Original Articles

# Water View Imagery: Perception and evaluation of urban waterscapes worldwide

Junjie Luo [a,b], Tianhong Zhao [b,c], Lei Cao [a], Filip Biljecki [b,d,*]

[a] *Department of Landscape Architecture, Tianjin University, Tianjin, China*
[b] *Department of Architecture, National University of Singapore, Singapore*
[c] *School of Architecture and Urban Planning, Shenzhen University, China*
[d] *Department of Real Estate, National University of Singapore, Singapore*

## ARTICLE INFO

## ABSTRACT

Gathering knowledge about physical settings and visual information of places has long been of interest to a wide variety of fields as they affect the experience of observers. Previous studies have relied on on-site surveys, low-throughput methods, and limited data sources, which especially hinder analyzing waterscape features. Thus, detecting the relationships between the human perception results of large-scale urban water areas and the waterfront features at high spatial resolutions remains challenging, and worldwide studies have not been conducted. We investigate an alternative: a data-driven waterscapes evaluation approach based on computer vision (CV) to analyze water view imagery (WVI) in 16 cities around the world and measure how people perceive scenes using virtual reality (VR). We bring attention to WVI – the counterpart of street view imagery (SVI) on water bodies, which is readily available for many cities thanks to the usual SVI services, but has been entirely overlooked in research hitherto. Specifically, a deep learning model, which has been trained with 500 segmented water-level photos, was developed to analyze them, achieving the mean pixel accuracy (MPA) of 94%, which advances state of the art. These panoramic images have been assessed through a virtual experience survey in which 60 participants indicated their perceptions across multiple dimensions. Afterwards, a series of statistical analyses were conducted to determine the visual indicators that drive perceptions, and the relationship between the people's subjective visual perceptions and objective waterscape environment as seen by machines has been established. The results take researchers and watercourse planners one step toward understanding the interactions of the perceptions and semantics of water areas globally. The large-scale dataset we produced in this research has been released openly as the first such instance of open segmented water view imagery, and it is intended to support future studies.

## 1. Introduction

Urban water areas (e.g. rivers, bays, lakes) supply diverse services and play important roles, including ecological, cultural, recreational, and aesthetic ones (Liu et al., 2021; Hua and Chen, 2019; Gr et-Regamey et al., 2016; Vollmer et al., 2015; Keten et al., 2020). A significant body of literature has highlighted the beneficial effects of blue space (i.e. urban riverscapes) (Bao et al., 2022; Xiao et al., 2022). The importance of the urban water environment in promoting the health and well-being of city dwellers cannot be overstated (Vert et al., 2020; De Vries et al., 2016). These areas also reflect the city's identity and qualities and often appealing to both residents and visitors. Water is, in fact, a key natural resource in urban regions, and it is inextricably linked to urban dwellers' lives, leisure, and commercial activities. As a distinctive aspect of urban water perception, the water tour line is a way for tourists to experience the city (Venverloo et al., 2021). To enhance urban water tourism, high-quality water level landscapes are crucial, and they can notably reflect the characteristics of a city (Li et al., 2021). As one of the most important approaches for the public to perceive the landscape, vision accounts for 76% on environment satisfaction (Krause, 2001; Jeon and Jo, 2020), and the water level perspective can provide residents and visitors with a unique visual experience. It is necessary to understand which waterscape features (percentage of landscape elements, ecology and cleanliness of the natural conditions, etc.) influence visitors' perceptions

---

favorably, for example, which variables have more significant impacts on waterfront quality and how the landscapes may be enhanced.

Most studies on river landscapes mainly start from the perspective of the river bank (Sun et al., 2021; Xin et al., 2022). Researchers perform on-site observations and/or take photos for remote perception. These approaches can only analyze the characteristics of the waterside, and cannot 'sense' on-water perspectives, which may differ substantially.

Only a few studies use the water-level perspective to observe waterscapes. For example, Li et al. (2021) has utilized images (perspective, not panorama) to evaluate visual quality from an on-water perspective of a part of the river in Guilin, China. However, the scale and scope of this research are limited because they only analyze the river-scape in the karst landform. Comparative analysis of multiple areas, especially across different cities, is often a neglected topic in the existing literature. Due to the fact that imagery collection on the water level is inconvenient, data obtained on a large scale is time-consuming and labour-intensive; hence, global urban water level studies are challenging to conduct. Further, the lack of open data hinders this research domain and limits our understanding of its current space and landscape, characteristics and distribution. Therefore, objective visual evaluation and efficient perception of large-scale water level landscapes remain a challenge.

Catalyzed by the rapid development of map services and volunteered geographic information (VGI), a massive volume of geo-tagged images have been compiled and made publicly available that portray nearly every corner of a city (Yan et al., 2020; Li et al., 2018; Taecharungroj and Mathayomchan, 2020; Richards and Friess, 2015). With the explosive growth of street view imagery (SVI) and crowdsourced photos, these data sources can offer a wealth of opportunities for geo-related studies and are commonly employed in urban environment analysis (Song et al., 2020; Li et al., 2015; Zhang et al., 2018). Many data sources for built environment studies are available through Google Street View (GSV), Baidu, Mapillary, and other platforms, which have been widely used in urban studies (Biljecki and Ito, 2021).

In this study, we investigate whether we can use SVI data, similar to how numerous studies examine urban streets environment using GSV or Baidu Total View (BTV) imagery, to evaluate the city's waterscapes widely following the little-known fact that — while the word *street* is prefixed to the term street view imagery — the data is not restricted to roads and streetscapes (Wang et al., 2022; Liu et al., 2016; Li et al., 2015; Li et al., 2018). After a global exploration, we identified dozens of cities with continuous linear water-level photos on GSV and Mapillary. As this kind of data is relatively unknown, there are virtually no use cases taking advantage of it. Scientific literature has not been documented that these types of SVI platforms may also have a water perspective for some locations. In fact, the awareness and use are so rare that — to the extent of our knowledge — there is not even a term for it. We name these on-water photos Water View Imagery (WVI), as a counterpart (or subset) of SVI collected on water-level (Fig. 1).

Simultaneously, advances in computer vision (CV), such as semantic segmentation and object detection, have accelerated ways to automatically and objectively analyse a large number of photos in the built environment (Ibrahim et al., 2020; Yao et al., 2021; Dang and Li, 2021; Biljecki and Ito, 2021). Some researchers have used human eye-level photographs and drone-based images to investigate landscape features using CV in the built environments (Seiferling et al., 2017; Wilkins et al., 2022; Luo et al., 2022; Dang and Li, 2021), thus CV is not new to landscape analysis, but CV application in the waterscape visual evaluations is few and far between. Only a few studies have looked at the river's visual landscape aspects (Li et al., 2021), but to a limited extent, and comprehensive and large-scale quantitative waterscape analysis is lacking. Another void in this research line is the lack of understanding of the differences in perceptions between ground and water-level perspectives. No CV and WVI combined large-scale successive subjective visual perception or objective evaluation study have been conducted yet, which is another gap we seek to bridge with this paper. Apart from that, studies employing CV on various types of urban imagery at the ground-level have relied on general datasets to train deep learning models to visually evaluate the environment (Cordts et al., 2016). However, for studies focused on waterscapes, these datasets may fall short, and there has been no processed water imagery dataset of the water landscapes available openly so far, which is another impediment of river environment analysis as it may be better suited for this research thread. We believe that it is beneficial to generate an open dataset to be shared with the scientific community to enhance the waterscape research line and provide reference for future landscape improvement by defining the relationship between environmental features and human perceptions.
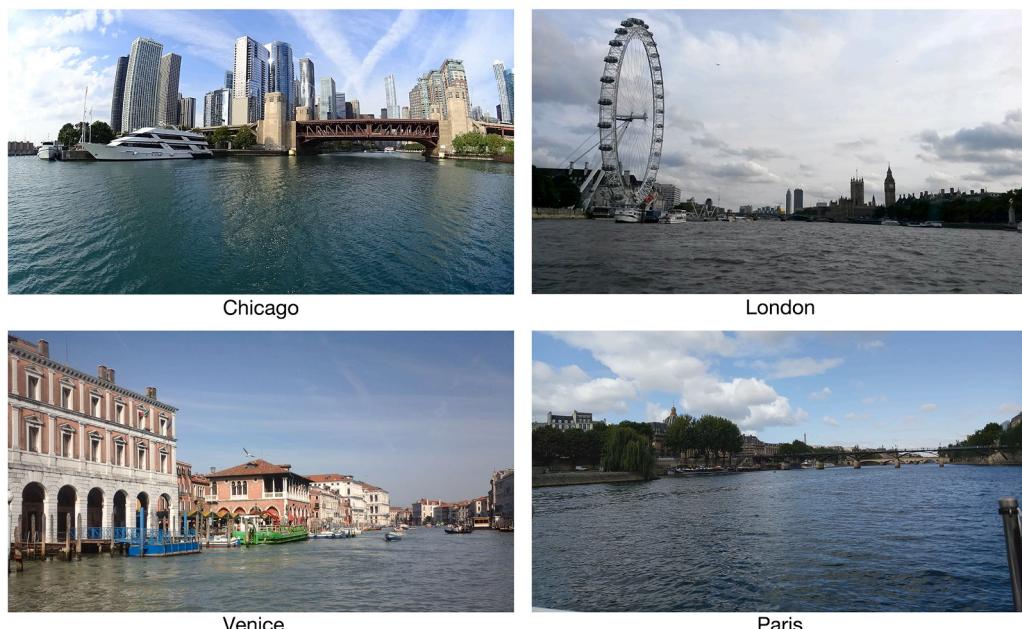


**Fig. 1.** Examples of water view imagery of four cities found in a crowdsourced street view imagery platform. This paper reveals the availability of such data and it investigates their usability. Source: (c) Mapillary contributors.

Considering the developments in CV and the massive amount of geo-tagged images on water, we believe that research marrying the two is needed and timely. In short, our study questions are threefold: (1) How can the continuous visual characteristics (e.g. green visibility and sky visibility) and human perceptions (e.g. beauty, depression, liveliness) of urban waterscapes be quantified and compared? (2) How will the physical urban waterscapes affect people's subjective perceptions when experiencing these water level perspectives? (3) Can we take advantage of existing water-level imagery in mainstream SVI services, bypassing the need for own data collection, to support analysis of multiple cities and to understand the characteristics of global waterscapes?.

## 2. Background and related work

### 2.1. Visual perception and evaluation of waterscapes

Visual perception and evaluation have become the mainstay for researchers, practitioners, and governments to understand the landscape quality of public regions (e.g. streets, rivers, parks) (Duchowski and Duchowski, 2017; Liao et al., 2022). For decades, the visual qualities of street canyons have been studied by a number of researchers. There have been in-depth analyses of the street green visibility (Cai et al., 2018), bikeability (Ito and Biljecki, 2021), walkability (Leslie et al., 2007), and so on among different cities, and the impact of street landscape features on people's sense of safety, comfort, attractiveness, and other topics (Zhang et al., 2018; Larkin and Hystad, 2019). In contrast, research on the visual characteristics of urban river canyons, lakes, bays, and other open water landscapes is limited, inhibiting landscape planning and future development of such places, even though the waterscape is one of the city's most essential natural landscape elements (Silva et al., 2013). Some research evaluated the landscape safety pattern from large-scale land-use types and environmental elements of the river corridor (Gargiulo et al., 2020). However, such large-scale analysis based on land-use types can only obtain the general spatial characteristics of river corridors, rather than the visual perceptions of human beings. To visually examine the spatial surroundings, some researchers used static discontinuous human perspective photographs to obtain multi-dimensional remote impression of aspects such as beauty, vitality, and ecology of urban rivers, lakes, and other waterscapes (Sun et al., 2021; Yamashita, 2002; Pflüger et al., 2010). These studies provide a reference for people to comprehend the distribution features of river space and landscape elements by objectively describing the visual characteristics of relevant places. However, these studies have not considered the continuity of visual perception, they can only statistically characterize one or few discontinuous perspectives rather than a sequence of continuous viewpoints.

To the extent of our knowledge, the paper of Li et al. (2021) is the most relevant to ours. It quantitatively analyzes the quality of riverscapes from an on-water perspective using ordinary 2D photographs in the karst landform area of Guilin, China. The continuous perception approach of waterscapes can provide an understanding of the linear landscape environment, which will serve as a foundation for landscape quality evaluation and future optimization and enhancement. In addition, other studies also adopt continuous visual perception means to help researchers understand the landscape traits; for example, Jin and Wang (2021) conducted virtual dynamic and continuous simulations of three sports modes on the Beijing Hangzhou Grand Canal: boating, running, and cycling, which can improve the understanding of landscape characteristics in the context of sports. Another continuous visual perception study also advocated that the scenery around the urban river can be perceived dynamically (Cheng and Wang, 2021). This research analyzed the landscape characteristics of riverside parks using pictures and analyzed people's physiological responses while perceiving these photos, which provides a better understanding of the riverside settings. Continuous visual perception, as opposed to single-point visual perception, is better suited to linear spaces such as rivers. Similar studies

can provide a strong grasp of riverfront scenery and are appropriate for small-scale evaluation (Cao and Zhang, 2020; Xin et al., 2022). However, relevant research efforts are time-consuming and labour-intensive, and scaling up approaches remains challenging. As a result, researchers are grappling with how to visually analyze waterscape elements across a wider range, and how to scientifically measure people's perceptions while observing the scenery.

### 2.2. Automated landscape analysis with computer vision

The efficiency of various research disciplines has substantially improved as a result of the continuous development of advanced technology, particularly the constant use of artificial intelligence and deep learning technologies in the built environment (Ibrahim et al., 2020; Ghermandi et al., 2022; Sun et al., 2022; Li et al., 2022; Wang and Biljecki, 2022). Collaboration between urban planning, geographic information science (GIS), computer science, and other disciplines can also help to improve the depth and breadth of knowledge about various urban places and landscape contexts (Biljecki and Ito, 2021; Kandt and Batty, 2021; Liu and Biljecki, 2022). The advancement of CV algorithms, as well as the increase in the number of relevant datasets, has aided in these large-scale and multi-regional studies (Badrinarayanan et al., 2017; Cordts et al., 2016; Luo et al., 2022). CV technology can be divided into three categories: image semantic segmentation, object recognition and instance segmentation (Ahmed et al., 2021; Li et al., 2022). The study scope of applying semantic segmentation to analyze urban landscape is currently highly broad, as are the data sources. For urban spatial studies and natural environment research, many scholars use readily available street view imagery offered by map suppliers or crowdsourced platforms such as Google, Baidu, and Mapillary (Xia et al., 2021; Hosseini et al., 2022). Because street view photographs can be taken on a wide scale in a short amount of time and contain a lot of information about ground elements, they offer a lot of potential in urban studies (Zhang et al., 2018; Yao et al., 2019; Chen et al., 2020; Kotowska et al., 2021). For example, using GSV data and a computer vision algorithm, the research by Seiferling et al. (2017) measures the vegetation covering proportion to quantify the green viewing rate of urban street spaces. This method achieved the ability to quantify the existence and distribution of trees from the perspective of human viewpoints. Aside from the human perspective, other research combines CV technology with drone photos, using UAV oblique perspective photographs to semantically separate diverse landscape aspects in a large-scale environment, improving our comprehension of landscapes (Meng et al., 2021; Osco et al., 2021; Lyu et al., 2020; Luo et al., 2022). The use of a combination of satellite images and CV to interpret urban roofscape information (Wu and Biljecki, 2021), building characteristics (Li et al., 2019; Wheeler and Karimi, 2020), green space distribution (Liu et al., 2019; Huerta et al., 2021), and vacant land identification (Mao et al., 2022) also became an important tool to understand landscapes.

However, there are certain barriers in the way. The absence of adequate semantic segmentation datasets to train the segmentation models, for example, has become a serious impediment to the use of computer vision in some specific fields. The Cityscapes dataset provides an opportunity to study urban SVI data (Cordts et al., 2016), Aerospaces (Nigam et al., 2018), Urban Drone Dataset (UDD) (Chen et al., 2018), Semantic Riverscapes Dataset (Luo et al., 2022) and other datasets provide the possibility to study urban landscape with perspectives of UAVs. These datasets are essential for deep learning model training. Although studies utilize computer vision to study the environment around the river, there is currently no publicly available imagery semantic segmentation dataset from the water-level perspective.

# 3. Methodology

## 3.1. Identifying water view imagery datasets

We manually checked major cities worldwide on the mainstream SVI platforms, and found that 27 cities on GSV and 10 cities on Mapillary have water view imagery. Because our paper is the one 'inaugurating' and bringing attention to the concept of WVI, we provide more information on the imagery identified in practice. By browsing the waterscape photos on the two platforms, we compared and analyzed the characteristics of WVI of different cities we found across four aspects: length of shooting on water level, lens obstacles, light condition, and field of view, which are features that are important for this research (Table 1).

After surveying WVI qualities in the identified cities, 16 diverse and geographically distributed urban water areas, including rivers, canals, and bays spanning Europe, North America, and Asia are selected as our study areas. These water areas are located in Amsterdam, Bangkok, Boston, Chicago, Istanbul, London, Memphis, Minneapolis, Paris, Phnom Penh, San Diego, San Francisco, Tokyo, Venice, Warsaw, and Washington. All of these selected study areas are located in cities (built environment areas); therefore, we believe that these areas are comparable with each other. We consider the overall quality of these water level panoramic imagery adequate for our subsequent analysis. We obtained WVI sampled each 50 m. In each study area, we randomly picked water routes of equal lengths (5 km; around 100 continuous panoramic photos) to minimize bias of variable lengths. As a result, we selected about 1600 panoramic images of 16 cities as our dataset, a mix of Google Street View and Mapillary. It is important to note that in the SVI domain, it is rare to have studies that include more than one data source, which may be considered as another contribution of this work.

## 3.2. Water view imagery dataset for semantic segmentation

As there is no open water level imagery semantic segmentation dataset, we capture and segment one, which is designed for semantic segmentation of water level scenarios and can assist waterscape-related segmentation applications. Out of the 1600 images, we selected 500 high-resolution images and segmented each. We named this dataset *Water View Imagery dataset* and released it openly, supporting future studies in the waterscape domain. Given the commercial nature and restrictive licence of Google Street View, all images in the released dataset come from Mapillary, which are subject to a liberal licence thanks to its crowdsourcing nature. These segmented photos are from 8 cities water areas (Amsterdam, Bangkok, Chicago, Istanbul, Japan, London, Paris, and Venice). We refer to the classification criteria of Cityscapes dataset and related research (Cordts et al., 2016; Li et al., 2021), and according to our research goals and the landscape elements characteristics of these images, a total of 15 categories were selected for annotation (i.e. semantic segmentation), namely: water, sky, terrain, traditional building, modern building, revetment, bridge, car, truck, bicycle, boat, tree, grass, people, and void. To separate vegetation around the river, we break down greens into tree and grass, based on the peculiarities of river landscapes. Similarly, we regard multiple groups of buildings: traditional buildings, and modern building. Each image was manually annotated at the pixel level with EISeg software (Hao et al., 2021).

We have released this dataset openly for public use, together with documentation. As far as we know, this is the first worldwide semantic segmentation open dataset using water view imagery, and it has been released[1] under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license (CC BY-NC-SA 4.0).

---

[1] The dataset is available at https://github.com/ualsg/Water-View-Imagery-dataset.

## 3.3. Objective visual evaluation

Many ready-to-use models, FCN, SegNet, U-net, PSP-net, SegFormer, etc., can detect objects and perform segmentation of imagery (Badrinarayanan et al., 2017; Xie et al., 2021). We select SegFormer, a cutting-edge Transformer framework that considers efficiency, accuracy, and robustness, for semantic segmentation (Xie et al., 2021), based on the characteristics of WVI data and applicability in waterscapes. The *Water View Imagery dataset* is used to train a CV model for image segmentation. To ensure the robustness of the reported model, we have adopted the common practice of randomly splitting the dataset into two portions: training (90%) and validation (10%). The training and validation processes were assessed using mean pixel accuracy (MPA), which is the ratio of correctly predicted pixels to the total pixels, and mean Intersection over Union (mIoU), the ratio of the intersection area of the predicted pixels and ground truth pixels to their union area. To ensure that the simulated visual field and sight are close to the real experience of people, we turned the panoramic imagery into ordinary perspective images in six directions, and the heading parameters were set to 0, 60, 120, 180, 240, and 300, respectively. Each image has a 60-degree visual field, which mimics human eyes view (Walker et al., 1990; Li et al., 2015; Zhou et al., 2019). Thus, we summarized the percentage of various visual elements in every image and, via Eq. 1 (where $P_i$ is the average proportion of every visual element and $V_i$ denotes the proportion of visual elements in one direction image), to calculate the average of each sampling point to represent the average status of the position. Finally, after training the SegFormer model using the WVI dataset, we can objectively examine the percentage of these categories of landscape features in various pictures by classifying the 15 types of elements in the panoramic water level imagery at the pixel level (Fig. 2).

$$P_i = \frac{\sum\limits_{i=1}^{6} V_i}{6} \{i \in (1, 2, \cdots, n)\} \tag{1}$$

Based on the previous research experience of visual landscapes (Li et al., 2015; Gong et al., 2018) and combined with the characteristics of river environment (Li et al., 2021), we adopt 6 visual indexes for waterscape evaluation, including green visibility factor (GVF), water visibility factor (WVF), sky visibility factor (SVF), hard revetment factor (HRF), dynamic factor (DF), and building visibility factor (BVF). We define Eqs. (2)–(7) to calculate the scores of the 6 evaluation indexes. Among them, vegetation is one of the most important landscape elements in the riverscapes, and the GVF includes tree and grass, which affects the ecology and natural degree of the watercourse (Richards and Friess, 2015). Water is the main element in riverscape; thus, WVF plays a substantial important role in vision. HRF and BVF are significant indicators reflecting the intensity of artificial construction in a river and surrounding areas. Among them, the HRF includes river revetment and bridges, and the BVF include traditional buildings, modern high-rise residential buildings, commercial office buildings, etc. SVF can measure the openness of the river space, and also has a great impact on people's vision, and DF can measure moving objects such as people, cars, and boats.

$$GVF_i = \frac{1}{n}\sum_{i=1}^{n} Tree_n + \frac{1}{n}\sum_{i=1}^{n} Grass_n$$
$$\{i \in (1, 2, \cdots, n)\} \tag{2}$$
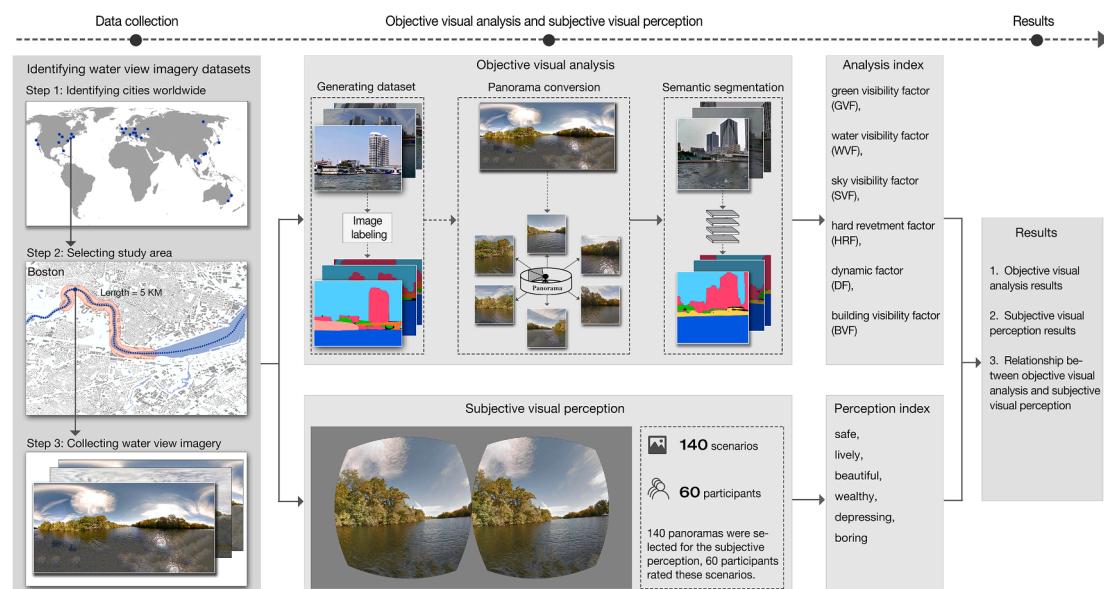
$$WVF_i = \frac{1}{n}\sum_{i=1}^{n} Water_n$$
$$\{i \in (1, 2, \cdots, n)\} \tag{3}$$

**Table 1**

Overview of cities in which we identified water view imagery in GSV or Mapillary.

| city | country | waters name | length of shooting | lens obstacles | light condition | field of view |
|---|---|---|---|---|---|---|
| | | | Google Street View | | | |
| Amsterdam | Netherlands | Amstel, etc. | over 30 km | light | good | panorama |
| Bangkok | Thailand | Chao Phraya | over 30 km | light | good | panorama |
| Belgrade | Serbia | Danube | over 30 km | heavy | good | panorama |
| Boston | USA | Charles River | 13 km | light | good | panorama |
| Bratislava | Slovakia | Danube | 11 km | heavy | good | panorama |
| Brisbane | Australia | Brisbane River | 28 km | light | medium | panorama |
| Budapest | Hungary | Danube | over 30 km | medium | good | panorama |
| Chicago | USA | Chicago River | over 30 km | no | good | panorama |
| Hong Kong | China | Victoria Harbour | 25 km | heavy | good | panorama |
| Istanbul | Turkey | Golden Horn | 23 km | light | good | panorama |
| London | UK | Thames | over 30 km | light | good | panorama |
| Memphis | USA | Mississippi | over 30 km | light | good | panorama |
| Miami | USA | Miami River | over 30 km | medium | medium | panorama |
| Minneapolis | USA | Mississippi | over 30 km | no | good | panorama |
| New York | USA | Hudson, etc. | over 30 km | heavy | good | panorama |
| Ottawa | Canada | Ottawa River | 7.7 km | medium | good | panorama |
| Paris | France | Seine | 19 km | light | good | panorama |
| Phnom Penh | Cambodia | Tonle Sap, etc. | over 30 km | light | good | panorama |
| San Diego | USA | San Diego Bay | over 30 km | light | good | panorama |
| San Francisco | USA | San Francisco Bay | over 30 km | light | good | panorama |
| Sydney | Australia | Sydney Harbour | 4 km | heavy | good | panorama |
| Tokyo | Japan | Ebitori, etc. | over 30 km | light | good | panorama |
| Venice | Italy | Grand Canal | 11 km | light | good | panorama |
| Voronezh | Russia | Voronezh River | 1 km | light | medium | panorama |
| Warsaw | Poland | Vistuna River | over 30 km | medium | good | panorama |
| Washington | USA | Potomac, etc. | 21 km | medium | good | panorama |
| Yakutst | Russia | Lena River | 5 km | heavy | bad | panorama |
| | | | Mapillary | | | |
| Amsterdam | Netherlands | Amstel, etc. | over 30 km | unstable | good | panorama & ordinary |
| Bangkok | Thailand | Chao Phraya | 10 km | unstable | unstable | panorama & ordinary |
| Chicago | USA | Chicago River | over 30 km | unstable | unstable | panorama & ordinary |
| Hong Kong | China | Victoria Harbour | 10 km | unstable | unstable | ordinary |
| Istanbul | Turkey | Golden Horn | 12 km | unstable | good | panorama & ordinary |
| London | UK | Thames | 8 km | unstable | unstable | ordinary |
| Macau | China | Canal de Shizimen | 16 km | medium | medium | panorama |
| Paris | France | Seine | over 30 km | unstable | unstable | panorama & ordinary |
| Tokyo | Japan | Ebitori River, etc. | 18 km | unstable | unstable | panorama & ordinary |
| Venice | Italy | Grand Canal | 12 km | unstable | unstable | panorama & ordinary |



**Fig. 2.** Workflow of this study. Sources of maps and images: (c) OpenStreetMap contributors and Google Street View.

$$SVF_i = \frac{1}{n}\sum_{i=1}^{n}Sky_n \tag{4}$$

$$\{i \in (1, 2, \cdots, n\}$$

$$HRF_i = \frac{1}{n}\sum_{i=1}^{n}Rev_n \tag{5}$$

$$\{i \in (1, 2, \cdots, n\}$$

$$DF_i = \frac{1}{n}\sum_{i=1}^{n}Car_n + \frac{1}{n}\sum_{i=1}^{n}Truck_n + \frac{1}{n}\sum_{i=1}^{n}Bic_n + \frac{1}{n}\sum_{i=1}^{n}Boat_n + \frac{1}{n}\sum_{i=1}^{n}Peo_n$$

$$\{i \in (1, 2, \cdots, n\}$$

$$\tag{6}$$

$$BVF_i = \frac{1}{n}\sum_{i=1}^{n}TB_n + \frac{1}{n}\sum_{i=1}^{n}MB_n \tag{7}$$

$$\{i \in (1, 2, \cdots, n\}$$

### 3.4. Subjective visual perception

People's evaluations of the physical environment are mostly obtained through 'subjective' or 'perceptual' approaches in built environment studies (Tabrizian et al., 2020; Li et al., 2022; Zhang et al., 2018). To capture human responses, several research frequently uses 2D pictures or panoramic photos as visual stimuli representing in situ experiences of the environment (Li et al., 2021; Ma et al., 2021; Zhang and Zhang, 2021). One of the most famous urban environmental perception case is 'Place Pulse' by MIT, which contains a wealth of information for landscape perception studies (Dubey et al., 2016). This dataset includes photos of city environment from around the world that can be used to measure six characteristics, including the safety, lively, boring, wealthy, depressing, and beautiful of photographic settings. However, there is not much information about waterscapes in this dataset, and it needs to be discussed further whether it can accurately reflect people's perceptual data in waterscape settings. We employ the same perceptual indicators as this type of urban landscape in our research, and we use virtual reality equipment to view the scenarios of the WVI waterscape panoramas from all angles and score them. The reason why we chose these six perceptual indicators is that, according to earlier research, they were representative of people's urban perception (including both positive and negative aspects of perception) (Dubey et al., 2016; Wang et al., 2019; Zhang et al., 2018; Yao et al., 2021).

We randomly chose 5 to 10 panoramas of each city for the subjective perception of water landscapes. To obtain subjective perception results, we invited 60 participants to evaluate these 140 scenarios in a laboratory environment using five tablets (iPad) we provided. Although such devices can only provide a non-immersive virtual environments for VR perception, it can avoid dizziness and nausea caused by the use of head-mounted displays (Lee et al., 2020; Zhang et al., 2020; Birenboim et al., 2019; Maffei et al., 2016; Liao et al., 2022). Thus, we use the iPad as our VR display and questionnaire survey device to allow more people to participate in this study. The mean age among the 60 participants was 23.7 years; 65% were male (n = 39) and 28.3% were graduate students (n = 17). The study protocol was approved by the Institutional Review Board of the National University of Singapore (reference code: NUS-IRB-2022–191).

To avoid the impact of fatigue caused by long-time viewing of these pictures, we randomly divided 140 WVI panoramic images from 16 cities into 10 groups with 14 images in each group. Therefore, each set of panoramic imagery is evaluated by 6 participants. Before the participants experience, they need to spend about one minute browsing all 14 WVI pictures to get an overall impression of the water-level scenes. Then, the participants perceived each panoramic image 360 degrees in turn; after experiencing each panorama, participants rated each through

multiple dimensions: safe, lively, boring, wealthy, depressing, and beautiful using the 7-point Likert scale (Likert, 1932). The final score of each scene is the average of the participants' scores. On average, the duration of each perception was 17 min.

## 4. Results

### 4.1. Objective visual analysis results

#### 4.1.1. Comparison of overall proportion of watersacpe elements among cities

With a MPA of 93.69% and a mIoU of 49.53%, our trained SegFormer model under the Transformer framework performs well in the imagery semantic segmentation task, meeting the experimental conditions. Fig. 3 illustrates the results of segmenting 15 elements of river landscapes.
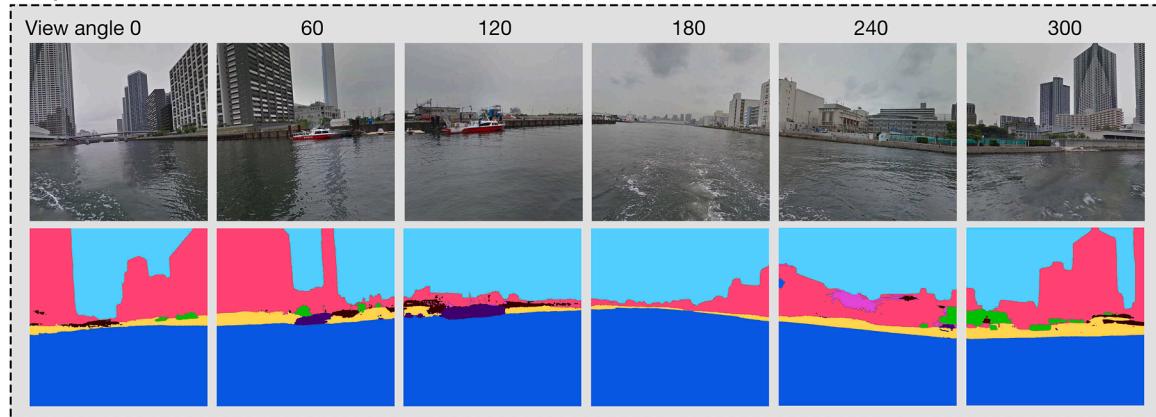
Following a statistical analysis of the visual evaluation findings of 16 waterscapes among cities (Table 2), we discovered that the proportion of water and sky elements, which make up the primary landscape elements of water areas, is relatively high, with an average of 46.58 % and 37.12 %, respectively. Among them, Venice has the smallest average proportion of water elements, accounting for 42.32%, while San Diego has the highest average proportion, accounting for 50.69%, and the standard deviation is 0.025. The water area with the highest average proportion of sky element is Phnom Penh, which is 46.32%, and the smallest is Chicago, with only 14.85%, and the standard deviation is 0.088. In contrast, the average proportions of tree, modern building, traditional building, revetment and bridge are relatively small. The average proportion of trees is 6.18%, the minimum value is 0.82%, located in San Diego, the maximum value is 16.23%, located in Minneapolis, and the standard deviation is 0.054. The average difference between modern building and traditional building is small, which are 3.00% and 2.97% respectively. Among them, the largest proportion of modern architecture is Chicago, accounting for 17.47%, and the least is Memphis, only 0.09. Traditional buildings account for the largest proportion of 23.28%, located in Venice, and Minneapolis, with the least proportion of 0.29%. The average proportion of revetment and bridge is less, only 2.24% and 1.2%. The maximum and minimum average proportion of revetment are 7.82% and 0.61%, respectively, in Chicago and Memphis, while the maximum and minimum average proportion of bridge are 4.83% and 0.21%, respectively. The average proportion of boat, grass, car, and other landscape elements is less than 1%.

Using the six visual evaluation indexes constructed to statistically analyze the overall water characteristics of 16 cities, including GVF, BVF, WVF, SVF, HRF, and DF (Fig. 4). GVF is made up of both trees and grass, and the highest average GVF value is 16.27 %, which is found in Minneapolis, while the lowest value is 0.82%, which is found in San Diego. BVF is composed of traditional building and modern architecture. The maximum average value of BVF is 24.06%, which is located in Venice, while the minimum average value is located in Minneapolis, which is only 0.58%. DF includes five categories: bicycle, car, truck, boat and people, so it can be used as the evaluation index of water space vitality. The maximum average value of DF is 2.85%, which is located in Amsterdam, and the minimum average value is only 0.11%, which is located in Memphis. The other three evaluation indexes of WVF, SVF and HRF are composed of water, sky and revetment respectively; therefore, their values are consistent.
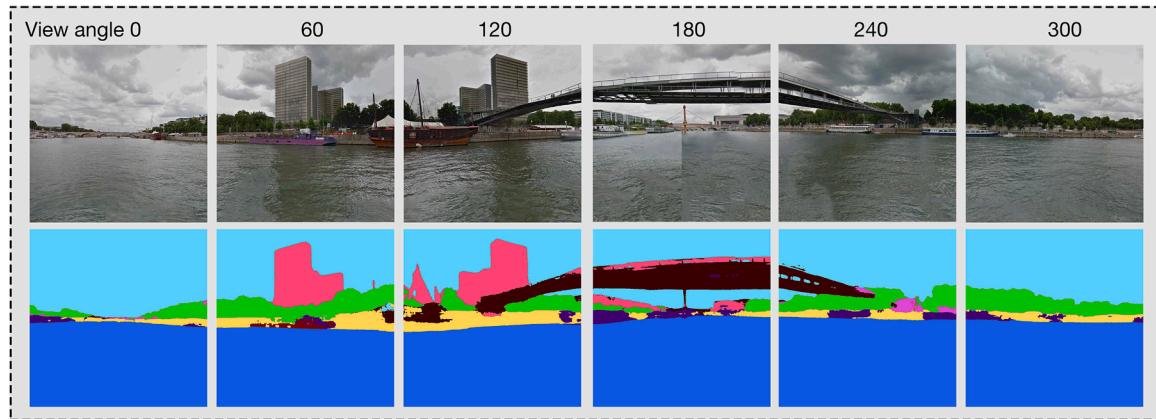
#### 4.1.2. Analyzing characteristics in each city

To facilitate understanding and to compare the waterscapes of different cities, we provide continuous plots of 16 cities, each with its own set of water features (Fig. 5). The waterscape spaces we investigated are primarily linear environments along the watercourse, with a number of indexes. Consequently, we choose to illustrate the waterscape characteristics of various locations using a stacked area plot. The waterscape dynamic features of Paris, Warsaw, and Washington are relatively similar, as can be inferred from the plot, and the six types of
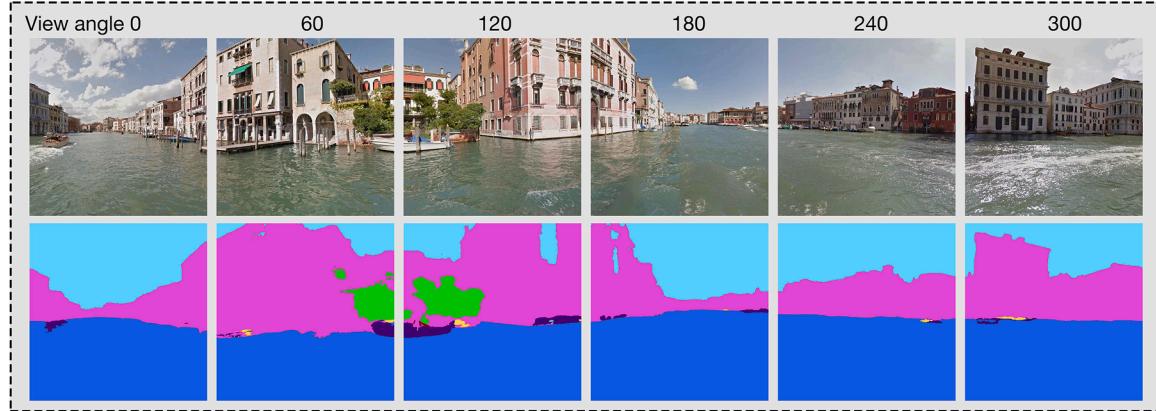
Tokyo



Paris



Venice



Class

water · sky · terrain · traditional building · modern building · revetment · bridge · car

truck · bicycle · boat · tree · grass · people · void

**Fig. 3.** Examples of water view imagery semantic segmentation performed in the work.

evaluation indicators demonstrate balanced distribution on the whole and varying in some locations. Among them, the average proportion of GVF in the three cities is comparable, about 9%. The proportion of greenery rate in these urban waterscapes is not high, but the overall distribution is relatively balanced, and the average proportion of WVF and SVF in the three cities is very large, exceeding 75%. Paris has a slightly greater average share of BVF, HRF, and DF than the other two cities, at 4.76%, 5.17%, and 1.21%, respectively. In the other two cities,

the average share of BVF, HRF, and DF is less than 2%. Boston and Minneapolis have a lot of similarities when it comes to their waterscapes. The average GVF of waterscape in these two cities, for example, is very high, reaching 14.91% and 16.27%, respectively, but the former's fluctuation in greenery is very noticeable in space, whilst the latter's change is considerably more muted. WVF and SVF share many of the same properties, while the spatial change characteristics of Boston are significantly lower than those of Minneapolis, the average

**Table 2**
Statistical analysis of visual elements of waterscapes in 16 cities.

| | water | sky | terrain | traditional-building | modern-building | revetment | bridge | car | truck | bicycle | boat | tree | grass | people | void |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| mean | 0.4658 | 0.3712 | 0.0000 | 0.0297 | 0.0300 | 0.0224 | 0.0120 | 0.0000 | 0.0006 | 0.0000 | 0.0062 | 0.0618 | 0.0002 | 0.0000 | 0.0001 |
| std | 0.0254 | 0.0878 | 0.0000 | 0.0561 | 0.0475 | 0.0193 | 0.0120 | 0.0001 | 0.0023 | 0.0000 | 0.0049 | 0.0539 | 0.0002 | 0.0000 | 0.0001 |
| min | 0.4232 | 0.1485 | 0.0000 | 0.0029 | 0.0009 | 0.0061 | 0.0021 | 0.0000 | 0.0000 | 0.0000 | 0.0011 | 0.0082 | 0.0000 | 0.0000 | 0.0000 |
| 25% | 0.4416 | 0.3368 | 0.0000 | 0.0060 | 0.0028 | 0.0113 | 0.0058 | 0.0000 | 0.0000 | 0.0000 | 0.0037 | 0.0216 | 0.0000 | 0.0000 | 0.0000 |
| 50% | 0.4720 | 0.4118 | 0.0000 | 0.0120 | 0.0123 | 0.0134 | 0.0071 | 0.0000 | 0.0000 | 0.0000 | 0.0050 | 0.0333 | 0.0001 | 0.0000 | 0.0001 |
| 75% | 0.4827 | 0.4318 | 0.0000 | 0.0233 | 0.0287 | 0.0297 | 0.0117 | 0.0000 | 0.0000 | 0.0000 | 0.0070 | 0.0908 | 0.0005 | 0.0000 | 0.0002 |
| max | 0.5069 | 0.4633 | 0.0001 | 0.2328 | 0.1747 | 0.0782 | 0.0483 | 0.0002 | 0.0093 | 0.0000 | 0.0192 | 0.1623 | 0.0007 | 0.0000 | 0.0005 |

proportion of these two sets of features is rather high, and the sum of the two is close to 80%. Other factors have proportions and spatial distributions that are similar to those of Warsaw and Washington. The features of Amsterdam and Chicago, on the other hand, are vastly different from those of the previous places. In Amsterdam and Chicago, the average share of SVF is small, at 23.34% and 14.85%, respectively, yet there is a noticeable variation in space. Both have a high overall proportion of WVI (47.12% and 43.31%, respectively), as well as a high geographical volatility. The average percentage of GVF in Amsterdam is roughly 15.65%, which is about three times that of Chicago, and both of them fluctuate in space. The same traits may be seen in DF, and Amsterdam's average DF is 2.85%, which is over three times that of Chicago (0.97 percent). HRF and BVF, on the other hand, have diametrically opposed features. Amsterdam's average HRF is only 3.34%, which is less than half of Chicago's (7.82 percent). The average BVF in Amsterdam is 6.09%, compared to 22.58% in Chicago; however both cities differ substantially in terms of space. Venice has great distinctiveness when compared to other cities' waterscape features. In Venice, there are few trees, and thus, their GVF proportion is tiny (1.14 percent), whereas the average ratio of BVF is very high (24.06 percent), with a huge spatial variation. Similarly, SVF, HRF, and WVF all display significant spatial shifts, although DF's changing features are less obvious. Other cities, on the other hand, exhibit the features that WVF and SVF account for a high share of the population on average, and the geographical change of six categories of assessment indicators is not apparent.

### 4.2. Subjective visual perception results

From the subjective perception result of different cities (Fig. 6), we observe that the average scores of 16 cities in safe, lively and beautiful are less varied, which are 4.43, 4.21 and 4.34, respectively. Boston has the highest average value in the three items (safe, lively, and beautiful), followed by Chicago, Amsterdam, Venice and Washington, with an average score of more than 5.0. Phnom Penh and Bangkok, on the other hand, were ranked last and second to last, respectively, with an average score of less than 3.0. Chicago has a very high wealthy value of 6.55, followed by Venice, Amsterdam, Tokyo, London, Paris, and San Diego, all of which have scores ranging from 4 to 5, while some cities have wealthy scores lower than the average of 3.32, including Phnom Penh, Warsaw, and Memphis. The average depression score in 16 cities is 3.03, with Chicago having the highest value of 5.86, which is significantly higher than Tokyo's (3.89), the second, while Bangkok, Phnom Penh, London, and other cities are close behind, with values of 3.89, 3.44, and 3.40, respectively. Boston (1.52), had the lowest depression score, and this demonstrates that its urban waterscape may provide a great sense of relaxation, and that individuals will not become depressed as a result. Phnom Penh (6.01) had the highest score in boring's rating, followed by Bangkok (5.36) and Memphis (4.95). In contrast, Venice scored the lowest, only 2.83, followed by San Diego (2.97), Amsterdam and Boston both scored 3.13.

### 4.3. Relationship between objective visual analysis and subjective visual perception

Pearson correlation analysis with objective visual analysis indexes was applied to validate the multicollinearity of the six indexes, and the pairwise comparison matrix of the correlation coefficients based on the bilateral significance test is shown in Table 3. The results indicated that most of the correlation coefficients of the indices have a low or negative correlation, which illustrates that the explanatory variables are relatively independent of each other.

We analyzed survey responses using a stepwise regression analysis, where the mean subjective perception ratings of the 140 scenes were the dependent variable, and the CV-based objective evaluation indexes were the independent variables. Stepwise regression is a method of fitting
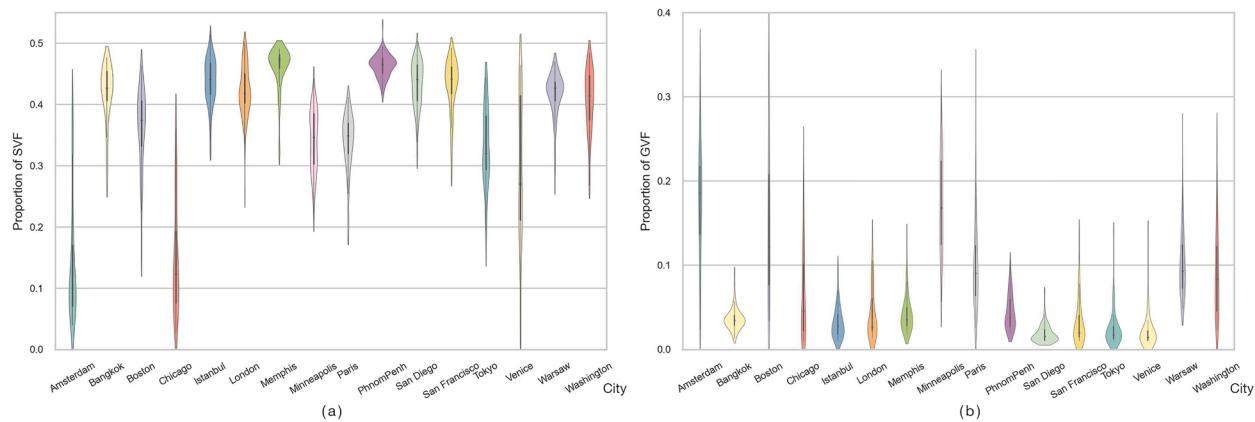
**Fig. 4.** Comparison of the distribution of two example indexes among cities: (a) proportion of SVF; and (b) proportion of GVF.

regression models in which the choice of predictive variables is carried out by an automatic procedure, and it can effectively reduce autocorrelation (Wu et al., 2020). In our study, we used a bidirectional elimination approach for stepwise regression.

Table 4 shows the results of regression analysis between subjective visual perception and objective visual analysis indicators. We can observe the influence degree of visual indicators from the value of betas — standardization coefficient and the * implies the significance level. Overall, we discovered that different objective visual indicators contribute significantly diverse amounts to perceived contents. The proportion of HRF and DF, for example, is positively connected with 'wealthy' scores, but SVF and GVF are negatively correlated. GVF has a slightly negative influence on 'wealthy', whereas SVF has a more significant negative impact. This demonstrates how a waterscape with a hard revetment and places with more boats and vehicles may easily convey a sense of wealth to people, while open places with a high proportion of sky and natural rivers with more vegetation cover, on the other hand, will not give a sense of wealth. HRF had a positive impact on 'depression' in this study, whereas SVF and GVF had a negative effect on 'depression' (Wang et al., 2019; Wang et al., 2019). As a result, too much artificial hard revetment in the water landscape will not only increase people's perception of prosperity, but also their feeling of 'depression'. Natural variables like SVF and GVF, as well as dynamic factors such as boats and cars, can successfully lessen people's feelings of 'depression'. GVF and DF have a beneficial influence on people's perceptions of 'beautiful', whereas SVF has a negative impact. Positive of 'beautiful' contributors include not only green vegetation like trees and grass, but also dynamic signs like boats on the water, cars and people along the shore. As a result, the beauty of a waterscape is the result of a mix of artificial and natural elements. At the same time, we can observe how excessive sky visibility or an open waterscape can detract from the beauty of the waterscape. This conclusion is remarkably comparable to the research by Zhang et al. (2018) and others' street perception studies. People's sense of safety is also very important for urban waterscape. In our research, we can find that BVF, GVF, HRF and DF all have a positive impact on the sense of 'safe'; among them, BVF and GVF have a strong positive impact on it. Furthermore, we can see that 'lively' is positively connected with GVF and DF, but negatively with SVF. 'Boring' was negatively correlated with DF, WVF, GVF, and BVF, and the influence degree of these four factors increased. This emphasizes the relevance of riparian vegetation and dynamic objects in improving the waterscape's vitality value. Simultaneously, the combination of these elements plus BVF can effectively alleviate waterscape boredom.

## 5. Discussion

### 5.1. The perspective of water view imagery

The city's waterscape is inextricably linked to the lives of its inhabitants. The area's diverse sorts of the physical environment will have an impact on human feelings, activities, and even physical and mental health (Bedla and Halecki, 2021). In the same way that urban streets, parks, and community environments have become research directions for multiple disciplines, the significance of the built environment's waterscape has been recognized by different fields of research. WVI data gives a unique view of urban watercourses from the water-level perspective, and it is an important data source for people to understand the waterscapes. It is beneficial for in-depth study of the characteristics of urban waterscape and the design of water tour routes through visual perception analysis of WVI data. We used computer vision to provide an impartial visual evaluation of the waterscape in 16 cities in six aspects in this context. We also employed virtual reality to gather people's subjective perceptions of these waterscapes, such as their feelings of safety, liveliness, and boredom when viewing these water scenes (Fig. 7). The findings of this research reveal a number of previously unknown phenomena. Venice and Amsterdam, for example, are both in Europe and have well-known canals and water tourism routes. There are several parallels and contrasts between the two cities' water features. Venice and Amsterdam are both quite visible in terms of architecture and dynamic elements, but there are significant disparities in terms of green visibility. Green vegetation is rarely visible on boat tours in Venice, but in Amsterdam, on the other hand, greenery is extensive. The subjective perceptions evoked by the two cities' waterscapes are strikingly similar. The two cities, for example, are in the top four in terms of safety, lively, beauty, and wealth, while their depression scores are in the middle. Venice came in bottom place in terms of boredom, while Amsterdam came in 12th. The Seine River in Paris and the Thames River in London are two of Europe's most well-known rivers. The two metropolitan waterscapes score in the top 7 in terms of architectural visibility, the proportion of hard revetment, and the percentage of dynamic factors, according to the objective visual evaluation results; Paris ranks fifth in green viewing rate, while London ranks tenth. In other words, touring on ships cannot see much more vegetation while appreciating London's waterscape. In people's six subjective perceptions, the two cities are almost equally ranked, and not particularly high. Furthermore, the waterscapes of Chicago, Phnom Penh, and other cities have distinct characteristics, a finding that could have been reached only through a global comparative study such as ours.

### 5.2. Impact mechanism

With respect to a related previous study (Li et al., 2021), our research

**Fig. 5.** Objective visual analysis results of each city.

covers a wider range, more cities and richer dimensions of people's subjective perception. This study brings attention to the concept of water view imagery, and to an extent, it doubles as a position paper. Further, the research quantitatively exposes the relationship between the physical environment of water perspectives and the human experience of the place using regression analysis. Specifically, the green viewing rate of on-water perspective, we discovered, has a favourable impact on several perceptual indexes such as safety, vitality, and beauty

of the urban waterscapes. At the same time, the greenery has been shown to reduce people's feelings of depression and boredom, which is consistent with the findings of many street view green viewing rates studies (Wang et al., 2019; Li et al., 2015; Kang et al., 2020). Therefore, it can be demonstrated that the influence of greenery on people's perception applies not only to streetscapes, but also to riverscapes. Furthermore, dynamic factors such as boats and cars are linked to people's perceptions of vitality. Our findings reveal that these dynamic
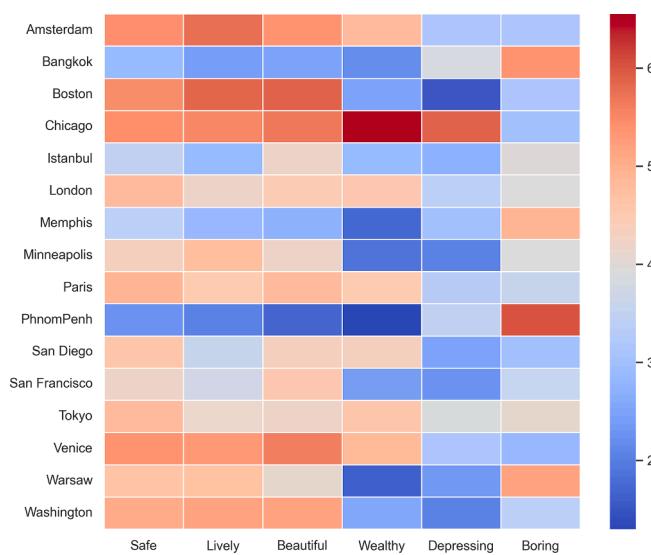
**Fig. 6.** The average subjective perception result of each city.

**Table 3**
Correlation analysis with visual evaluation indexes.

|  | GVF | BVF | WVF | SVF | HRF | DF |
|---|---|---|---|---|---|---|
| GVF | 1 |  |  |  |  |  |
| BVF | −0.362** | 1 |  |  |  |  |
| WVF | 0.001 | −0.745** | 1 |  |  |  |
| SVF | −0.189* | −0.754** | 0.571** | 1 |  |  |
| HRF | −0.023 | 0.345** | −0.258** | −0.639** | 1 |  |
| DF | −0.096 | 0.406** | −0.342** | −0.428** | 0.227** | 1 |

\* p<0.05 \*\* p<0.01

elements contribute to a sense of safety and beauty in the waterscapes and are positively connected with wealth and adversely correlated with depression and boredom. The SVF or openness of the urban waterscape, in addition to DF and GVF, is a significant factor. People's perceptions of waterscape liveliness will be considerably diminished if sceneries are too open. As a result, improving the vegetation coverage on the river bank, attracting ships, cars, and choosing a route with less openness for water tourism are all valuable strategies for enhancing the vitality of urban waterscapes. Urban inhabitants' mental health is also a critical research topic, and many studies have shown that waterscape can help urban residents feel less stressed (Völker et al., 2018; Subiza-Pérez et al., 2020). More vegetation, more open waterscape and dynamic elements, and less hard revetment on the river bank, according to our research, are all beneficial to lessening people's feelings of depression, and this could serve as a guide for creating restorative waterscapes and choosing water tour routes. Besides, our study samples are sufficiently representative because these water areas come from different continents and countries; thus, the relationship to perceptions is much clearer than before. In essence, people's feelings toward ground objects are reflected in urban perception; therefore, policymakers should pay attention to the visual proportion of various ground elements.

## 5.3. Limitations, challenges and future directions

This study faces some challenges and limitations. Firstly, although we have found WVI data from dozens of cities, the WVI imagery quality of different places varies, and there is still a lack of global sequential waterscape photographs with a consistent angle. For example, we could not find any WVI panoramic image from South America and Africa on different platforms. Some city's water level panoramic images have obstacles, which will have a certain impact on the comprehensive understanding of the waterscapes. Therefore, although we found a total of more than 30 cities with continuous WVI data on the two platforms (Table 1), we finally selected only 16 of them as the research areas. Data sources have a significant impact on urban waterscape research, but

**Table 4**
Stepwise regression analysis results.

| Dependent Variables | Independent Variables | Unstandardized Coefficients | | Standardized Coefficients | t | p | VIF | $R^2$ | Adjusted $R^2$ | F |
|---|---|---|---|---|---|---|---|---|---|---|
|  |  | B | Std. Error | Beta |  |  |  |  |  |  |
| Wealthy | Constant | 5.366 | 0.53 | – | 10.125 | 0.000** | – | 0.601 | 0.589 | F (4,135)=50.856,p = 0.000 |
|  | GVF | −6.203 | 1.317 | −0.272 | −4.709 | 0.000** | 1.129 |  |  |  |
|  | HRF | 18.144 | 4.33 | 0.303 | 4.19 | 0.000** | 1.775 |  |  |  |
|  | DF | 24.8 | 10.127 | 0.151 | 2.449 | 0.016* | 1.29 |  |  |  |
|  | SVF | −6.404 | 1.167 | −0.443 | −5.486 | 0.000** | 2.204 |  |  |  |
| Boring | Constant | 9.389 | 1.766 | – | 5.315 | 0.000** | – | 0.253 | 0.23 | F (4,135)=11.401,p = 0.000 |
|  | GVF | −5.466 | 1.391 | −0.348 | −3.93 | 0.000** | 1.416 |  |  |  |
|  | BVF | −5.242 | 1.246 | −0.572 | −4.206 | 0.000** | 3.343 |  |  |  |
|  | WVF | −9.429 | 3.464 | −0.337 | −2.722 | 0.007** | 2.764 |  |  |  |
|  | DF | −31.021 | 9.223 | −0.275 | −3.363 | 0.001** | 1.204 |  |  |  |
| Depressing | Constant | 4.703 | 0.425 | – | 11.075 | 0.000** | – | 0.613 | 0.602 | F (4,135)=53.534,p = 0.000 |
|  | GVF | −8.094 | 1.056 | −0.436 | −7.668 | 0.000** | 1.129 |  |  |  |
|  | SVF | −4.411 | 0.935 | −0.375 | −4.716 | 0.000** | 2.204 |  |  |  |
|  | HRF | 21.726 | 3.47 | 0.446 | 6.262 | 0.000** | 1.775 |  |  |  |
|  | DF | −24.451 | 8.115 | −0.183 | −3.013 | 0.003** | 1.29 |  |  |  |
| Beautiful | Constant | 5.491 | 0.384 | – | 14.304 | 0.000** | – | 0.326 | 0.311 | F (3,136)=21.883,p = 0.000 |
|  | GVF | 3.695 | 1.376 | 0.197 | 2.685 | 0.008** | 1.08 |  |  |  |
|  | SVF | −4.79 | 0.961 | −0.402 | −4.986 | 0.000** | 1.309 |  |  |  |
|  | DF | 25.988 | 10.751 | 0.192 | 2.417 | 0.017* | 1.275 |  |  |  |
| Lively | Constant | 5.335 | 0.368 | – | 14.511 | 0.000** | – | 0.398 | 0.385 | F (3,135)=29.799,p = 0.000 |
|  | GVF | 6.009 | 1.318 | 0.316 | 4.557 | 0.000** | 1.081 |  |  |  |
|  | SVF | −5.057 | 0.92 | −0.419 | −5.498 | 0.000** | 1.304 |  |  |  |
|  | DF | 22.175 | 10.361 | 0.161 | 2.14 | 0.034* | 1.271 |  |  |  |
| Safe | Constant | 3.45 | 0.156 | – | 22.067 | 0.000** | – | 0.389 | 0.37 | F (4,135)=21.447,p = 0.000 |
|  | GVF | 5.982 | 1.017 | 0.428 | 5.881 | 0.000** | 1.17 |  |  |  |
|  | BVF | 3.69 | 0.674 | 0.453 | 5.474 | 0.000** | 1.51 |  |  |  |
|  | HRF | 5.527 | 2.657 | 0.151 | 2.08 | 0.039* | 1.162 |  |  |  |
|  | DF | 14.785 | 7.449 | 0.147 | 1.985 | 0.049* | 1.213 |  |  |  |

| Which city looks greener ? | Which city has more buildings ? | Which can see more water ? |
|---|---|---|
| 1 Minneapolis | 1 Venice | 1 San Diego |
| 2 Amsterdam | 2 Chicago | 2 San Francisco |
| 3 Boston | 3 Tokyo | 3 Washington |
| 4 Warsaw | 4 Amsterdam | 4 Memphis |
| 5 Paris | 5 London | 5 Istanbul |
| 6 Washington | 6 Bangkok | 6 Phnom Penh |
| 7 Chicago | 7 Paris | 7 Bangkok |
| 8 Phnom Penh | 8 Istanbul | 8 Tokyo |
| 9 Memphis | 9 San Diego | 9 Amsterdam |
| 10 London | 10 San Francisco | 10 Minneapolis |
| 11 Bangkok | 11 Boston | 11 London |
| 12 San Francisco | 12 Warsaw | 12 Warsaw |
| 13 Istanbul | 13 Memphis | 13 Boston |
| 14 Tokyo | 14 Phnom Penh | 14 Chicago |
| 15 Venice | 15 Washington | 15 Paris |
| 16 San Diego | 16 Minneapolis | 16 Venice |

(Green visibility) (Building visibility) (Water visibility)

| Which can see more sky ? | Which can see more hard revetment ? | Which can see more dynamic elements ? |
|---|---|---|
| 1 Phnom Penh | 1 Chicago | 1 Amsterdam |
| 2 Memphis | 2 Paris | 2 Venice |
| 3 Istanbul | 3 Tokyo | 3 Paris |
| 4 San Diego | 4 Amsterdam | 4 Chicago |
| 5 San Francisco | 5 London | 5 Warsaw |
| 6 London | 6 Warsaw | 6 Istanbul |
| 7 Bangkok | 7 Bangkok | 7 London |
| 8 Warsaw | 8 Boston | 8 Tokyo |
| 9 Washington | 9 Istanbul | 9 Bangkok |
| 10 Boston | 10 San Francisco | 10 San Diego |
| 11 Minneapolis | 11 Venice | 11 Boston |
| 12 Paris | 12 Minneapolis | 12 Phnom Penh |
| 13 Tokyo | 13 San Diego | 13 San Francisco |
| 14 Venice | 14 Washington | 14 Minneapolis |
| 15 Amsterdam | 15 Phnom Penh | 15 Washington |
| 16 Chicago | 16 Memphis | 16 Memphis |

(Sky visibility) (Hard revetment visibility) (Dynamic elements visibility)

| Which city looks safer ? | Which city looks more lively ? | Which city looks more beautiful ? |
|---|---|---|
| 1 Boston | 1 Boston | 1 Boston |
| 2 Chicago | 2 Amsterdam | 2 Chicago |
| 3 Amsterdam | 3 Chicago | 3 Venice |
| 4 Venice | 4 Venice | 4 Amsterdam |
| 5 Washington | 5 Washington | 5 Washington |
| 6 Paris | 6 Minneapolis | 6 Paris |
| 7 London | 7 Warsaw | 7 San Francisco |
| 8 Tokyo | 8 Paris | 8 London |
| 9 Warsaw | 9 London | 9 San Diego |
| 10 San Diego | 10 Tokyo | 10 Minneapolis |
| 11 Minneapolis | 11 San Francisco | 11 Tokyo |
| 12 San Francisco | 12 San Diego | 12 Istanbul |
| 13 Istanbul | 13 Istanbul | 13 Warsaw |
| 14 Memphis | 14 Memphis | 14 Memphis |
| 15 Bangkok | 15 Bangkok | 15 Bangkok |
| 16 Phnom Penh | 16 Phnom Penh | 16 Phnom Penh |

(Safe) (Lively) (Beautiful)

| Which city looks more wealthy ? | Which makes you more depressed ? | Which city looks more boring ? |
|---|---|---|
| 1 Chicago | 1 Chicago | 1 Phnom Penh |
| 2 Venice | 2 Tokyo | 2 Bangkok |
| 3 Amsterdam | 3 Bangkok | 3 Warsaw |
| 4 Tokyo | 4 Phnom Penh | 4 Memphis |
| 5 London | 5 London | 5 Tokyo |
| 6 Paris | 6 Paris | 6 Istanbul |
| 7 San Diego | 7 Venice | 7 London |
| 8 Istanbul | 8 Amsterdam | 8 Minneapolis |
| 9 Washington | 9 Memphis | 9 Paris |
| 10 Boston | 10 Istanbul | 10 San Francisco |
| 11 San Francisco | 11 San Diego | 11 Washington |
| 12 Bangkok | 12 Warsaw | 12 Amsterdam |
| 13 Minneapolis | 13 San Francisco | 13 Boston |
| 14 Memphis | 14 Washington | 14 Chicago |
| 15 Warsaw | 15 Minneapolis | 15 San Diego |
| 16 Phnom Penh | 16 Boston | 16 Venice |

(Wealthy) (Depressing) (Boring)

**Fig. 7.** Ranking of objective visual analysis and subjective visual perception of different cities.

with the continued expansion of crowdsourcing data, additional study options will be available in the near future. Secondly, as it is the case with most street view studies, the images may have been taken during different seasons. We realized this heterogeneity and verified the images of different regions to avoid the study being influenced by seasonal changes, such as the difference between winter and summer vegetation in high latitudes. Thirdly, this is the first study on the connection between visual elements in waterscape imagery and human perception at a large-scale. Human vision is one of the primary means by which we perceive our surroundings. Yet, we acknowledge that human perception of the water environment is not solely visual; it is also influenced by the culture people experience, the sounds they hear, and the activities they engage in, which are difficult to portray through visual imagery (Li et al., 2021). The research on human perception of various water landscapes may be expanded in the future to include more dimensions such as point of interest (POI), soundscape analysis, and so on. In addition, as a comprehensive comparative study of multiple cities, this study rarely considers the difference of characteristics in different sections of the same city, and subsequent studies can choose different sections of the same river for self-comparison.

High quality image semantic segmentation dataset is the principal fuel of computer vision related research. We collected 500 photographs of various types of water features from eight cities on Mapillary and labeled them at pixel level. This is the first open semantic segmentation dataset from the standpoint of water that we are aware of. With the help of this dataset for understanding these water landscapes semantically, the findings of this study can be used to directly assist urban waterscape design concepts and projects, as well as the selection of water tour routes in various cities. Furthermore, using rendering technology, the results of this study can be used to analyze the future scenes of an urban waterscape and predict perceptions of future scenarios. For example, the generative adversarial nets (GAN) can be used to show a specified proportion of visual factors in order to establish various urban scenes (Sun et al., 2022; Wu and Biljecki, 2022; Wu et al., 2022), as well as to give information for the transformation and optimization of the urban water areas. This research also paves the way for the enhancement of urban digital twins, which may incorporate human perception data into the scene knowledge of urban waterscapes and create a two-way coupling between the digital system, the physical environment, and human perception (Luo et al., 2022).

## 6. Conclusion

Rivers, lakes, bays and other waterscapes in cities are important ecological resources. Waterfronts also serve a multitude of purposes, including leisure and enjoyment for city dwellers and visitors. Large-scale quantitative assessment of waterscape characteristics of different areas and measuring human subjective perception of diverse waterscapes have always been difficult due to a lack of proper data and processing technologies. The contributions of this paper are as follows.

Firstly, this study brings attention to water view imagery (WVI) — a counterpart of ground-level street view imagery that is taken on water, but unlike its ground-level sibling, it is virtually unknown and not taken advantage of. We promote the concept, identify continuous waterscape imagery of different cities in Asia, Europe, North America, and Oceania using GSV and Mapillary, and assess their characteristics and quality in dozens of cities. With the developed methodology and obtained results, we demonstrate that they can be used for a particular (and novel) purpose in urban planning. We hope that by giving currency to this kind of data, and thanks to the dataset we created and released openly, we will witness the development of further use cases relying on this overlooked urban dataset. Another contribution is general and to research arena relying on street view imagery: to maximise the scope of the study, we tap into multiple datasets (Google Street View and Mapillary), which is a rarity as virtually all SVI-driven studies use a single data source (Biljecki and Ito, 2021).

Secondly, we apply computer vision technology to analyze the proportion of visual elements in different waterscape photos to understand the information of different scenes, achieve the automatic and efficient visual analysis of waterscape environments, and quantitatively evaluate the objective visual features of 16 cities around the world. One of the findings is that some cities have more heterogeneous riverscapes than others that tend to be monotonous along the paths we have focused. Furthermore, this study used VR equipment to obtain human perception results in six dimensions: safe, lively, beautiful, wealthy, depressing, and boring, based on a comprehensive survey.

Finally, it quantitatively analyzes the relationship between the objective visual analysis results and the subjective visual perception results of different scenes, and discusses the objective visual factors affecting human subjective visual perceptions through a regression analysis. Specifically, green vegetation can induce positive perception, while hard revetment will drive feelings of depression, and perception of liveliness will be considerably reduced if the scene is too empty, which is compatible with the literature in related fields focusing on other types of landscapes. We also discovered that, according to results in SVI studies focused on the ground, these relationships have similarities with those of streetscapes, such as greenery playing an important role in reducing feelings of depression and boredom.

The findings of this study can be used to promote research and practice in urban waterscape design, water transportation routes selection, as well as demonstrate the utility of employing computer vision to do visual evaluations of large-scale waterscapes. At the same time, this research may serve as a guide for scholars interested in learning more about the objective visual qualities and subjective perception outcomes of various urban waterscapes.

## CRediT authorship contribution statement

**Junjie Luo:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Tianhong Zhao:** Software, Formal analysis, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Lei Cao:** Resources, Supervision, Project administration, Funding acquisition. **Filip Biljecki:** Formal analysis, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Heartbeats to Cities, which is supported by the Singapore Ministry of Education Academic Research Fund Tier 1.

## References

Ahmed, I., Jeon, G., Chehri, A., Hassan, M.M., 2021. Adapting gaussian yolov3 with transfer learning for overhead view human detection in smart cities and societies. Sustain. Cities Soc. 70, 102908.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 2481–2495.

Bao, C., Wang, H., Sun, S., 2022. Comprehensive simulation of resources and environment carrying capacity for urban agglomeration: A system dynamics approach. Ecol. Ind. 138, 108874.

Bedla, D., Halecki, W., 2021. The value of river valleys for restoring landscape features and the continuity of urban ecosystem functions–a review. Ecol. Ind. 129, 107871.

Biljecki, F., Ito, K., 2021. Street view imagery in urban analytics and gis: A review. Landscape Urban Plann. 215, 104217.

Birenboim, A., Dijst, M., Ettema, D., de Kruijf, J., de Leeuw, G., Dogterom, N., 2019. The utilization of immersive virtual environments for the investigation of environmental preferences. Landscape Urban Plann. 189, 129–138.

Cai, B.Y., Li, X., Seiferling, I., Ratti, C., 2018. Treepedia 2.0: Applying deep learning for large-scale quantification of urban tree cover. In: 2018 IEEE International Congress on Big Data (BigData Congress), IEEE, pp. 49–56.

Cao, J., Zhang, M., 2020. Evaluation of urban waterfront landscape quality based on semantic differential method: a case of zhonghuamen section of qinhuai river in nanjing. J. Nanjing Forestry Univ. 44, 221.

Chen, J., Zhou, C., Li, F., 2020. Quantifying the green view indicator for assessing urban greening quality: An analysis based on internet-crawling street view data. Ecol. Ind. 113, 106192.

Chen, Y., Wang, Y., Lu, P., Chen, Y., Wang, G., 2018. Large-scale structure from motion with semantic constraints of aerial images. In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Springer, pp. 347–359.

Cheng, S., Wang, J., 2021. Visual landscape research on dynamic viewing evaluation: A case study of nanjing riverside public viewing space. Chin. Landscape Architect. 37, 57–62.

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. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3213–3223.

Dang, H., Li, J., 2021. The integration of urban streetscapes provides the possibility to fully quantify the ecological landscape of urban green spaces: A case study of xi'an city. Ecol. Ind. 133, 108388.

De Vries, S., Ten Have, M., van Dorsselaer, S., van Wezep, M., Hermans, T., de Graaf, R., 2016. Local availability of green and blue space and prevalence of common mental disorders in the netherlands. BJPsych Open 2, 366–372.

Dubey, A., Naik, N., Parikh, D., Raskar, R., Hidalgo, C.A., 2016. Deep learning the city: Quantifying urban perception at a global scale. European conference on computer vision, Springer 196–212.

Duchowski, A.T., Duchowski, A.T., 2017. Eye tracking methodology: Theory and practice. Springer.

Gargiulo, I., Garcia, X., Benages-Albert, M., Martinez, J., Pfeffer, K., Vall-Casas, P., 2020. Women's safety perception assessment in an urban stream corridor: Developing a safety map based on qualitative gis. Landscape Urban Plann. 198, 103779.

Ghermandi, A., Depietri, Y., Sinclair, M., 2022. In the ai of the beholder: A comparative analysis of computer vision-assisted characterizations of human-nature interactions in urban green spaces. Landscape Urban Plann. 217, 104261.

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. Build. Environ. 134, 155–167.

Grêt-Regamey, A., Weibel, B., Vollmer, D., Burlando, P., Girot, C., 2016. River rehabilitation as an opportunity for ecological landscape design. Sustain. Cities Soc. 20, 142–146.

Hao, Y., Liu, Y., Wu, Z., Han, L., Chen, Y., Chen, G., Chu, L., Tang, S., Yu, Z., Chen, Z., et al., 2021. Edgeflow: Achieving practical interactive segmentation with edge-guided flow. arXiv preprint arXiv:2109.09406.

Hosseini, M., Miranda, F., Lin, J., Silva, C.T., 2022. Citysurfaces: City-scale semantic segmentation of sidewalk materials. Sustain. Cities Soc. 103630.

Hua, J., Chen, W.Y., 2019. Prioritizing urban rivers' ecosystem services: An importance-performance analysis. Cities 94, 11–23.

Huerta, R.E., Yépez, F.D., Lozano-García, D.F., Guerra Cobián, V.H., Ferriño Fierro, A.L., de León Gómez, H., Cavazos González, R.A., Vargas-Martínez, A., 2021. Mapping urban green spaces at the metropolitan level using very high resolution satellite imagery and deep learning techniques for semantic segmentation. Remote Sens. 13, 2031.

Ibrahim, M.R., Haworth, J., Cheng, T., 2020. Understanding cities with machine eyes: A review of deep computer vision in urban analytics. Cities 96, 102481.

Ito, K., Biljecki, F., 2021. Assessing bikeability with street view imagery and computer vision. Transp. Res. Part C: Emerg. Technol. 132, 103371.

Jeon, J.Y., Jo, H.I., 2020. Effects of audio-visual interactions on soundscape and landscape perception and their influence on satisfaction with the urban environment. Build. Environ. 169, 106544.

Jin, X., Wang, J., 2021. Assessing linear urban landscape from dynamic visual perception based on urban morphology. Front. Architect. Res. 10, 202–219.

Kandt, J., Batty, M., 2021. Smart cities, big data and urban policy: Towards urban analytics for the long run. Cities 109, 102992.

Kang, Y., Zhang, F., Gao, S., Lin, H., Liu, Y., 2020. A review of urban physical environment sensing using street view imagery in public health studies. Ann. GIS 26, 261–275.

Keten, A., Eroglu, E., Kaya, S., Anderson, J.T., 2020. Bird diversity along a riparian corridor in a moderate urban landscape. Ecol. Ind. 118, 106751.

Kotowska, D., Pärt, T., Żmihorski, M., 2021. Evaluating google street view for tracking invasive alien plants along roads. Ecol. Ind. 121, 107020.

Krause, C.L., 2001. Our visual landscape: Managing the landscape under special consideration of visual aspects. Landscape Urban Plann. 54, 239–254.

Larkin, A., Hystad, P., 2019. Evaluating street view exposure measures of visible green space for health research. J. Exposure Sci. Environ. Epidemiol. 29, 447–456.

Lee, V.W., Hodgson, P., Chan, C.S., Fong, A., Cheung, S.W., 2020. Optimising the learning process with immersive virtual reality and non-immersive virtual reality in an educational environment. Int. J. Mobile Learn. Organis. 14, 21–35.

Leslie, E., Coffee, N., Frank, L., Owen, N., Bauman, A., Hugo, G., 2007. Walkability of local communities: using geographic information systems to objectively assess relevant environmental attributes. Health Place 13, 111–122.

Li, D., Zhou, X., Wang, M., 2018. Analyzing and visualizing the spatial interactions between tourists and locals: A flickr study in ten us cities. Cities 74, 249–258.

Li, W., He, C., Fang, J., Zheng, J., Fu, H., Yu, L., 2019. Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source gis data. Remote Sens. 11, 403.

Li, X., Li, L., Wang, X., Lin, Q., Wu, D., Dong, Y., Han, S., 2021. Visual quality evaluation model of an urban river landscape based on random forest. Ecol. Ind. 133, 108381.

Li, X., Li, Y., Jia, T., Zhou, L., Hijazi, I.H., 2022. The six dimensions of built environment on urban vitality: fusion evidence from multi-source data. Cities 121, 103482.

Li, X., Ratti, C., Seiferling, I., 2018. Quantifying the shade provision of street trees in urban landscape: A case study in boston, usa, using google street view. Landscape Urban Plann. 169, 81–91.

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 Green. 14, 675–685.

Li, Y., Yabuki, N., Fukuda, T., 2022. Exploring the association between street built environment and street vitality using deep learning methods. Sustain. Cities Soc. 79, 103656.

Liao, B., van den Berg, P.E., van Wesemael, P.J., Arentze, T.A., 2022. Individuals' perception of walkability: Results of a conjoint experiment using videos of virtual environments. Cities 125, 103650.

Likert, R., 1932. A technique for the measurement of attitudes. Archives of psychology.

Liu, L., Zhou, B., Zhao, J., Ryan, B.D., 2016. C-image: city cognitive mapping through geo-tagged photos. GeoJournal 81, 817–861.

Liu, P., Biljecki, F., 2022. A review of spatially-explicit geoai applications in urban geography. Int. J. Appl. Earth Obs. Geoinf. 112, 102936 https://doi.org/10.1016/j.jag.2022.102936.

Liu, S., Lai, S.Q., Liu, C., Jiang, L., 2021. What influenced the vitality of the waterfront open space? a case study of huangpu river in shanghai, china. Cities 114, 103197.

Liu, W., Yue, A., Shi, W., Ji, J., Deng, R., 2019. An automatic extraction architecture of urban green space based on deeplabv3plus semantic segmentation model. In: 2019 IEEE 4th International Conference on Image. Vision and Computing (ICIVC), IEEE, pp. 311–315.

Luo, J., Liu, P., Cao, L., 2022. Coupling a physical replica with a digital twin: A comparison of participatory decision-making methods in an urban park environment. ISPRS Int. J. Geo-Inform. 11, 452.

Luo, J., Zhao, T., Cao, L., Biljecki, F., 2022. Semantic riverscapes: Perception and evaluation of linear landscapes from oblique imagery using computer vision. Landscape Urban Plann. 228, 104569.

Lyu, Y., Vosselman, G., Xia, G.S., Yilmaz, A., Yang, M.Y., 2020. Uavid: A semantic segmentation dataset for uav imagery. ISPRS J. Photogramm. Remote Sens. 165, 108–119.

Ma, X., Ma, C., Wu, C., Xi, Y., Yang, R., Peng, N., Zhang, C., Ren, F., 2021. Measuring human perceptions of streetscapes to better inform urban renewal: A perspective of scene semantic parsing. Cities 110, 103086.

Maffei, L., Masullo, M., Pascale, A., Ruggiero, G., Romero, V.P., 2016. Immersive virtual reality in community planning: Acoustic and visual congruence of simulated vs real world. Sustain. Cities Soc. 27, 338–345.

Mao, L., Zheng, Z., Meng, X., Zhou, Y., Zhao, P., Yang, Z., Long, Y., 2022. Large-scale automatic identification of urban vacant land using semantic segmentation of high-resolution remote sensing images. Landscape Urban Plann. 222, 104384.

Meng, C., Song, Y., Ji, J., Jia, Z., Zhou, Z., Gao, P., Liu, S., 2021. Automatic classification of rural building characteristics using deep learning methods on oblique photography. Building Simulation, Springer 1–14.

Nigam, I., Huang, C., Ramanan, D., 2018. Ensemble knowledge transfer for semantic segmentation. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, pp. 1499–1508.

Osco, L.P., Junior, J.M., Ramos, A.P.M., de Castro Jorge, L.A., Fatholahi, S.N., de Andrade Silva, J., Matsubara, E.T., Pistori, H., Gonçalves, W.N., Li, J., 2021. A review on deep learning in uav remote sensing. Int. J. Appl. Earth Obs. Geoinf. 102, 102456.

Pflüger, Y., Rackham, A., Larned, S., 2010. The aesthetic value of river flows: An assessment of flow preferences for large and small rivers. Landscape Urban Plann. 95, 68–78.

Richards, D.R., Friess, D.A., 2015. A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. Ecol. Ind. 53, 187–195.

Seiferling, I., Naik, N., Ratti, C., Proulx, R., 2017. Green streets- quantifying and mapping urban trees with street-level imagery and computer vision. Landscape Urban Plann. 165, 93–101.

Silva, J., da Graça Saraiva, M., Loupa Ramos, I., Bernardo, F., 2013. Improving visual attractiveness to enhance city–river integration—a methodological approach for ongoing evaluation. Planning Practice Res. 28, 163–185.

Song, X.P., Richards, D.R., He, P., Tan, P.Y., 2020. Does geo-located social media reflect the visit frequency of urban parks? a city-wide analysis using the count and content of photographs. Landscape Urban Plann. 203, 103908.

Subiza-Pérez, M., Vozmediano, L., San Juan, C., 2020. Green and blue settings as providers of mental health ecosystem services: Comparing urban beaches and parks and building a predictive model of psychological restoration. Landscape Urban Plann. 204, 103926.

Sun, C., Zhou, Y., Han, Y., 2022. Automatic generation of architecture facade for historical urban renovation using generative adversarial network. Build. Environ. 108781.

Sun, D., Li, Q., Gao, W., Huang, G., Tang, N., Lyu, M., Yu, Y., 2021. On the relation between visual quality and landscape characteristics: a case study application to the waterfront linear parks in shenyang, china. Environ. Res. Commun. 3, 115013.

Sun, M., Zhang, F., Duarte, F., Ratti, C., 2022. Understanding architecture age and style through deep learning. Cities 128, 103787.

Tabrizian, P., Baran, P.K., Van Berkel, D., Mitasova, H., Meentemeyer, R., 2020. Modeling restorative potential of urban environments by coupling viewscape analysis of lidar data with experiments in immersive virtual environments. Landscape Urban Plann. 195, 103704.

Taecharungroj, V., Mathayomchan, B., 2020. The big picture of cities: Analysing flickr photos of 222 cities worldwide. Cities 102, 102741.

Venverloo, T., Duarte, F., Benson, T., Bitran, Q., Beldad, A.D., Alvarez, R., Ratti, C., 2021. Evaluating the human experience of autonomous boats with immersive virtual reality. J. Urban Technol. 28, 141–154.

Vert, C., Gascon, M., Ranzani, O., Márquez, S., Triguero-Mas, M., Carrasco-Turigas, G., Arjona, L., Koch, S., Llopis, M., Donaire-Gonzalez, D., et al., 2020. Physical and mental health effects of repeated short walks in a blue space environment: A randomised crossover study. Environ. Res. 188, 109812.

Völker, S., Heiler, A., Pollmann, T., Claßen, T., Hornberg, C., Kistemann, T., 2018. Do perceived walking distance to and use of urban blue spaces affect self-reported physical and mental health? Urban Forestry Urban Green. 29, 1–9.

Vollmer, D., Prescott, M.F., Padawangi, R., Girot, C., Grêt-Regamey, A., 2015. Understanding the value of urban riparian corridors: considerations in planning for cultural services along an indonesian river. Landscape Urban Plann. 138, 144–154.

Walker, H.K., Hall, W.D., Hurst, J.W., 1990. Clinical methods: the history, physical. and laboratory examinations.

Wang, F., Yuan, C., Li, J., Liu, B., 2022. What makes a place special? research on the locality of cities in the yellow river and rhine river basins based on street view images. Indoor Built Environ. 31, 435–451.

Wang, J., Biljecki, F., 2022. Unsupervised machine learning in urban studies: A systematic review of applications. Cities 129, 103925. https://doi.org/10.1016/j.cities.2022.103925.

Wang, R., Liu, Y., Lu, Y., Zhang, J., Liu, P., Yao, Y., Grekousis, G., 2019. Perceptions of built environment and health outcomes for older chinese in beijing: A big data approach with street view images and deep learning technique. Comput. Environ. Urban Syst. 78, 101386.

Wang, R., Lu, Y., Zhang, J., Liu, P., Yao, Y., Liu, Y., 2019. The relationship between visual enclosure for neighbourhood street walkability and elders' mental health in china: using street view images. J. Transp. Health 13, 90–102.

Wheeler, B.J., Karimi, H.A., 2020. Deep learning-enabled semantic inference of individual building damage magnitude from satellite images. Algorithms 13, 195.

Wilkins, E.J., Van Berkel, D., Zhang, H., Dorning, M.A., Beck, S.M., Smith, J.W., 2022. Promises and pitfalls of using computer vision to make inferences about landscape preferences: Evidence from an urban-proximate park system. Landscape Urban Plann. 219, 104315.

Wu, A.N., Biljecki, F., 2021. Roofpedia: Automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability. Landscape Urban Plann. 214, 104167.

Wu, A.N., Biljecki, F., 2022. GANmapper: geographical data translation. Int. J. Geogr. Inform. Sci. 36, 1394–1422.

Wu, A.N., Stouffs, R., Biljecki, F., 2022. Generative Adversarial Networks in the Built Environment: A Comprehensive Review of the Application of GANs across Data Types and Scales. Build. Environ. 223, 109477 https://doi.org/10.1016/j.buildenv.2022.109477.

Wu, Z., Yao, L., Zhuang, M., Ren, Y., 2020. Detecting factors controlling spatial patterns in urban land surface temperatures: A case study of beijing. Sustain. Cities Soc. 63, 102454.

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 Green. 59, 126995.

Xiao, R., Guo, Y., Zhang, Z., Li, Y., 2022. A hidden markov model based unscented kalman filtering framework for ecosystem health prediction: A case study in shanghai-hangzhou bay urban agglomeration. Ecol. Ind. 138, 108854.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P., 2021. Segformer: Simple and efficient design for semantic segmentation with transformers. Advances in Neural Information Processing Systems 34.

Xin, L., Danzi, W., Liang, L., Xiangrong, W., 2022. Research on visual perception evaluation of urban riverside greenway landscape based on deep learning. J. Beijing Forestry Univ. 20 (43), 93–104.

Yamashita, S., 2002. Perception and evaluation of water in landscape: use of photo-projective method to compare child and adult residents' perceptions of a japanese river environment. Landscape Urban Plann. 62, 3–17.

Yan, Y., Feng, C.C., Huang, W., Fan, H., Wang, Y.C., Zipf, A., 2020. Volunteered geographic information research in the first decade: a narrative review of selected journal articles in giscience. Int. J. Geogr. Inform. Sci. 34, 1765–1791.

Yao, L., Liu, T., Qin, J., Lu, N., Zhou, C., 2021. Tree counting with high spatial-resolution satellite imagery based on deep neural networks. Ecol. Ind. 125, 107591.

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. Int. J. Geogr. Inform. Sci. 33, 2363–2384.

Yao, Y., Wang, J., Hong, Y., Qian, C., Guan, Q., Liang, X., Dai, L., Zhang, J., 2021. Discovering the homogeneous geographic domain of human perceptions from street view images. Landscape Urban Plann. 212, 104125.

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H.H., Lin, H., Ratti, C., 2018. Measuring human perceptions of a large-scale urban region using machine learning. Landscape Urban Plann. 180, 148–160.

Zhang, R.X., Zhang, L.M., 2021. Panoramic visual perception and identification of architectural cityscape elements in a virtual-reality environment. Future Gener. Comput. Syst. 118, 107–117.

Zhang, Y., Yang, Y., Feng, S., Qi, J., Li, W., Yu, J., 2020. The evaluation on visual fatigue and comfort between the vr hmd and the ipad. International Conference on Applied Human Factors and Ergonomics, Springer 213–219.

Zhou, H., He, S., Cai, Y., Wang, M., Su, S., 2019. Social inequalities in neighborhood visual walkability: Using street view imagery and deep learning technologies to facilitate healthy city planning. Sustain. Cities Soc. 50, 101605.