Landscape and Urban Planning 228 (2022) 104569

Contents lists available at ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

Research Paper

Semantic Riverscapes: Perception and evaluation of linear landscapes from oblique imagery using computer vision

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

HIGHLIGHTS

- Visual perception and evaluation of landscapes are important in large-scale river analysis.
- A new approach using UAV oblique imagery and computer vision.
- A comprehensive perception study of riverscapes with bifurcated experiments.
- The method is automated and scalable in other geographies.
- The open dataset supports future studies.

ARTICLE INFO

Keywords:
Riverside
Open data
GeoAI
Aerial surveys
Drones
Virtual reality

ABSTRACT

Traditional approaches for visual perception and evaluation of river landscapes adopt on-site surveys or assessments through photographs. The former is expensive, hindering large-scale analyses, and it is conducted only on street-level or top-down imagery. The latter only reflects the subjective perception and also entails a laborious process. Addressing these challenges, this study proposes an alternative: a novel workflow for visual analysis of urban river landscapes by combining unmanned aerial vehicle (UAV) oblique photography with computer vision (CV) and virtual reality (VR). The approach is demonstrated with an experiment on a section of the Grand Canal in China where UAV oblique panoramic imagery has been processed using semantic segmentation for visual evaluation with an index system we designed. Concurrent surveys, immersive and non-immersive VR, are used to evaluate these photos, with a total of 111 participants expressing their perceptions across multiple dimensions. Then, the relationship between the people's subjective visual perception and the river landscape environment as seen by computers has been established. The results suggest that using this approach, rivers and surrounding landscapes can be analyzed automatically and efficiently, and the mean pixel accuracy (MPA) of the developed model is 90%, which advances state of the art. The results of this study can benefit urban planners in formulating riverside development policies, analyzing the perception of plans for a future scenario before an area is redeveloped, and the method can also aid relevant parties in having a macro understanding of the overall situation of the river as a basis for follow-up research. Due to simplicity, accuracy and effectiveness, this workflow is transferable and cost-effective for large-scale investigations of riverscapes and linear heritage. We openly release Semantic Riverscapes—the dataset we collected and processed, bridging another gap in the field.

1. Introduction

Human development is closely related to river landscapes worldwide, and therefore it is necessary to consider how people perceive, value, and interact with river landscapes in various ways (Garau, Torralba, & Pueyo-Ros, 2021; Verbrugge & van den Born, 2018; Portela et al., 2021; Gottwald & Stedman, 2020; Guo, Fu, Wang, Xu, & Liu, 2021). As one of the most important means for the public to perceive the

* Corresponding author.
E-mail addresses: luojunjie4669@tju.edu.cn (J. Luo), filip@nus.edu.sg (F. Biljecki).

https://doi.org/10.1016/j.landurbplan.2022.104569
Received 8 February 2022; Received in revised form 24 August 2022; Accepted 2 September 2022
Available online 23 September 2022

landscape, vision accounts for 76% on environment satisfaction (Krause, 2001; Jeon & Jo, 2020). Visual perception and evaluation have become the mainstay for researchers, practitioners, and governments to understand the landscape quality of urban streets, parks, scenic spots, and rivers (Qi et al., 2020; Jin & Wang, 2021). River visual perception and evaluation refers to the analysis of the characteristics and functions of the research area based on specific purposes, combined with qualitative and quantitative approaches. Traditional visual analysis methods of river landscapes involve on-site visits and field photography, which are labour-intensive, time-consuming, and often restricted by factors such as obstacles, topography and climate (Mouratidis & Hassan, 2020). Some visual studies use 2D images for virtual perception (Sun et al., 2021; Li et al., 2021), but such an approach has limitations in terms of interactivity, virtual immersion and field of view, and it is often a tedious process. Therefore, objective visual evaluation and efficient perception of large-scale linear river landscapes remain underexplored and challenging, especially in locations where field experiments are impossible or dangerous.

With the rapid development of unmanned aerial vehicle (UAV) technology, UAV has been widely used in large-scale landscape analysis including the field of riverscapes (Woodget, Austrums, Maddock, & Habit, 2017; Rusnák, Sládek, Kidová, & Lehotský, 2018; Torgersen et al., 2021; Rivas Casado, Ballesteros Gonzalez, Wright, & Bellamy, 2016; Miřijovskỳ & Langhammer, 2015). In comparison with the small sensing range of the traditional ground view and the limitation that the satellite perspective is on the nadir, UAV offers middle ground with an optimal perspective—it can take oblique panoramic images at different heights in addition to taking ordinary photos in three modes of the oblique, top view and horizontal (Brumana, Barazzetti, Oreni, & Roncoroni, 2013), and it overcomes the limitation of ground view and satellite nadir, as cameras capture images from different angles and can obtain both the top information and facade textures of the research area in the same shot (Lyu, Vosselman, Xia, Yilmaz, & Yang, 2020). According to recent papers, UAV oblique panoramic images are now entrenched as novel geospatial data characterized by the superiority of full perspective, virtual reality (VR), and high realism (Li, Karim, & Qin, 2022; Zhang et al., 2020). The direction of large-scale landscape visual perception is moving towards the use of UAVs combined with a variety of cutting-edge technologies (Harknett et al., 2022; Meng et al., 2022). For example, the combination of UAV panorama images and VR technologies allows visualizing the surroundings of the landscape, which is more beneficial to the public's omnidirectional perception of a location (Lan et al., 2016; Santos, Henriques, Mariano, & Pereira, 2018). The VR technologies are particularly useful in areas where fieldwork is impossible, dangerous, or expensive. VR can further improve the interactive experience in the process of visual evaluation and bridge the gaps of limited shooting angle and poor interactivity of traditional photos (Feng, 2021; Birenboim et al., 2019). Meanwhile, virtual landscape perception through UAV and VR is proliferating on the internet with multiple social media (Facebook, Twitter, DJI Forum, etc.) and video platforms (e.g. YouTube, TikTok, BiliBili).[1] This virtual aerial tour and visual perception type have developed into a crucial tool for displaying a location's overall landscape qualities and as a vital basis for determining if a location is worthwhile for travel. Thus, such an approach allows the extensive visual perspective of the landscape, with favourable aerial positions that cannot be obtained by satellite or ground observers (Papadopoulou, Papakonstantinou, Zouros, & Soulakellis, 2021). Therefore, the combination of UAV and VR has clear advantages in the research of visual perception of large-scale landscapes.

In parallel, coupling the UAV oblique photography and computer vision (CV) has become an important method for quantifying vast urban landscapes (Lyu et al., 2020). Thanks to the fast development of CV, such as semantic segmentation and object detection, studies on visual quality evaluation based on such trending techniques are proliferating (Wu, Li, Hong, Tao, & Du, 2021; Garg et al., 2021; Wilkins et al., 2022; Wu & Biljecki, 2021; Ito & Biljecki, 2021). CV can process the profusion of images automatically, objectively and efficiently, and it is not entirely new to riverscapes either (Sharma, Isha, & Vashisht, 2021). For example, the study by Li et al. (2021) has used semantic segmentation to evaluate visual qualities of urban rivers from an on-water perspective. Wawrzyniak and Stateczny (2018) and Ming, Ya-duan, Lin-kai, Peng, and Qi-mei (2017) have used object detection to identify vessels on rivers. However, there is no existing study on visual perception or evaluation of riverscapes that combines CV and UAV, which is a gap we seek to bridge in this paper.

Apart from that, studies employing CV on various types of urban imagery at the ground level (e.g. street view imagery and photos taken by tourists and residents) have relied on general datasets such as MS-COCO (Lin et al., 2014), Cityscapes (Cordts et al., 2016), and Pascal VOC (Everingham, Van Gool, Williams, Winn, & Zisserman, 2010; Shetty, 2016) to train deep learning models to visually evaluate the environment (Biljecki & Ito, 2021; Hosseini, Miranda, Lin, & Silva, 2022; Seiferling, Naik, Ratti, & Proulx, 2017; Verma, Jana, & Ramamritham, 2019; Ibrahim, Haworth, & Cheng, 2020). However, for our studies, these datasets may fall short, and there is no openly available processed UAV oblique dataset for the river landscapes by the time of writing this paper, which indicates the importance of our study to fill in such a gap.

Considering the developments in computer vision and virtual reality and the convenience of UAV oblique photography, we believe that research marrying the three is needed and timely. In this study, we aim to build a visual analysis workflow for large-scale river landscapes based on UAV oblique panoramas. By using CV and VR, we seek to assess people's subjective visual perceptions and the proportion of physical environment elements effectively. Taking the Tianjin section of the Grand Canal in China as a case study, this study proposes an objective visual evaluation approach to river landscapes based on the combination of UAV oblique photography and CV so as to achieve flexible and efficient visual analysis of rivers and surrounding areas. In the subjective visual perception study of this research, the UAV panoramic photos are displayed through two VR experiments. One is the immersive virtual reality (IVR) approach using head-mounted displays, while the other one relies on non-immersive virtual reality (nIVR), which uses tablets, smartphones and so on. In both, study participants can have a remote virtual experience of the river landscapes and will provide how they feel about the tranquillity, pleasure, beauty and other dimensions of the study area. We validate the virtual experience outcomes by cross-validation of the two experimental results. The research questions are as follows:

- How to construct a workflow for a visual analysis of large-scale river landscapes based on UAV oblique panoramas?
- Taking the south canal and the north canal in China as examples, how are their objective visual characteristics different, and what are the differences in people's subjective perception of various riverscapes?
- What is the relationship between objective visual analysis results and subjective visual perception results?

2. Background and related work

2.1. The way of oblique: UAV photos compared with satellite images and SVI

UAV, aerial/satellite, and street view imagery (SVI) are essential for understanding landscapes (Meinen & Robinson, 2020; Rouse, Tabaldiev, & Matuzeviciute, 2021; Hritz, 2014; del Río-Mena, Willemen, Tesfamariam, Beukes, & Nelson, 2020; Kim, Lee, Hipp, & Ki, 2021; Biljecki & Ito, 2021; Li, Ratti, & Seiferling, 2018; Luo, Liu, & Cao, 2022). These

[1] Example link: https://www.youtube.com/watch?v=L_tqK4eqelA.

three types have their own characteristics, and each plays an instrumental role in spatial information sciences, producing significant volumes of data contributing to a wide range of domains and use cases (Fig. 1). The increasing production of imagery can be partly explained by the democratization of UAVs and SVI due to the decreasing cost of exploitation (Sun & Scanlon, 2019), the increase in the number of deployed satellites (Ghamisi et al., 2019), and the growing coverage of commercial services such as Google Street View, Baidu Maps, and volunteered geographic information (Yan et al., 2020; Ito & Biljecki, 2021).

Satellite images have high temporal and global coverage; however, they are not without limitations (Pettorelli et al., 2018; Sheffield et al., 2018). One shortcoming of satellites is that their viewing angle is fixed, as they can only acquire nadir imagery without clear facade information of the study areas; thus, it may not be suitable for visual perception of scenery (Emilien, Thomas, & Thomas, 2021; Ding, Zhou, Meng, & Long, 2021; Tian, Shao, Ouyang, & Shen, 2021). In contrast, a camera mounted on a UAV can record flexibly, obtaining oblique, nadir, and even panoramic imagery (Brumana et al., 2012; Che et al., 2020), including video footage (Sun et al., 2021). UAV oblique and panoramic imagery can include the side textures of the viewing area, with a wider field of vision and richer content, facilitating field perception and evaluation (Santos et al., 2018). Other shortages of satellites are control and the lack of general flexibility—one cannot launch their own satellite, and the spatial data cannot be acquired easily on specific dates or at a specific time, as the data acquisition depends upon the satellite's revisit or temporal resolution (Bhardwaj, Sam, Martín-Torres, & Kumar, 2016). In comparison with satellite remote sensing, UAV allows a flexible flight schedule, its entry barriers are low (it is low-cost and easy to use), and the interval of repeated access may be shorter (Shao et al., 2021), which makes it possible to quickly analyze the landscapes of specific locations during particular time (Ashilah et al., 2021; Hervouet, Dunford, Piégay, Belletti, & Trémélo, 2011). Image resolution is another limitation of the satellites (Khaliq et al., 2019; Iizuka et al., 2018). The flight altitude of UAVs is low and flexible (barring local flight regulations), meaning that it is generally below clouds (Watkins et al., 2020), and allows a very high ground resolution of imagery and video (Guerra-Hernández, Díaz-Varela, Ávarez-González, & Rodríguez-González, 2021; Miraki, Sohrabi, Fatehi, & Kneubuehler, 2021; Qu et al., 2021). Therefore, we believe that it is not only more suited with respect to the perspective, but also it is visually clearer, and thus, more appropriate for accurate and reliable analysis in this particular context.

In the same environment, UAV also has many advantages over the other end of the spectrum—SVI, which has been increasingly seen as a useful resource that enables researchers to measure urban landscapes precisely and thus examine the effects of the environment on residents' well-being more effectively (Biljecki & Ito, 2021; Li, Santi, Courtney, Verma, & Ratti, 2018; Seiferling et al., 2017; Zhou, He, Cai, Wang, & Su, 2019). However, despite the growing coverage of data, many off-road places, such as riversides, parks, villages, and other areas with rugged ground conditions, remain out of reach in street view surveys (Verma et al., 2019) and many of those urban objects that are captured remain obscured (cf. Fig. 1) (Pang & Biljecki, 2022). Further drawbacks include seasonal and time variability, and infrequent updates (Kim et al., 2021). In contrast, drones can collect data almost anywhere, and the shooting time is virtually unlimited (Nex et al., 2022), enabling a focused study and ensuring proper attention to capturing the required data. The operating height is another advantage as UAV can fly at different altitudes to provide a more suitable perspective and appropriate sight coverage. Flying higher, UAV can observe a wider range of scenes, which is conducive to large-scale scene perception (Lytkin & Syromyatnikov, 2021; Schenone et al., 2021); while operating closer to the ground, more details can be observed, an unparalleled benefit. Therefore, UAV has become an important research tool in the fields of environmental detection (Youme, Bayet, Dembele, & Cambier, 2021), building facade inspection (Chen, Reichard, Xu, & Akanmu, 2021), agricultural monitoring (Kerkech, Hafiane, & Canals, 2020), disaster rescue (Erdelj & Natalizio, 2016) and it also has been applied to city traffic, cultural heritage, and other disciplines (Ahmed, Ngoduy, Adnan, & Baig, 2021; Beg, Qureshi, Sheltami, & Yasar, 2021; Cai, Fang, Zhang, & Chen, 2021; Castrignanò et al., 2021; Baranwal, Raghvendra, Tiwari, & Pande, 2021;

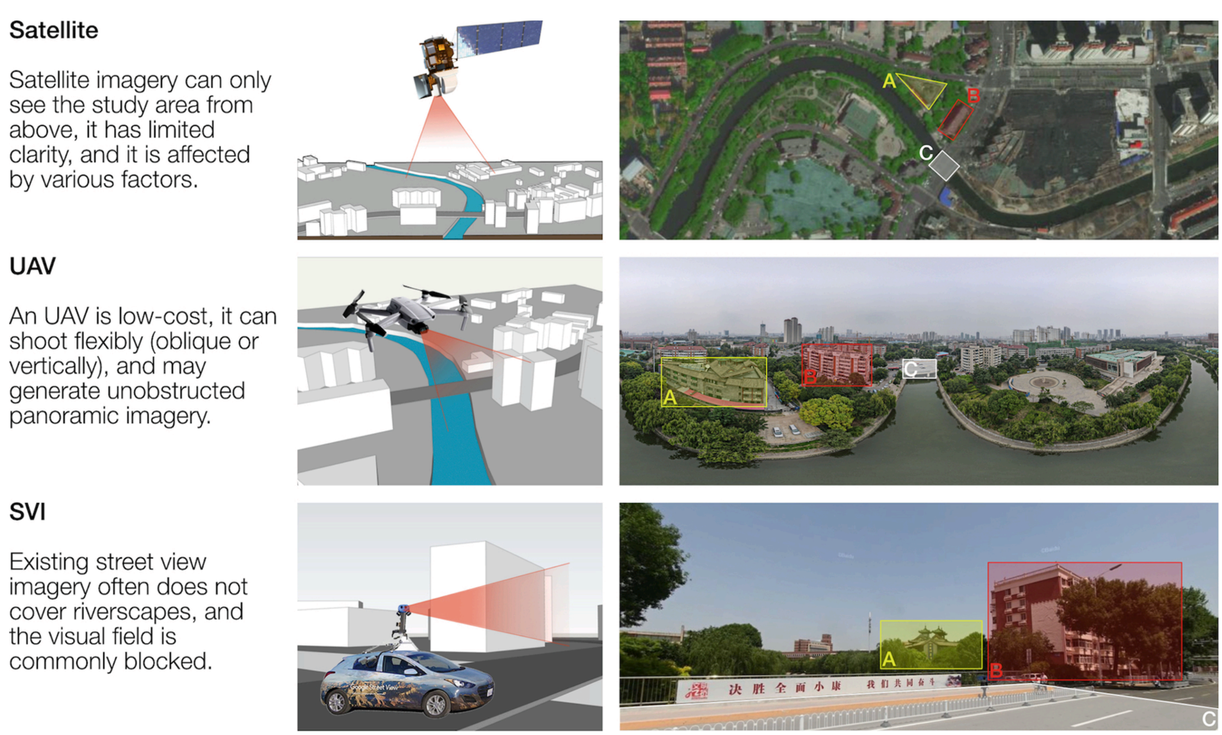

Fig. 1. Comparison of the three key types in sensing landscapes. The objects seen by the three types are outlined in different colours (e.g. two buildings – A and B, and a bridge over the river – C). Satellite imagery can only provide an understanding from nadir, while the street view perspective may not be able to fully perceive the riverscape. Source of the satellite image (top right) and SVI image (bottom right): Baidu Maps.

Karthik et al., 2021; Munawar, Ullah, Qayyum, & Heravi, 2021). Small, low-cost, and portable UAV will likely remain the key instrument of many data acquisition campaigns in the future, and this paper explores their usability for evaluating riverscapes.

2.2. Role of CV and the landscape of UAV data

With the continuous development of deep learning, especially the improvement of CV, techniques such as semantic segmentation have been gradually introduced into the research of landscape visual evaluation (Ma et al., 2021; Fong, Ong, & Nee, 2009; Song, Ning, Ye, Chandana, & Wang, 2022). CV has improved the ability to automatically and efficiently process a large volume of imagery for quantitatively analysis (Hu, Zhang, Gong, Ratti, & Li, 2020; He, Páez, & Liu, 2017). At present, there are some open segmentation datasets that can be used for a variety of studies in analysing the urban environment (Cordts et al., 2016; Lin et al., 2014). However, such datasets have not been developed specifically for UAV oblique scenes and there are only few semantic segmentation datasets of UAV aerial imagery (Lyu et al., 2020), a challenge that hampers a variety of studies relying on UAV as training CV models (Nex et al., 2022). Training datasets obtained with UAV would be beneficial, as in comparison with nadir photography, oblique counterparts have a broader perspective, contain a variety of objects, and may have more complex semantic information.

In this section, to understand related work that may support our study, we provide an overview of UAV open semantic segmentation datasets available so far to the extent of our knowledge (Table 1). These mainly include nadir data: ICG Drone Dataset (Sun, Yang, Zhang, & Zhang, 2021), FloodNet (Rahnemoonfar et al., 2021), and the Urban Drone Dataset (UDD) (Chen, Wang, Lu, Chen, & Wang, 2018). The ICG Drone Dataset focuses on semantic understanding of urban scenes for increasing the safety of autonomous drone flight and landing procedures (Sun et al., 2021). The imagery depicts more than 20 houses from nadir views acquired at an altitude of 5 to 30 meters above the ground. FloodNet dataset focuses on the post-disaster damage assessments, and it poses several challenges, including detection of flooded roads and buildings and distinguishing between natural water and flooded water (Chowdhury & Rahnemoonfar, 2021). UDD is collected by DJI-Phantom 4 UAV at altitudes between 60 and 100 m and contains most part of nadir imagery and a few oblique imagery (Chen et al., 2018). It has 160 images and contains 4 semantic classes: vegetation, building, car and free space for urban scene understanding (Wei, Wang, Yi, Chen, & Wang, 2020; Xiang, Xia, & Zhang, 2018).

Apart form that, UAV oblique datasets mainly include Aeroscapes (Nigam, Huang, & Ramanan, 2018) and UAVid (Lyu et al., 2020). The Aeroscapes semantic segmentation dataset includes imagery captured from an altitude range of 5 to 50 m using a commercial UAV. This dataset provides 3269 720p imagery and labels for 11 classes: person, bike, car, drone, boat, animal, obstacle, construction, vegetation, road and sky. The UAVid dataset has 300 oblique imagery. It is an urban street scene semantic segmentation dataset, and it has 8 object categories considered: building, road, static car, tree, low vegetation, human, moving car and background clutter. There are also some UAV aerial datasets, including video datasets that can support the analysis of the urban environment (Sun et al., 2021), understand the transportation problems (Mandal, Kumar, & Vipparthi, 2020), detect vehicles (Zhang, Liu, Chang, & Song, 2020) and so on, but most of them are not available openly and cannot be used for river scene segmentation.

Table 1
Overview of existing open UAV datasets and our newly introduced contribution.

Datasets	Classes	Images	Shooting height	Shooting angle
Aeroscapes	11	3269	5–50 m	nadir, oblique
UAVid	8	300	unknown	oblique
FloodNet	9	2343	60 m	nadir
ICG Drone Dataset	20	400	5–30 m	nadir
UDD	5	160	60–100 m	nadir
Semantic Riverscapes *(our contribution)*	14	400	30–60 m	oblique

2.3. UAV and virtual reality

Virtual reality technology is frequently employed in built environment studies, allowing users to gain a comprehensive awareness of environmental aspects (Van Leeuwen, Hermans, Jylhä, Quanjer, & Nijman, 2018). Immersive virtual reality and non-immersive virtual reality perception modalities can be distinguished (Okeil, 2010; Isaacs, Gilmour, Blackwood, & Falconer, 2011). The IVR simulated environments typically completely surround the participant through the use of VR glasses (head-mounted display), while nIVR environments can be viewed directly on smart phones, iPads or computer screens (Paes, Irizarry, & Pujoni, 2021; Xu, Oberman, Aletta, Tong, & Kang, 2020). Both approaches have their own set of benefits and drawbacks. Using VR glasses to create an IVR experience will make participants feel more real than nIVR perception; however, some participants experience after long exposures to IVR glasses may have some negative side effects (or "VR sickness"), such as nausea, headache, and disorientation (Birenboim et al., 2019). In contrast, additional equipment, such as head-mounted displays, is not required for a nIVR experience. The realism of nIVR will be less than immersive perception, but the negative side effects will be minimal.

People can have a large-scale immersive landscape environment perception experience with the combination of UAV panoramas and virtual reality technology, especially in locations with terrible ground conditions. In addition, UAV oblique photography modelling can be used to render 3D real-world scenes, which can then be coupled with virtual reality (Schmohl, Tutzauer, & Haala, 2020). These have been widely employed in news and sports event broadcasting, environmental monitoring, urban space management and so on (Keil, Edler, Schmitt, & Dickmann, 2021; Bakirman et al., 2020; Pavlik, 2020; Esposito, Mastrorocco, Salvini, Oliveti, & Starita, 2017). The first benefit of combining the two is that it is easy for individuals to view the landscapes and monitor buildings (Kikuchi, Fukuda, & Yabuki, 2022; Bacco et al., 2020). With the help of VR and drones, people can get a comprehensive view of the study area. The broad viewpoint allows urban planners, governments, journalists, and residents to gain a macro understanding of places (Pavlik, 2020). When compared to earlier means of observing large-scale landscape elements from many angles from high-rise buildings, observation platforms on high mountains, or employing helicopters, UAV marrying VR is unquestionably more convenient. The second distinguishing aspect is the high level of interactivity (Elghaish et al., 2020). Viewers can enjoy these UAV panoramic photographs or videos based on their preferences and even enlarge some areas of interest to learn more about the research region in more detail. High-precision geotagged data is the third characteristic. The UAV's panoramic image includes high-precision longitude, latitude, and altitude information, and it enables people to create map-based immersive imagery. People can clearly know their specific location and height details when they remotely perceive these panoramic pictures with geographical labels, which has become an important data source for them to watch and understand the research location information, which is conducive to people understanding and analyzing the spatial characteristics of a specific place. The fourth feature is the ability to achieve augmented reality (AR) visualization (Kikuchi et al., 2022; Lindner, Ortwein, Staar, & Rienow, 2021). The advancement of 3D modelling technology based on UAV oblique photography has greatly improved people's ability to virtual perceive landscapes, and it is now widely used in the fields of cultural heritage protection, landscape perception, building inspection and so on (Liu, Xia, Chen, & Li, 2021; Smaczyński & Horbiński, 2021; Al-Bahri, Al Kishri, & Dharamshi, 2021).

3. Methods and materials

3.1. Study area

China's Grand Canal is one of the most famous man-made rivers in the world. Its length is 1794 km, it flows through 21 major cities (including Tianjin, Beijing, and Hangzhou), and connects five major rivers (Qiantang, Huai, Yangtze, Yellow, and Hai) (Wen, Xiao, & Zhang, 2017; Li, Zhang, & Sun, 2020). The study area is located in the Tianjin section of the Grand Canal, which is composed of the north canal and the south canal, with a total length of about 24 km (Fig. 2). The main reasons for choosing this particular section of this nationally important river as the research area are as follows. First, it has played a monumental role in the local economic and cultural development, and the Chinese government is preparing to build the Grand Canal National Cultural Park, which has attracted much attention worldwide (Li et al., 2021; Zhao, Yan, & Hou, 2021), and which encompasses the study area. The landscape visual evaluation and perception of this river section can provide information support for the construction of the National Cultural Park. Second, this river connects the southern and northern suburbs and the downtown of Tianjin, intersecting the daily life of residents, including providing open spaces for citizens for leisure activities and others. However, the visual characteristics of different areas are not clear at present, which is worth investigating. Third, there are no light-drone flight restrictions in this area, which enables us some flexibility and experiment with different scenarios of data acquisition. In addition, there is no complex electromagnetic interference in this urban area, which ensures the flight safety of UAVs, so we can use small drones easily for aerial oblique photography.

3.2. Data collection

In this study, a DJI Mavic Air 2 UAV (Lan & Lee, 2021) was used to obtain geo-tagged aerial oblique imagery for objective visual analysis and subjective visual perception (Fig. 3). This drone is equipped with an image sensor with 1/ 2-inch CMOS, an angle of view of 84°, an equivalent focal length of 24 mm, and an effective resolution of 48 million pixels. We checked the clarity of this configuration and found that we can distinguish people, vehicles, shrubs, and other minor objects on the ground at a height of 60 m. As a result, we believe that this drone meets the requirements of this research. The aerial photography data collection took place over four days from 10 am to 6 pm during the period from July to September 2021 under stable light conditions. We set photo acquisition points every 300–500 m in the 24 km long linear research area, take panoramic oblique pictures, and number them successively from south to north (Fig. 2).

With the change of UAV flight altitude, the shape and size of objects will change roughly in proportion (Xiang et al., 2018). The increase of the observation height brings a broader vision, but vehicles, people, vegetation and other objects in aerial images will become smaller, which brings challenges to the recognition of semantic information (Lyu et al., 2020). In addition, low flying altitude also means a smaller field of vision. In some complex scenes, there may be potential safety hazards, such as electric towers, wires and branches, which may affect the flight (Watkins et al., 2020). Therefore, after comparing the data of four heights of 30 m, 60 m, 90 m and 120 m, we chose 60 m height as a compromise among safety, large field of vision and ground clarity.

3.3. Semantic Riverscapes dataset

To address the first research question, we start by describing the steps carried out to construct the Semantic Riverscapes dataset. After our literature review, we found that there is no openly available UAV oblique imagery semantic segmentation dataset that focuses on the river environment. Therefore, we acquire a large dataset containing UAV oblique images and segment them, which is tailored for the semantic segmentation of river scenes and can support applications such as the comprehensive visual analysis in river landscapes. In addition to collecting oblique panoramic imagery required for visual perception and evaluation, we also took a series of oblique photos along the river to construct a dataset for imagery semantic segmentation. After data

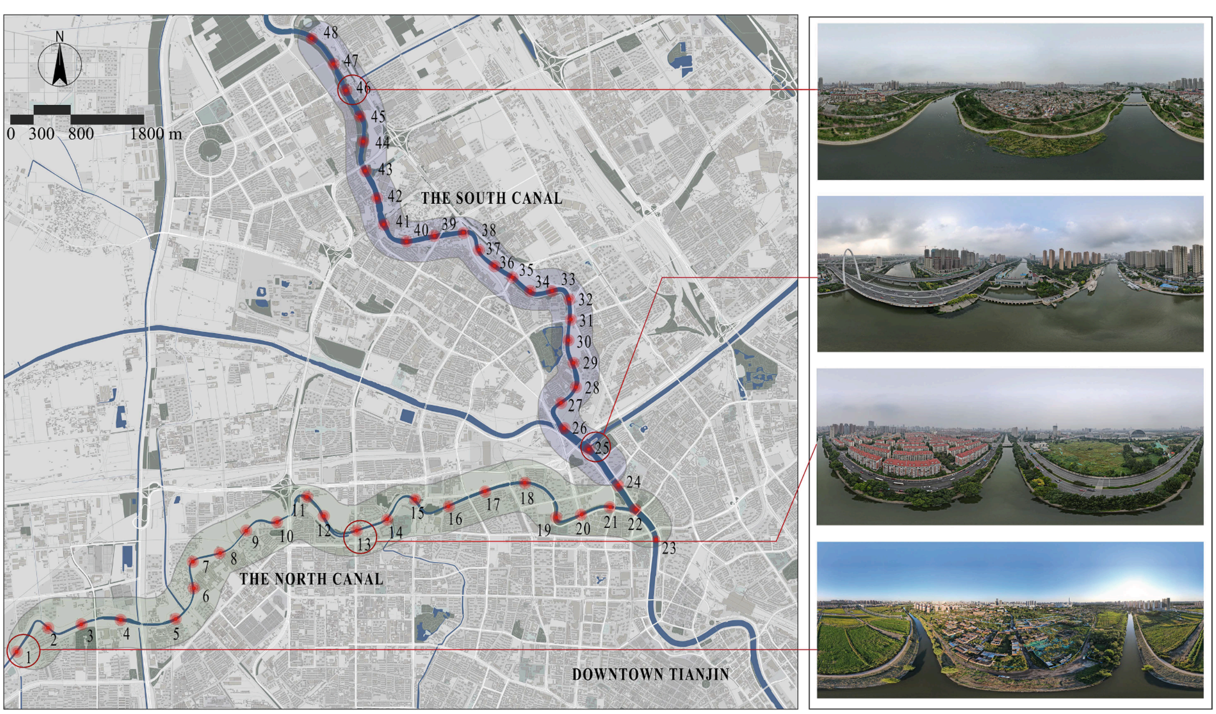

Fig. 2. Study area and data collection points. Field panoramic oblique photographs of the four selected mapping points show the surveyed river landscapes. Source of the base map: Amap.

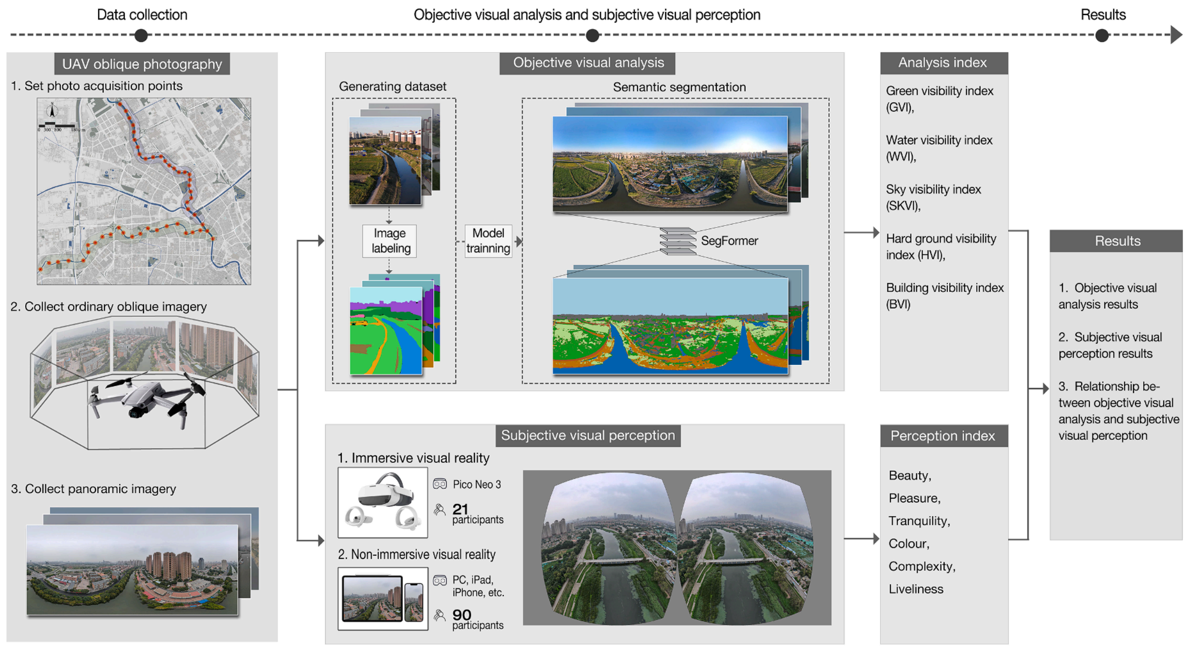

Fig. 3. Overview of the workflow. Step 1: generating the original sample data. Step 2: visual evaluation and perception methods. Step 3: visualization of evaluation and perception results and their correlation analysis.

acquisition and processing, we derived the dataset with 400 high-resolution images spanning the river and surrounding areas, each with a size of 1800 x 1480 pixels. When creating this dataset, we took careful consideration of the shooting conditions and referred to the characteristics of other datasets in order to make it more universal (Nigam et al., 2018; Sun et al., 2021). In terms of height, the shooting height of these images has changed from 30 m to 60 m, which is similar to the height coverage of several other UAV semantic segmentation datasets (Nigam et al., 2018; Sun et al., 2021), in order to meet more research needs in the later stage, and it also contains the data of 60 m height that we use for visual perception and evaluation. We also consider the lighting conditions, and the pictures in the dataset include cloudy days and sunny days. Each image was manually labelled. We labelled the imagery into 14 categories, namely: building, cottage, under construction place, tree, grass, water grass, soil, hard ground, water, sky, human, car, boat, and void, a relatively comprehensive segmentation (cf. Table 1), which will be applicable to the river landscapes in most parts of the world. According to the characteristics of river landscapes, we have decided to break down greenery into trees, grass, and water grass, to distinguish the water plants and the vegetation around the river. Similarly, we regard multiple groups of buildings: regular buildings, cottages, and construction sites. Our aerial images have been labelled at pixel level with EISeg software (Xian, Xu, Cheng, Zhang, & Ding, 2016; Hao et al., 2021), which was developed based on PaddlePaddle (Ma, Yu, Wu, & Wang, 2019), which covers the majority of high-quality segmentation models in different directions, namely general scenarios, portrait, remote sensing, medical treatment, etc., providing convenience to the rapid annotation of semantic and instance labels with reduced cost (Hao et al., 2021).

3.4. Objective visual analysis

3.4.1. Automated image segmentation

The manually annotated dataset is used to develop a CV model for image segmentation. Many ready-to-use models, such as FCN, SegNet, U-net, PSP-net, and SegFormer, can detect objects and perform segmentation of an image (Badrinarayanan, Kendall, & Cipolla, 2017; Zhao, Shi, Qi, Wang, & Jia, 2017; Xia, Yabuki, & Fukuda, 2021). Considering the characteristics of UAV data and operability in river landscapes, we select SegFormer, a cutting-edge Transformer framework for semantic segmentation that jointly considers efficiency, accuracy, and robustness for image semantic segmentation (Xie et al., 2021). 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%). Two metrics were used to evaluate the training and validation process: mean pixel accuracy (MPA) and mean Intersection over Union (mIoU). The former is the ratio of correctly predicted pixels to the total pixels, the latter is a common and effective evaluation metric used for image semantic segmentation tasks, and it is the ratio of the intersection area of the predicted pixels and ground truth pixels to their union area. It is also commonly used in urban analytics and spatial information sciences (Wu & Biljecki, 2022). After using the UAV riverscapes dataset to train the SegFormer model, we classify the 14 types of elements in the panoramic oblique imagery at the pixel level and can objectively analyze the proportion of these types of landscape elements in different imagery.

3.4.2. Index system—characterising the view

Both the natural elements (e.g. greenery, water, and sky) and the artificial elements (e.g. building and hard ground) have a considerable impact on the visual quality and aesthetic cognition of landscapes (Jahani & Saffariha, 2020). In work engaging image segmentation to extract indicators of the built environment, researchers often computed one or more indexes to quantify the view from the semantic point of view (Ki & Lee, 2021; Li et al., 2015; Li, 2021). Based on the previous research experience of visual landscapes and combined with the characteristics of river environment (Li et al., 2021), we extend existing indexing approaches for river landscape visual evaluation, adopting the green visibility index (GVI), water visibility index (WVI), and sky visibility index (SKVI), and introducing two new measures: the hard ground visibility index (HVI) and building visibility index (BVI) (Table 2). Among them, vegetation is one of the most important landscape elements in the river landscapes (Xin, Xiangrong, Liang, & Danzi, 2021), and the GVI includes trees, grass and water plants, which affect the ecology and natural degree of the river space. Water is the main element in riverscape; thus, WVI plays a substantial important role in vision.

Table 2
Description of the objective indexes.

Dimension	Parameter	Parameter description	Parameter equation
Natural	Green Visibility Index (GVI)	The proportion of vegetation pixels (tree, grass, water grass) in the image	$GVI = (A_{tr\_i} + A_{gr\_i} + A_{wg\_i})/A_{total\_i}$
	Water Visibility Index (WVI)	The proportion of water pixels in the image	$WVI = A_{wa\_i} / A_{total\_i}$
	Sky Visibility Index (SKVI)	The proportion of sky pixels in the image	$SKVI = A_{sk\_i} / A_{total\_i}$
Artificial	Hard ground Visibility Index (HVI)	The proportion of hard ground pixels (includes not only carriageways and sidewalks, but also bridges, community squares, etc.) in the image	$HVI = A_{hg\_i} / A_{total\_i}$
	Building Visibility Index (BVI)	The proportion of building pixels (high-rise residential buildings, cottage, commercial office buildings, buildings under construction, etc.) in the image	$BVI = (A_{bu\_i} + A_{co\_i} + A_{uc\_i})/A_{total\_i}$

SKVI can measure the openness of river space, and also has a great impact on people's vision. HVI and BVI are significant indicators reflecting the intensity of artificial construction in a river channel and surrounding areas. These two indicators, which are different from those in previous studies (Li et al., 2021; Gong et al., 2018), are the contents viewed from UAV oblique perspectives, and contain a wider range of semantic information. Among them, the HVI includes not only hard pavement, driveway and sidewalk, but also bridges and community squares, and it can reflect the hard condition of the ground in an area. Buildings (apartments, office buildings, residential buildings, etc.), tiny cottages, and rural dwellings, as well as under-construction places, etc., are all included in the BVI, which is useful for portraying the percentage of buildings in an area in three-dimensional panoramas.

A_{total_i} is the total number of pixels in image i, A_{tr_i} is the number of tree pixels in image i, A_{gr_i} is the number of grass pixels in image i, A_{wg_i} is the number of water grass pixels in image i, A_{wa_i} is the number of water pixels in image i, A_{sk_i} is the number of sky pixels in image i, A_{hg_i} is the number of hard ground pixels in image i, A_{bu_i} is the number of building pixels in image i, A_{co_i} is the number of cottage pixels in image i, A_{uc_i} is the number of under construction place pixels in image i.

3.5. Subjective visual perception

Although UAV aerial photography has been widely used recently, most of the relevant research focuses on ordinary photos taken from a single perspective, which cannot fully display all the characteristics of the shooting area in combination with advanced virtual interactive equipment such as VR glasses. The panoramic image shows the surrounding environment centred on the position of the UAV itself and can provide participants with an immersive virtual feeling by using VR glasses (Newman, Gatersleben, Wyles, & Ratcliffe, 2022). It can also provide non-immersive virtual perception through iPads, smartphones with gyroscopes and accelerometers and other devices so as to achieve remote virtual display and reproduce the real environment. So the UAV panoramic images effectively bridge this defect. Compared with the traditional image-based evaluation methods, the use of both IVR and nIVR technologies for landscape visual perception can bring more intuitive experience (Birenboim et al., 2019).

A previous UAV-related landscape perception study used pleasure, tranquillity, colour, complexity, etc. as indicators (Yang, Gao, Li, & Van Eetvelde, 2020). Relevant urban environmental studies have analyzed the types of human perception, such as safety, beauty, colour, liveliness, boredom and depression (Ma, Hauer, Xu, & Li, 2021; Dubey, Naik, Parikh, Raskar, & Hidalgo, 2016; Yao et al., 2019; Yao et al., 2021; Zhang, Fan, Kang, Hu, & Ratti, 2021). Adopting the previous experience in the state of the art of visual perception and the characteristics of river landscapes, this study takes beauty, pleasure, tranquillity, colour, complexity and liveliness as perception indexes, and uses these six indexes to analyze the subjective visual perception of river landscapes. Beauty estimation is a common way for landscape visual quality assessment and can describe public aesthetic preferences (Sun, Shao, Li, Huang, & Yang, 2018; Li, Shen, & Ding, 2020). Pleasure, tranquillity, and liveliness are also used as the landscape perceptual analysis contents (Yang et al., 2020; Ma et al., 2021). The colour richness and visual complexity, as perceptual quality indexes, are related to the affective appraisal of the landscape (Berlyne, 1970; Cavalcante et al., 2014; Yang et al., 2020). Assessing these perception types can help understand participants' feelings about the river environments.

The ethical aspects of this study have been reviewed, and the experiment was approved by the Institutional Review Board of the National University of Singapore. The survey was divided into two groups: immersive virtual perception group and non-immersive perception group, and the data obtained from the two groups of experiments can be cross-verified. It took place in January and April 2022. The immersive virtual environment was presented via the lenses of a Pico Neo 3 head-mounted display, and the non-immersive virtual environment was presented via iPads, smartphones, and PCs. The participants were students and staff from the National University of Singapore and Tianjin University, adding diversity to the demographics and including also participants who are not residents of Tianjin. The immersive VR perception group of participants who took part in the experiment comprised 21 individuals with a mean age of 27.1, 16 (76.2%) were females, and 15 (71.4%) were students, and the non-immersive VR perception group of participants who took part in the experiment comprised 90 individuals with a mean age of 25.6, 53 (58.9%) were females, and 78 (86.7%) were students. Participants who took part in the nIVR experiment were involved in this visual perception process through a web questionnaire. The 720 yun platform was used for virtual display of panoramic photos, so the participants could conduct nIVR experience online.

For participants to fully understand the content of each panoramic image, each participant needed to look around each scene and browse for no less than 40 s. To avoid the negative side effects (or "VR sickness"), such as dizziness and nausea, caused by the long exposures to head-mounted displays and the influence of fatigue on the score, we divided the panoramic images of 48 mapping locations into 3 groups using an equal difference sequence, and each group experienced 16 locations. Both the two perception groups of participants only watched 16 panoramic pictures, and their experience time was no more than 20 min (Park & Lee, 2020; Birenboim et al., 2019). Therefore, each mapping point (cf. Fig. 2) has IVR scores of 7 participants and nIVR scores of 30 participants. After experiencing each panoramic image, participants rated it through multiple dimensions: beauty, pleasure, tranquillity, colour, complexity, and liveliness using the 7-point Likert scale (Likert, 1932) (e.g. with 1 referring to 'It is not tranquil at all' to 7 indicating that it appears to be very much tranquil). The final score of each scene is the average of the participants' scores of the two groups.

4. Results

4.1. Objective visual evaluation results

4.1.1. Proportion of visual elements

With an MPA of 90% and a mIoU of 47%, our trained SegFormer model under the Transformer framework performs well in the imagery semantic segmentation task, meeting the experimental conditions. Fig. 4 shows the results of successfully segmenting 14 elements of river landscapes.

The findings of pixel-level semantic segmentation of panoramic oblique images of 48 mapping points we obtained using this model are

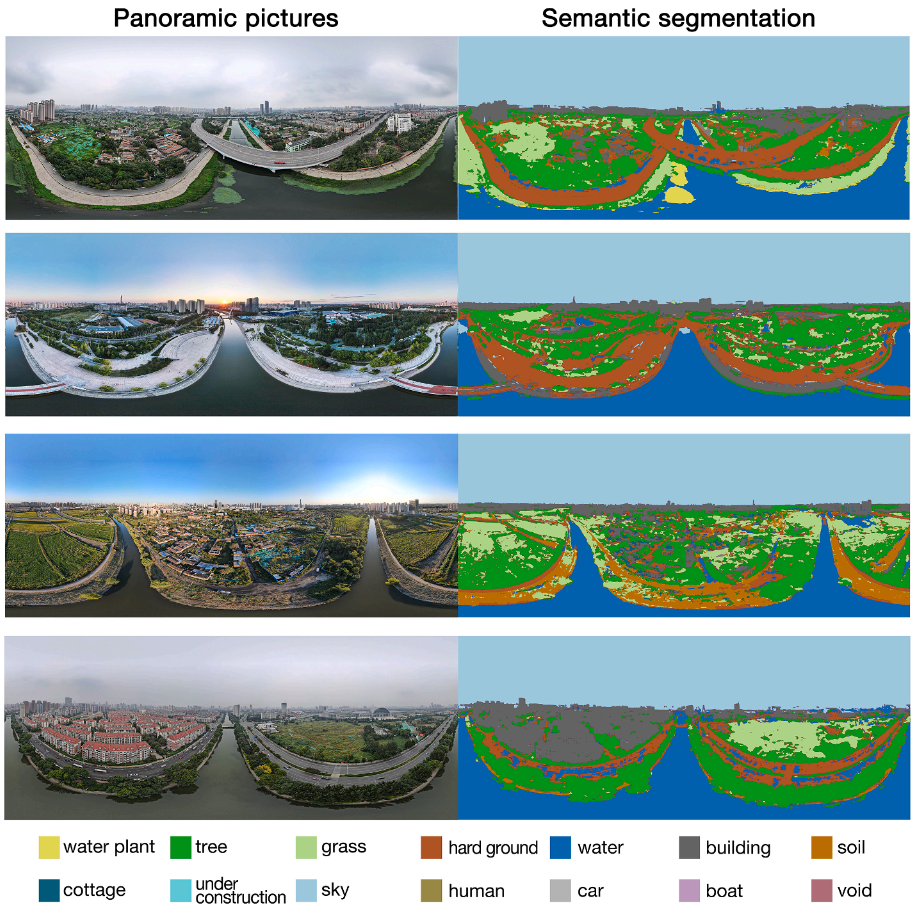

Fig. 4. UAV panoramic oblique images and their semantic segmentation results.

helpful in analyzing river sceneries from various perspectives. After counting the 14 semantic segmentation contents in all panoramic images, we discover that, in general, the proportion of sky, water and tree is 75%, which forms the leading skeleton of the river landscape. Among them, the sky accounts for 38%, the water accounts for 23%, and the tree accounts for 14%. This part of the Grand Canal's vision is dominated by these features, which form the primary visual style. Hard ground, buildings, and grassland make up a smaller percentage of the total, with 9% of hard ground, 7% of buildings, and 6% of grassland. The proportions of soil, cottage, automobile, boat, and other components, on the other hand, are tiny, with the proportion of soil being 3% and the proportion of other elements being less than 1%.

4.1.2. Evaluation results

To answer the second research question, we compared the objective and subjective visual characteristics of 48 locations of the two rivers by classifying and counting pixel ratios of landscape elements in the panoramic images (Fig. 5). The geographical distribution of water, trees, and grass is noticeably unequal. The distribution of buildings, hard ground, and other objects, on the other hand, is rather uniform, whereas the variation in sky is smaller. Specifically, the visible area of the water shows the characteristics of more in the middle, less on both sides, and less in the south part of the studied portion of the canal than in the northern one. The visible area of the water at the canal intersection is substantially larger than the south canal and the north canal. The observable surface of water between mapping locations 19–26 is very large, whereas the average water area between mapping points 26–50 is higher than that of mapping sites 1–18, based on the placement of mapping points. In some locations, the distribution of trees reveals the characteristics of considerable changes. The visible area of the tree is higher between mapping locations 11–18 than it is in other regions, whereas the visible area of mapping points 32–38 and 6–10 is slightly lower than that of mapping points 1–5 and 40–48. The overall spatial alteration of the building elements is minor. As a total, the proportion of buildings indicates a slight declining tendency from south to north. Among them, the proportion of buildings between mapping points 14 and 19 of the south canal is relatively high, while the proportion of buildings between 1 and 10 is fairly low. The proportion of buildings in the north canal, on the other hand, is lower overall.

By adding the proportions of different landscape elements, we obtained the spatial distribution of five indexes. GVI is composed of trees, grass and aquatic plants, and it presents the characteristics of less in the middle and more on both sides in space. Specifically, the GVI of mapping sites 19–24 is particularly low, whereas the GVI of locations 1–18 and 32–48 is relatively high. The spatial green visibility of this part of the Grand Canal is directly proportional to the distance from the mapping locations to the central urban area, indicating that the higher the green visibility, the further away from the urban centre. BVI is composed of building, cottage and under construction place, and it changes little in space. Locations 12–21 have a slightly higher BVI, while locations 22–40 have a comparatively low BVI. SKVI, WVI and HVI are separately composed of sky, water and hardground, so they are consistent with the spatial distribution characteristics of these three elements.

4.2. Subjective visual perception results

We cross-verified the results of the immersive VR and non-immersive VR perception experiments, and analyzed the correlation between them. The scores from the immersive VR experiment were highly correlated with non-immersive VR experiment scores in six perceptual indicators:

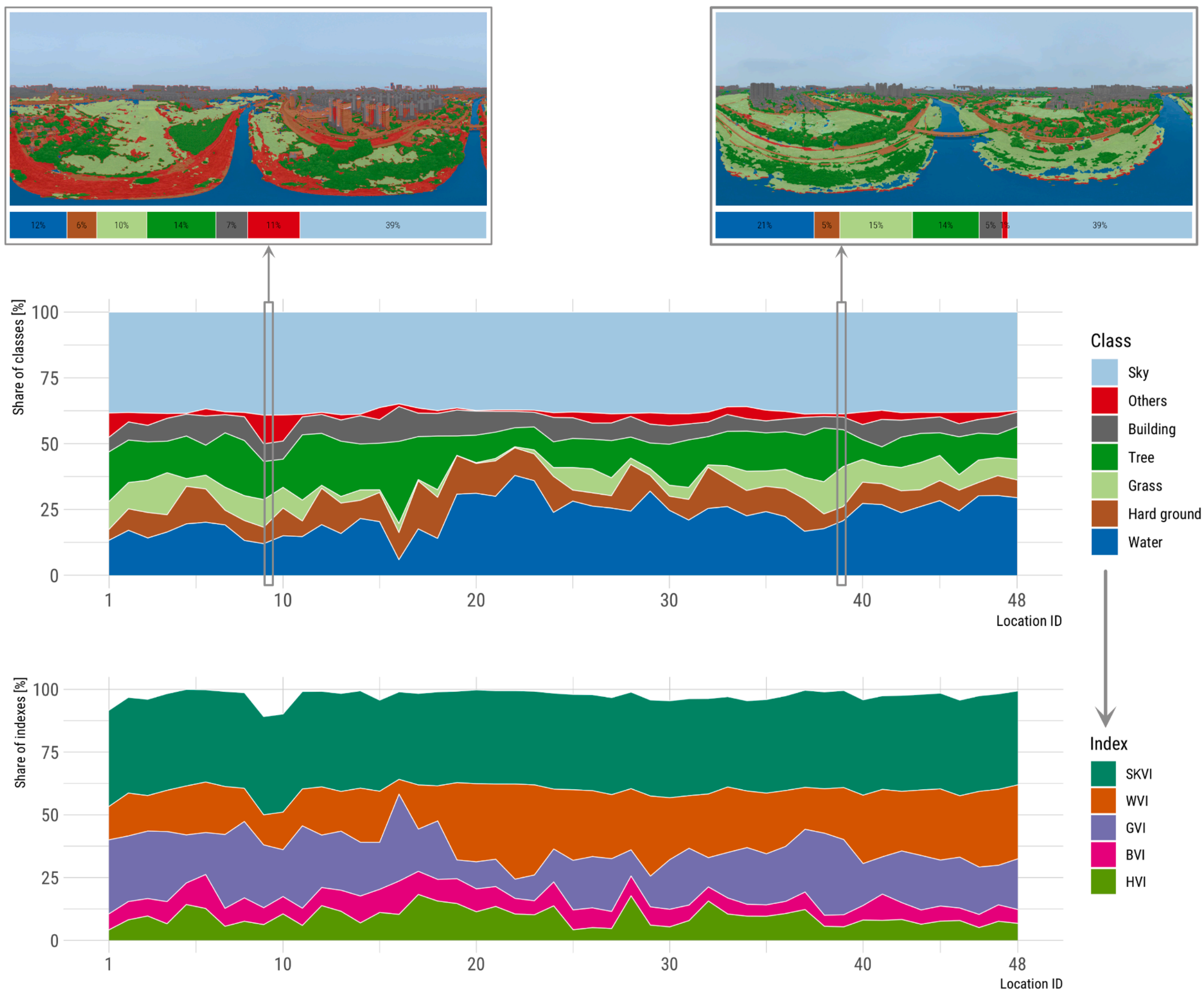

Fig. 5. Visual evaluation of the results: the top images portray the segmentation results of panoramas at two locations in the sequence. The middle image indicates the sequential distribution of visual elements throughout the observed points (the most common 6 classes are included). The bottom image illustrates the results characterised by the index system (derived from the classes visible in the middle plot). The indexes, which are mutually exclusive, do not add up to 100% because not all classes are part of them.

beauty (Pearson correlation coefficient $r = 0.841$, $p < 0.01$), pleasure ($r = 0.822$, $p < 0.01$), tranquillity ($r = 0.890$, $p < 0.01$), colour ($r = 0.731$, $p < 0.01$), complexity ($r = 0.757$, $p < 0.01$), and liveliness ($r = 0.675$, $p < 0.01$).

Six types of visual perception indexes of UAV panoramic photographs were quantitatively studied in 48 sites in this study (Fig. 6), the values of the six indexes were the average of the immersive VR and non-immersive VR experiments. On the whole, the average value of beauty of river landscape in the research region is relatively high, which is 3.949, the maximum value is 5.920, which appears at point 14, and

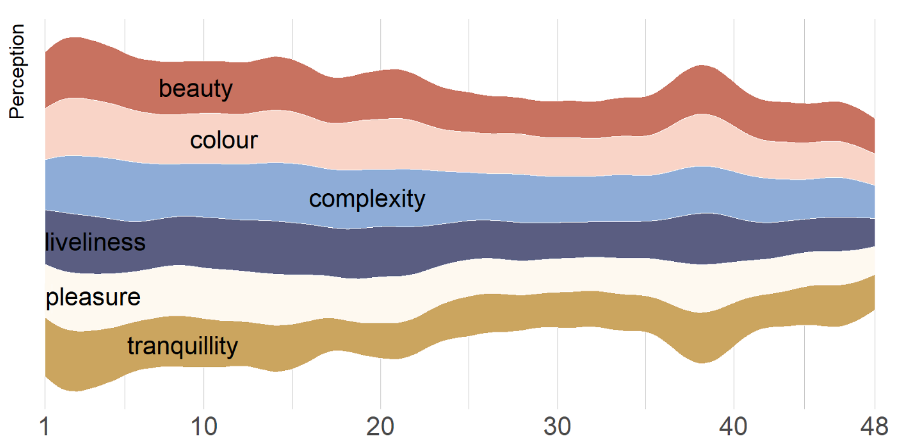

Fig. 6. Visual perception results throughout the linear study area, visualized as a streamgraph.

the minimum value is 2.170 at point 32. We observe that points in the range 1–14 and 37–41 have greater ratings, whereas points 26–35 have lower values. The mean value of notion of pleasure is the lowest, standing at 3.648, while the maximum value is 5.910, which occurs in location 5. The minimum value is 1.925, which appears at point 32. The average value of tranquillity is 3.657, the maximum value is 5.980 (location 2), and the minimum value is 1.425 (location 18). Overall, the tranquillity score of the south canal is slightly higher than that of the north canal, while the score of the middle position (16–24) is lower. The average value of colour is 3.893, the maximum value is 5.725, which occurs in point 2, and the minimum value is 2.175 at location 32. The average value of complexity is the highest, 4.350, the maximum value is 5.970, which appears in point 19, and the minimum value is 2.620, in position 44. The average value of liveliness is 3.826, the maximum value is 6.125, which occurs at position 2, and the minimum value is 2.130, which occurs at position 35.

4.3. Correlation analysis

To answer the third research question, we quantitatively explored the relationship between five visual evaluation indexes and six visual perception indexes based on the selected river landscape assessment and visual perception results.

The relationships, visualised in Fig. 7a, indicate that the GVI and beauty, pleasure, tranquillity, colour and liveliness are all significant, and the correlation coefficient values are 0.55, 0.52, 0.64, 0.38 and 0.4, respectively, all of which are positive and in the moderate range, indicating that there is an association between the green vegetation and these five perception indexes. Simultaneously, the correlation coefficient between GVI and complexity is close to 0, showing that GVI and complexity do not exhibit a relationship. The correlation coefficient between SKVI and tranquillity is 0.32, which means there is a positive correlation between sky and tranquillity. In contrast, the correlation coefficient between SKVI and beauty, pleasure, colour, complexity and liveliness is around 0, indicating no relationship between sky and these indexes. There is a significant correlation between BVI and complexity, and the correlation coefficient is 0.47, which means that there is a positive correlation between buildings and complexity. However, the correlation coefficient between BVI and beauty, pleasure, tranquillity, colour and liveliness is close to 0, indicating that there is no clear correlation. The correlation coefficient between HVI and tranquillity and complexity is significant. Specifically, the correlation coefficient between HVI and tranquillity is −0.49, indicating a significant negative correlation between hard ground and tranquillity. Between HVI and complexity, the correlation coefficient is 0.41 and shows the significance of a 0.05 level, which shows a significant positive correlation between the two indexes. In addition, the correlation between HVI and beauty, pleasure, colour and liveliness is not significant ($p > 0.05$), which means no correlation between hard ground and these four indexes. The WVI and beauty, pleasure, tranquillity, colour and liveliness all show a significant correlation, and the correlation coefficient values are −0.46, −0.46, −0.44, −0.39 and −0.45, respectively, all of which are less than 0, which means a moderate negative correlation between the water and beauty, pleasure, tranquillity, colour and liveliness. At the same time, there is no significant relationship between WVI and complexity, and the correlation coefficient is close to 0, suggesting no correlation between water and complexity. It can be seen that the water conditions in the study area are not pleasant, which will produce negative emotions for people. Finally, Fig. 7b indicates the correlations among the indexes. HVI and BVI, the two new indexes introduced in this paper, are not strongly correlated with any other index, affirming their uniqueness and contribution, and thus, we propound that they complement existing indexes.

5. Discussion

5.1. UAV perspective and the Semantic Riverscapes dataset

River-related landscape design and construction continue to account for a significant portion of the overall environmental development. Therefore, it is crucial to investigate the current characteristics of river landscapes, and understanding this issue visually remains central to both the government and research institutions. However, most existing approaches heavily rely on field survey workflow, including the investigation of riversides on the ground, which is time-consuming, labour intensive and costly; therefore, these means could benefit from introducing new technologies (Yamashita, 2002; Sun et al., 2021). During our literature study, we discovered that neither satellite imagery nor SVI is optimal for visual perception and evaluation of large-scale riverscapes. The UAV, on the other hand, offers significant benefits for these operations, but we discovered that no research had been done on river subjective visual perception using drone oblique imagery and VR and objective visual evaluation utilizing CV. In this study, we proposed to use UAV oblique photography to assess river landscapes, which can obtain a larger perspective and more content than a human viewpoint, and has become an important auxiliary tool and method for overall

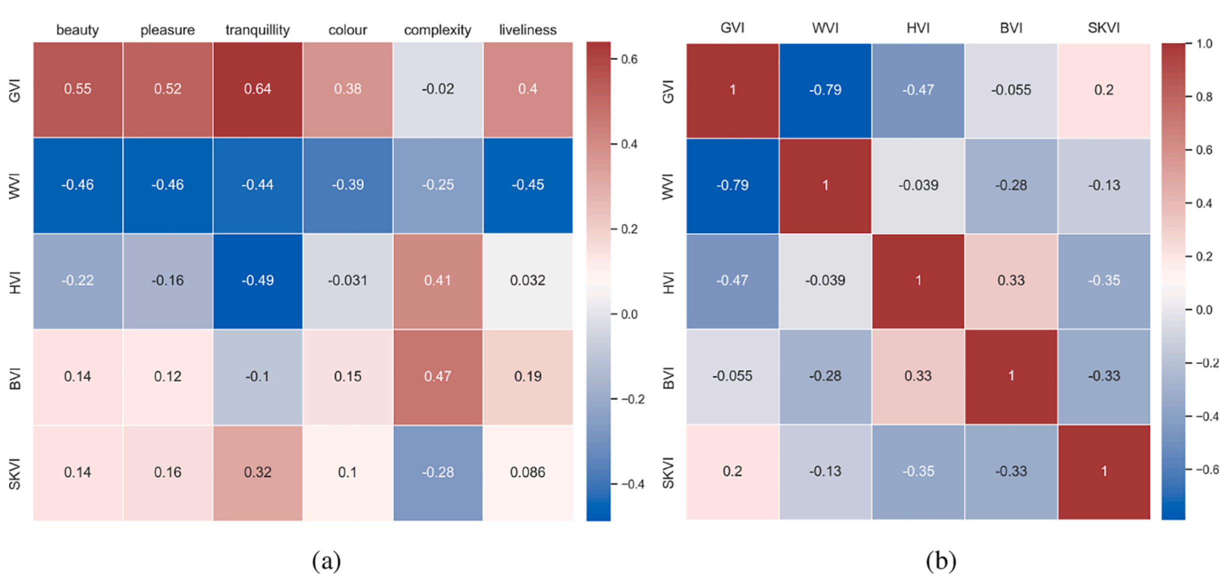

Fig. 7. Correlation coefficients among (a) the visual perception results and evaluation results; and (b) the indexes.

understanding of large-scale landscapes (Meng et al., 2021). We used an immersive sensing device (head-mounted display) and non-immersive sensing equipment (iPad, PC, etc.) to achieve the remote perception of river landscapes from a UAV perspective. The immersive VR perception brings people a high-quality sense of presence, while non-immersive VR perception can enable more individuals to participate in this visual perception experiment remotely. We cross-verified the perception results of IVR and nIVR experiments and found a high correlation between them, and we believe these remote visual evaluation approaches can provide a reference for the follow-up study of UAV-based VR perception.

The Semantic Riverscapes dataset is created in this study as a novel semantically annotated dataset of UAV oblique photography to aid in the comprehension of large-scale river landscapes and enrich the landscape of open UAV datasets which are scarce, and none of these hitherto includes rivers and the surrounding context. The 14 categories that have been regarded in the semantic segmentation (e.g. building, cottage, tree, grass) can be detected by training the deep learning model, with MPA reaching 90 percent, compensating for the current research flaws in this domain. We can accurately batch process river landscape photos of both urban and rural locations with this dataset and semantic segmentation model, and it is applicable. On this basis, the workflow we proposed can quickly obtain evaluation results for the general condition of river landscapes, as well as analyze and compare GVI, BVI, HVI, and other river indexes, allowing environmental management institutions and relevant public bodies to better understand the environmental characteristics of rivers and provide data support for improving the spatial quality of river scenery. We have released this dataset openly for public use, together with documentation. The dataset has been released under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license (CC BY-NC-SA 4.0) on Github,[2] filling the aforementioned void in the field and complementing existing datasets (cf. Table 1). With this open dataset, practitioners and researchers can use it to conduct a large number of river scene-related studies, which we hope will promote the development of this field. Further contributions of this open dataset are: (i) it is a linear dataset and one that is focused on heritage, potentially benefiting research on other linear landscapes and types of heritage; (ii) it represents a study area in Asia and it contains oblique imagery, while existing open UAV datasets (Section 2.2) are mostly focused on other locations and other perspectives; and (iii) it contains a relatively large number of classes benefiting other types of research.

5.2. Riverscape characteristics

The environment features in most river landscapes remain ambiguous, and there is a rising conflict between the needs of riverside environment understanding and river visual perception and evaluation. This is the first study to associate subjective VR perceptions of large-scale urban riverscapes from UAV oblique imagery in conjunction with a computer vision technique. Through the classified statistics of the pixel proportion of different landscape elements in the oblique photography panoramic pictures of dozens of locations in the research area, we accurately analyzed the proportion of different landscape elements in different areas and obtained the visual evaluation results. The physical setting of a place will affect people's subjective visual perceptions of the site (Tabrizian, Baran, Van Berkel, Mitasova, & Meentemeyer, 2020). Using the common perception indexes of beauty, pleasure, tranquillity, colour, complexity, and liveliness, the subjective VR perceptions of river scenery were quantitatively examined to produce the visual perception findings, which is in line with related work examining other dimensions of urban landscapes. After analyzing the correlation between objective visual evaluation results and subjective visual perception results, we found that GVI exhibited an obvious positive correlation with beauty, pleasure, tranquillity, colour and liveliness, which is similar to the results of street-level GVI analysis (Zhang et al., 2018; Ma et al., 2021). Therefore, it can be proved that the influence of GVI on people's perception is not only applicable to urban street landscapes but also applicable to riverscapes from the oblique viewpoint, like the perspective of UAV. The HVI had a negative correlation with tranquillity, and HVI, BVI and complexity showed a positive correlation, similar to related research conclusions (Li et al., 2021; Kerebel, Gélinas, Déry, Voigt, & Munson, 2019). In other words, plants are conducive to improving the visual quality of river landscapes, while artificial objects such as buildings and roads will affect and reduce people's perception of beauty and pleasure. However, this study found that there is a negative correlation between water and beauty, pleasure, tranquillity, colour and liveliness, which is different from the previous research results (Li et al., 2021), as the cited research highlights that the water quality in different regions and other influencing factors will affect the overall visual quality of river landscapes. To understand why the water body is negatively correlated with the perception indicators (beauty, pleasure, etc.), we examined the water body in these panoramic photographs and discovered that the colour of the water is not a pleasant blue, but rather a dark grey, making it unappealing. We further consulted the water quality information of these rivers and found that there is a lack of water resources in Tianjin, accompanied by severe water pollution (Cao et al., 2021). Therefore, poor water quality can have a negative impact on people's visual experience, which is also confirmed by previous research conclusions (Li, Chen, Hu, & Cho, 2021). The overhead viewpoint of high-rise residential buildings is comparable to that of UAVs. Li et al. (2021) has found a link between the visual characteristics (water visibility rate, green visibility rate, etc.) of urban rivers and housing values. The high green viewing rate of urban high-rise residential buildings and the river view with good water quality can raise house prices, whereas a low-quality river environment will lower house prices and affect people's environmental perception, which is similar to our correlation conclusion.

5.3. Limitations, challenges and future directions

Although we demonstrate that we can engage UAV oblique photography data and deep learning to analyze the characteristics of river landscapes instead of manual analysis, there are still some issues to be solved and this work leaves opportunities for further investigations.

- Firstly, image semantic segmentation can be more accurate with further efforts. At present, our dataset can identify green plants, but in higher latitudes, most of the vegetation in winter lacks green leaves and mostly exists in the form of branches, so it can not be identified as vegetation. In follow-up research, we plan to mark more images in different seasons, and then use it to train the existing model to obtain improved image segmentation.
- Secondly, while we had experimented with a few heights (Section 3.2), we chose 60 m as the altitude of UAV to obtain data, which maintains the consistency of data, but we have not studied the data at other altitudes extensively. Therefore, more data at different heights will be considered in the future to deepen the understanding of spatial features.
- Thirdly, we used manual flight to obtain data; thus, the number of UAV aerial survey locations is still relatively limited. The survey areas can be further expanded to lay a foundation for larger quantitative analysis studies of visual perception and evaluation. In the later stage, we will consider using a full-automatic program (e.g. GeoAI-empowered approaches (Liu & Biljecki, 2022)) to control the UAV to obtain spatial data so as to further expand the research area and reduce the workload and shorten the time interval for obtaining data.

[2] The dataset is available at https://github.com/ualsg/semantic-riverscapes-dataset.

- Finally, in the age of 'Metaverse' and digital twin, perceptions using UAV and virtual reality are widely adopted. The best way for humans to view these bird-eye-level scenes is via virtual and distant methods because it is practical in this study and future landscape evaluation and perception works (Pavlik, 2020). However, it is worth noting that the evaluation based on the means of virtual reality may not reflect the ratings in the real world. Compared with the real-world visual perception, the remote virtual rating may face bias. Therefore, our proposed method of remote visual perception can only provide a reference for future studies which adopt the same approaches as ours. Additionally, we adopted six common perceptual indexes (beauty, pleasure, tranquillity, etc.) according to our study aims. Future works can explore more specific indicator systems for visual perception analysis.

With the further development of UAV autopilot technologies, the efficiency of acquiring research data will be improved in the coming years. The accumulation of drone oblique photographic images and the assistance of automatic analysis technologies such as CV, UAV-related data will become a useful tool for large-scale spatial analysis and monitoring. Furthermore, without the need for surveyors to study the river environment, extensive data can be gathered. As a result, it has a great potentiality in regions with poor field conditions. Through the standardized UAV data processing process, the results of this study can not only facilitate promotion in different regions, but also meet the needs of iterative data updates in the same area, and it will also help to analyze the dynamic change characteristics of landscapes on a time scale, so as to improve the refinement and efficiency of spatial management, which will be used by urban planners, environmental managers and other researchers.

This study can be used to investigate large-scale river landscapes, provide a reference for the authorities to formulate riverside development policies, and can also be used to guide river planning projects. At the same time, the results of this study can benefit the construction of the National Cultural Park of the Grand Canal, and the methods can also help the government (and others leading similar projects elsewhere) to have a macro understanding of the overall situation of the river as a basis for follow-up works. For future work, we also plan to investigate whether we can render simulated scenarios of future redevelopments and predict the perception of each of these proposed scenarios to assist in decision-making. In addition, we intend to investigate the application of segmented 3D city models to enrich our approach, e.g. using other openly released datasets, complementing ours (Gao, Nan, Boom, & Ledoux, 2021), and to infuse soundscape into the models to better understand the built environment (Edler, Kühne, Keil, & Dickmann, 2019; Hruby, 2019). For future instances of the dataset, we also plan to include an additional urban area.

6. Conclusion

We developed a visual analysis workflow based on UAV oblique panoramas for understanding macro river landscapes by combining subjective visual perceptions and objective visual evaluation through automated CV approaches, a novelty in this domain. Our method relies on concurrent experiments involving immersive and non-immersive experiences, a rarity. Satellite imagery has dominated related analyses in the built environment, and the rise of street view imagery has been pivoting and revolutionary. Still, these two types are often out of reach—in terms of coverage, clarity, access to the data, and acquisition flexibility. We show that UAVs are the middle ground with unique advantages, and they provide a new perspective that cannot be rivalled by the aforementioned types. By introducing UAV oblique photography, a standardized workflow of UAV mapping, oblique image semantic segmentation, immersive VR and non-immersive VR experiences are constructed to achieve the automatic landscape evaluation and people's perception effectively and remotely. Besides a novel application of UAV oblique imagery in this research line, there are several key contributions of this study. First, we generated Semantic Riverscapes, an open semantic segmentation dataset of UAV oblique photography images based on river landscapes. Using this dataset and CV algorithms, rivers and surrounding landscapes can be analyzed automatically and efficiently, which overcomes the shortcomings of the state of the art. Second, we obtained 48 oblique panoramic images and quantitatively analyzed the proportion of 14 landscape elements such as buildings, trees and water in different locations of the river by using computer vision. The index system of river visual evaluation was extended with two novel instances, presenting a versatile set of several indexes. According to five of them, the river landscape was visually evaluated, and the evaluation results of the research area were obtained. Third, we used VR to visualize panoramic images, and had more than a hundred of participants in a non-immersive VR remote virtual experience and in an immersive VR perception of the river landscapes, and obtained their subjective visual perception of six dimensions (beauty, pleasure, tranquillity, colour, complexity and liveliness) through a systematic questionnaire. Also, we compared the two approaches, discovering their relationships. Fourth, we analyzed the correlation between the visual evaluation data of image semantic segmentation and human perception data and found the relationship between people's visual perception and landscape environment; further, we explored the possible reasons for the correlation findings, which indicate that the variables 'vegetation' exhibited a positive correlation with beauty, pleasure, tranquillity, colour and liveliness, consists with the results of the street-level analysis. Our results also indicate that the variable 'water' had a negative correlation with these perceptual indicators, which is different from the previous research results, and we explored the possible reasons. Therefore, our findings and proposed workflow can help planners to gather a macro understanding of the overall situation of the river and prompt authorities to formulate riverside development policies, which are beneficial to the river-related environment.

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.

Acknowledgements

We gratefully acknowledge the participants of the experiment. We thank the members of the NUS Urban Analytics Lab for the discussions, in particular, Pengyuan Liu for his comments, and Zexin Lei and Wenzheng Zhang for the help with the UAV data collection. The Institutional Review Board of the National University of Singapore has reviewed and approved the ethical aspects of this research (reference code NUS-IRB-2021–941). The first author has been supported by the China Scholarship Council under the State Scholarship Fund (202106250086). The second author has been supported by the Department of Education of Guangdong Province, China. This research is part of the projects (i) Research on the theory and digital technologies of the Grand Canal's cultural heritage protection, which is supported by the National Social Science Foundation of China (19ZDA193); (ii) The Technical Key Project of Shenzhen Science and Technology Innovation Commission Grant JSGG20201103093401004; and (iii) Large-scale 3D Geospatial Data for Urban Analytics, which is supported by the National University of Singapore under the Start Up Grant R-295-000-171–133.

References

Ahmed, A., Ngoduy, D., Adnan, M., & Baig, M. A. U. (2021). On the fundamental diagram and driving behavior modeling of heterogeneous traffic flow using uav-based data. *Transportation Research Part A: Policy and Practice, 148*, 100–115.

Al-Bahri, M., Al Kishri, W., & Dharamshi, R. R. (2021). Implementation of augmented reality and drones to serve smart cities. *Artificial Intelligence & Robotics Development Journal*, 147–157.

Ashilah, Q. P., Hernina, R., et al. (2021). Urban slum identification in bogor tengah sub-district, bogor city using unmanned aerial vehicle (uav) images and object-based image analysis. In *IOP Conference Series: Earth and Environmental Science* (p. 012133). IOP Publishing.

Bacco, M., Barsocchi, P., Cassará, P., Germanese, D., Gotta, A., Leone, G. R., Moroni, D., Pascali, M. A., & Tampucci, M. (2020). Monitoring ancient buildings: real deployment of an iot system enhanced by uavs and virtual reality. *IEEE Access, 8*, 50131–50148.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence, 39*, 2481–2495.

Bakirman, T., Bayram, B., Akpinar, B., Karabulut, M. F., Bayrak, O. C., Yigitoglu, A., & Seker, D. Z. (2020). Implementation of ultra-light uav systems for cultural heritage documentation. *Journal of Cultural Heritage, 44*, 174–184.

Baranwal, E., Raghvendra, S., Tiwari, P. S., & Pande, H. (2021). Health monitoring and assessment of the cultural monument through unmanned aerial vehicle (uav) image processing. *Advances in Systems Engineering. Springer*, 145–160.

Beg, A., Qureshi, A. R., Sheltami, T., & Yasar, A. (2021). Uav-enabled intelligent traffic policing and emergency response handling system for the smart city. *Personal and Ubiquitous Computing, 25*, 33–50.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & psychophysics, 8*, 279–286.

Bhardwaj, A., Sam, L., Martín-Torres, F. J., Kumar, R., et al. (2016). Uavs as remote sensing platform in glaciology: Present applications and future prospects. *Remote sensing of environment, 175*, 196–204.

Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning, 215*, Article 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 and Urban Planning, 189*, 129–138.

Brumana, R., Barazzetti, L., Oreni, D., & Roncoroni, F. (2013). Uav panoramic images for waterfront landscape analysis and topographicdb texturing. *International Conference on Computational Science and Its Applications, Springer.*, 328–343.

Brumana, R., Oreni, D., Alba, M., Barazzetti, L., Cuca, B., & Scaioni, M. (2012). Panoramic uav views for landscape heritage analysis integrated with historical maps atlases. *Geoinformatics FCE CTU, 9*, 39–50.

Cai, Z., Fang, C., Zhang, Q., & Chen, F. (2021). Joint development of cultural heritage protection and tourism: the case of mount lushan cultural landscape heritage site. *Heritage Science, 9*, 1–16.

Cao, Q., Yu, G., Sun, S., Dou, Y., Li, H., & Qiao, Z. (2021). Monitoring water quality of the haihe river based on ground-based hyperspectral remote sensing. *Water, 14*, 22.

Castrignanò, A., Belmonte, A., Antelmi, I., Quarto, R., Quarto, F., Shaddad, S., Sion, V., Muolo, M. R., Ranieri, N. A., Gadaleta, G., et al. (2021). A geostatistical fusion approach using uav data for probabilistic estimation of xylella fastidiosa subsp. pauca infection in olive trees. *Science of The Total Environment, 752*, Article 141814.

Cavalcante, A., Mansouri, A., Kacha, L., Barros, A. K., Takeuchi, Y., Matsumoto, N., & Ohnishi, N. (2014). Measuring streetscape complexity based on the statistics of local contrast and spatial frequency. *PloS one, 9*, Article e87097.

Che, Y., Wang, Q., Xie, Z., Zhou, L., Li, S., Hui, F., Wang, X., Li, B., & Ma, Y. (2020). Estimation of maize plant height and leaf area index dynamics using an unmanned aerial vehicle with oblique and nadir photography. *Annals of botany, 126*, 765–773.

Chen, K., Reichard, G., Xu, X., & Akanmu, A. (2021). Automated crack segmentation in close-range building façade inspection images using deep learning techniques. *Journal of Building Engineering, 43*, Article 102913.

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

Chowdhury, T., & Rahnemoonfar, M. (2021). Attention for damage assessment. *UMBC Faculty Collection.*

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 *in: Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3213–3223).

Ding, L., Zhou, J., Meng, L., & Long, Z. (2021). A practical cross-view image matching method between uav and satellite for uav-based geo-localization. *Remote Sensing, 13*, 47.

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.

Edler, D., Kühne, O., Keil, J., & Dickmann, F. (2019). Audiovisual cartography: Established and new multimedia approaches to represent soundscapes. *KN-Journal of Cartography and Geographic Information, 69*, 5–17.

Elghaish, F., Matarneh, S., Talebi, S., Kagioglou, M., Hosseini, M. R., & Abrishami, S. (2020). Toward digitalization in the construction industry with immersive and drones technologies: a critical literature review. In *Smart and Sustainable Built Environment*.

Emilien, A. V., Thomas, C., & Thomas, H. (2021). Uav & satellite synergies for optical remote sensing applications: A literature review. *Science of Remote Sensing, 100019*.

Erdelj, M., & Natalizio, E. (2016). Uav-assisted disaster management: Applications and open issues. In *in: 2016 international conference on computing, networking and communications (ICNC), IEEE* (pp. 1–5).

Esposito, G., Mastrorocco, G., Salvini, R., Oliveti, M., & Starita, P. (2017). Application of uav photogrammetry for the multi-temporal estimation of surface extent and volumetric excavation in the sa pigada bianca open-pit mine, sardinia, italy. *Environmental Earth Sciences, 76*, 1–16.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International journal of computer vision, 88*, 303–338.

Feng, W. Z. (2021). Implementation of aerial panoramic photography for environmental studies through vr experiences. *Journal of Environmental Science Studies, 4*, 1.

Fong, W., Ong, S. K., & Nee, A. Y. (2009). Computer vision centric hybrid tracking for augmented reality in outdoor urban environments, in. In *Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry* (pp. 185–190).

Gao, W., Nan, L., Boom, B., & Ledoux, H. (2021). SUM: A benchmark dataset of semantic urban meshes. *ISPRS Journal of Photogrammetry and Remote Sensing, 179*, 108–120. https://doi.org/10.1016/j.isprsjprs.2021.07.008. URL:https://doi.org/10.1016%2Fj.isprsjprs.2021.07.008.

Garau, E., Torralba, M., & Pueyo-Ros, J. (2021). What is a river basin? assessing and understanding the sociocultural mental constructs of landscapes from different stakeholders across a river basin. *Landscape and Urban Planning, 214*, Article 104192.

Garg, P., Chakravarthy, A. S., Mandal, M., Narang, P., Chamola, V., & Guizani, M. (2021). Isdnet: Ai-enabled instance segmentation of aerial scenes for smart cities. *ACM Transactions on Internet Technology (TOIT), 21*, 1–18.

Ghamisi, P., Rasti, B., Yokoya, N., Wang, Q., Hofle, B., Bruzzone, L., Bovolo, F., Chi, M., Anders, K., Gloaguen, R., et al. (2019). Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art. *IEEE Geoscience and Remote Sensing Magazine, 7*, 6–39.

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.

Gottwald, S., & Stedman, R. C. (2020). Preserving ones meaningful place or not? understanding environmental stewardship behaviour in river landscapes. *Landscape and Urban Planning, 198*, Article 103778.

Guerra-Hernández, J., Díaz-Varela, R. A., Ávarez-González, J. G., & Rodríguez-González, P. M. (2021). Assessing a novel modelling approach with high resolution uav imagery for monitoring health status in priority riparian forests. *Forest Ecosystems, 8*, 1–21.

Guo, Y., Fu, B., Wang, Y., Xu, P., & Liu, Q. (2021). Identifying spatial mismatches between the supply and demand of recreation services for sustainable urban river management: a case study of jinjiang river in chengdu, china. *Sustainable Cities and Society, 103547*.

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. In *in: Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1551–1560).

Harknett, J., Whitworth, M., Rust, D., Krokos, M., Kearl, M., Tibaldi, A., Bonali, F., Wyk, Van, de Vries, B., Antoniou, V., Nomikou, P., Reitano, D., Falsaperla, S., Vitello, F., & Becciani, U. (2022). The use of immersive virtual reality for teaching fieldwork skills in complex structural terrains. *Journal of Structural Geology, 104681*. https://doi.org/10.1016/j.jsg.2022.104681

He, L., Páez, A., & Liu, D. (2017). Built environment and violent crime: An environmental audit approach using google street view. *Computers, Environment and Urban Systems, 66*, 83–95.

Hervouet, A., Dunford, R., Piégay, H., Belletti, B., & Trémélo, M. L. (2011). Analysis of post-flood recruitment patterns in braided-channel rivers at multiple scales based on an image series collected by unmanned aerial vehicles, ultra-light aerial vehicles, and satellites. *GIScience & Remote Sensing, 48*, 50–73.

Hosseini, M., Miranda, F., Lin, J., & Silva, C. T. (2022). Citysurfaces: City-scale semantic segmentation of sidewalk materials. *Sustainable Cities and Society, 103630*.

Hritz, C. (2014). Contributions of gis and satellite-based remote sensing to landscape archaeology in the middle east. *Journal of Archaeological Research, 22*, 229–276.

Hruby, F. (2019). The sound of being there: Audiovisual cartography with immersive virtual environments. *KN-Journal of Cartography and Geographic Information, 69*, 19–28.

Hu, C. B., Zhang, F., Gong, F. Y., Ratti, C., & Li, X. (2020). Classification and mapping of urban canyon geometry using google street view images and deep multitask learning. *Building and Environment, 167*, Article 106424.

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

Iizuka, K., Itoh, M., Shiodera, S., Matsubara, T., Dohar, M., & Watanabe, K. (2018). Advantages of unmanned aerial vehicle (uav) photogrammetry for landscape analysis compared with satellite data: A case study of postmining sites in indonesia. *Cogent Geoscience, 4*, 1498180.

Isaacs, J. P., Gilmour, D. J., Blackwood, D. J., & Falconer, R. E. (2011). Immersive and non immersive 3d virtual city: decision support tool for urban sustainability. *Journal of Information Technology in Construction., 16*, 151–162.

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

Jahani, A., & Saffariha, M. (2020). Aesthetic preference and mental restoration prediction in urban parks: an application of environmental modeling approach. *Urban Forestry & Urban Greening, 54*, Article 126775.

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. *Building and Environment, 169*, Article 106544.

Jin, X., & Wang, J. (2021). Assessing linear urban landscape from dynamic visual perception based on urban morphology. *Frontiers of Architectural Research, 10*, 202–219.

Karthik, M., Usha, S., Predeep, B., Saran, G. S., Sridhar, G., & Theeksith, R. (2021). Design and development of solar powered unmanned aerial vehicle (uav). In *for surveying, mapping and disaster relief, in: AIP Conference Proceedings, AIP Publishing LLC* (p. 140025).

Keil, J., Edler, D., Schmitt, T., & Dickmann, F. (2021). Creating immersive virtual environments based on open geospatial data and game engines. *KN-Journal of Cartography and Geographic Information, 71*, 53–65.

Kerebel, A., Gélinas, N., Déry, S., Voigt, B., & Munson, A. (2019). Landscape aesthetic modelling using bayesian networks: Conceptual framework and participatory indicator weighting. *Landscape and Urban Planning, 185*, 258–271.

Kerkech, M., Hafiane, A., & Canals, R. (2020). Vine disease detection in uav multispectral images using optimized image registration and deep learning segmentation approach. *Computers and Electronics in Agriculture, 174*, Article 105446.

Khaliq, A., Comba, L., Biglia, A., Ricauda Aimonino, D., Chiaberge, M., & Gay, P. (2019). Comparison of satellite and uav-based multispectral imagery for vineyard variability assessment. *Remote Sensing, 11*, 436.

Ki, D., & Lee, S. (2021). Analyzing the effects of green view index of neighborhood streets on walking time using google street view and deep learning. *Landscape and Urban Planning, 205*, Article 103920.

Kikuchi, N., Fukuda, T., & Yabuki, N. (2022). Future landscape visualization using a city digital twin: integration of augmented reality and drones with implementation of 3d model-based occlusion handling. *Journal of Computational Design and Engineering, 9*, 837–856.

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

Krause, C. L. (2001). Our visual landscape: Managing the landscape under special consideration of visual aspects. *Landscape and Urban planning, 54*, 239–254.

Lan, G., Sun, J., Li, C., Ou, Z., Luo, Z., Liang, J., & Hao, Q. (2016). Development of uav based virtual reality systems. In *in: 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), IEEE* (pp. 481–486).

Lan, J. K. W., & Lee, F. K. W. (2021). Drone forensics: A case study on dji mavic air 2, in. In *2021 23rd International Conference on Advanced Communication Technology (ICACT), IEEE* (pp. 291–296).

Li, C., Shen, S., & Ding, L. (2020). Evaluation of the winter landscape of the plant community of urban park green spaces based on the scenic beauty esitimation method in yangzhou, china. *PloS one, 15*, Article e0239849.

Li, J., Zhang, H., & Sun, Z. (2020). Spatiotemporal variations of land urbanization and socioeconomic benefits in a typical sample zone: A case study of the beijing-hangzhou grand canal. *Applied Geography, 117*, Article 102187.

Li, X. (2021). Examining the spatial distribution and temporal change of the green view index in new york city using google street view images and deep learning. *Environment and Planning B: Urban Analytics and City Science, 48*, 2039–2054.

Li, X., Chen, W. Y., Hu, F. Z. Y., & Cho, F. H. T. (2021). Homebuyers' heterogeneous preferences for urban green–blue spaces: A spatial multilevel autoregressive analysis. *Landscape and Urban Planning, 216*, Article 104250.

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. *Ecological Indicators, 133*, Article 108381.

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 and Urban Planning, 169*, 81–91.

Li, X., Santi, P., Courtney, T. K., Verma, S. K., & Ratti, C. (2018). Investigating the association between streetscapes and human walking activities using google street view and human trajectory data. *Transactions in GIS, 22*, 1029–1044.

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*, 675–685.

Li, Y., Karim, M. M., & Qin, R. (2022). *A virtual-reality-based training and assessment system for bridge inspectors with an assistant drone*. IEEE Transactions on Human-Machine Systems.

Li, Y., Wu, L., Han, Q., Wang, X., Zou, T., & Fan, C. (2021). Estimation of remote sensing based ecological index along the grand canal based on pca-ahp-topsis methodology. *Ecological Indicators, 122*, Article 107214.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 22*, 1–55.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. *European conference on computer vision, Springer.*, 740–755.

Lindner, C., Ortwein, A., Staar, K., & Rienow, A. (2021). Different levels of complexity for integrating textured extra-terrestrial elevation data in game engines for educational augmented and virtual reality applications. *KN-Journal of Cartography and Geographic Information, 71*, 253–267.

Liu, D., Xia, X., Chen, J., & Li, S. (2021). Integrating building information model and augmented reality for drone-based building inspection. *Journal of Computing in Civil Engineering, 35*, 04020073.

Liu, P., & Biljecki, F. (2022). A review of spatially-explicit GeoAI applications in urban geography. *International Journal of Applied Earth Observation and Geoinformation, 112*, Article 102936. https://doi.org/10.1016/j.jag.2022.102936

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 International Journal of Geo-Information, 11*. https://doi.org/10.3390/ijgi11080452. URL:https://www.mdpi.com/2220-9964/11/8/452.

Lytkin, L., & Syromyatnikov, I. (2021). In *Application of an unmanned aerial vehicle for large-scale mapping of thermokarst landforms, in: IOP Conference Series: Earth and Environmental Science* (p. 062030). IOP Publishing.

Lyu, Y., Vosselman, G., Xia, G. S., Yilmaz, A., & Yang, M. Y. (2020). Uavid: A semantic segmentation dataset for uav imagery. *ISPRS journal of photogrammetry and remote sensing, 165*, 108–119.

Ma, B., Hauer, R. J., Xu, C., & Li, W. (2021). Visualizing evaluation model of human perceptions and characteristic indicators of landscape visual quality in urban green spaces by using nomograms. *Urban Forestry & Urban Greening, 65*, Article 127314.

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*, Article 103086.

Ma, Y., Yu, D., Wu, T., & Wang, H. (2019). Paddlepaddle: An open-source deep learning platform from industrial practice. *Frontiers of Data and Domputing, 1*, 105–115.

Mandal, M., Kumar, L. K., & Vipparthi, S. K. (2020). Mor-uav: A benchmark dataset and baselines for moving object recognition in uav videos. In *in: Proceedings of the 28th ACM International Conference on Multimedia* (pp. 2626–2635).

Meinen, B. U., & Robinson, D. T. (2020). Mapping erosion and deposition in an agricultural landscape: Optimization of uav image acquisition schemes for sfm-mvs. *Remote Sensing of Environment, 239*, Article 111666.

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.

Meng, C., Song, Y., Ji, J., Jia, Z., Zhou, Z., Gao, P., & Liu, S. (2022). Automatic classification of rural building characteristics using deep learning methods on oblique photography. *Building Simulation, Springer.*, 1161–1174.

Ming, Y., Ya-duan, R., Lin-kai, C., Peng, Z., & Qi-mei, C. (2017). New video recognition algorithms for inland river ships based on faster r-cnn. *Journal of Beijing University of Posts and Telecommunications, 40*, 130.

Miraki, M., Sohrabi, H., Fatehi, P., & Kneubuehler, M. (2021). Individual tree crown delineation from high-resolution uav images in broadleaf forest. *Ecological Informatics, 61*, Article 101207.

Miřijovský, J., & Langhammer, J. (2015). Multitemporal monitoring of the morphodynamics of a mid-mountain stream using uas photogrammetry. *Remote sensing, 7*, 8586–8609.

Mouratidis, K., & Hassan, R. (2020). Contemporary versus traditional styles in architecture and public space: A virtual reality study with 360-degree videos. *Cities, 97*, Article 102499.

Munawar, H. S., Ullah, F., Qayyum, S., & Heravi, A. (2021). Application of deep learning on uav-based aerial images for flood detection. *Smart Cities, 4*, 1220–1242.

Newman, M., Gatersleben, B., Wyles, K., & Ratcliffe, E. (2022). The use of virtual reality in environment experiences and the importance of realism. *Journal of Environmental Psychology, 79*, Article 101733.

Nex, F., Armenakis, C., Cramer, M., Cucci, D., Gerke, M., Honkavaara, E., Kukko, A., Persello, C., & Skaloud, J. (2022). Uav in the advent of the twenties: Where we stand and what is next. *ISPRS Journal of Photogrammetry and Remote Sensing, 184*, 215–242. https://doi.org/10.1016/j.isprsjprs.2021.12.006. URL:https://www.sciencedirect.com/science/article/pii/S0924271621003282.

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

Okeil, A. (2010). Hybrid design environments: immersive and non-immersive architectural design. *Journal of Information Technology in Construction (ITcon), 15*, 202–216.

Paes, D., Irizarry, J., & Pujoni, D. (2021). An evidence of cognitive benefits from immersive design review: Comparing three-dimensional perception and presence between immersive and non-immersive virtual environments. *Automation in Construction, 130*, Article 103849.

Pang, H. E., & Biljecki, F. (2022). 3D building reconstruction from single street view images using deep learning. *International Journal of Applied Earth Observation and Geoinformation, 112*, Article 102859.

Papadopoulou, E. E., Papakonstantinou, A., Zouros, N., & Soulakellis, N. (2021). Scale-variant flight planning for the creation of 3d geovisualization and augmented reality maps of geosites: The case of voulgaris gorge, lesvos, greece. *Applied Sciences, 11*, 10733.

Park, S., & Lee, G. (2020). Full-immersion virtual reality: Adverse effects related to static balance. *Neuroscience letters, 733*, Article 134974.

Pavlik, J. V. (2020). Drones, augmented reality and virtual reality journalism: Mapping their role in immersive news content. *Media and Communication, 8*, 137–146.

Pettorelli, N., Schulte to Bühne, H., Tulloch, A., Dubois, G., Macinnis-Ng, C., Queirós, A. M., Keith, D.A., Wegmann, M., Schrodt, F., Stellmes, M., et al., 2018. Satellite remote sensing of ecosystem functions: opportunities, challenges and way forward. Remote Sensing in Ecology and Conservation 4, 71–93.

Portela, P., Vieira, C., Carvalho-Santos, C., Gonçalves, J., Durance, I., Honrado, J., 2021. Regional planning of river protection and restoration to promote ecosystem services and nature conservation.

Qi, Y., Chodron Drolma, S., Zhang, X., Liang, J., Jiang, H., Xu, J., & Ni, T. (2020). An investigation of the visual features of urban street vitality using a convolutional neural network. *Geo-spatial Information Science, 23*, 341–351.

Qu, C., Calyam, P., Yu, J., Vandanapu, A., Opeoluwa, O., Gao, K., Wang, S., Chastain, R., & Palaniappan, K. (2021). Dronecoconet: Learning-based edge computation offloading and control networking for drone video analytics. *Future Generation Computer Systems, 125*, 247–262.

Rahnemoonfar, M., Chowdhury, T., Sarkar, A., Varshney, D., Yari, M., & Murphy, R. R. (2021). Floodnet: A high resolution aerial imagery dataset for post flood scene understanding. *IEEE Access, 9*, 89644–89654.

del Río-Mena, T., Willemen, L., Tesfamariam, G. T., Beukes, O., & Nelson, A. (2020). Remote sensing for mapping ecosystem services to support evaluation of ecological

restoration interventions in an arid landscape. *Ecological indicators, 113*, Article 106182.
Rivas Casado, M., Ballesteros Gonzalez, R., Wright, R., & Bellamy, P. (2016). Quantifying the effect of aerial imagery resolution in automated hydromorphological river characterisation. *Remote Sensing, 8*, 650.
Rouse, L. M., Tabaldiev, K., & Matuzeviciute, G. M. (2021). Exploring landscape archaeology and uav-based survey in the kochkor valley, kyrgyzstan. *Journal of Field Archaeology*, 1–22.
Rusnák, M., Sládek, J., Kidová, A., & Lehotský, M. (2018). Template for high-resolution river landscape mapping using uav technology. *Measurement, 115*, 139–151.
Santos, I., Henriques, R., Mariano, G., & Pereira, D. I. (2018). Methodologies to represent and promote the geoheritage using unmanned aerial vehicles, multimedia technologies, and augmented reality. *Geoheritage, 10*, 143–155.
Schenone, S., Azhar, M., Ramírez, C. A. V., Strozzi, A. G., Delmas, P., & Thrush, S. F. (2021). Mapping the delivery of ecological functions combining field collected data and unmanned aerial vehicles (uavs). *Ecosystems*, 1–12.
Schmohl, S., Tutzauer, P., & Haala, N. (2020). Stuttgart city walk: A case study on visualizing textured dsm meshes for the general public using virtual reality. PFG–Journal of Photogrammetry. *Remote Sensing and Geoinformation Science, 88*, 147–154.
Seiferling, I., Naik, N., Ratti, C., & Proulx, R. (2017). Green streets- quantifying and mapping urban trees with street-level imagery and computer vision. *Landscape and Urban Planning, 165*, 93–101.
Shao, H., Song, P., Mu, B., Tian, G., Chen, Q., He, R., & Kim, G. (2021). Assessing city-scale green roof development potential using unmanned aerial vehicle (uav) imagery. *Urban Forestry & Urban Greening, 57*, Article 126954.
Sharma, C., Isha, I., & Vashisht, V. (2021). Water quality estimation using computer vision in uav. In *in: 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE* (pp. 448–453).
Sheffield, J., Wood, E. F., Pan, M., Beck, H., Coccia, G., Serrat-Capdevila, A., & Verbist, K. (2018). Satellite remote sensing for water resources management: Potential for supporting sustainable development in data-poor regions. *Water Resources Research, 54*, 9724–9758.
Shetty, S., 2016. Application of convolutional neural network for image classification on pascal voc challenge 2012 dataset. arXiv preprint arXiv:1607.03785.
Smaczyński, M., & Horbiński, T. (2021). Creating a 3d model of the existing historical topographic object based on low-level aerial imagery. *KN-Journal of Cartography and Geographic Information, 71*, 33–43.
Song, Y., Ning, H., Ye, X., Chandana, D., & Wang, S. (2022). *Analyze the usage of urban greenways through social media images and computer vision*. Environment and Planning B: Urban Analytics and City Science, 23998083211064624.
Sun, A. Y., & Scanlon, B. R. (2019). How can big data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. *Environmental Research Letters, 14*, Article 073001.
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. *Environmental Research Communications, 3*, Article 115013.
Sun, F., Yang, G., Zhang, A., Zhang, Y., et al. (2021). Circle-u-net: An efficient architecture for semantic segmentation. *Algorithms, 14*, 159.
Sun, L., Shao, H., Li, S., Huang, X., & Yang, W. (2018). Integrated application of eye movement analysis and beauty estimation in the visual landscape quality estimation of urban waterfront park. *International Journal of Pattern Recognition and Artificial Intelligence, 32*, 1856010.
Sun, L., Wang, J., Yang, K., Wu, K., Zhou, X., Wang, K., Bai, J., 2021c. Aerial-pass: Panoramic annular scene segmentation in drone videos. arXiv preprint arXiv: 2105.07209.
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 and Urban planning, 195*, Article 103704.
Tian, X., Shao, J., Ouyang, D., & Shen, H. T. (2021). Uav-satellite view synthesis for cross-view geo-localization. *IEEE Transactions on Circuits and Systems for Video Technology*.
Torgersen, C. E., Le Pichon, C., Fullerton, A. H., Dugdale, S. J., Duda, J. J., Giovannini, F., Tales, É., Belliard, J., Branco, P., Bergeron, N. E., et al. (2021). Riverscape approaches in practice: perspectives and applications. *Biological Reviews*.
Van Leeuwen, J. P., Hermans, K., Jylhä, A., Quanjer, A. J., & Nijman, H. (2018). Effectiveness of virtual reality in participatory urban planning: A case study. In *in: Proceedings of the 4th Media Architecture Biennale Conference* (pp. 128–136).
Verbrugge, L., & van den Born, R. (2018). The role of place attachment in public perceptions of a re-landscaping intervention in the river waal (the netherlands). *Landscape and Urban Planning, 177*, 241–250.
Verma, D., Jana, A., & Ramamritham, K. (2019). Machine-based understanding of manually collected visual and auditory datasets for urban perception studies. *Landscape and Urban Planning, 190*, Article 103604.
Watkins, S., Burry, J., Mohamed, A., Marino, M., Prudden, S., Fisher, A., Kloet, N., Jakobi, T., & Clothier, R. (2020). Ten questions concerning the use of drones in urban environments. *Building and Environment, 167*, Article 106458.
Wawrzyniak, N., & Stateczny, A. (2018). Automatic watercraft recognition and identification on water areas covered by video monitoring as extension for sea and river traffic supervision systems. Polish Maritime. *Research*.
Wei, Z., Wang, Y., Yi, H., Chen, Y., & Wang, G. (2020). Semantic 3d reconstruction with learning mvs and 2d segmentation of aerial images. *Applied Sciences, 10*, 1275.
Wen, H., Xiao, Y., & Zhang, L. (2017). Spatial effect of river landscape on housing price: An empirical study on the grand canal in hangzhou, china. *Habitat International, 63*, 34–44.
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 and Urban Planning, 219*, Article 104315.
Woodget, A. S., Austrums, R., Maddock, I. P., & Habit, E. (2017). Drones and digital photogrammetry: from classifications to continuums for monitoring river habitat and hydromorphology. *Wiley Interdisciplinary Reviews: Water, 4*, Article e1222.
Wu, A. N., & Biljecki, F. (2021). Roofpedia: Automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability. *Landscape and Urban Planning, 214*, Article 104167.
Wu, A. N., & Biljecki, F. (2022). GANmapper: geographical data translation. *International Journal of Geographical Information Science, 36*, 1394–1422.
Wu, X., Li, W., Hong, D., Tao, R., Du, Q., 2021. Deep learning for uav-based object detection and tracking: A survey. arXiv preprint arXiv:2110.12638.
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*, Article 126995.
Xian, M., Xu, F., Cheng, H. D., Zhang, Y., & Ding, J. (2016). Eiseg: Effective interactive segmentation. In *in: 2016 23rd International Conference on Pattern Recognition (ICPR), IEEE* (pp. 1982–1987).
Xiang, T.Z., Xia, G.S., Zhang, L., 2018. Mini-uav-based remote sensing: techniques, applications and prospectives. arXiv preprint arXiv:1812.07770.
Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). *Segformer: Simple and efficient design for semantic segmentation with transformers*. arXiv preprint arXiv:2105.15203.
Xin, L., Xiangrong, W., Liang, L., & Danzi, W. (2021). Visual perception evaluation of urban waterfront greenway landscape based on deep learning. *Journal of Beijing Forestry University, 43*, 1–12.
Xu, C., Oberman, T., Aletta, F., Tong, H., & Kang, J. (2020). Ecological validity of immersive virtual reality (ivr) techniques for the perception of urban sound environments. *Acoustics, MDPI.*, 11–24.
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 and Urban Planning, 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. *International Journal of Geographical Information Science, 34*, 1765–1791. https://doi.org/10.1080/13658816.2020.1730848. URL:https://doi.org/10.1080%2F13658816.2020.1730848.
Yang, D., Gao, C., Li, L., & Van Eetvelde, V. (2020). Multi-scaled identification of landscape character types and areas in lushan national park and its fringes, china. *Landscape and Urban Planning, 201*, Article 103844.
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*, 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 and Urban Planning, 212*, Article 104125.
Youme, O., Bayet, T., Dembele, J. M., & Cambier, C. (2021). Deep learning and remote sensing: Detection of dumping waste using uav. *Procedia Computer Science, 185*, 361–369.
Zhang, F., Fan, Z., Kang, Y., Hu, Y., & Ratti, C. (2021). "perception bias: Deciphering a mismatch between urban crime and perception of safety. *Landscape and Urban Planning, 207*, Article 104003. https://doi.org/10.1016/j.landurbplan.2020.104003. URL:https://doi.org/10.1016%2Fj.landurbplan.2020.104003.
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 and Urban Planning, 180*, 148–160.
Zhang, W., Liu, C., Chang, F., & Song, Y. (2020). Multi-scale and occlusion aware network for vehicle detection and segmentation on uav aerial images. *Remote Sensing, 12*, 1760.
Zhang, X., Zhao, P., Hu, Q., Ai, M., Hu, D., & Li, J. (2020). A uav-based panoramic oblique photogrammetry (pop) approach using spherical projection. *ISPRS Journal of Photogrammetry and Remote Sensing, 159*, 198–219.
Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network, in. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2881–2890).
Zhao, Y., Yan, J., & Hou, S. (2021). In *"scenes" updating the cultural landscape of villages along the grand canal under the all media communication, in: IOP Conference Series: Earth and Environmental Science* (p. 012133). IOP Publishing.
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. *Sustainable Cities and Society, 50*, Article 101605.