020). An empirical analysis for the street quality for a city-wide Pudong would provide essential implications for urban renewal. The data includes (1) SVIs collected from Baidu Street View API, (2) POI data from DaZhongDianPing and AutoNavi Map, and (3) shapefile of road networks from Open Street Map (OSM).

3.2   Calculating Subjective Qualities

3.2.1   Downloading Baidu SVIs

SVIs were downloaded from Baidu Street View Static API with consistent camera settings. The 'heading' was set using the street angle; image size was 600×300 pixels. The FOV (the horizontal field of view) was $120°$. The 'pitch' which specifies the up or down angle of the camera was $0°$. To ensure our training images would cover most urban area types, 300 images were randomly sampled across Shanghai region (Fig. 2).
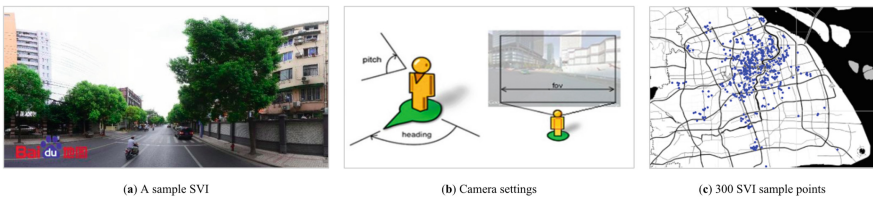

(a) A sample SVI

(b) Camera settings

(c) 300 SVI sample points

Fig. 2. Downloading Baidu SVIs (a) A typical SVI downloaded for this study. (b) The camera settings were controlled by "heading", "FOV", "pitch" and "resolution". (c) SVI training samples.

3.2.2   Collecting Public Perceptions as Training Labels

To collect people's preferences on street scenes as the training labels, we developed an online questionnaire platform where people can select the image preferred in pairwise comparisons regarding the four perceptual qualities (Fig. 3a). During a one-week period, we collected 3,120 valid entries from 23 volunteers who are mostly architecture students in Shanghai. In average, an image was compared to 10 other images, which is sufficient to lead the results converge (Naik et al. 2014).

These preferences were then translated to ranked scores with TrueSkill Algorithm (Microsoft 2005) which has also been applied to rank perceived safety (Naik et al. 2014). The ranked scores were normalized into 0–10 scale. People seemed to favor streetscapes with less sky exposure, more trees, and more pedestrians (Fig. 3b). These 300 labelled images become our training data.
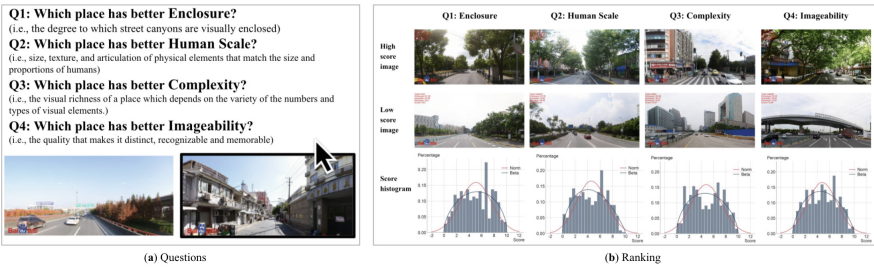

Fig. 3. Collecting the collaborative image of streetscape with an online survey platform. (a) Our online survey system asking participants to click on one of pairwise SVIs in response to evaluative questions. (b) High score, low score example images, and the histogram of score distribution, for each of the four perceived street qualities.

3.2.3   Physical Feature Classification

Pyramid Scene Parsing Network (PSPNet) is an image segmentation algorithm to produce reliable results on the scene parsing task (Zhao et al. 2016). We used PSPNet to extract and calculate the pixel ratios of individual features as view indices from SVIs. 35 kinds of streetscape elements have been detected (Fig. 4a). For the quantity of cars, peoples, signs, street furniture, the pixel ratio makes less sense, therefore, we applied MASKRCNN (He et al. 2017) to count the amounts (Fig. 4b).

3.2.4   Predicting Subjective Scores

We then applied several ML algorithms such as K-nearest neighbors (KNN), support vector machine (SVM) and random forest (RF) to predict the four perceptions. Mean Absolute Error (MAE) was set as the loss function, resulting in best models with an average MAE of 1.83, which is acceptable. With a scoring system of 0–10, an error of 1.83 will not alter the interpretation of a quality. We then applied the best performance models to all the downloaded 14,274 Baidu SVIs and derived the four subjective scores for Pudong Area.
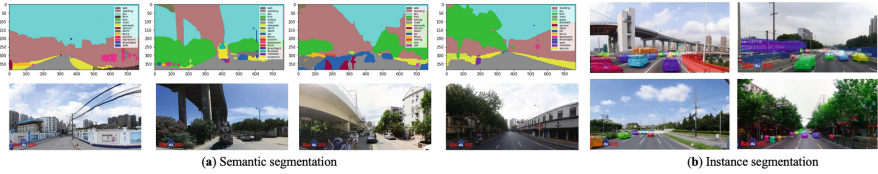
(a) Semantic segmentation   **(b)** Instance segmentation

Fig. 4. CV segmentation results (a) Pairwise PSPNet semantic segmentation results with its raw input (b) Mask R-CNN instance segmentation results counting objects

3.3 Correlation Test and Result Validation

Meanwhile, a logistic regression analysis was conducted among four qualities to check their correlations. The result shows that the degree of 'imageability' is significantly and positively correlated with 'enclosure' and 'complexity' (Fig. 5a). Furthermore, we crossed reference complexity score to the POI density (using food & beverage, entertainment, and recreation). Higher complexity score is correlated with more POIs, indicating the predicted complexity score effectively captures the impacts of urban amenities and services (Fig. 5b).
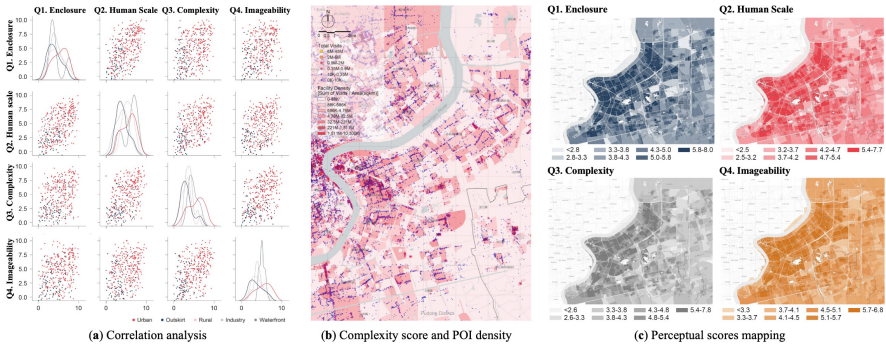

(a) Correlation analysis   **(b)** Complexity score and POI density   **(c)** Perceptual scores mapping

Fig. 5. Validation of results using (a) correlations test between four scores and (b) cross-reference to actual POIs density (c) Cognitive maps of four perceptual qualities

3.4 Global Comparison with Other Cities

To validate the generalizability of our framework and to inform what kind of environment facilitate urban innovation, we selected five renowned innovative districts, namely Cambridge Kendall Square, London Knowledge Quarter, Manhattan Wall Street, San Francisco Downtown and Seattle South Lake Union as the benchmark. The scores of Zhangjiang High-Tech Park were compared to that of the five benchmarks. Implications for urban design and renewal for Pudong and Zhangjiang were discussed based on comparison results.

4   Results and Findings

4.1   Spatial Distribution of Perception Qualities

Figure 5c provides the first comprehensive cognitive maps for Pudong District. The distributions of four perceptual qualities are heterogeneous, with downtown area (i.e., Lujiazui) conceived highest. The result indicates that when considering allocating renewal resources, more could be invested in the periphery residential areas and industrial parks where street qualities are conceived low, but with large residential population and employments, such as Zhangjiang High-Tech Park.

4.2   Comparison with Other Cities

Pudong's street qualities fall behind global best practices. Zhangjiang has the lowest average score compared to other five best practices (Fig. 6b), indicating more urban design implementations could be considered to improve the overall appreciation of street environment, as good street environment facilitate innovation. Second, five global districts have smaller variance in scores, while the scores in Zhangjiang are highly polarized, implying its uneven development (Fig. 6a). It suggests future study to investigate whether such uneven distribution have posed inequitable issues to specific population segments (Salesses et al. 2013). Last, the result confirms our method is applicable to a wide range of regions.
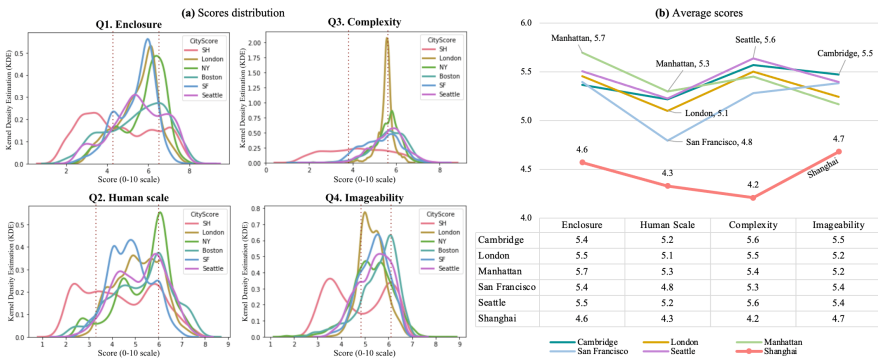

	Enclosure	Human Scale	Complexity	Imageability
Cambridge	5.4	5.2	5.6	5.5
London	5.5	5.1	5.5	5.2
Manhattan	5.7	5.3	5.4	5.2
San Francisco	5.4	4.8	5.3	5.4
Seattle	5.5	5.2	5.6	5.4
Shanghai	4.6	4.3	4.2	4.7

Fig. 6. Comparing six cities' perceptual qualities. (a) Score distributions (b) Averaged scores

4.3   Cross-reference with Zoning Metrics

To provide actionable policy suggestions, we cross-referenced perception scores with objective metrics of urban form and density, such as the average block size and floor area ratio (FAR) (Fig. 7). Zhangjiang has the widest roads but the lowest density measured, which explains its lowest perceived enclosure, since lower building heights and wider streets lead to less enclosure (Yin and Wang 2016). Less enclosure limits the neighborhood walkability and results in less walking behaviors, which is confirmed by the pedestrian counts from SVIs.
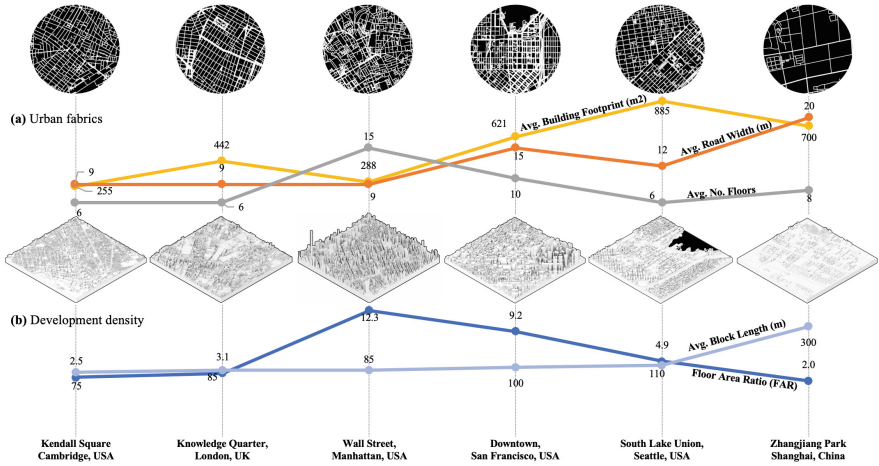
Fig. 7. Comparing (a) urban fabrics and block metrics, (b) development density and metrics

5   Conclusion

5.1   Effectiveness of Subjective Measures Using SVI and AI

While this method may not immediately replace the long-existing techniques in urban environment auditing, it offers many merits. For example, being closely related to the pedestrians' perspective, low-cost, requesting nothing from proprietary software or methods, and is commonly applied to where SVI dataset is available. The proposed method provides a useful alternative for planners and policymakers.

First, the cross-study of six global urban cores including Pudong district confirms the generalizability of our proposed framework. The method is reproducible and consistently predict perceptions from open-source SVI dataset that widely exists. Second, subjective measures capture more comprehensive and subtle human perceptions than using individual view indices. All the four important human perceptions suggested by urban design theory have been operationalized and the accuracy rates have been improved comparing to prior works (Ewing and Handy 2009). Third, although measured from simply images, perceptual scores capture many urban space qualities and characteristics that traditionally viable through objectively measured urban metrics. For example, the FAR, street width, building height, block size, and amenity density. We find a significant correlation between the complexity score and the POI density, as well as the enclosure with urban form and density metrics including FAR and street width. While the objective urban metrics must be measured using massive POIs and urban 3D model data with complicated workflows with ArcGIS and Rhino, our framework can stand alone without any licensed programs and software. All information needed are open-source. Therefore, compared to objective measures of urban form, our proposed framework is more accessible and higher throughput. Lastly, the cross-study indicates the polarized and uneven urban development in Pudong District. Unlike other benchmark cities, Pudong have large variances and lower average scores within all four perceptions, which suggests more equable allocation of urban design efforts and investment resources.

5.2   Limitations

First, our segmentation only used pre-trained models. Future studies could train specific models to fulfil more tailored tasks, such as to extract façades and windows which significantly affect many perceptions. Second, our training data was limited by the scarce of volunteer raters, and raters were not randomly selected. Third, further investigation could be done to address the divergence and coherence between subjective and objective measures of urban perceptions.

References

Cervero, R., Kockelman, K.: Travel demand and the 3Ds: density, diversity, and design. Transp. Res. Part Transp. Environ. **2**, 199–219 (1997). https://doi.org/10.1016/S1361-9209(97)00009-6

Chen, L., et al.: Measuring impacts of urban environmental elements on housing prices based on multisource data—a case study of Shanghai. China ISPRS Int. J. Geo.-Inf. **9**, 106 (2020). https://doi.org/10.3390/ijgi9020106

Ewing, R., Cervero, R.: Travel and the built environment: a meta-analysis. J. Am. Plan. Assoc. **76**, 265–294 (2010)

Ewing, R., Handy, S.: Measuring the unmeasurable: urban design qualities related to walkability. J. Urban Des. **14**, 65–84 (2009)

Glaeser, E.L., Kominers, S.D., Luca, M., Naik, N.: Big data and big cities: the promises and limitations of improved measures of urban life: big data and big cities. Econ. Inq. **56**, 114–137 (2018). https://doi.org/10.1111/ecin.12364

He, K., Georgia Gkioxari, Dollár, P., Girshick, R.: Mask R-CNN (2017)

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

Lin, L., Moudon, A.V.: Objective versus subjective measures of the built environment, which are most effective in capturing associations with walking? Health Place **16**, 339–348 (2010). https://doi.org/10.1016/j.healthplace.2009.11.002

Lynch, K.: The Image of the City. The MIT Press, Cambridge (1960)

Ma, X., et al.: Measuring human perceptions of streetscapes to better inform urban renewal: a perspective of scene semantic parsing. Cities **110**, 103086 (2021). https://doi.org/10.1016/j.cities.2020.103086

Microsoft: TrueSkillTM ranking system. Microsoft Res (2005). https://www.microsoft.com/en-us/research/project/trueskill-ranking-system/. Accessed 13 May 2019

Naik, N., Philipoom, J., Raskar, R., Hidalgo, C.: Streetscore – predicting the perceived safety of one million streetscapes. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Presented at the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 793–799. IEEE, Columbus (2014). https://doi.org/10.1109/CVPRW.2014.121

Qiu, W., Li, W., Zhang, Z., Li, X., Liu, X., Huang, X.: Subjective and objective measures of streetscape perceptions: relationships with property value in Shanghai (2021). https://doi.org/10.20944/preprints202103.0506.v1

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

Salesses, P., Schechtner, K., Hidalgo, C.A.: The collaborative image of the city: mapping the inequality of urban perception. PLoS ONE **8**, e68400 (2013). https://doi.org/10.1371/journal.pone.0068400

Yin, L., Cheng, Q., Wang, Z., Shao, Z.: 'Big data' for pedestrian volume: exploring the use of Google street view images for pedestrian counts. Appl. Geogr. **63**, 337–345 (2015). https://doi.org/10.1016/j.apgeog.2015.07.010

Yin, L., Wang, Z.: Measuring visual enclosure for street walkability: using machine learning algorithms and Google street view imagery. Appl. Geogr. **76**, 147–153 (2016). https://doi.org/10.1016/j.apgeog.2016.09.024

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network (2016)