A room with a view: Automatic assessment of window views for high-rise high-density areas using City Information Models and deep transfer learning

Maosu Li[1], Fan Xue[2*], Yijie Wu[3], and Anthony G. O. Yeh[4]

Highlights

- Four Window View Indices (WVIs) were defined for measuring outside greenery, water-body, sky, and construction views.
- WVIs complemented existing view indices from the ground, aircraft, and satellites for urban computing.
- City Information Model (CIM)-based view images were trustworthy data sources for WVIs.
- Automatic WVI assessment based on deep transfer learning with an ML regression layer was performed.
- Highly satisfactory ($R^2 > 0.95$) and fast (3.08 s/view) assessment results from experimental tests were obtained.

Abstract

Every windowed room has a view, which reflects the visibility of nature and landscape and has a strong influence on the health, living satisfaction, and housing value of inhabitants. Thus, automatic accurate window view assessment is vital in examining neighborhood landscape and optimizing the social and physical settings for sustainable urban development.

[1] Maosu LI, PhD Candidate.
Department of Urban Planning and Design, The University of Hong Kong, Pokfulam, Hong Kong SAR, China; Email: maosulee@connect.hku.hk, : https://orcid.org/0000-0002-9970-4053

[2] Fan XUE, Assistant Professor.
Department of Real Estate and Construction, The University of Hong Kong, Pokfulam, Hong Kong SAR, China; Email: xuef@hku.hk, : http://orcid.org/0000-0003-2217-3693
*: Corresponding author, Tel: +852 3917 4174, Fax: +852 2559 9457; Email: xuef@hku.hk

[3] Yijie WU, PhD Student.
Department of Real Estate and Construction, The University of Hong Kong, Pokfulam, Hong Kong SAR, China; Email: yijiewu@connect.hku.hk, : https://orcid.org/0000-0003-1441-1583

[4] Anthony G. O. YEH, Chair Professor.
Department of Urban Planning and Design, The University of Hong Kong, Pokfulam, Hong Kong SAR, China; Email: hdxugoy@hku.hk, : http://orcid.org/0000-0002-0587-0588

However, existing methods are labor-intensive, inaccurate, and non-scalable to assess window views in high-rise, high-density cities. This study aims to assess Window View Indices (WVIs) quantitatively and automatically by using a photo-realistic City Information Model (CIM). First, we define four WVIs to represent the outside (i) greenery, (ii) water-body, (iii) sky, and (iv) construction views quantitatively. Then, we proposed a deep transfer learning method to estimate the WVIs for the window views captured in the CIM. Preliminary experimental tests in Wan Chai District, Hong Kong confirmed that our method was highly satisfactory ($R^2 > 0.95$) and fast (3.08 s per view), and the WVIs were accurate (RMSE < 0.042). The proposed approach can be used in computing city-scale window views for landscape management, sustainable urban planning and design, and real estate valuation.

Keywords: Window view; View quality index; High-rise buildings; City information model; Deep learning; Urban computing.

1 Introduction

High-quality views can promote the physical and mental health, satisfaction, restoration, and productivity of inhabitants as shown by studies on psychology, physiology, and urban health (Ulrich 1984; Lottrup et al. 2015; Waczynska et al. 2021). In general, high-quality views often involve considerable proportions of natural features, such as greenery, sky, and water body, which are preferred by people (Hellinga 2013). Although the world population is migrating to cities (UNPD 2014), urban planners and citizens find it challenging to optimize the visibility of nature and landscape for windows in cities, especially in high-rise, high-density areas. The Covid-19 pandemic recently has limited people's physical access to nature in many places, further amplifying the benefits of high-quality window views. Consequently, high-quality window views, as a scarce resource, have been found to have considerable influence on real estate values and sustainable urban development in terms of neighborhood satisfaction, psychological and physical well-being, and urban planning and design (Benson et al. 1998; Bishop et al. 2004; Jim & Chen 2009; Baranzini & Schaerer 2011).

Researchers have developed a plethora of urban indices and computational methods to assess various urban views from different dimensions and perspectives. For example, on the global scale, satellite images can produce overhead view indices (Tucker 1979; McFeeters 1996), such as the Normalized Difference Vegetation Index (NDVI) for vegetation (Liu et al.

2016) and the Normalized Difference Water Index for blue space exposure (Helbich et al. 2019). At the city and neighborhood scale, photographs and videos taken by vehicle-borne cameras can assess the street views (Li et al. 2015; Shen et al. 2017; Dong et al. 2018; Lu 2018). These quantified view indices considerably contribute to human-built-environment studies, such as urban depression symptoms (Helbich et al. 2019). However, cities, especially high-rise, high-density ones, are not flat, so overhead and street-view assessment methods cannot correctly represent window views (Li et al. 2015). High-rise, high-density areas, like Hong Kong urban areas, have characteristics of high-rise buildings, narrow compacted street canyons, high-level plot ratios, and high building densities (Gong et al. 2018). Within this kind of context, the view from the window of a 30/F of an apartment may be completely different from that of a 3/F one.

Window view quality is receiving increased attention from researchers in the fields of architecture, urban health, and property valuation. For example, an ideal architectural design tends to assess the indoor design and the outdoor views holistically (Ko et al. 2021; Li & Samuelson 2020). Urban health researchers often use survey, interview, or questionnaire methods to classify the window view qualitatively (Lottrup et al. 2015; Masoudinejad & Hartig 2020). Qualitative descriptions of housing quality such as "with deluxe sea view" and "with hill view" have been popular in the housing and hostel market of urban areas such as Hong Kong (Jim & Chen 2009), and Mediterranean coastal cities (Fleischer 2012). However, the existing methods are challenging for the assessment of window views at the city level. First, too many window views exist in a city to be represented and preprocessed (Li et al. 2015). Second, conventional methods are too laborious to assess millions of window views for a city, and manual assessments are prone to various errors, such as preconceived notions in questionnaires and subjective judgments in valuation (Helbich et al. 2019). Thus, an automatic accurate assessment of window views can contribute to large-scale landscape and urban studies, as well as related disciplines and industries, for billions of urban inhabitants. The quantified views serving as a vertical view information hub can facilitate developers, urban planners, and other decision-makers to make well-informed decisions in real estate valuation, sustainable urban planning, e.g., green space planning for prioritized buildings, and new town design for balanced natural view acquisition and high-quality landscape view conservation, especially in high-rise, high-density cities.

This study aims to present a series of Window View Indices (WVIs) together with an automatic assessment method based on City Information Models (CIMs) and deep transfer learning. A CIM is a digital representation of the physical and functional characteristics of a city, and it can serve as a shared-knowledge resource (Song et al. 2017; Xue et al. 2021). With advanced remote sensing technologies, photo-realistic 3D CIMs become increasingly accurate in geometry and affordable in price. Recently, researchers have applied virtual cameras to CIMs to generate realistic images of 3D window views as needed (Li & Samuelson 2020; Li et al. 2020). The proposed method in the presented study extends the existing work with deep transfer learning to quantify massive quantities of view images.

The main contributions of this study are thus twofold:

i. From a theoretical perspective, the WVIs and assessment methods in this study extend the knowledge on computing window views in cities, especially in high-rise, high-density areas. The WVIs complement the existing studies on overhead and street-level urban views.

ii. For urban planning and design, the assessment results of this study are automatic and accurate for any window (or 3D viewport), thanks to the up-to-date CIM and deep transfer learning model pre-trained on other urban datasets. The output WVIs can facilitate planners, architects, and other decision-makers in optimizing the neighborhood landscape, urban planning and design, and property valuation for sustainable urban development.

The remainder of this study is organized as follows. The related work in literature is reviewed in Section 2. The WVI definitions and the automatic assessment method are presented in Section 3. Section 4 describes preliminary experiments and the results. The discussion and conclusion are presented in Sections 5 and 6, respectively.

2 Literature review

2.1 Urban views

Numerous studies have been conducted to compute and analyze urban views, e.g., impacts on human response (Roe et al. 2013) and economic development (Bishop et al. 2004; Jim & Chen 2009). Examples are green, water, sky, and construction views. Such views are not only of the interest in landscape and urban planning, but also attracting researchers in other disciplines such as psychology, physiology, urban health, and real estate.

First, greenery is of great significance to urban dwellers' psychological and physical health. Classical theories such as the stress reduction theory (Ulrich 1983) and attentional restorative theory (Kaplan S. 1995) have already shown this. For instance, green views can reportedly heighten positive effects such as performance and vitality (Van den Berg et al. 2016), reduce fears (Ulrich 1984), and block stressful thoughts (Roe et al. 2013). Other studies have shown that people's accessibility to greenery can increase their restorative potential (Pazhouhanfar & Kamal 2014) and thus influence the recovery from surgery (Ulrich 1984), and promote productivity and job satisfaction (Kaplan R. 2001; Lottrup et al. 2015). The green view impacts have been more extensively related to topics such as mental fatigue, depression (Helbich et al. 2019), and potential for violence and crime (Kuo & Sullivan 2001).

Water and sky views as blue elements enable housing to enhance human healthcare and property value. High-quality water bodies benefit people by having better aesthetic enjoyment and restorative potential (White et al. 2010), whereas viewing the sky offers occupants the sight and feeling of openness and spaciousness (Kaya & Erkip 2001). They found that exposures to water and sky, similar to green views, benefit health and well-being, such as stress reduction (Ulrich 1981), increased physical activity (Gascon et al. 2017), high restorative potential (Masoudinejad & Hartig 2020), and promotion of positive mood and satisfaction (Kaplan R. 2001; Gascon et al. 2017). Meanwhile, as precious attributes of the aesthetic landscape, water and sky views are of great value, especially in high-rise, high-density areas. As a result, both are influential on the property price (Baranzini & Schaerer 2011; Fleischer 2012).

For construction views from buildings, streets, and roads, their aesthetics and scarcity affect the preferences of humans. For instance, features such as constructed landmarks are desirable in window views (Baranzini & Schaerer 2011; Damigos & Anyfantis 2011). By contrast, studies also demonstrated that urban views with natural features are preferred by occupants over plain and dull construction scenes (Ulrich 1981; Grinde & Patil 2009). In summary, the four types of view features are worthy of assessment for windows in high-rise, high-density areas.

2.2 Assessments of window views

Generally, window-view quality can be assessed by two methods, namely, subjective and objective. First, numerous studies have utilized mostly a window view assessment according to the participants' subjective judgments on views (Lottrup et al. 2015; Li & Samuelson 2020; Masoudinejad & Hartig 2020). The window views are presented by physical forms, such as photographs and virtual forms (e.g., virtual reality). Researchers and practitioners first collect window views according to their research objects. Then, participants assess or rank the window views by using interview forms and questionnaire tables. The assessment results are not concrete owing to fuzzy scales and criteria. The assessment methods on the participants' subjective answers are also time-consuming (Helbich et al. 2019; Labib et al. 2021). Thus, subjective methods are limited to a small scale and cannot practically form common standards to coalesce the window view information objectively and automatically.

Objective methods and indices have emerged in the last decade for quantifying vertical views. An example is a simulation-based view index harnessing the power of techniques in the Geographic Information System, Remote Sensing, and 3D modeling (Yu et al. 2016; Labib et al. 2021). A traditional method, namely 3D visibility analysis, has been used to examine neighborhood amenities at the site and ground levels (Turan et al. 2019; Labib et al. 2021). Particularly, Yu et al.'s (2016) method measures floor-level greenery view based on the NDVI metric in a high-rise, high-density context, though the oversimplified 2.5D greenery can lead to errors. Alternatively, view photography method can effectively compute and analyze the real profile view of landscapes (Li et al. 2015; Shen et al. 2017; Dong et al. 2018). Recently, the method has also been used in Li et al.'s (2020) two-class window view classification, i.e., “nature” and “construction,” based on *Apriori* rules and a transfer learning model. However, Li et al.'s (2020) method relies on rigid classification rules and has only two types of features. Thus, next-generation objective assessment methods should be able to adapt to more urban scenes, with up-to-date machine learning (ML) technologies.

2.3 Deep learning and applications in urban studies

Deep learning is a group of multi-layer artificial neural networks involving multiple levels of representation learning (LeCun et al. 2015). Deep learning models have shown strengths in general pattern recognition tasks (LeCun et al. 2015). For instance, SegNet as one of the best deep convolutional network models has been used in visible landscape segmentation and quantification tasks (Liang et al. 2017; Shen et al. 2017). To study the relationship of natural features including greenery and water with geriatric depression in Beijing, China, a fully convolutional neural network (FCN-8s) was used to segment the street view images into green parts and blue parts (Helbich et al. 2019).

Deep transfer learning adopts a pre-trained network, inductively or transductively, from a source domain to the target domain on the basis of a mapping mechanism (Pan & Yang 2009). A small training dataset in the target domain can effectively map the variables' relationships and transfer the pre-trained network. Deep transfer learning has become prevalent for saving the time and resource costs in labeling training data with negligible performance downgrades from the original model. Thus, it is widely used in the semantic understanding of urban research, such as environmental management (Chen et al., "Looking beneath the surface": A visual-physical feature hybrid approach for unattended gauging of construction waste composition 2021; 2022), urban morphology (Middel et al. 2019), and perception (Yao et al. 2019; Li et al. 2020). For instance, fed by the street view images, FCN-8s pre-trained on the ADE20K dataset was transferred in water and greenery extraction of streetscape (Helbich et al. 2019). All previous studies have confirmed that deep transfer learning can be a versatile and inexpensive instrument from one domain to a similar domain application. Thus, for large-scale window view quality assessment and applications, deep transfer learning can provide cost-effective support for the semantic segmentation of the view.

In summary, large-scale window view assessment, especially the automatic method, has previously been a conundrum owing to the poor availability of window data and immature window view reconstruction and processing. Meanwhile, textured CIMs, deep transfer learning, and other learning technologies may open a window of opportunity to improve the automatic window view assessment for high-rise, high-density areas significantly.

3 Research methods

Figure 1 shows the proposed method as an Icam DEFinition for Function modeling (IDEF0) diagram, which is a public-domain flowchart-like methodology for modeling processes and functions (Colquhoun et al. 1993). The legend in Figure 1 explains the inputs, methods and tools, control parameters, and final outputs of each sub-process. The proposed automatic window view assessment method comprises three steps: (i) batch generation, (ii) semantic segmentation of pixels, and (iii) estimation of view indices. Each step employs a specific method and control parameters. Overall, the main inputs are a 3D photo-realistic CIM and corresponding 2D building footprints in this study. The output is a set of quantified WVIs. Finally, post-processing enriches the input CIM with the WVIs for smart decision-making for landscape and urban planning and related disciplines. A practitioner can follow the same methods and tools in Figure 1 and adjust the control parameters for specific application scenarios.

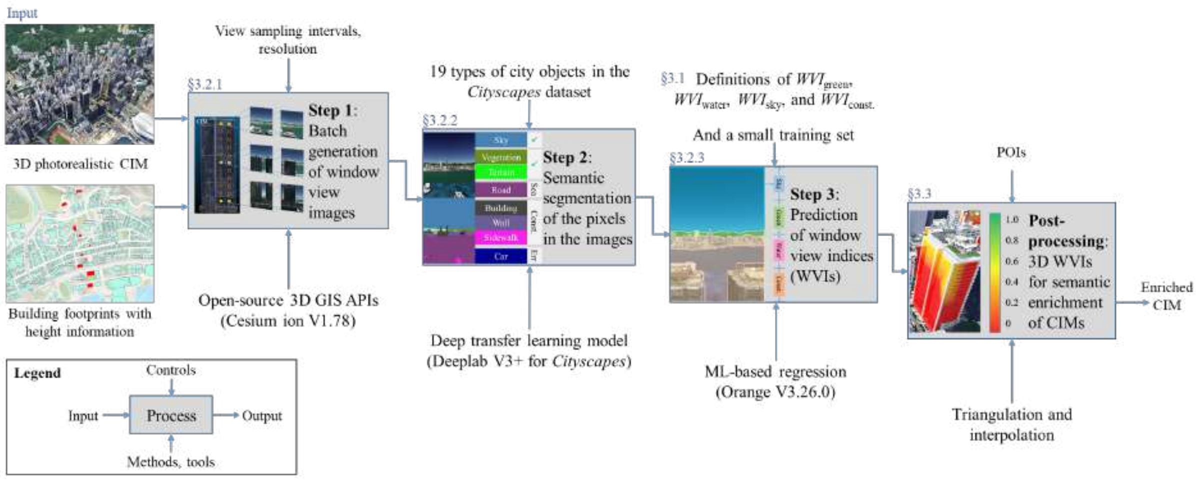

Figure 1. IDEF0 (Colquhoun et al. 1993) diagram of the proposed method for assessing window views.

3.1 Definitions of WVIs

3.1.1 Window view index

This study defines the WVI as the ratio of pixels for each view type. Given a view image $v = \{ p_{ij} \mid 1 \leq i \leq M, 1 \leq j \leq N \}$ of $M \times N$ pixels and a finite set L of views, as shown in Figure 2, the WVI in an input window view image is the ratio:

$$WVI_l = \frac{|\{p \mid p \in v, \lambda(p) = l\}|}{M \times N}, \quad l \in L, \qquad (1)$$

where $\lambda(p) = l$ is the semantic label of a pixel p, e.g., “green” or “waterbody”, and $| \cdot |$ is the cardinality operator indicating the total number of pixels. Thus, all WVIs are scalars bounded between 0 and 1:

$$WVI_l \in [0, 1], \quad l \in L. \qquad (2)$$

We select the four major types of window views as summarized in Section 2.1. That is, L = {‘green’, ‘waterbody (water)’, ‘sky’, ‘construction (const.)’}, as shown in Figure 2. Table 1 lists the common city objects’ mapping to L. For instance, the “green” view type covers all kinds of greenery, including trees, bushes, and grasses. Four symbols, namely, WVI_{green}, WVI_{water}, WVI_{sky}, and $WVI_{const.}$, represent the scalar values, respectively. Despite the presence of other possible city objects such as pedestrians, pets, vehicles, and aircraft, the four major types are dominant in window views in our experiment, i.e., $WVI_{green} + WVI_{water} + WVI_{sky} + WVI_{const.} \approx 1$, as shown in Figure 2. Furthermore, the ratio-based definition is consistent and robust for the comparison of views from different window sizes across districts and cities, which is helpful for the proof-of-concept purpose in this study, e.g., a window with $WVI_{green} = 0.8$ owns more proportions of greenery and can thus be regarded as a totally green-view window, compared with another having $WVI_{green} = 0.4$.

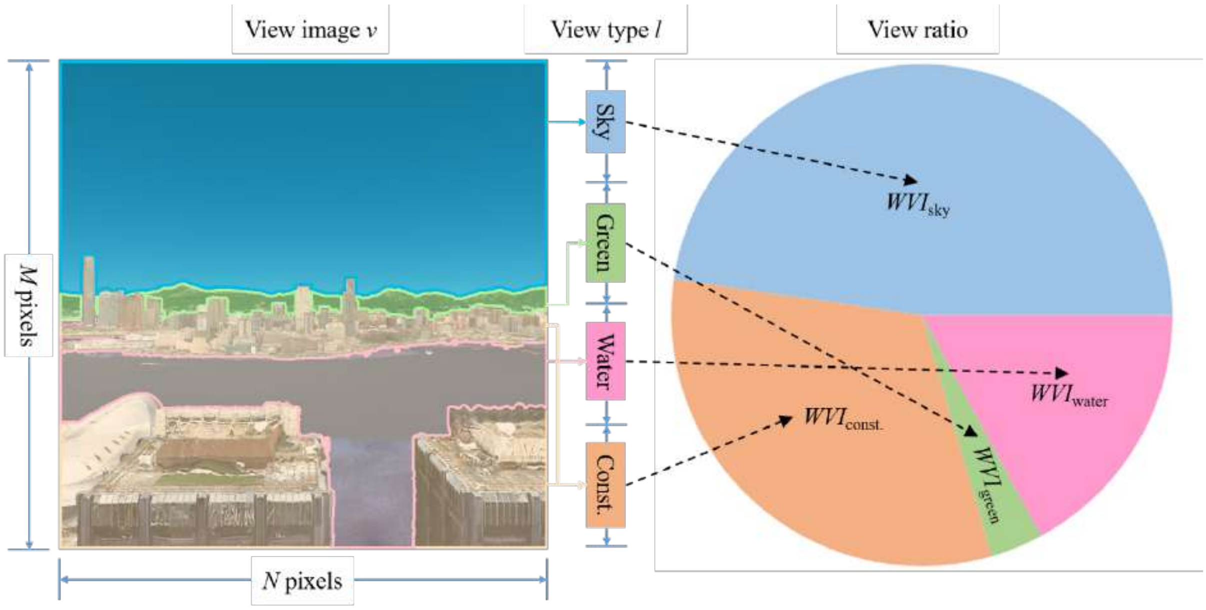

Figure 2. Examples of the four Window View Indices (WVIs).

Table 1. List of types of views and associated common city objects.

Type	Symbol	Example objects
Green	WVI_{green}	Trees, bushes, and grasses
Waterbody	WVI_{water}	Sea, lakes, ponds, and rivers
Sky	WVI_{sky}	Sky, clouds, and fog
Construction	$WVI_{const.}$	Building facades, roofs, walls, streets, houses, and roads

3.1.2 Window view ranking

Furthermore, the relative window view ranking (WVR) of a window's WVI within a high-rise, high-density area A can be defined as the percentile to the maximum WVI of the context:

$$WVR_l^A = \frac{WVI_l}{\max(WVI_l^A)} = \begin{cases} \text{Very high, } WVR_l^A \in [0.8, 1.0] \\ \text{High, } WVR_l^A \in [0.6, 0.8) \\ \text{Average, } WVR_l^A \in [0.4, 0.6) \\ \text{Low, } WVR_l^A \in [0.2, 0.4) \\ \text{Very low, } WVR_l^A \in [0.0, 0.2) \end{cases}, l \in L. \quad (3)$$

Therefore, the WVR classifies all the windows in an area into five isometric groups. WVR can resolve the issue of inconsistent upper bounds of different WVIs, which enables an inter-view-type comparison. For instance, although max(WVI_{sky}) is roughly 0.5 and max(WVI_{green}) is 1.0 theoretically, max(WVR_{sky}) and max(WVR_{green}) can still reach 1.0. Thus, one window with absolute WVI_{sky} = 0.5 and WVI_{green} = 0.5 can be tagged as a "very high"-level sky view but an "average"-level green view within the context. People's decision-making is expected to be associated more with WVRs than WVIs, e.g., in property valuation.

3.2 Proposed assessment method

3.2.1 Batch generation of window view images

The first step aims to generate the window view images in an urban area in batch. The image extraction process, as shown in Figure 3, is automatic on 3D GIS platforms with camera functions, such as Cesium (Cesium GS 2022). Figure 3a shows a window's 3D geolocation (*lng*, *lat*, *height*) and *heading* direction are computed on the facade of extruded footprints by building height information, where the *heading* direction is assumed perpendicular to the facade at (*lng*, *lat*). The field of view is set to 60° to represent the normal human field of vision (FoV) (Tara et al. 2021), while the pose of the virtual camera is set on the window with *tilt* = 0 and *pitch* = 0 to capture views. The image extraction extends Li et al. (2020) as the camera's view of the photo-realistic CIM's textured appearance. The difference from Li et al. (2020) is the full automation for massive windows using a JavaScript program as shown in Figure 3b.

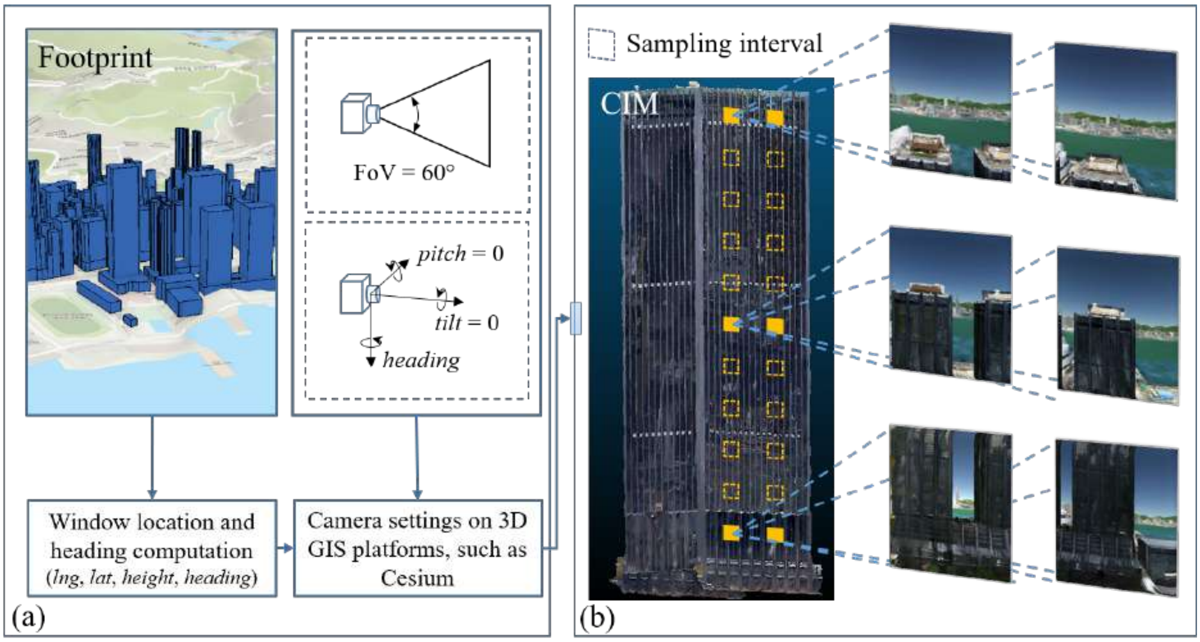

Figure 3. Batch generation of window view images. (a) Window location computation, camera settings, and (b) image generation process.

However, neighboring windows on the same facade often share similar views. Thus, sampling the facade with certain intervals, e.g., every 10 or 20 m, is a cost-effective method, as shown in Figure 3b, which can considerably save computational effort without losing notable WVI accuracy. Based on the efficient sampling and GIS-based view visualization, the batch generation can extract view images for the windows of a high-rise, high-density area. Learned from experiments and sensitivity analysis results in Section 4, we used 20 and 5 m to obtain a location matrix of view sites within the large and small facades, respectively.

3.2.2 Deep transfer learning-based semantic segmentation

This step classifies every pixel in an input image to a semantic view label through deep transfer learning. One of the most relevant deep learning datasets is the *Cityscapes* benchmarking dataset (Cordts et al. 2016), which comprises 25,000 urban views annotated as 19 pixel-level labels from 50 cities in Germany. According to the study of Pan & Yang (2009), the models trained in Germany can potentially be transferred to other areas like Hong Kong. Table 2 lists the labels for *Cityscapes* in seven groups. Apparently, three types of views, i.e., green, sky, and construction, can be directly mapped from *Cityscapes*' definitions.

Table 2. List of labels for the *Cityscapes* dataset and for WVIs in this study.

Group	Labels for *Cityscapes*	Labels for WVIs
Nature	Vegetation, Terrain [a]	Green
Sky	Sky	Sky
Construction	Building, Wall, Fence	Const.
Paved	Road, Sidewalk	Const.
Object	Pole, Traffic sign, Traffic light	Const.
Human	Rider, Person	– [b]
Vehicle	Car, Truck, Bus, Motorcycle, Bicycle, On rails	– [b]

a: Including all kinds of horizontal vegetation in Cityscapes; *b*: Negligible in this study.

A Deeplab (Ver. 3+ with the Xception_65 backbone) model pre-trained on *Cityscapes* (Chen et al. 2018; Xia et al. 2021) is transductively transferred to the segmentation of captured window view images to the labels in Table 2. The off-the-shelf Deeplab model is one of the top open-source deep learning models for urban views, where the training parameters can be referred to (Chollet 2017) and (Chen et al. 2018). Xue et al. (2021) showed that transductive transferring Deeplab leads to an efficient and low-cost semantic segmentation of view images, even though the training and target datasets are from different contexts. As shown in

Figure 4, the incorporated version of Deeplab has a network architecture consisting of two parts, i.e., an encoder and a decoder (Chen et al. 2018). The encoder mainly includes an Atrous Spatial Pyramid Pooling Module (ASPP) for concatenated features from a low-level Atrous convolution (Chollet 2017), while the decoder concatenates the ASPP outputs and low-level features with convolution and upsampling.

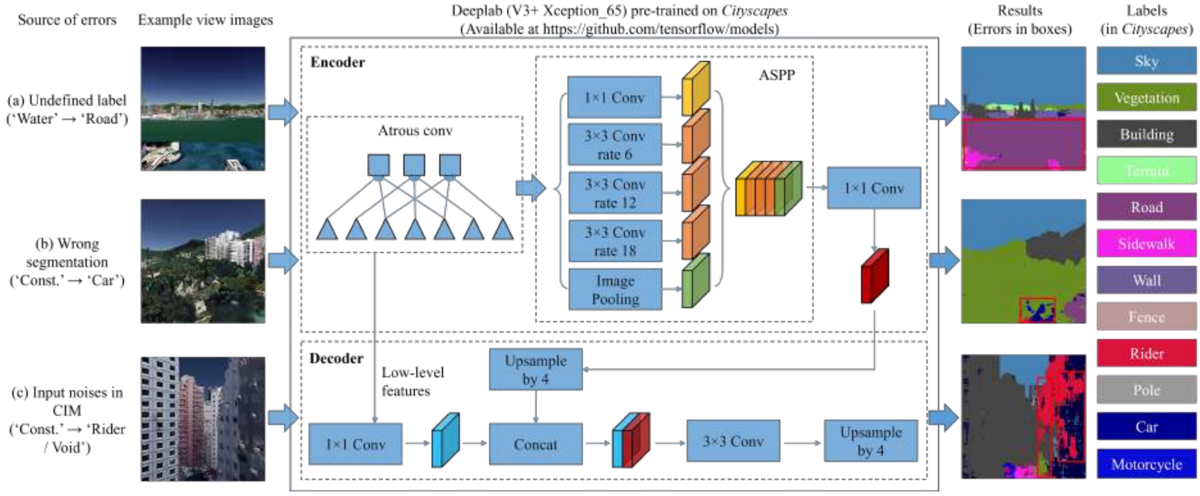

Figure 4. Three types of semantic segmentation errors in direct transductive transferring of Deeplab. (a) Undefined labels, (b) segmentation errors, and (c) from input noises.

However, as shown in Figure 4, the segmentation results of a direct transductive transferring were erroneous and unsatisfactory for WVIs in the study area. The primary source of errors was from the inconsistent labels, e.g., the water body, between the training dataset *Cityscapes* and our view images. Besides, minor errors resulted from the segmentation and input noises. Therefore, deep transfer learning can deliver pixel-level semantic segmentation with relevant labels for window view images, but the results must be corrected for the errors to improve the accuracy in computing WVIs and WVRs using the ML-based WVI regression layer described in Section 3.2.3 below.

3.2.3 ML-based regression for WVIs

This step applies an ML-based WVI regression, as shown in Figure 5, to correct the errors from deep transfer learning for computing WVIs. The input features to the regression are *Cityscapes* labels in terms of proportions of pixels segmented by Deeplab in Figure 4. The outputs are the four WVIs, i.e., WVI_{green}, WVI_{water}, WVI_{sky}, and WVI_{const}. We annotate a small set of window view images with five labels, i.e., green, waterbody, sky, construction, and others (e.g., terrain and vehicles), which provide ground truth WVIs for the training process. The candidate ML models include Decision Trees, Linear Regression, Support Vector Machines (SVMs), *k*NN, Artificial Neural Network (ANN), Random Forest, and Adaboost. A standard train-compare-finetune pipeline is applied to select appropriate ML models to estimate the WVIs through cross-validations. For each type of WVI, the most accurate ML model (together with its parameters) is selected for the regression layer. As a result, the first two types of errors shown in Figure 4 can be considerably reduced.

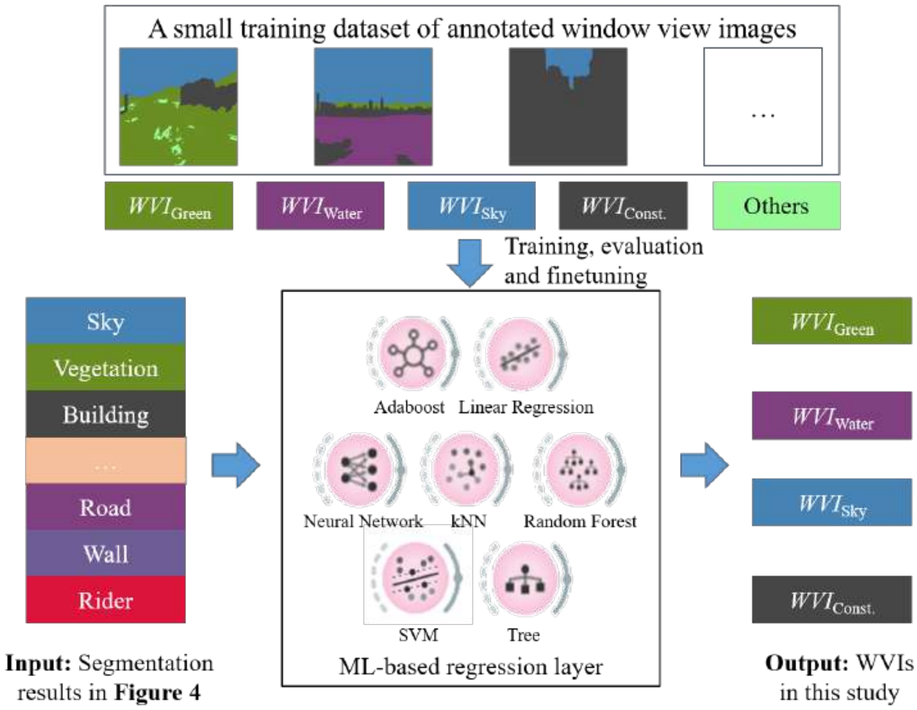

Figure 5. ML-based regression layer for estimating WVIs.

The results of ML training are compared with the actual values of the four view types from view image annotation using root-mean-squared error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{l \in L}(Pred_l - WVI_l)^2}{n}}, \quad (4)$$

where WVI_l indicates the actual value for the view type l, $Pred_l$ is the estimated value, and n denotes the number of window view images. The ML model trained with the minimum RMSE is selected for WVI estimation. We utilize 10-fold cross-validation for unbiased RMSEs.

3.3 Post-processing for semantic enrichment of CIM

The estimated WVIs are post-processed to enrich the semantics of 3D CIM, which can conveniently support future applications in related domains as a common knowledge platform. The detailed workflow is as follows. First, geocoded view sites with WVIs are registered at the 3D globe. Then, regarding view sites within the same facade as a group of vertices, we triangulate them to reconstruct the building facade through a classic Delaunay method (Lee & Schachter 1980). Thereafter, a linear interpolation-based 3D rendering (Akenine-Möller et al. 2019) of WVIs visualizes the whole building facades in a mesh surface. The interpolation result also estimates the WVIs of all locations of the building facades. Finally, the CIM is enriched with the WVI semantics for a spectrum of applications in landscape management, sustainable urban planning and design, and real estate valuation.

4 Experimental tests

4.1 Experimental area and settings

The study area was Wan Chai in Hong Kong, as shown in Figure 6a. Wan Chai is one of the highest residential density zones according to the Hong Kong Planning Standards and Guidelines (HKPlanD 2018). The average of building heights is 35.5 m and the 75th percentile is over 48 m. The study area owns a plot ratio at 8.0 and building density at 0.29 (the ratio of building site area to land area). Although the area enjoys considerable sky, sea, and greenery view contents, the visibility of natural features is often blocked by other buildings. The 2D footprint data with building height information were extracted from the iB1000 digital topographic map of Hong Kong (HKLandsD 2014) as shown in Figure 6b, and converted into the GeoJSON (Butler et al. 2016) format for batch attribute computations of view sites' locations and headings. The data source of 3D photorealistic CIM was produced and freely shared by the Planning Department of Hong Kong (2019) as shown in Figure 6c. We calibrated the CIM as 3D tiles to the correct geographical locations on the WGS-84 globe. Then, 2D and 3D datasets were loaded and registered in an open-source 3D GIS platform named "Cesium ion." Ten buildings with typical different built environments from the seaside to the mountain area were selected as case studies to examine the proposed approach, as shown in Figure 6d.

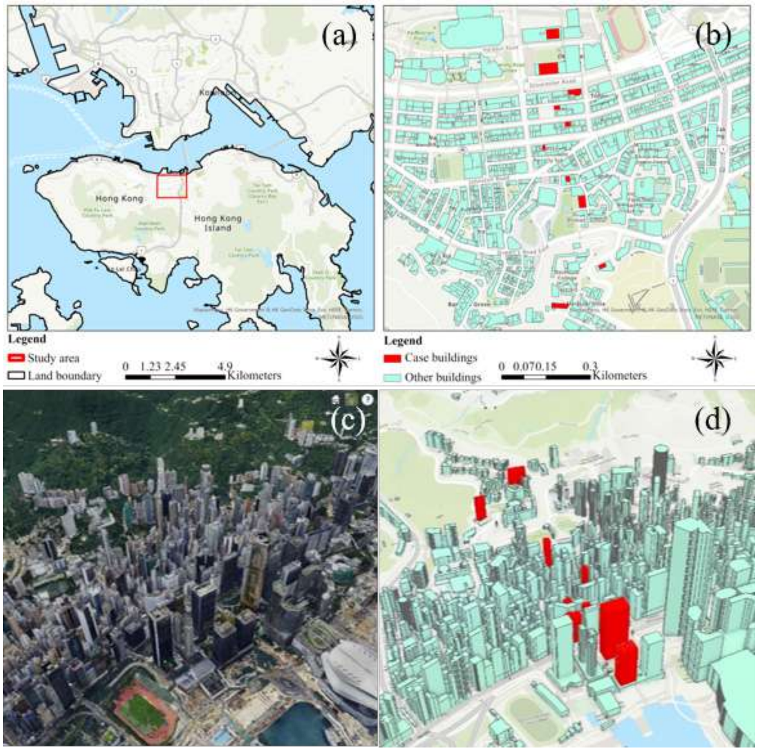

Figure 6. Study area of Wan Chai, Hong Kong. (a) Location of Wan Chai, (b) building footprints, (c) input CIM, and (d) location of 10 case study buildings.

The computational experiments were set up as follows. The workstation comprised an Intel i7-10700 CPU (2.90GHz, 16 cores), 128 GB memory, one Nvidia GeForce RTX 2070 graphic card, and Ubuntu 20.04 (64-bit) operating system. Sample window views were collected on the Cesium platform (ver. 1.75). Deep transfer learning was in the environment of Tensorflow (ver. 2.4) and Python (ver. 3.6). We adopted the seven ML models implemented on Orange (ver. 3.26), a Python ML platform. From the case study buildings, 110 training examples were selected for unbiased representation of diversified window views and manually annotated with the WVIs for training the ML models. The one-off annotation work consumed about 10 person-hours. The size of training examples satisfied the requirements of deep transfer learning. We set each view image with 900 × 900 pixels to represent the view features seen from the window.

4.2 Results

Results showed that the proposed method is automatic and efficient, as shown in Table 3. The first step of batch generation returned 1,416 window view images from the 10 selected buildings for the case study. The average time for generating one view image was 2.00 s. The deep transfer learning processed the view images at an average time of 1.08 s in the second step. The ML-based regression estimated the WVIs in <0.001 s on average for each image.

Table 3. Computational time of the proposed method for a window view image.

Step	Processing	Software library	Average time (s)
1	CIM-based batch generation	Cesium (ver. 1.75)	2.00[#]
2	Deep transfer learning	Deeplab (ver. 3+)	1.08
3	ML-based regression	Orange (ver. 3.26)	0.00[*]
		Total	3.08

#: A pre-set value that can be fine-tuned by workstation performance; *: Less than 0.001 s.

The WVIs' assessments of the proposed method were also satisfactory. Table 4 shows that for the best model of the four view indices' estimation, the R^2 values were 0.952, 0.965, 0.978, and 0.977 respectively, which represented more than 95% of the variance in the dependent variables. The RMSEs of the four training models were 0.021, 0.022, 0.025, and 0.042, respectively. The optimal parameter of each best model was as follows. For WVI_{green}, the Linear Regression model was trained with Lasso (L1) regularization and strength at 0.0001. For WVI_{water}, the SVM model performed the best, with kernel = RBF, C = 0.9, gamma = 0.05. For WVI_{sky}, a Linear Regression model with an elastic net regularization (L1:L2=0.50:0.50) was utilized with the best accuracy of estimation, whereas for $WVI_{\text{const.}}$, the best estimation was observed from a Linear Regression model with a Ridge (L2) regularization (Alpha = 0.003).

Table 4. Training errors and time of the best model for four WVIs.

WVI	Best model	Parameters	RMSE	R^2	Training time (s)
Green	Linear Regression	L1 = 0.0001	0.021	0.952	0.077
Water	SVM	Kernel = RBF, C = 0.9, gamma = 0.05	0.022	0.965	0.154
Sky	Linear Regression	L1:L2 = 0.50:0.50	0.025	0.978	0.070
Const.	Linear Regression	L2 =0.003	0.042	0.977	0.091

WVRs were computed from the WVIs by the best model. Table 5 shows three typical window views and their WVIs and WVRs. In Table 5, a WVR is represented in an array of stars, showing the level from "very low" to "very high" in Eq. 3. The highest WVRs correctly reflected the given dominant features for all the samples.

Table 5. Sample WVIs and WVRs for typical sample window views.

View images							
Dominant feature		Sky		Green		Construction	
Feature	**Max.**	**WVI**	**WVR**	**WVI**	**WVR**	**WVI**	**WVR**
Green	0.5421	0.0165	★	0.4867	★★★★★	0.0130	★
Water	0.4375	0.3352	★★★★	0.0024	★	0.0000	★
Sky	0.5505	0.4682	★★★★★	0.3236	★★★	0.0928	★
Const.	1.0000	0.1870	★	0.1704	★	0.9057	★★★★★

4.3 Post-processing for enriching CIMs

In the post-processing, the estimated WVIs and WVRs were registered for enriching the semantics of input CIM. Figure 7 shows the 3D mesh model of the regional WVIs in the study area. Generally, most rooms of the buildings owned a high $WVI_{\text{const.}}$ in this area as shown in Figure 7d. Figure 7b shows that only windows facing the seaside in the high-rise buildings near the harbor can have high-level WVI_{water} values in Wan Chai. Great sky views were scattered across the rooms with the high storeys as shown in Figure 7c. Figure 7a shows the generally low and fluctuated $\text{WVI}_{\text{green}}$, reflecting the varied amount of the surrounding greenery at different locations. In summary, the disparity of possession of natural view resources, i.e., greenery, water, and sky, is significant in the study area. The quantified disparity can help the urban planners to make a more accurate and specific decision for future landscape management and urban planning, e.g., prioritized greenery planning for buildings without any nature views.

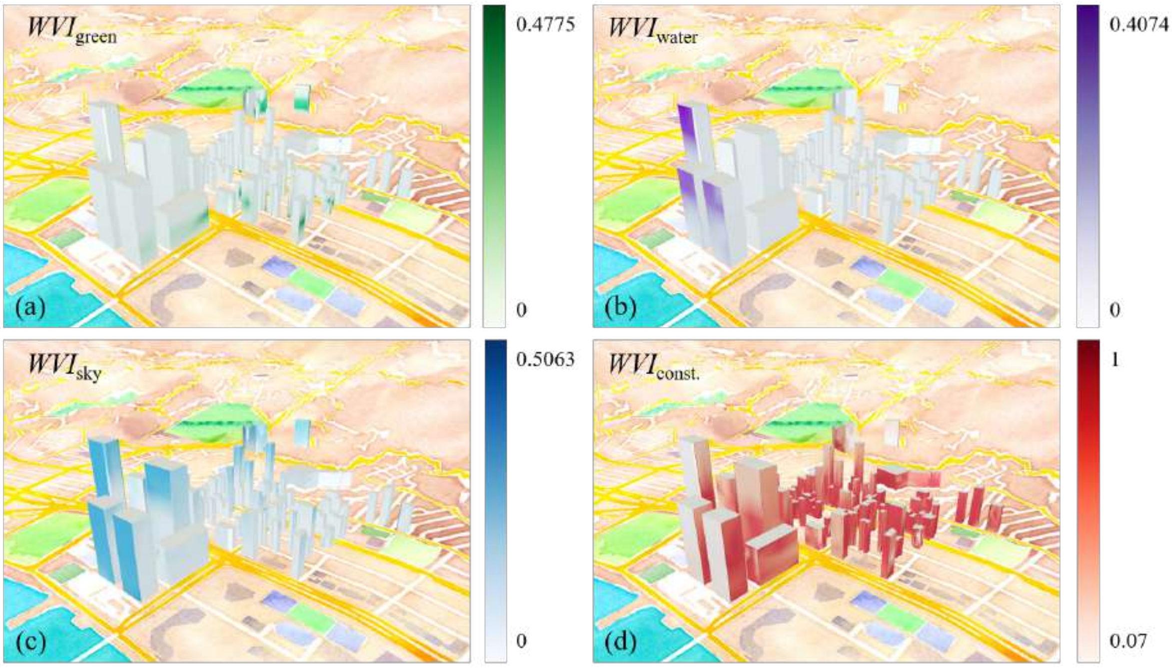

Figure 7. Regional patterns of WVIs. (a) WVI_{green}, (b) WVI_{water}, (c) WVI_{sky}, and (d) $WVI_{const.}$.

Figure 8 shows a WVR-enriched comparison of two example north-facing facades, one nearby and the other far away from the seafront, of which the locations are marked in Figure 8e. Holistically, water and sky views of the first facade were above the "average" levels in the study area (≥ 40%), as shown in Figures 8b and 8c; in contrast, the levels of those views of the second facade were consistently lower due to the inter-building obstruction. Figure 8a shows the green views were both at a "very low" level (< 20%) due to the less visible greenery. The construction view patterns of the two facades varied as shown in Figure 8d, where construction views dominated the second facade. In comparison with WVI values, the relativity in such WVR results is more convenient for certain applications such as real estate valuation, since the levelization of the window view such as "very high" and "very low" can intuitively inform developers and occupants of the room view quality within the local context.

Figure 8. WVR patterns of two example building facades. (a) WVR_{green}, (b) WVR_{water}, (c) WVR_{sky}, (d) $WVR_{const.}$, and (e) their general locations.

4.4 Sensitivity analysis

4.4.1 View sampling interval in Step 1

A trade-off existed between processing time cost and accuracy when applying the view sampling interval in Step 1. A sensitivity analysis was conducted to identify a cost-effective sampling plan. In the experiments, the case was a facade area (120 m × 60 m) of the China Resources Building, as shown in Figure 9a. The benchmark was set to the result of a 5 m sampling interval. We tested a range of sampling intervals from 10 m to 60 m in an approximately exponential increment. Figure 9b shows the example of WVI_{sky} estimation results resampled back to the 5 m scale through linear interpolation to compare the accuracies in terms of RMSE. We found that with increased sampling interval, the time consumption of the window view image processing from generation to estimation witnessed a sharp decline, whereas the RMSEs of four WVIs increased accordingly, as shown in Figure 9c. From the observation, the sample interval of 20 m can be a “sweet point,” in which an efficient and accurate estimation of WVI (RMSE<0.015) was obtained without excessive processing time. Thus, for the view image processing of case buildings, we used 20 m as the sampling interval for large facades. For a building facade whose length or width was less than 20 m, 5 m was used.

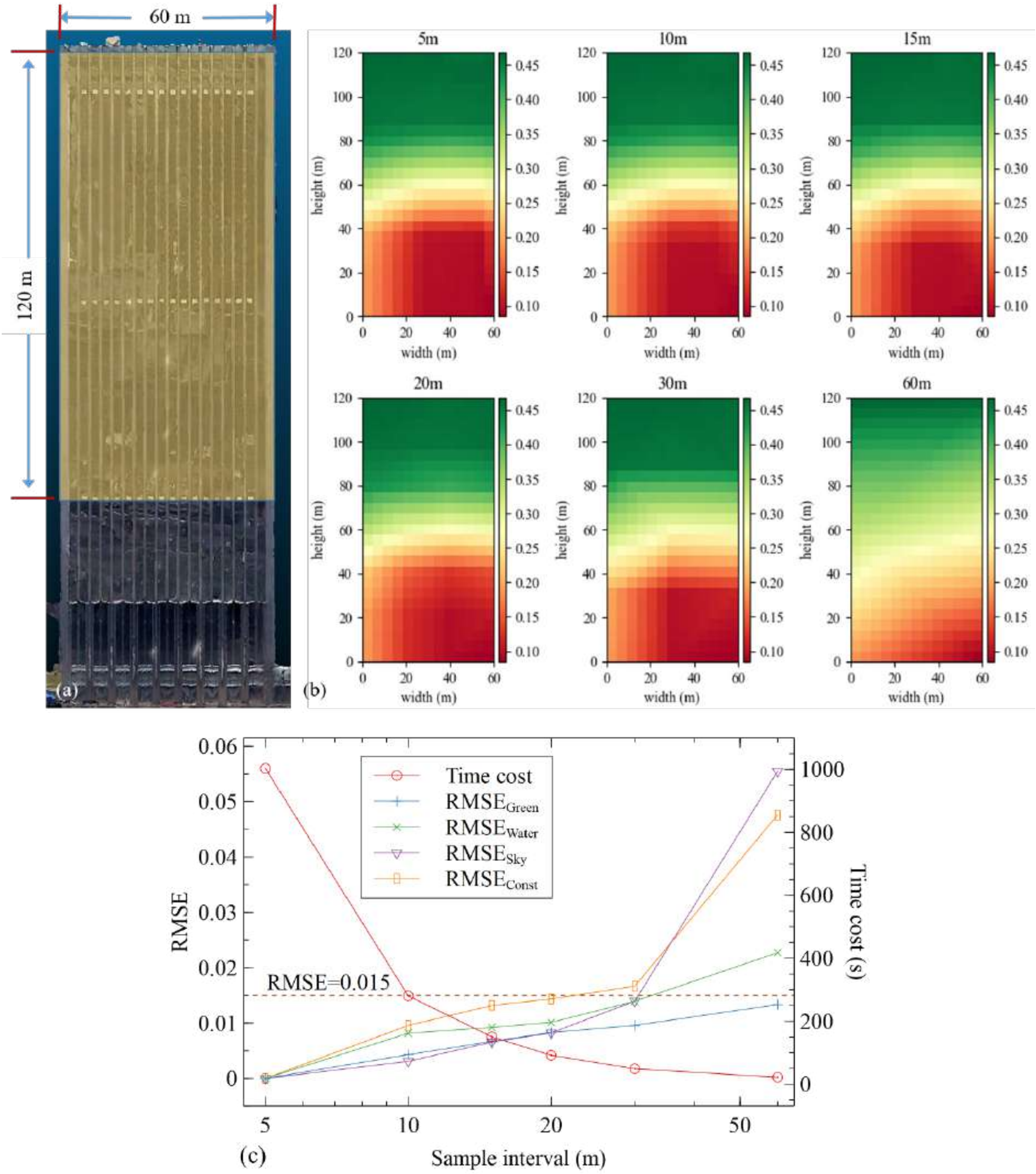

Figure 9. Sensitivity analysis of sampling intervals. (a) A case facade, (b) estimated WVI_{sky} at different sampling intervals, and (c) trade-off between time cost and four WVIs' accuracy.

4.4.2 Input CIM in Step 1

Figure 10 compares view segmentation results using two different CIMs. The appearances of the two 3D models were close but clearly distinguishable. First, the color contrast of Google Earth's CIM was softer than the model adopted in this study, and the low contrast resulted in the misclassification of constructions and greenery highlighted in Figure 10a. Second, the model fidelity also affected the stimulation effects. Figure 10b shows that some parts of the vegetation view (as highlighted in the rectangles), which were wrongly segmented using our CIM, can be corrected using Google Earth's model. This finding was due to the higher quality of Google Earth's in expressing the vegetation features, especially in close range. Lastly, the distortions in CIMs affected the segmentation accuracy. As shown in the red rectangles in Figure 10c, the blurred facades in the left column resulted in inaccurate segmentation, whereas the distortions in Google Earth's model led to the wrong detection of buildings to vegetation.

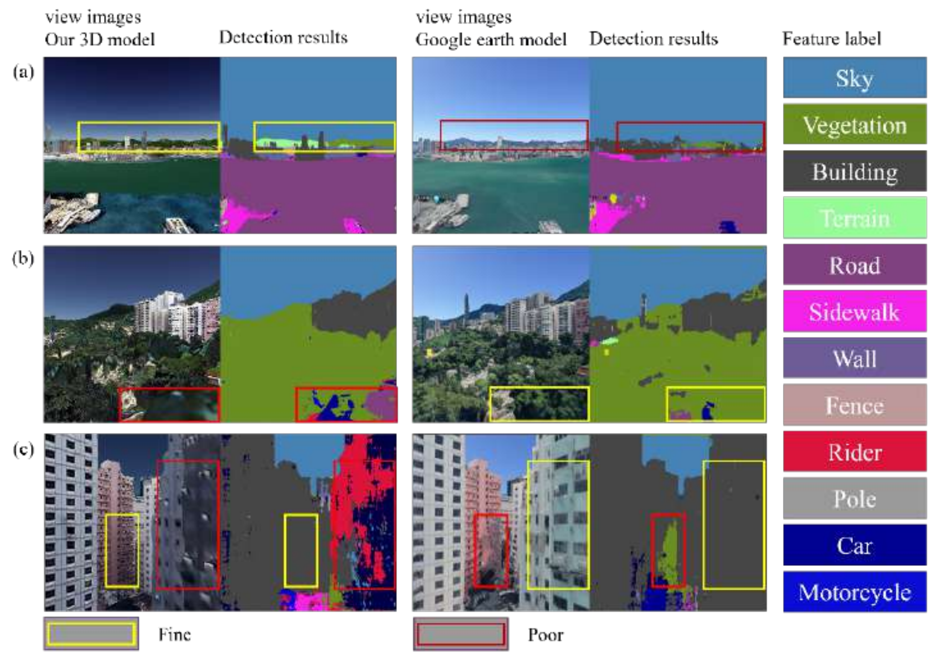

Figure 10. Comparison of window view image segmentation (Step 2) against different CIMs. (a) View color, (b) view fidelity, and (c) view distortions.

4.4.3 ML models for regression in Step 3

Based on the R^2, the performance of trained models is examined, and results are shown in Figure 11. For the estimation of the four WVIs, all ML models had R^2 values greater than 0.7. For three types of WVIs, i.e., WVI_{green}, WVI_{sky}, and $WVI_{const.}$, the best models were produced by Linear Regression. For the WVI_{water} estimation, the best model was SVM, whereas the Linear Regression returned $R^2 > 0.93$. The satisfactory results from Linear Regression might echo the assumption that four window view types could be mapped directly from the urban street view features in high-rise, high-density areas.

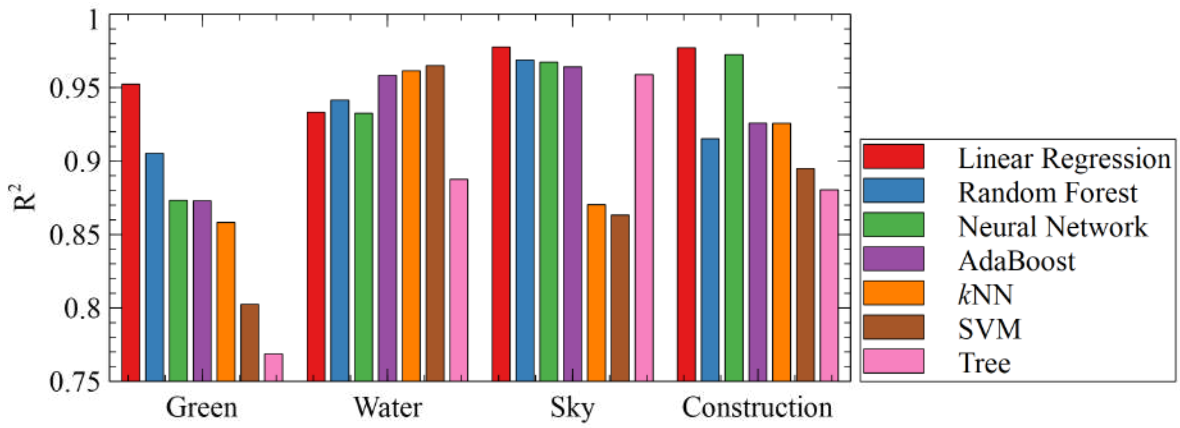

Figure 11. Comparison of R^2 performances of the seven ML models.

5 Discussion

5.1 Significance

Large-scale window view assessment has a great potential to support many smart city applications. The window view quality is of great significance for residents in high-rise, high-density areas. In the post-Covid-19 era, window view plays an important role in accessing nature as people have to stay longer in their houses or offices. The quantitative window view quality assessment at the city scale can provide an intuitive understanding of environmental inequality. Planners can use the results to prioritize improvements of the poor living environments, such as prioritized provision of more green space for neighborhoods with poor window views. And government sectors and policymakers can make the regulations, e.g., minimum acquisition of nature views in the future sustainable urban development. The results can also facilitate urban and architectural design by quantifying the window view quality at a relatively low cost. Designers can integrate the quantified view results for more comprehensive generative designs of building spaces (Laovisutthichai et al. 2021) and new towns. In addition, the method can serve as a new indicator for the housing market and thus has a great potential to the architecture, engineering, and construction development.

In the past, surveyors had to enter real rooms of buildings to capture the window views. Owing to this time-consuming, labor-intensive task, the window view dataset is always limited (Labib et al. 2021). Furthermore, accessing all window views manually at a large scale becomes impossible in terms of cost, labor force, and privacy (Helbich et al. 2019). Nowadays, with the advancement of remote sensing, photogrammetry, and digital twin technology, mature 3D CIMs with high-quality textured appearances are becoming increasingly available for detecting multiple groups of view features. CIM-based simulated window views for the real world have been validated effectively (Li & Samuelson 2020; Li et al. 2020). However, for an urban-scale window view quality evaluation, processing a large number of views manually remains laborious and expensive for surveyors. The proposed window view quality assessment method can free humans from repetitive and time-consuming tasks, and provide a set of quantifiable indicators to support fundamental and derivative applications in window view quality evaluation.

The proposed automatic assessment method can effectively generate four major view indices for quantifying and analyzing the urban-scale window views. First, this study makes full use of volumetric landscapes from 3D photo-realistic CIMs to further enrich the CIM with four WVIs, thereby enabling many window-view-based digital twin city applications, such as 3D city living environment assessment and housing scenic quality comparison. From a practitioners' point of view, the method is easy-to-use, low-cost, and accurate. For example, the automation process can be implemented without considerable prior knowledge. The pre-trained Deeplab model was shared freely. Based on the transfer learning theory, only a small dataset is required for a satisfactory WVI assessment. Moreover, the experimental results confirmed a high accuracy of assessing the window views ($R^2 > 0.95$). In summary, the proposed method contributes to window view assessment using CIM and AI, and also provides relatively low-cost and high-accuracy WVIs for applications in urban planning and design, and property valuation.

5.2 Limitations and future work

Nevertheless, a few limitations exist in the work presented in this study. First, the assessed window view quality in this study only involved limited contents, including greenery, sky, water body, and construction. Movable city objects e.g., pedestrian, car, and rare urban features e.g., bare soil surface were not involved. Other view elements exerting influence on indoor living satisfaction and outdoor environment perception such as aesthetic and environmental quality, view distance, and layer were not considered. Second, the horizontal view was set to compute the WVIs, which might miss visible features from other directions, e.g., the ground level. Next, another limitation was the high workload of 2D image segmentation involving repeated computation. For instance, similar view images from neighboring windows were independent without reusing the intermediate segmentations. The computation cost could be slightly higher for irregular buildings due to more view samples and processing. Last, the window sampling and interpolation also led to possible accuracy losses.

Future directions to improve the presented study are as follows. The first is extending the 2D image format of window views to incorporate high-dimensional factors (e.g., fine-scale classified view features, view distance that influences residents' feeling of spaciousness, and aesthetics and environmental quality attributes that influence living satisfaction) for holistic quality and optimization. More FoVs, such as 360-views, can extend the WVIs assessed in the 60° horizontal views in this study. Well-labelled CIM for landscapes is proven effective for large-scale view quantification (Yu et al. 2016). Thus, a 3D segmented CIM may eliminate the repetitive and redundant 2D image segmentation and save considerable costs of training and applying deep transfer learning, especially for irregular buildings. Another direction is to identify the accurate 3D location and orientation for each physical window in the CIM so that the assessed WVIs and WVRs can be associated with windows and rooms.

6 Conclusion

A high-quality window view with enough features such as greenery, sky, and water not only has a good impact on residents' health, well-being, and performance, but also can enrich the value of the house, especially in high-rise, high-density areas. Traditional window view assessment methods have common problems such as subjectivity, scalability, and efficiency. To address these limitations, this study uses an automatic method for the large-scale window view quality assessment through the use of CIM-based window view images of city buildings.

This study defines an indicator named Window View Index (WVI) including four sub-indices i.e. Green view index, Water view index, Sky view index, and Construction view index, which are measured at one time efficiently. By implementing a fast-sampling method, outside views are captured at each view site of the 3D CIM at the initial stage. Then, a pre-trained deep transfer learning model is used to classify view images into multiple features efficiently. To construct the regression between detected features and the WVI, seven traditional machine learning models are tuned to achieve the best performance. Our method achieved highly satisfactory results in estimating the WVIs for the high-rise, high-density area, in Wan Chai, Hong Kong. The RMSEs of estimation did not exceed 0.042, whereas the average time of processing each window was 3.08 s.

The proposed method provides intuitive indicators of the window view quality for high-rise, high-density areas. The automatic, accurate method is scalable to the urban scale, thereby enabling many window view-based applications in landscape management, sustainable urban planning and design, and real estate valuation, which would benefit residents' health, urban optimization, and the housing industry. Future work includes extending the view indices, 3D semantic segmentation of CIM, and mapping the WVIs to physical windows and rooms.

Acknowledgment

The work presented in this study was supported in part by the Hong Kong Research Grants Council (RGC) under the Early Career Scheme (ECS) No. 27200520 and the Guangdong–Hong Kong-Macau Joint Laboratory for Smart Cities of the 2020 Guangdong New Innovative Strategic Research Fund, Guangdong Science and Technology Department (Project No.: 2020B1212030009)

References

Akenine-Möller, T., Haines, E. & Hoffman, N. (2019). *Real-Time Rendering* (3rd ed.). New York: A K Peters/CRC Press. doi:10.1201/9781315365459

Baranzini, A. & Schaerer, C. (2011). A sight for sore eyes: Assessing the value of view and land use in the housing market. *Journal of Housing Economics, 20*(3), 191-199. doi:10.1016/j.jhe.2011.06.001

Benson, E. D., Hansen, J. L., Schwartz, A. L. & Smersh, G. T. (1998). Pricing residential amenities: the value of a view. *The Journal of Real Estate Finance and Economics, 16*(1), 55-73. doi:10.1023/A:1007785315925

Bishop, I. D., Lange, E. & Mahbubul, A. M. (2004). Estimation of the influence of view components on high-rise apartment pricing using a public survey and GIS modeling. *Environment and Planning B: Planning and Design, 31*(3), 439-452. doi:10.1068/b3042

Butler, H., Daly, M., Doyle, A., Gillies, S., Hagen, S. & Schaub, T. (2016). *The geojson format.* California, USA: Internet Engineering Task Force (IETF).

Cesium GS. (2022). *The Cesium Platform.* Philadelphia, USA: Cesium GS, Inc. Retrieved from https://cesium.com/platform/

Chen, J., Lu, W. & Xue, F. (2021). "Looking beneath the surface": A visual-physical feature hybrid approach for unattended gauging of construction waste composition. *Journal of Environmental Management, 286*, 112233. doi:10.1016/j.jenvman.2021.112233

Chen, J., Lu, W., Yuan, L., Wu, Y. & Xue, F. (2022). Estimating construction waste truck payload volume using monocular vision. *Resources, Conservation and Recycling, 177*, 106013. doi:10.1016/j.resconrec.2021.106013

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 801-818). Munich, Germany: Springer Science. doi:10.1007/978-3-030-01234-2_49

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1251-1258). Honolulu, Hawaii: IEEE. doi:10.1109/CVPR.2017.195

Colquhoun, G. J., Baines, R. W. & Crossley, R. (1993). A state of the art review of IDEFO. *International Journal of Computer Integrated Manufacturing, 6*(4), 252-264. doi:10.1080/09511929308944576

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3213-3223). Las Vegas, Nevada: IEEE. doi:10.1109/CVPR.2016.350

Damigos, D. & Anyfantis, F. (2011). The value of view through the eyes of real estate experts: A Fuzzy Delphi Approach. *Landscape and Urban Planning, 101*(2), 171-178. doi:10.1016/j.landurbplan.2011.02.009

Dong, R., Zhang, Y. & Zhao, J. (2018). How green are the streets within the sixth ring road of Beijing? An analysis based on tencent street view pictures and the green view index. *International Journal of Environmental Research and Public Health, 15*(7), 1367. doi:10.3390/ijerph15071367

Fleischer, A. (2012). A room with a view—A valuation of the Mediterranean Sea view. *Tourism Management, 33*(3), 598-602. doi:10.1016/j.tourman.2011.06.016

Gascon, M., Zijlema, W., Vert, C., White, M. P. & Nieuwenhuijsen, M. J. (2017). Outdoor blue spaces, human health and well-being: A systematic review of quantitative studies. *International Journal of Hygiene and Environmental Health, 220*(8), 1207-1221. doi:10.1016/j.ijheh.2017.08.004

Gong, F.-Y., Zeng, Z.-C., Zhang, F., Li, X., Ng, E. & Norford, L. K. (2018). Mapping sky, tree, and building view factors of street canyons in a high-density urban environment. *Building and Environment, 134*, 155-167. doi:10.1016/j.buildenv.2018.02.042

Grinde, B. & Patil, G. G. (2009). Biophilia: does visual contact with nature impact on health and well-being? *International Journal of Environmental Research and Public Health, 6*(9), 2332-2343. doi:10.3390/ijerph6092332

Helbich, M., Yao, Y., Liu, Y., Zhang, J., Liu, P. & Wang, R. (2019). Using deep learning to examine street view green and blue spaces and their associations with geriatric depression in Beijing, China. *Environment International, 126*, 107-11. doi:10.1016/j.envint.2019.02.013

Hellinga, H. I. (2013). *Daylight and View: The influence of windows on the visual quality of indoor spaces.* Architectural Engineering and Technology. Delft, Netherlands: Delft University of Technology. doi:10.4233/uuid:2daeb534-9572-4c85-bf8f-308f3f6825fd

HKLandsD. (2014). *iB1000 Digital Topographic Map.* Hong Kong: Lands Department, Goverment of Hong Kong SAR.

HKPlanD. (2018). *Hong Kong Planning Standards and Guidelines.* Hong Kong: Planning department, Hong Kong SAR. doi:https://www.pland.gov.hk/pland_en/tech_doc/hkpsg/full/pdf/ch2.pdf

HKPlanD. (2019). *3D Photo-realistic Model.* Hong Kong: Planning Department, Goverment of Hong Kong SAR. Retrieved from https://www.pland.gov.hk/pland_en/info_serv/3D_models/download.htm

Jim, C. Y. & Chen, W. Y. (2009). Value of scenic views: Hedonic assessment of private housing in Hong Kong. *Landscape and Urban Planning, 91*(4), 226-234. doi:10.1016/j.landurbplan.2009.01.009

Kaplan, R. (2001). The nature of the view from home: Psychological benefits. *Environment and Behavior, 33*(4), 507-542. doi:10.1177/00139160121973115

Kaplan, S. (1995). The restorative benefits of nature: Toward an integrative framework. *Journal of Environmental Psychology, 15*(3), 169-182. doi:10.1016/0272-4944(95)90001-2

Kaya, N. & Erkip, F. (2001). Satisfaction in a dormitory building: The effects of floor height on the perception of room size and crowding. *Environment and Behavior, 33*(1), 35-53. doi:10.1177/00139160121972855

Ko, W. H., Kent, M. G., Schiavon, S., Levitt, B. & Betti, G. (2021). A window view quality assessment framework. *Leukos*, 1-26. doi:10.1080/15502724.2021.1965889

Kuo, F. E. & Sullivan, W. C. (2001). Environment and crime in the inner city: Does vegetation reduce crime? *Environment and Behavior, 33*(3), 343-367. doi:10.1177/0013916501333002

Labib, S. M., Huck, J. J. & Lindley, S. (2021). Modelling and mapping eye-level greenness visibility exposure using multi-source data at high spatial resolutions. *Science of the Total Environment, 755*, 143050. doi:10.1016/j.scitotenv.2020.143050

Laovisutthichai, V., Li, M., Xue, F., Lu, W., Tam, K. & Yeh, A. G. (2021). CIM-enabled quantitative view assessment in architectural design and space planning. *38th International Symposium on Automation and Robotics in Construction (ISARC 2021).* Dubai.

LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444. doi:10.1038/nature14539

Lee, D.-T. & Schachter, B. J. (1980). Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer & Information Sciences, 9*(3), 219-242. doi:10.1007/BF00977785

Li, M., Xue, F., Yeh, A. G. & Lu, W. (2020). Classification of photo-realistic 3D window views in a high-density city: The case of Hong Kong. *25th International Symposium on Advancement of Construction Management and Real Estate* (pp. 1339-1350). Singapore: Springer Nature.

Li, W. & Samuelson, H. (2020). A new method for visualizing and evaluating views in architectural design. *Developments in the Built Environment, 1*, 100005. doi:10.1016/j.dibe.2020.100005

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. doi:10.1016/j.ufug.2015.06.006

Liang, J., Gong, J., Sun, J., Zhou, J., Li, W., Li, Y., Liu, J. & Shen, S. (2017). Automatic sky view factor estimation from street view photographs—A big data approach. *Remote Sensing, 9*(5), 411. doi:10.3390/rs9050411

Liu, Y., Meng, Q., Zhang, J., Zhang, L., Jancso, T. & Vatseva, R. (2016). An effective Building Neighborhood Green Index model for measuring urban green space. *International Journal of Digital Earth, 9*(4), 387-409. doi:10.1080/17538947.2015.1037870

Lottrup, L., Stigsdotter, U. K., Meilby, H. & Claudi, A. G. (2015). The workplace window view: a determinant of office workers' work ability and job satisfaction. *Landscape Research, 40*(1), 57-75. doi:10.1080/01426397.2013.829806

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. doi:10.3390/ijerph15081576

Masoudinejad, S. & Hartig, T. (2020). Window view to the sky as a restorative resource for residents in densely populated cities. *Environment and Behavior, 52*(4), 401-436. doi:10.1177/0013916518807274

McFeeters, S. K. (1996). Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing, 17*(7), 1425-1432. doi:10.1080/01431169608948714

Middel, A., Lukasczyk, J., Zakrzewski, S., Arnold, M. & Maciejewski, R. (2019). Urban form and composition of street canyons: A human-centric big data and deep learning approach. *Landscape and Urban Planning, 183*, 122-132. doi:10.1016/j.landurbplan.2018.12.001

Pan, S. J. & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering, 22*(10), 1345-1359. doi:10.1109/TKDE.2009.191

Pazhouhanfar, M. & Kamal, M. (2014). Effect of predictors of visual preference as characteristics of urban natural landscapes in increasing perceived restorative potential. *Urban Forestry & Urban Greening, 13*(1), 145-151. doi:10.1016/j.ufug.2013.08.005

Roe, J. J., Thompson, C. W., Aspinall, P. A., Brewer, M. J., Duff, E. I., Miller, D., Mitchell, R. & Clow, A. (2013). Green space and stress: evidence from cortisol measures in deprived urban communities. *International Journal of Environmental Research and Public Health, 10*(9), 4086-4103. doi:10.3390/ijerph10094086

Shen, Q., Zeng, W., Ye, Y., Arisona, S. M., Schubiger, S., Burkhard, R. & Qu, H. (2017). StreetVizor: Visual exploration of human-scale urban forms based on street views. *IEEE Transactions on Visualization and Computer Graphics, 24*(1), 1004-1013. doi:10.1109/TVCG.2017.2744159

Song, Y., Wang, X., Tan, Y., Wu, P., Sutrisna, M., Cheng, J. C. & Hampson, K. (2017). Trends and opportunities of BIM-GIS integration in the architecture, engineering and construction industry: a review from a spatio-temporal statistical perspective. *ISPRS International Journal of Geo-Information, 6*(12), 397. doi:10.3390/ijgi6120397

Tara, A., Lawson, G. & Renata, A. (2021). Measuring magnitude of change by high-rise buildings in visual amenity conflicts in Brisbane. *Landscape and Urban Planning,* *205*, 103930. doi:10.1016/j.landurbplan.2020.103930

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment, 8*(2), 127-150. doi:10.1016/0034-4257(79)90013-0

Turan, I., Reinhart, C. & Kocher, M. (2019). Evaluating spatially-distributed views in open plan work spaces. *Proceedings of the IBPSA International Building Simulation Conference*, (pp. 1098-1105). Rome. doi:10.26868/25222708.2019.210755

Ulrich, R. S. (1981). Natural versus urban scenes: Some psychophysiological effects. *Environment and Behavior, 13*(5), 523-556. doi:10.1177/0013916581135001

Ulrich, R. S. (1983). Aesthetic and affective response to natural environment. *Behavior and the Natural Environment* (pp. 85-125). Boston: Springer. doi:10.1007/978-1-4613-3539-9_4

Ulrich, R. S. (1984). View through a window may influence recovery from surgery. *Science, 224*(4647), 420-421. doi:10.1126/science.6143402

UNPD. (2014). *World Urbanization Prospects: The 2014 Revision, Highlights.* Population Division, Department of Economic and Social Affairs, ST/ESA/SER.A/352. United Nations. Retrieved from https://population.un.org/wup/Publications/Files/WUP2014-Highlights.pdf

Van den Berg, M., Poppel, M. v., Kamp, I. v., Andrusaityte, S., Balseviciene, B., Cirach, M. & Danileviciute, A. e. (2016). Visiting green space is associated with mental health and vitality: A cross-sectional study in four european cities. *Health & Place, 38*, 8-15. doi:10.1016/j.healthplace.2016.01.003

Waczynska, M., Sokol, N. & Martyniuk-Peczek, J. (2021). Computational and experimental evaluation of view out according to European Standard EN17037. *Building and Environment, 188*, 107414. doi:10.1016/j.buildenv.2020.107414

White, M., Smith, A., Humphryes, K., Pahl, S., Snelling, D. & Depledge, M. (2010). Blue space: The importance of water for preference, affect, and restorativeness ratings of natural and built scenes. *Journal of Environmental Psychology, 30*(4), 482-493. doi:10.1016/j.jenvp.2010.04.004

Xia, Y., Yabuki, N. & Fukuda, T. (2021). Development of a system for assessing the quality of urban street-level greenery using street view images and deep learning. *Urban Forestry & Urban Greening , 59*, 126995. doi:10.1016/j.ufug.2021.126995

Xue, F., Li, X., Lu, W., Webster, C. J., Chen, Z. & Lin, L. (2021). Big data-driven pedestrian analytics: unsupervised clustering and relational query based on Tencent Street View photographs. *ISPRS International Journal of Geo-Information, 10*(8), 561. doi:10.3390/ijgi10080561

Xue, F., Wu, L. & Lu, W. (2021). Semantic enrichment of building and city information models: A ten-year review. *Advanced Engineering Informatics, 47*, 101245. doi:10.1016/j.aei.2020.101245

Yao, Y., Liang, Z., Yuan, Z., Liu, P., Bie, Y., Zhang, J., Wang, R., Wang, J. & Guan, Q. (2019). A human-machine adversarial scoring framework for urban perception assessment using street-view images. *International Journal of Geographical Information Science, 33*(12), 2363-2384. doi:10.1080/13658816.2019.1643024

Yu, S., Yu, B., Song, W., Wu, B., Zhou, J., Huang, Y., Wu, J., Zhao, F. & Mao, W. (2016). View-based greenery: A three-dimensional assessment of city buildings' green visibility using Floor Green View Index. *Landscape and Urban Planning , 152*, 13-26. doi:10.1016/j.landurbplan.2016.04.004