International Journal of Geo-Information

MDPI

Article

Unveiling Urban River Visual Features Through Immersive Virtual Reality: Analyzing Youth Perceptions with UAV Panoramic Imagery

Yunlei Shou [1,2], Zexin Lei [3], Jiaying Li [4] and Junjie Luo [1,2,*]

[1] Zhejiang Province Key Think Tank: Institute of Ecological Civilization, Zhejiang A&F University, Hangzhou 311300, China; jasminesyl@163.com
[2] College of Landscape Architecture, Zhejiang A&F University, Hangzhou 311300, China
[3] College of Architecture and Art, Taiyuan University of Technology, Taiyuan 030024, China; leizexin@tyut.edu.cn
[4] School of Architecture, Tianjin University, Tianjin 300072, China; lijiaying_925@tju.edu.cn
* Correspondence: luojunjie@zafu.edu.cn

Abstract: The visual evaluation and characteristic analysis of urban rivers are pivotal for advancing our understanding of urban waterscapes and their surrounding environments. Unmanned aerial vehicles (UAVs) offer significant advantages over traditional satellite remote sensing, including flexible aerial surveying, diverse perspectives, and high-resolution imagery. This study centers on the Haihe River, South Canal, and North Canal in Tianjin China, employing UAVs to capture continuous panoramic image data. Through immersive virtual reality (VR) technology, visual evaluations of these panoramic images were obtained from a cohort of young participants. These evaluations encompassed assessments of scenic beauty, color richness, vitality, and historical sense. Subsequently, computer vision techniques were utilized to quantitatively analyze the proportions of various landscape elements (e.g., trees, grass, buildings) within the images. Clustering analysis of visual evaluation results and semantic segmentation outcomes from different study points facilitated the effective identification and grouping of river visual features. The findings reveal significant differences in scenic beauty, color richness, and vitality among the Haihe River, South Canal, and North Canal, whereas the South and North Canals exhibited a limited sense of history. Six landscape elements—water bodies, buildings, trees, etc.—comprised over 90% of the images, forming the primary visual characteristics of the three rivers. Nonetheless, the uneven spatial distribution of these elements resulted in notable variations in the visual features of the rivers. This study demonstrates that the visual feature analysis method based on UAV panoramic images can achieve a quantitative evaluation of multi-scene urban 3D landscapes, thereby providing a robust scientific foundation for the optimization of urban river environments.

Keywords: visual perception; UAV remote sensing; immersive virtual reality; semantic segmentation; riverscapes

check for updates

Citation: Shou, Y.; Lei, Z.; Li, J.; Luo, J. Unveiling Urban River Visual Features Through Immersive Virtual Reality: Analyzing Youth Perceptions with UAV Panoramic Imagery. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 402. https://doi.org/10.3390/ijgi13110402

Academic Editors: Wolfgang Kainz and Florian Hruby

Received: 10 September 2024
Revised: 23 October 2024
Accepted: 1 November 2024
Published: 7 November 2024

Copyright: © 2024 by the authors. Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The development of human civilization is closely intertwined with rivers, which are often the cradle of urban centers where water is abundant. Urban rivers not only provide essential spaces for citizens' recreation and exercise but also play a crucial ecological role, such as regulating the urban climate and providing habitats for various species [1–3]. Consequently, the study of urban rivers and their adjacent spaces has become a vital part of built environment research, laying the foundation for enhancing the social, cultural, and ecological values of river corridors [4–6].

The visual characteristics of riverscapes are increasingly recognized as a crucial component of urban environmental quality. These characteristics significantly enhance the

aesthetic appeal of urban areas and influence public perception and engagement with these spaces [7]. Previous studies have examined several dimensions of riverscapes, including their ecological functions, social significance, and contributions to urban sustainability [8,9]. For example, Garau et al. (2021) explored the sociocultural and ecological constructs of riverscapes, emphasizing their essential role in shaping urban experiences [10]. In another study, researchers analyzed canalized rivers in Beijing, using panoramic imagery captured from riverbanks and applying deep learning and virtual reality technologies to derive both objective visual characteristics and subjective visual perceptions of river landscapes [1]. Their analysis delved into the mechanisms linking these objective features with subjective perceptions. Similarly, Luo et al. (2022) focused on the aesthetic evaluation of urban waterscapes, underscoring the importance of scenic beauty in shaping public perception [11]. Research in this area has also explored continuous water landscape imagery along rivers like the Thames, utilizing computer vision techniques to analyze the visual elements of riverscapes [12]. This method, which involves perceiving the landscape from a water-level or boat-level perspective, allows for a more detailed understanding of the river and its surrounding environment. It provides valuable insights for assessing water landscape quality and optimizing landscape design. Additionally, studies have highlighted the importance of aerial perspectives in understanding the visual perception of urban riverscapes [13–15]. Aerial views offer a more comprehensive and holistic understanding of urban rivers and their surrounding environments, capturing features such as construction intensity, ecological characteristics, and the distribution of landscape elements. The findings from these perspectives serve as critical references for decision-makers and planners involved in urban riverscape planning and redevelopment [16].

The integration of advanced technologies, such as immersive virtual reality (IVR), and computer-aided techniques, has further broadened the scope of riverscape analysis. For instance, Kim et al. (2024) explored the visual impact of urban waterfront development on mountainous backgrounds, a contextually significant aspect of riverscapes that also reflects broader concerns about preserving the visual integrity of urban natural settings [17]. Their study, which utilized immersive VR experiments, provided scientific justification for urban waterfront visual impact regulations, particularly in terms of maintaining mountain visibility against the backdrop of high-rise developments. This work underscores the importance of considering the interaction between built environments and natural landscapes in visual impact assessments, a principle that is equally relevant to riverscape studies. Additionally, Puspitasari and Kwon (2019) analyzed the visual quality of riverfront skylines, focusing on the height and spatial arrangement of tall buildings. Their research highlighted how the configuration of built elements along a river can significantly influence the overall visual appeal of an urban skyline when viewed from the water [18]. This study's emphasis on skyline aesthetics and its methodical approach to quantifying visual quality through building characteristics offer valuable insights that can be applied to the analysis of riverscapes, particularly in understanding how built structures contribute to or detract from the visual harmony of urban rivers.

However, despite the growing body of literature on riverscapes, several challenges remain. Traditional methods of riverscape assessment, such as on-site surveys and 2D photography, often fall short in capturing the dynamic and three-dimensional nature of these environments [1,12]. These methods typically offer a limited perspective, failing to account for the complexity and variability of riverscapes across different spatial and temporal scales [19,20]. Furthermore, the subjective nature of visual assessments can lead to inconsistencies, complicating efforts to establish standardized criteria for evaluating riverscape quality [11,21].

Recent advancements in technology offer promising solutions to these challenges. Unmanned aerial vehicles (UAVs) have revolutionized environmental monitoring by providing high-resolution, multi-perspective imagery that captures the full scope of riverscapes [22–24]. UAVs are highly flexible, capable of collecting environmental data from almost any location and time, except in restricted flight zones [25]. Increasing the flight

altitude of UAVs expands the range of vision, aiding in a better understanding of large-scale scenes [11]. In addition to capturing conventional image data in tilt, nadir, and horizontal modes, UAVs can also take panoramic photos at different heights [26–28]. Compared to satellite remote sensing imagery, UAV aerial surveying offers richer perspectives, higher data resolution, and shorter revisiting intervals, providing opportunities for rapid monitoring and multi-perspective analysis of urban rivers and other environmental characteristics [29–32]. Immersive virtual reality (IVR) technology further enhances the ability to perceive and evaluate riverscapes by allowing users to experience these environments in three dimensions [33–35]. This technology, widely applied in urban environment studies, offers a more comprehensive and interactive approach to visual evaluation, enabling a deeper understanding of the spatial characteristics and visual appeal of riverscapes [12]. The combination of UAVs and IVR technology addresses the limitations of traditional methods, offering a more accurate and detailed analysis of large-scale urban landscapes [36–39]. Additionally, the rapid development of computer vision technology has significantly improved the accuracy and efficiency of image semantic information recognition and analysis [40,41]. Techniques such as image semantic segmentation and object recognition in computer vision have supported the visual quantitative analysis of urban landscapes [42–44]. Some studies have begun to use computer vision technology for quantitative analysis of UAV aerial survey data [11,45]. For instance, datasets such as UAVid [46] and AeroScapes [47], along with semantic segmentation models like SegNet [48] and SegFormer [49], have enabled pixel-level precise identification of various objects in UAV oblique photography images [11]. The quantitative analysis results of related studies have become important data for researchers to understand the elements of large-scale landscapes, enhancing the accuracy of visual environment analysis.

Based on these advancements, the current study proposes a method for analyzing the visual characteristics of urban river landscapes using UAV panoramic images combined with immersive VR perception technology. The aim is to achieve a flexible and efficient visual evaluation of large-scale rivers and their surrounding areas. Using computer vision technology, pixel-level semantic segmentation of panoramic images allows for the identification and quantitative analysis of various landscape elements. Cluster analysis is utilized to comprehensively summarize the characteristics of different river areas. This study conducts an empirical analysis of the Haihe River, South Canal, and North Canal, which flow through the central urban area of Tianjin. The research questions include the following: (1) What are the spatial variation characteristics of the immersive VR perception results of the Haihe River, South Canal, and North Canal? (2) What similarities and differences exist in the proportion of visual landscape elements among the three rivers? (3) How can VR perception results and the characteristics of landscape element proportions be used to comprehensively describe the spatial features of the rivers?

2. Research Area and Data Source

2.1. Study Area

This study focuses on three river sections located in the central urban area of Tianjin City: the Haihe River, the North Canal, and the South Canal. These sections have lengths of 6 km, 10 km, and 12 km, respectively, totaling 28 km (Figure 1). The selection of these river segments as the study area is primarily justified for the following reasons: (1) These rivers traverse the central urban district of Tianjin, flowing through regions that include numerous waterfront parks, tourist attractions, and residential areas. They are closely related to the daily lives of the populace, representing the urban rivers of the city. (2) In recent years, Tianjin has introduced a series of renovation and enhancement projects targeting the cultural and tourism belt of the Grand Canal and the Haihe River. However, the current visual characteristics of these rivers and their adjacent riverside spaces remain unclear, lacking a "bottom-up" evaluation of river space perception based on the public's viewpoint. (3) This area is free of UAV flight restrictions and electromagnetic interference, ensuring the safe operation of UAVs for data collection.

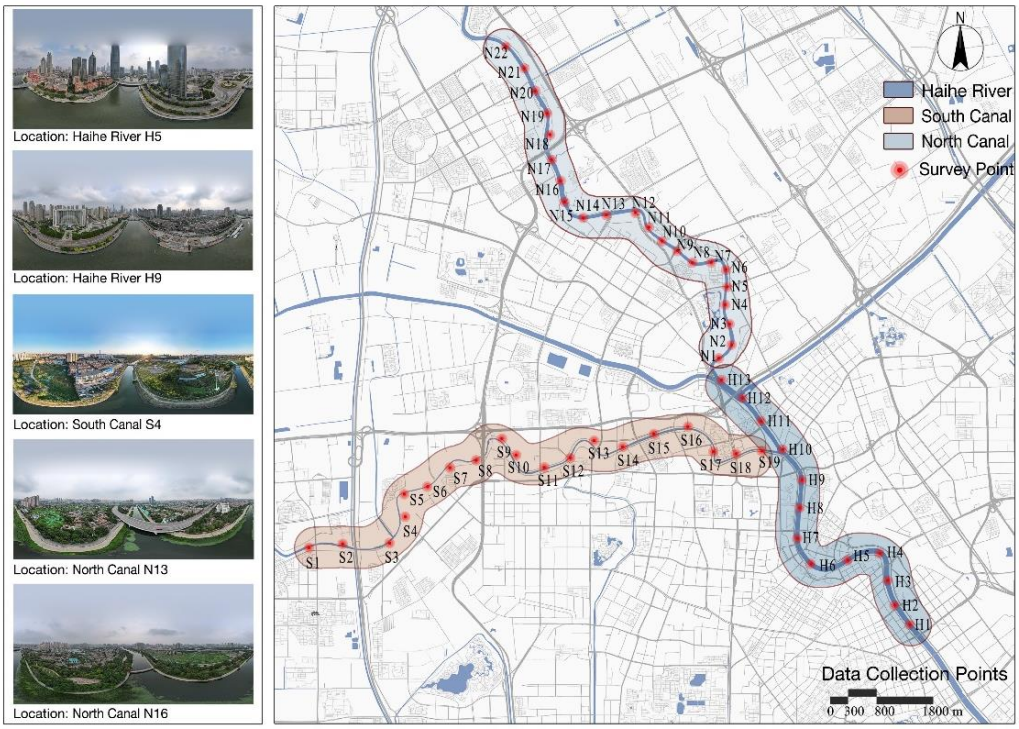


Figure 1. Study area and examples of panoramic imagery.

2.2. Data Sources

The study utilized the DJI Mavic Air 2S UAV to acquire aerial photography data. Within the 28 km linear study area, a panoramic image data collection point was set up approximately every 500 m, totaling 54 points. Specifically, the Haihe section had 13 mapping points (labeled H1–H13), the South Canal had 19 (labeled S1–S19), and the North Canal had 22 (labeled N1–N22).

Choosing the appropriate altitude for UAV data collection was crucial: On one hand, increasing the shooting height could offer a broader perspective, capturing a larger area of the riverscape, but at the cost of reducing the clarity of smaller objects such as vehicles, people, and plants. This reduction in resolution poses challenges for accurate semantic segmentation and feature recognition within the images. On the other hand, lower flight heights provide more detailed imagery of ground elements, but they limit the field of view and increase the risk of safety issues due to potential obstacles such as wires, tree branches, and other flight hazards [50]. To determine the optimal flight height, this study conducted a comparative analysis of panoramic images captured at four altitudes—30 m, 60 m, 90 m, and 120 m (Figure 2). The images taken at 30 m provided exceptional detail of ground elements; however, the limited field of view at this height made it difficult to capture the broader spatial distribution of landscape features. Conversely, the 120 m altitude allowed for a comprehensive perspective of the landscape but at the expense of detail, with smaller elements becoming indistinct, affecting the accuracy of semantic segmentation. The 60 m altitude was ultimately chosen as the optimal height, striking a balance between these trade-offs. At 60 m, the images provided a sufficiently wide field of view while maintaining the clarity needed for effective semantic segmentation and accurate feature extraction. As demonstrated in the comparison results, the 60 m images offered better visibility of both built and natural features, such as riverside vegetation and pedestrian areas, without sacrificing resolution or safety. Therefore, UAV operators manually flew 60 m above the river mid-lines based on real-time flight data to capture panoramic imagery for this study. This decision ensured both data quality and operational safety across the 54 mapping points.

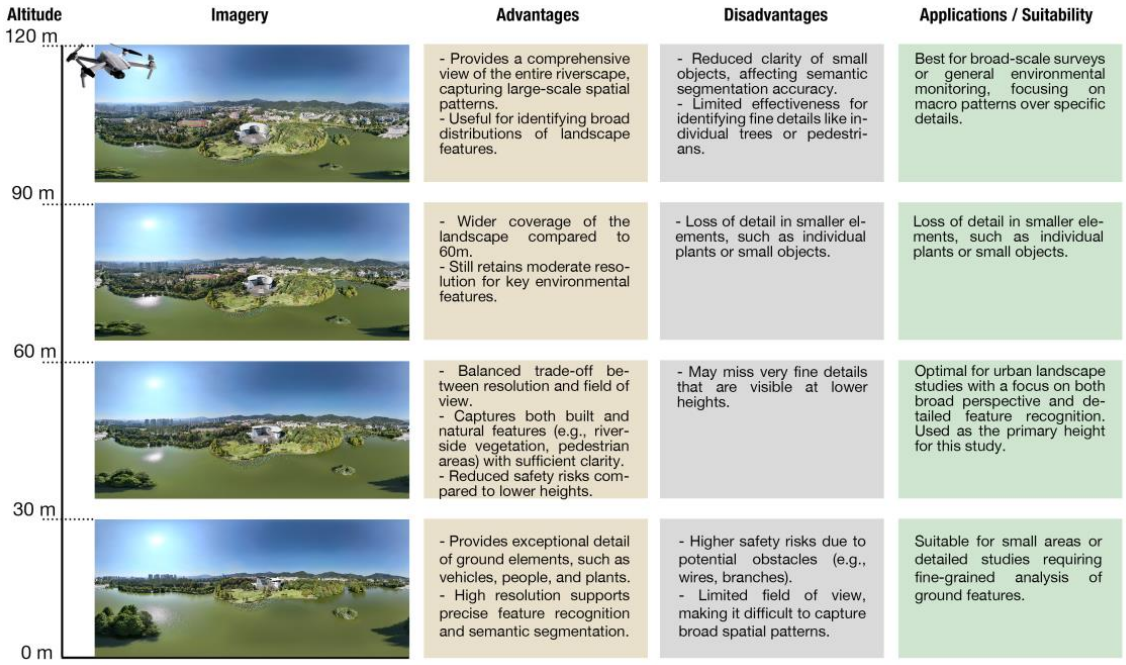


Figure 2. Comparison of aerial imagery at different altitudes.

To ensure the quality of the photography, the shooting was scheduled during the hours with optimal lighting and weather conditions, from 10 a.m. to 6 p.m. The aerial photography was conducted over non-consecutive days between July and September 2022. The high-resolution panoramic photos obtained by this study reveal very clear details of the ground, such as pavement, vehicles, plants, etc., meeting the research needs. It is important to note that the images used in this study were panoramic, each composed of 32 individual 2D photos stitched together. Therefore, the proportion of landscape elements in these images is consistent and unaffected by the camera's gimbal angle during capture [51,52]. This approach ensures the objectivity and reproducibility of the study, as the camera angle does not influence the final image composition.

3. Methodology

3.1. Immersive Virtual Reality

While numerous studies have begun to utilize UAV aerial imagery for visual perception, they predominantly focus on ordinary photographs captured from a single perspective, which fails to fully showcase all features of the photographed area; nor do they integrate advanced virtual interactive devices such as VR. UAV panoramic imagery effectively bridges this gap [46]. Compared to traditional assessment methods based on 2D pictures, utilizing virtual reality technology for landscape visual perception offers a more interactive visual experience [53]. Some studies have evaluated the scenic beauty and vitality of study areas using panoramic pictures [54,55].

Drawing upon previous experiences in visual evaluation research and considering the characteristics of the three rivers involved in this study, scenic beauty, vitality, color richness, and a sense of history were selected as evaluation indicators to capture people's visual impressions of river landscapes. Scenic beauty is one of the most commonly used indicators in visual evaluation, allowing the analysis of the public's landscape preferences. Scenic beauty refers to the aesthetic appeal of a landscape, capturing how visually pleasant and attractive it appears to participants. It is often influenced by the balance, harmony, and composition of natural and built elements in the environment [56]. Vitality reflects the sense of liveliness conveyed by the environment. It measures the visual cues that indicate an active, engaging space, influenced by architectural forms, pedestrian and vehicular movement, and other urban activities [57,58]. Color richness measures the variety and

quality of colors present in the landscape, reflecting the character and vibrancy of the environment. It plays a crucial role in urban aesthetics by contributing to the identity of the space [59,60]. Sense of history refers to the visual connection with the past, as perceived through historical districts, monuments, or architectural elements with significant cultural heritage. This indicator measures how well the landscape reflects a cultural narrative or historical value [61].

To ensure the reliability and consistency of visual evaluations, this study used statistical measures, including Pearson correlation coefficients, Cronbach's alpha, and the Kaiser–Meyer–Olkin (KMO) measure, to assess participants' scoring consistency. The Pearson correlation coefficient was employed to evaluate the strength and direction of the linear relationship between the initial evaluation scores and verification scores for the four visual indicators. The inclusion of Pearson correlations in this study ensures that participant assessments remain consistent, validating the immersive VR-based evaluation as a robust tool for landscape perception studies. Cronbach's alpha was used to evaluate the internal consistency of participants' responses across the four visual indicators. We also employed the KMO measure to assess whether the dataset was adequate for correlation analysis.

The technological approach of this study is illustrated in Figure 3. In immersive virtual reality visual perception research, selecting the appropriate VR device is crucial. Common VR devices include head-mounted displays (HMDs), stereoscopic screens, and multi-projection systems. HMDs typically consist of a display and associated sensors, projecting images directly in front of the user's eyes to provide an immersive experience [62,63]. Stereoscopic screens, often large or surround displays, offer 3D visual effects, requiring users to wear 3D glasses for a realistic VR experience. Multi-projection systems, composed of several projectors, create multiple images in an area, offering a surround 3D visual experience.

For this study, considering the strengths and weaknesses of various VR devices, along with the study's objectives, experimental convenience, and reliability, we selected the PICO NEO 3 head-mounted display as the primary perception device [64]. This device was chosen for its combination of portability, high resolution, and affordability, which meets the study's need for immersive visual assessments in various environments. The key parameters of the PICO NEO 3 HMD are summarized in Table 1. These features ensure that participants receive a high-quality immersive experience, supporting the accuracy of visual assessments in the IVR environment [65].

Table 1. The key parameters of the PICO NEO 3 HMD.

Parameter	Specification
Resolution	1920 × 3664 pixels (4K)
Pixels per inch (PPI)	773
Refresh Rate	90 Hz
Field of View (FoV)	98 degrees
Weight	350 g
Battery Life	3 h
Controllers	2 handheld controllers with 6DoF

A total of 36 volunteers were invited to participate in the immersive VR perception experiment. The perception experiment consisted of two parts: visual evaluation and result verification. In the visual evaluation, 27 volunteers used immersive VR devices to score the four evaluation indicators across different scenes; in the verification part, an additional 9 volunteers re-evaluated these panoramic image data for visual perception scoring. Verification scoring was used for correlation analysis with the first part of the visual evaluation. Among the volunteers participating in the first part of the visual evaluation, 70.4% (19 people) were female, 85.2% (23 people) were students, and the average age was 23.2 years. In the verification group, 55.6% (5 people) were female, 77.8% (7 people) were students, and the average age was 24.3 years.

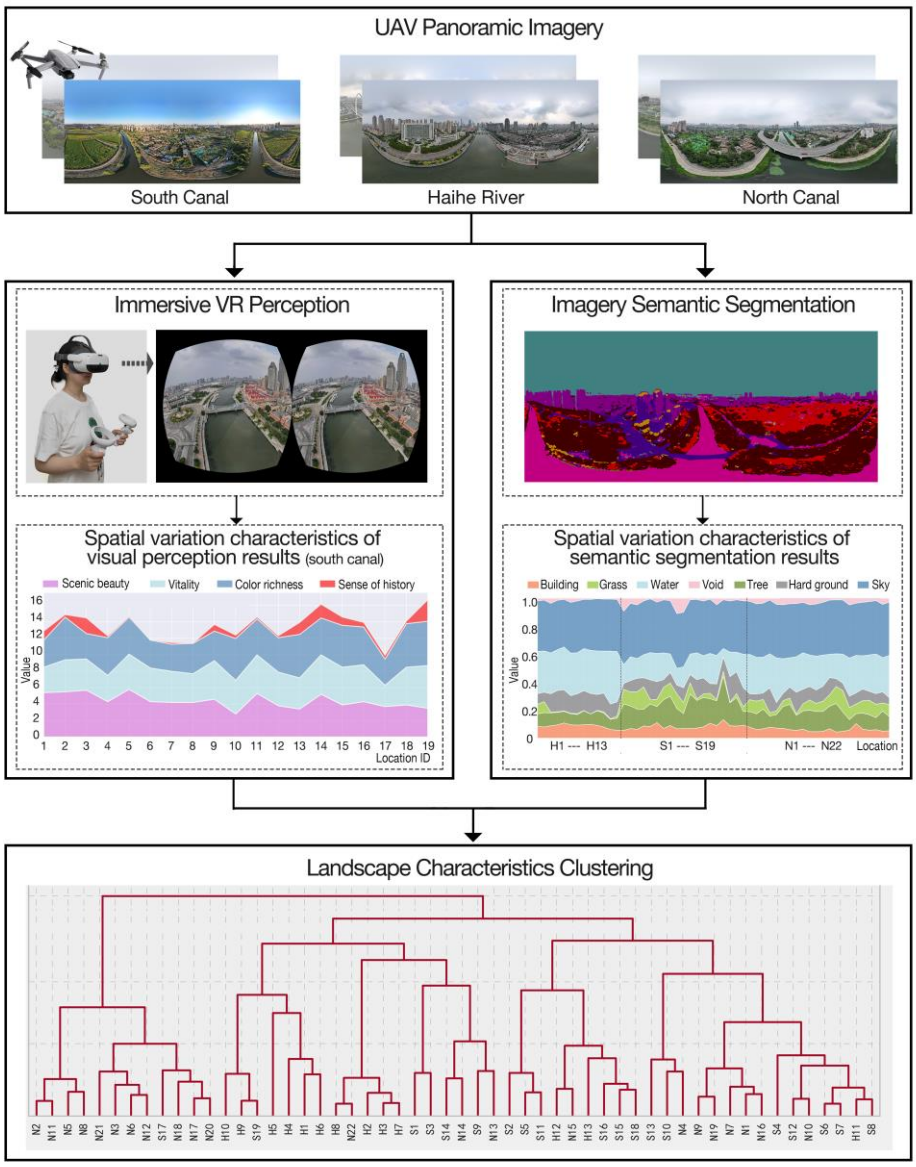


Figure 3. Workflow of this study.

Over two-thirds of the 36 volunteers had lived in the study area for more than two years, making them very familiar with the region. Additionally, these volunteers had a high level of education, which ensured that they were proficient in using the head-mounted VR devices, contributing to the smooth execution of the study. Given the participants' extended residency in the study area and their academic backgrounds, their lifestyles and environmental perspectives likely reflect a deep familiarity with the local environment and an informed understanding of landscape aesthetics and sustainability. This familiarity and educational background may have influenced their visual evaluations, particularly in terms of their sensitivity to environmental quality and historical context.

The participation process was as follows: First, researchers explained the purpose and steps of the study to all participating volunteers and allowed each participant to browse through the panoramic pictures as a whole. Next, to ensure participants fully understood the content of each panoramic photo, they were required to look around each scene for at least a full circle, with a viewing time of no less than 40 s. To prevent fatigue from wearing VR glasses for a long time from affecting the scoring, the panoramic photos of the 54 mapping points were divided into three groups in an arithmetic sequence, each containing 18 panoramic pictures. Thus, each mapping point had scores from 9 volunteers

and verification scores from 3 volunteers. The Likert scale was used to rate the four evaluation indicators on 7 levels (with 1 being the lowest and 7 the highest). After viewing each scene's panoramic picture, researchers sequentially asked volunteers for their scores on scenic beauty, vitality, color richness, and a sense of history for the viewed scene. The final score for each scene was the average of the visual evaluation scores from 9 participants. This study was reviewed and approved by the ethics committee of the authors' institution. All participants signed informed consent forms before the experiment, ensuring they understood the purpose and procedure of the study and agreed to participate.

3.2. Imagery Semantic Segmentation

In recent years, computer vision technologies, particularly image semantic segmentation, have made significant strides [66–70]. Semantic segmentation involves grouping and segmenting image pixels based on their semantic meanings, allowing for precise identification and categorization within an image. Models such as FCN, SegNet, and SegFormer are widely used for this purpose, effectively detecting and segmenting objects in images [48,49]. SegFormer is built on a Transformer framework that excels in capturing long-range dependencies within an image. This is particularly advantageous in riverine environments, where both fine-grained local features (e.g., vegetation) and broader global patterns (e.g., spatial layout of water bodies and built structures) are important. Compared to FCN and SegNet, SegFormer offers a lighter decoder architecture that ensures faster inference times and lower computational demands, while still maintaining high accuracy.

After our literature review, we found that there is no openly available UAV oblique imagery semantic segmentation dataset that focuses on the river environment. Thus, 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 of river landscapes. We derived the dataset with 400 high-resolution images spanning the river and surrounding areas, each with a size of 1800 $\times$ 1480 pixels. We categorized the imagery into 14 groups, namely: building, cottage, under construction place, tree, grass, water grass, soil, hard ground, water, sky, human, car, boat, and void. This is a relatively comprehensive segmentation that will be applicable to river landscapes in most parts of the world.

Given the characteristics of UAV data and the operational complexity of riverine environments, we selected SegFormer, a state-of-the-art semantic segmentation model based on the Transformer framework. SegFormer offers a compelling balance of efficiency, accuracy, and robustness in handling image segmentation tasks [71]. Compared to traditional convolutional neural networks (CNNs), the Transformer-based architecture of SegFormer excels in capturing long-range dependencies and processing both fine-grained details and global features. Below, we present the detailed workflow involved in applying the SegFormer model for this study.

1. Data Preprocessing: The UAV-obtained imagery was annotated at the pixel level to fit the 14 predefined categories. The annotations were stored in GeoJSON format to ensure compatibility with the segmentation model. The dataset was split into training (80%) and validation (20%) subsets. We applied data augmentation techniques, such as flipping, scaling, and cropping, to improve the generalization capacity of the model and mitigate overfitting.
2. Model Training: SegFormer leverages a Transformer-based encoder and a lightweight decoder, facilitating both accurate pixel-wise classification and efficient inference. The model was trained on an NVIDIA RTX 3090 GPU over 100 epochs with a batch size of 16. Following each epoch, the model's performance was evaluated on the validation dataset to ensure stability and convergence.
3. Model Evaluation: We employed mean pixel accuracy (MPA) and mean intersection over union (MIoU) as the primary evaluation metrics. During the prediction phase, the model generated pixel-level classifications aligned with the original UAV imagery.

The implementation of the SegFormer model in this study offers insights that extend beyond river landscape analysis. The segmentation results provide valuable data for urban planners and landscape architects, enabling them to analyze, categorize, and optimize urban landscapes in a systematic manner. Additionally, the ability to automate the analysis of large-scale imagery facilitates environmental monitoring, such as tracking changes in land use, vegetation coverage, and water quality over time. In the context of riverscapes, the customized dataset developed in this study provides a significant resource for comprehensive visual analysis. The segmentation results allow practitioners to assess the spatial distribution of landscape elements and evaluate their visual impact on the environment. This approach offers a standardized framework for evaluating and improving the visual quality of riverscapes, making it applicable to various regions and supporting sustainable landscape management on a global scale.

3.3. Landscape Feature Clustering

Cluster analysis is a technique used to identify similarities or relationships among multiple research objects, grouping them into clusters based on these similarities. Objects within the same cluster share strong similarities, while those in different clusters show significant differences [72,73]. This method is particularly useful in exploratory research, providing multiple potential solutions and allowing researchers to interpret the grouping results according to the study's objectives [74].

In this study, we used Python's SciPy library to perform hierarchical cluster analysis, which generated clear dendrograms for visualizing the results [75]. To unify the data from immersive virtual reality visual evaluations and computer vision statistics, we normalized both datasets to a 0–1 scale. This ensured that subjective visual scores and objective environmental measurements contributed equally to the clustering process. The combined feature matrix, with each row representing a river section, was used as input for the cluster analysis.

The cluster analysis provided a nuanced understanding of how different landscape elements are distributed across the study area, allowing us to identify spatial locations with similar visual characteristics. For instance, certain river sections were grouped together based on their high scenic beauty and vegetation density, while others were clustered due to their higher concentration of built structures.

These clustering results are not only descriptive but also have practical implications for landscape management. Urban planners and landscape architects can use this information to prioritize areas for aesthetic improvement or conservation. For example, clusters with lower scenic beauty but high development potential could be targeted for landscape enhancements, while clusters with valuable natural features might be designated for preservation. Moreover, the approach used in this study can be applied to other geographic contexts, providing a replicable method for analyzing landscape features in urban and rural environments. By adapting the clustering criteria to specific research needs, practitioners can gain insights into the spatial distribution of landscape features and their impact on visual quality, ultimately guiding more informed decision-making in environmental planning and management.

4. Results

4.1. Immersive Virtual Reality Perception Results

4.1.1. Evaluation Analysis

In this study, volunteers used immersive VR devices to evaluate four visual indicators—scenic beauty, vitality, color richness, and historical sense—across various river scenes. The results showed strong correlations between the initial evaluations and the verification scores, indicating consistency and reliability in participants' perceptions. Specifically, the Pearson correlation coefficients for scenic beauty ($r = 0.821$), vitality ($r = 0.942$), color richness ($r = 0.906$), and historical sense ($r = 0.885$) were all significant at $p < 0.01$. The reliability and validity of the data were confirmed using Cronbach's alpha (0.682) and the

Kaiser–Meyer–Olkin (KMO) measure (0.643), both exceeding the acceptable thresholds, affirming the robustness of the evaluation process.

These findings suggest that the immersive VR setup used in this study is effective in generating consistent and reliable visual assessments, making it a valuable tool for future studies in urban landscape analysis. The high correlation values indicate that participants' visual experiences in the VR environment closely mirror their natural perceptions, which is crucial for ensuring the validity of the results in applied urban planning contexts.

4.1.2. Comprehensive Comparison of Virtual Perception Results

Statistical analysis of the visual evaluation results across four dimensions—scenic beauty, color richness, vitality, and historical sense (Figure 4)—revealed that the South Canal had the highest average scenic beauty score of 4.4, slightly higher than the Haihe River's score of 4.0 and significantly higher than the North Canal's average of 3.4. The highest scenic beauty score of 5.8 was found in the South Canal, with both the Haihe River and North Canal reaching a maximum of 5.3. The lowest scenic beauty scores for all three rivers were 2.3, located in the North Canal; the South Canal and Haihe River had slightly higher minimum scores of 2.8 and 3.4, respectively. Moreover, the Haihe River had the smallest standard deviation in scenic beauty, only 0.60, compared to 0.86 and 0.75 for the South Canal and North Canal, respectively. Comparing the average, maximum and minimum values, and standard deviations of scenic beauty across the three rivers shows that the South Canal ranked highest in average visual attractiveness; the Haihe River had generally good visual attractiveness with relatively small differences between its various mapping points; the locations with the highest and lowest scenic beauty were in the South Canal and North Canal, respectively, indicating significant differences in scenic beauty between the South and North Canals.

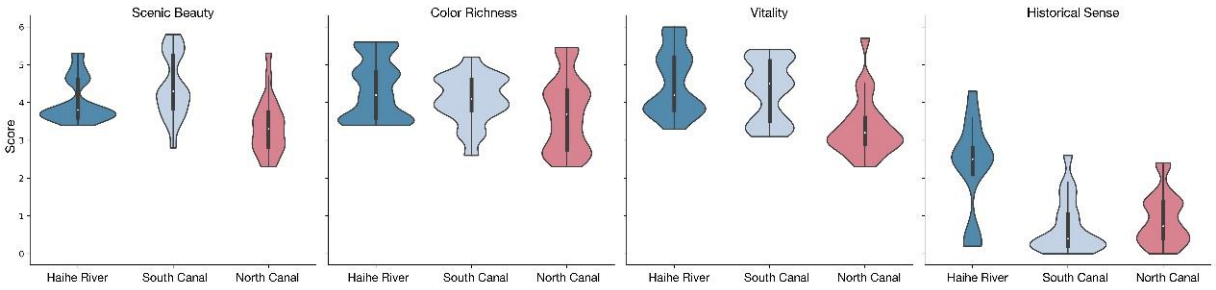


Figure 4. Comparison of visual perception results.

Comparing the color richness data of the three rivers revealed that the Haihe River had the highest average color richness score (4.4), maximum score (5.6), and minimum score (3.4), all higher than the other two rivers. The lowest color richness score was only 2.3, located in the North Canal. The South Canal's highest score for this item (5.2) was slightly lower than that of the North Canal (5.5), but its minimum score (2.6) was slightly higher than the North Canal. Additionally, the standard deviation for color richness in the South Canal was 0.66, lower than in the North Canal (0.99) and the Haihe River (0.83). Meanwhile, the scoring results for vitality and color richness across the three rivers showed similar characteristics. The Haihe River had the highest average vitality score (4.5), maximum score (6.0), and minimum score (3.3), all higher than those of the South and North Canals. The lowest scores for all three rivers were in the North Canal, only 2.3, with its standard deviation being 0.79, also lower than that of the South Canal (0.81) and the Haihe River (0.88).

The scoring for the sense of history differed significantly from the other three indicators. The overall scores for the sense of history were not high across all three rivers, with the Haihe River having an average score of 2.3; in comparison, both the South and North Canals scored less than 1, with scores of 0.7 and 0.9, respectively. The highest scores for the sense of history were in the Haihe River at 4.3, with the maximum values for the South and

North Canals significantly lower at 2.6 and 2.4, respectively. The minimum scores for the sense of history in both the South and North Canals were 0, while the Haihe River had a slightly higher minimum score of 0.2. The standard deviation for the sense of history in the Haihe River was 1.27, higher than in the South Canal (0.73) and the North Canal (0.69). This evaluation indicates that the Haihe River conveys a stronger sense of history visually compared to the South and North Canals. In other words, the sections of the South and North Canals studied provide a very limited sense of history.

4.1.3. Spatial Distribution Characteristics of Virtual Perception Results

By correlating the immersive virtual reality evaluation results with the spatial locations of the three rivers, this study identified distinct spatial patterns in the visual perception results across the three rivers (Figure 5). The North Canal exhibited higher evaluation scores in the central sections (N12–N16), while the South Canal's scores were more consistent, with occasional peaks (e.g., S16–S18). The Haihe River displayed a bimodal distribution, with higher scores around H4–H6 and H9–H11. Notably, the South Canal consistently had high scenic beauty scores, while the Haihe River excelled in historical sense.

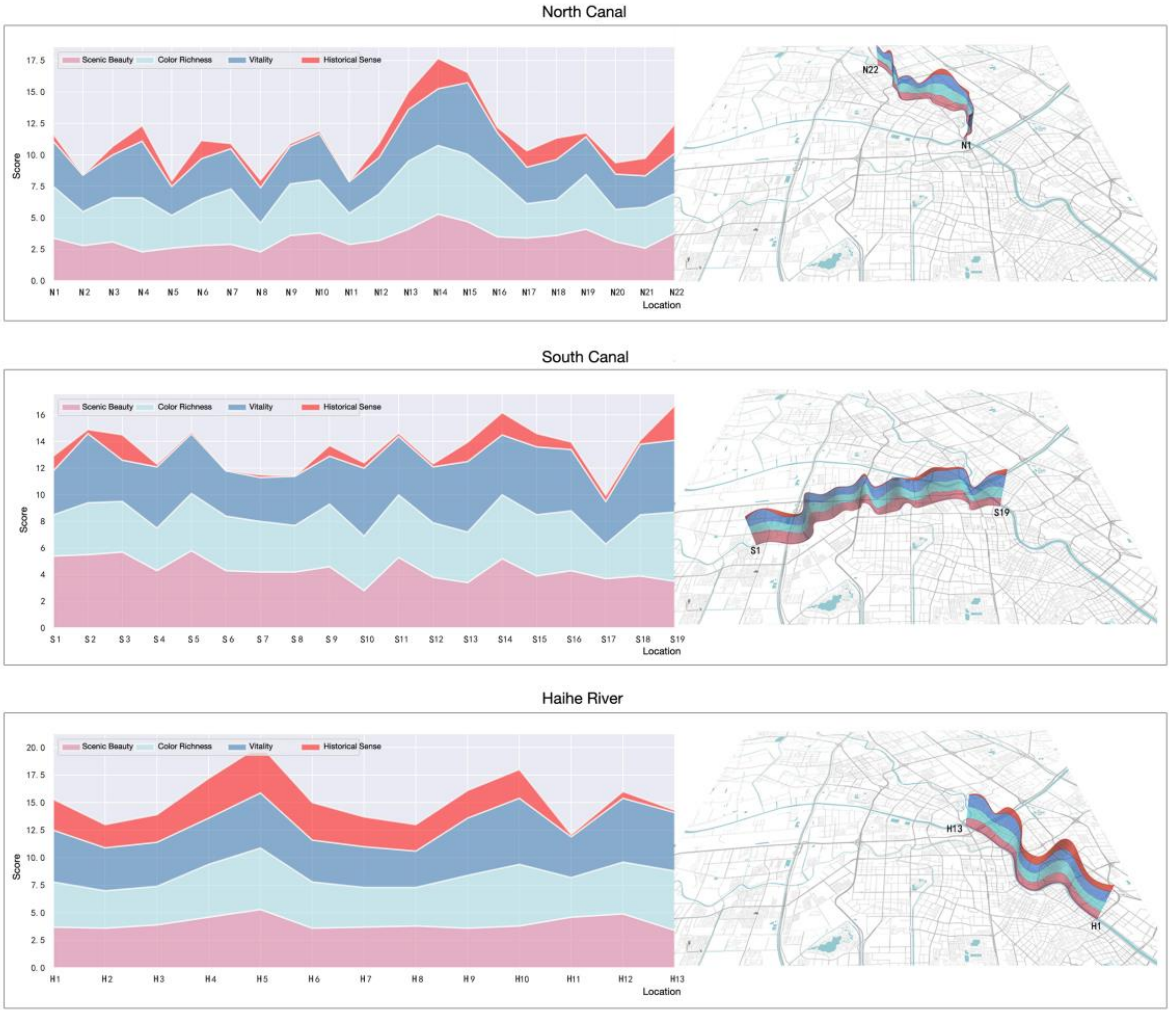


Figure 5. Spatial variation characteristics of visual perception results.

These findings align with previous research on urban riverscapes, which also noted that central urban sections often receive higher aesthetic evaluations due to denser vegetation and better-maintained infrastructure [17]. Additionally, the observed variability in historical sense scores across different sections of the rivers underscores the importance of integrating cultural heritage into urban design, a point emphasized by Puspitasari and

Kwon (2019) in their analysis of skyline aesthetics [18]. These results suggest that targeted interventions in specific river sections could significantly enhance overall visual appeal and cultural value.

4.2. Computer Vision Analysis Results

By training the SegFormer computer vision semantic segmentation model, this study conducted pixel-level semantic segmentation on panoramic images from a total of 54 mapping points across the Haihe River, South Canal, and North Canal. It classified and quantified the pixel proportions of different landscape elements within the images. The semantic segmentation model achieved a mean pixel accuracy (MPA) of 89.72% and a Mean Intersection over Union (MIoU) of 47.04%, enabling relatively precise identification and classification of landscape elements such as buildings, trees, grasslands, and water bodies in the panoramic images. This aids in the quantitative analysis of the visual features at different river mapping points (Figure 6).

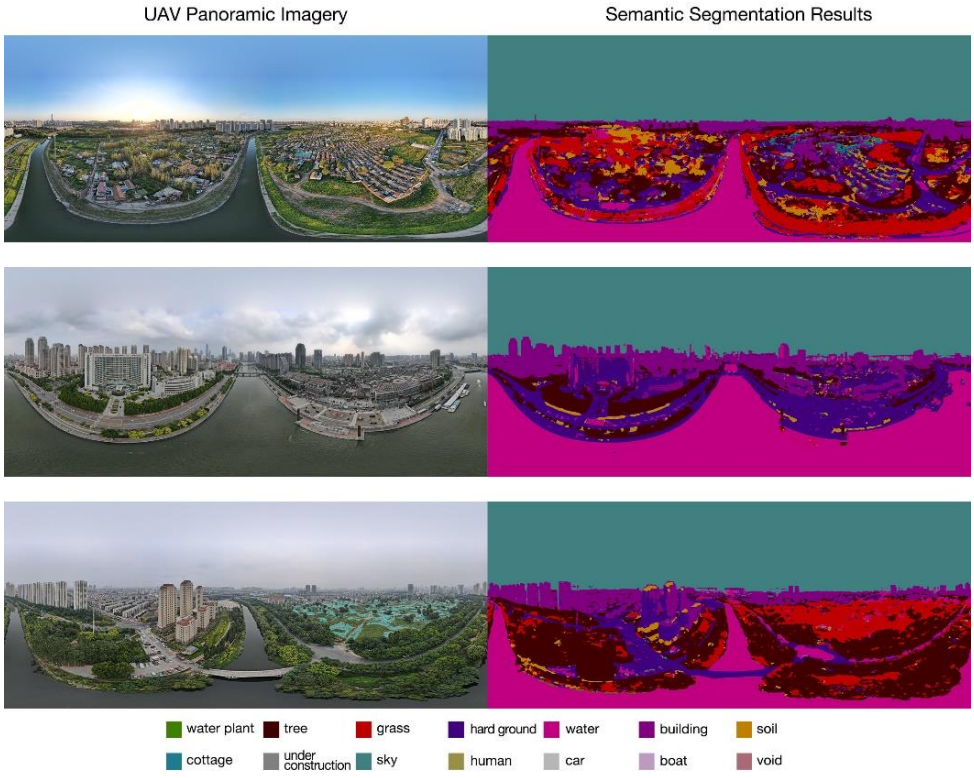


Figure 6. Examples of UAV panoramic images and their semantic segmentation results.

By analyzing the proportions of 14 types of landscape elements across the Haihe River, South Canal, and North Canal, it was found that the proportions of six primary visual landscape elements—sky, water bodies, buildings, trees, grasslands, and hard surfaces—exceed 90%, forming the main visual characteristics of the three rivers. Analyzing the spatial distribution characteristics of these six main visual landscape elements is beneficial for understanding the variations in objective visual elements at different locations along the rivers. As depicted in Figure 7, the distribution of water bodies, trees, grasslands, and hard surfaces among these four elements is not uniform across the three rivers. Notably, the visible area of water bodies shows significant differences among the three rivers, with the Haihe River's visible water area far exceeding that of the South Canal and surpassing most areas of the North Canal. The visible water area of the South Canal and North Canal fluctuates noticeably, while that of the Haihe River is relatively stable.

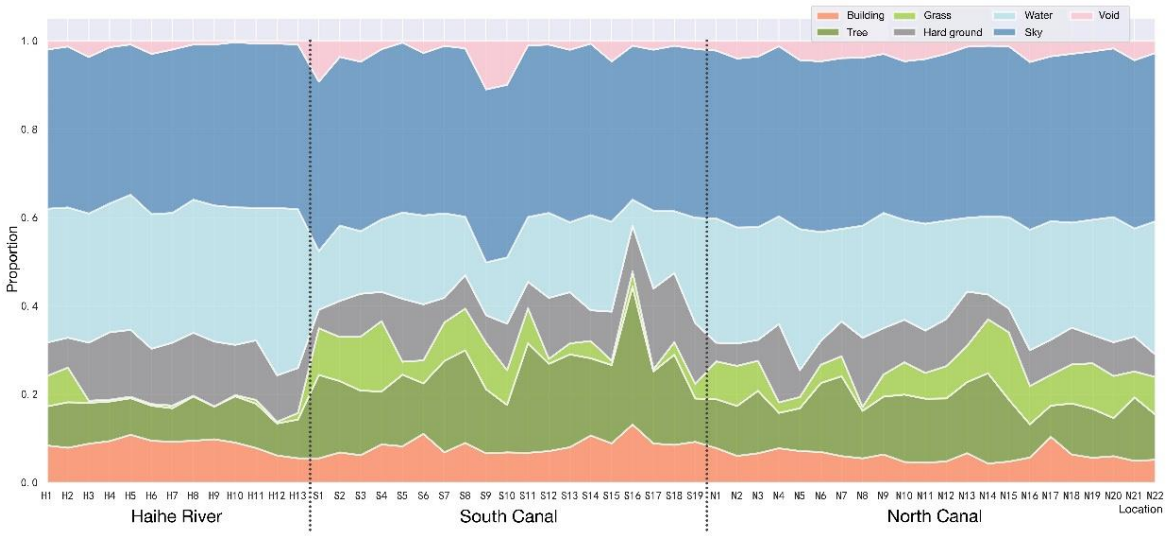


Figure 7. Spatial distribution of semantic segmentation results.

Trees and grasslands present a distribution pattern opposite to that of water bodies. The proportion of trees in the South Canal is higher than in the North Canal, with significant fluctuations in both rivers across different sections, while the overall proportion in the Haihe River is relatively lower and exhibits smaller fluctuations across locations. Similarly, the visibility rate of grasslands in the South Canal and North Canal is higher than in the Haihe River; however, there is significant spatial fluctuation in all three rivers. For instance, the grassland visibility rate is higher in sections S1–S4 of the South Canal and N13–N20 of the North Canal, with other areas having lower rates. The spatial distribution characteristics of hard surfaces are similar to those of water bodies, with the Haihe River exceeding the South Canal and North Canal, and the distribution across space being uneven. In contrast, the spatial distribution of buildings and sky is relatively uniform across the three rivers. Specifically, the proportion of buildings is slightly higher in sections H4–H10 of the Haihe River, S4–S7 and S14–S17 of the South Canal, and N6–N18 of the North Canal compared to other locations; the sky proportion in the Haihe River is slightly lower than in the South Canal and North Canal.

4.3. Landscape Feature Clustering Analysis Results

Using Python's SciPy scientific computing library, this study conducted hierarchical cluster analysis on four types of visual evaluation results obtained from the immersive VR perception experiment and six types of landscape elements acquired through computer vision technology (Figure 8). Visual analysis of the clustering results revealed that dividing into four clusters provided a satisfactory differentiation and characterization of the mapping locations along the Haihe River, South Canal, and North Canal. These clusters are as follows: Cluster 1, urban rivers with low visual value; Cluster 2, urban rivers with high visual value; Cluster 3, suburban rivers with low visual value; and Cluster 4, suburban rivers with high visual value.

The first cluster includes a total of eleven research points, with eight located on the South Canal, two on the North Canal, and one on the Haihe River. It is defined as urban rivers with lower visual value because this area has lower scenic beauty and color richness and almost no perceptible historical atmosphere; moreover, the area has a relatively higher proportion of buildings and hard surfaces, indicating a higher level of urbanization. Additionally, this area exhibits an overall low rate of green visibility. Despite a higher proportion of trees at some research points, the overall proportions of trees and grasslands are lower in this area. The second cluster contains 13 research points, all but one (S19 on the South Canal) located on the Haihe River. This cluster is defined as urban river areas with high visual value because these areas have higher levels of vitality and color richness;

a larger proportion of buildings and hard surfaces indicates a high level of urbanization; simultaneously, these areas convey a strong sense of history. Additionally, these areas have a larger visible area of water bodies and relatively less vegetation. The third cluster represents suburban rivers with low visual value, including all 17 research points located on the North Canal. Such research areas have very low scenic beauty, color richness, and vitality, and they do not convey a strong sense of history; the proportion of buildings and hard surfaces is also low. While this cluster has a higher proportion of grasslands than the first and second clusters, the proportion of trees is not high, resulting in an overall low green visibility rate. The fourth cluster is predominantly composed of research points on the South Canal, defined in this study as suburban rivers with high visual value. Research areas in this cluster have higher scenic beauty and color richness; a relatively lower proportion of buildings and hard surfaces; a higher visibility rate of the sky, indicating better openness; and a smaller visible water area, suggesting narrower river channels. Additionally, a weaker sense of history is also a distinctive feature of this cluster.

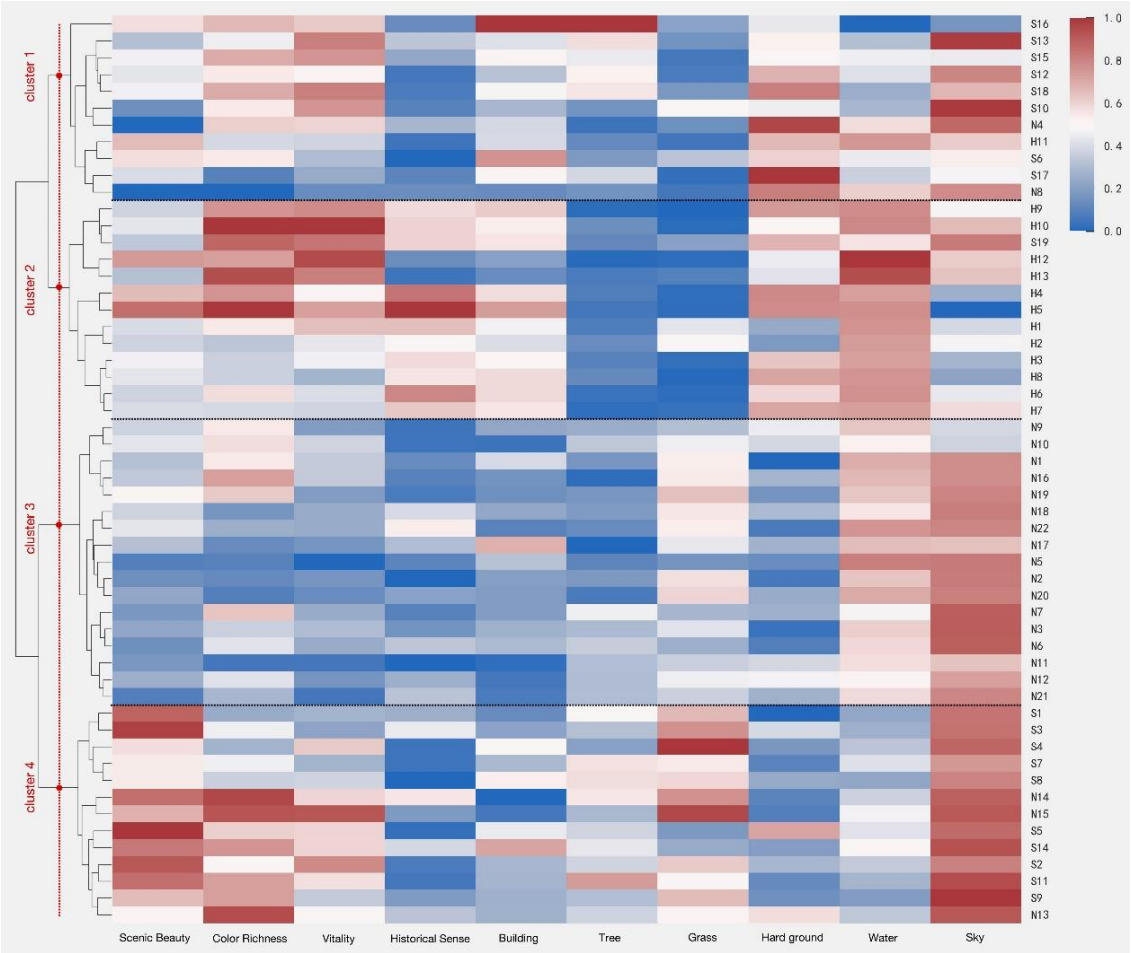


Figure 8. Cluster analysis of landscape features.

5. Discussion

5.1. Key Findings and Contributions

The rapid development of unmanned aerial vehicle technology has provided opportunities for quick monitoring and multi-perspective analysis of spatial characteristics in built environments. Compared to two-dimensional satellite remote sensing imagery, UAV-acquired image data offer higher resolution, enabling three-dimensional perception of large-scale scenes [37]. Moreover, UAV panoramic images offer a broader viewing angle and richer three-dimensional spatial information than traditional ground-level panoramic

images, becoming an important data source for understanding the spatial characteristics of large-scale landscapes [11,76].

This study combines UAV panoramic image data with immersive virtual reality perception technology to construct a method for analyzing the visual characteristics of large-scale urban river landscapes in three-dimensional space. By integrating the landscape characteristics of the study area and related research experiences, it proposes evaluating the visual landscape perception from four dimensions: scenic beauty, color richness, vitality, and historical sense. Additionally, using computer vision technology for pixel-level high-precision semantic segmentation of panoramic images, this approach accurately identifies various landscape elements within the images and quantitatively analyzes the spatial distribution characteristics of major landscape elements like buildings, trees, and water bodies in different river panoramas. Furthermore, by clustering analysis of immersive VR visual evaluation data and computer vision semantic segmentation results, this paper groups and characterizes different sections of the Haihe River, South Canal, and North Canal. The results show that many areas of the Haihe River belong to the second cluster, possessing high vitality and color richness, and conveying a strong sense of history. However, compared to the fourth cluster, primarily the South Canal, its scenic beauty is lower. Targeted enhancement of scenic beauty in the rivers and surrounding areas is key to further improving their landscape visual value. The South Canal and North Canal are important sections of the Grand Canal, a World Cultural Heritage site; compared to the Haihe River, this study's sections hardly convey any sense of history. Knowing how to enhance the historical perception of the South and North Canals is crucial for improving their visual value. Researchers found that the architectural facade styles might impact historical perception after comparing panoramic images of the South and North Canals with the Haihe River and consulting with volunteers who participated in the visual evaluation. However, the specific mechanisms of this impact remain unclear, and future research could delve into this issue. Additionally, there are certain differences in scenic beauty and vitality between the North Canal and South Canal, with most areas of the North Canal belonging to the third cluster, i.e., suburban rivers with low visual value. Enhancing the scenic beauty and vitality of such river sections is key to improving the visual value of the canals.

5.2. Limitations, Challenges, and Future Directions

Despite its contributions, this study faces several limitations and challenges that suggest directions for future research:

Seasonal constraints in data collection: Data were collected only in the summer, limiting the study's ability to account for seasonal variations in landscape characteristics. This is especially relevant in mid- to high-latitude regions, where environmental elements such as vegetation cover and scenic beauty change significantly across seasons. Future studies should explore multi-season data collection to better understand the spatiotemporal dynamics of landscape features.

Manual UAV flight operations: The UAV flights conducted in this study were manual, which required significant time and effort. While this approach ensured precise data collection, it is not scalable for larger or more frequent studies. Automated drone control systems could streamline future operations, reducing human workload and shortening data collection intervals.

Historical perception and architectural styles: Although the study identified differences in historical perception between the Haihe River and the South and North Canals, the underlying mechanisms affecting these perceptions remain unclear. Future research could investigate the influence of architectural styles and facade designs to better understand how historical value is visually conveyed, offering insights for heritage-based landscape enhancements.

Volunteer sample composition and subjective bias: Many of the volunteers who participated in the immersive VR evaluation were students and had lived in the study area for an extended period. While this sample provided valuable insights, it introduces

potential biases that may affect the generalizability of the results. The homogeneity of the sample in terms of age, background, and occupation may limit the ability to fully capture the diverse perspectives of the general population. Additionally, the volunteers' familiarity with the area may have influenced their evaluations, potentially causing them to rely more on their pre-existing perceptions than on the immersive virtual scenarios presented, thus affecting the objectivity of the scoring process. To mitigate these biases, future studies should aim to recruit a more diverse participant pool, including individuals from different age groups, occupations, and backgrounds.

This study has successfully identified and grouped the visual characteristics of different river sections, providing data support for a precise understanding of urban river landscape features. The methodology facilitates a quantitative evaluation of large-scale urban 3D landscapes and provides practical insights for landscape management and urban planning. By addressing these limitations and future challenges, future research can further improve the scalability, generalizability, and seasonal adaptability of the proposed framework, ensuring that it remains relevant for long-term landscape monitoring and urban development strategies.

5.3. Addressing Research Gaps

This study specifically addresses a key research gap in the field by integrating UAV imagery with immersive VR for environmental assessment, a combination that has been underexplored in previous studies. While UAVs are commonly used for aerial photography, few studies have incorporated UAV-based panoramic imagery into VR environments to enable immersive landscape assessments. By combining subjective visual evaluations through VR with objective computer vision techniques, this study introduces a novel approach to analyzing the spatial distribution of landscape features and evaluating visual quality. This research not only provides a replicable framework for urban landscape analysis but also offers valuable insights for urban planners and environmental managers, enabling them to better understand how different landscape elements contribute to the overall aesthetic and ecological value of urban environments.

In summary, this study offers a practical, scalable solution for evaluating landscape quality and provides a pathway for future research to incorporate automated UAV operations and seasonal variability analysis to further enhance the effectiveness of landscape management practices.

6. Conclusions

This study utilized UAVs to obtain panoramic image data of three rivers in Tianjin City; acquired dozens of participants' visual evaluation results on the scenic beauty, color richness, vitality, and sense of history of different scenes through immersive VR devices; and validated these evaluation results. It analyzed the spatial proportion of various landscape elements such as trees, grasslands, and buildings using computer vision semantic segmentation algorithms. By conducting hierarchical cluster analysis on VR visual perception data and computer vision semantic segmentation data from different research points, it realized the effective identification and grouped description of the visual features of different river sections. The study found the following: (1) There are significant differences in scenic beauty, color richness, and vitality among the three rivers. The South Canal had the highest average scenic beauty within the study area; the Haihe River's average color richness, vitality, and sense of history were higher than the other two rivers; the South Canal and North Canal provided a very limited sense of history. (2) Analysis of the landscape element proportions at different mapping points using computer vision technology found that the proportions of sky, water bodies, buildings, trees, grasslands, and hard surfaces exceed 90%, forming the main visual characteristics of the three rivers; among these, water bodies, trees, grasslands, and hard surfaces showed significant distribution differences across different river sections, while buildings and sky were relatively evenly distributed across the three rivers. (3) Hierarchical cluster analysis grouped and characterized mapping points with

different landscape features, revealing that many points in the South Canal belong to high visual value suburban river and low visual value urban river categories; the North Canal primarily comprised low visual value suburban rivers, while many points in the Haihe River fell into the high visual value urban river category.

By analyzing the multidimensional visual features of different urban rivers, the results of this study provide data support for urban planners, environmental researchers, and experts in other fields. It is also beneficial for the construction of the Grand Canal National Cultural Park and the enhancement of river landscape quality in Tianjin City. Meanwhile, the methodology proposed in this study can be widely applied to the research of river landscapes, offering opportunities for a better understanding of the river landscape environment.

Author Contributions: Conceptualization, Yunlei Shou and Junjie Luo; methodology, Zexin Lei and Junjie Luo; software, Yunlei Shou and Junjie Luo; validation, Zexin Lei and Jiaying Li; formal analysis, Yunlei Shou, Zexin Lei and Junjie Luo; investigation, Jiaying Li and Junjie Luo; resources, Yunlei Shou and Junjie Luo; data curation, Zexin Lei and Junjie Luo; writing—original draft preparation, Yunlei Shou and Junjie Luo; writing—review and editing, Junjie Luo; visualization, Junjie Luo; supervision, Junjie Luo; project administration, Junjie Luo; funding acquisition, Yunlei Shou and Junjie Luo. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Education Humanities and Social Sciences Research Special Task Project, grant number 24JDSZ3122; Zhejiang Province Philosophy and Social Sciences Planning Project, grant number 23ZK123Z, and Zhejiang A&F University Scientific Research Development Fund grant number 2023LFR076.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: We gratefully acknowledge the participants of the experiment.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, X.; Wang, X.; Jiang, X.; Han, J.; Wang, Z.; Wu, D.; Lin, Q.; Li, J.; Zhang, S.; Dong, Y. Prediction of riverside greenway landscape aesthetic quality of urban canalized rivers using environmental modeling. *J. Clean. Prod.* **2022**, *367*, 133066. [CrossRef]
2. Pradilla, G.; Hack, J. An urban rivers renaissance? Stream restoration and green–blue infrastructure in Latin America–Insights from urban planning in Colombia. *Urban Ecosyst.* **2024**, *27*, 2245–2265. [CrossRef]
3. Zhou, M.; Wang, F. The driving factors of recreational utilization of ecological space in urban agglomerations: The perspective of urban political ecology. *Ecol. Indic.* **2024**, *158*, 111409. [CrossRef]
4. Haeri, S.; Masnavi, M.R. Analyzing and Developing Strategies for the Ecological Restoration of Urban Rivers in the Framework of Ecological Urbanism. *MANZAR Sci. J. Landsc.* **2023**, *15*, 54–71.
5. Grzyb, T.; Kulczyk, S. How do ephemeral factors shape recreation along the urban river? A social media perspective. *Landsc. Urban Plan.* **2023**, *230*, 104638. [CrossRef]
6. D'Orey, A.; de Sousa Aragão, N.K.T.; Costa, C.S.; Duarte, T. A review of urban river restoration projects: Practices and challenges. *Default J.* **2023**.
7. Wantzen, K.M. River culture: How socio-ecological linkages to the rhythm of the waters develop, how they are lost, and how they can be regained. *Geogr. J.* **2024**, *190*, e12476. [CrossRef]
8. Li, X.; Li, Y.; Zhang, S.; Lin, R.; Chen, M.; Feng, L. Driving effects of land use and landscape pattern on different spontaneous plant life forms along urban river corridors in a fast-growing city. *Sci. Total Environ.* **2023**, *876*, 162775. [CrossRef] [PubMed]
9. Bush, J.; Doyon, A. Planning a just nature-based city: Listening for the voice of an urban river. *Environ. Sci. Policy* **2023**, *143*, 55–63. [CrossRef]
10. Garau, E.; Torralba, M.; Pueyo-Ros, J. What is a river basin? Assessing and understanding the sociocultural mental constructs of landscapes from different stakeholders across a river basin. *Landsc. Urban Plan.* **2021**, *214*, 104192. [CrossRef]
11. Luo, J.; Zhao, T.; Cao, L.; Biljecki, F. Semantic Riverscapes: Perception and evaluation of linear landscapes from oblique imagery using computer vision. *Landsc. Urban Plan.* **2022**, *228*, 104569. [CrossRef]
12. Luo, J.; Zhao, T.; Cao, L.; Biljecki, F. Water View Imagery: Perception and evaluation of urban waterscapes worldwide. *Ecol. Indic.* **2022**, *145*, 109615. [CrossRef]
13. Chiciudean, V.; Florea, H.; Beche, R.; Blaga, B.; Oniga, F.; Nedevschi, S. Data augmentation for environment perception with unmanned aerial vehicles. *IEEE Trans. Intell. Veh.* **2024**. [CrossRef]

14. Çınar, E.; Tuncal, A. The Future of UAVs in Urban Air Mobility: Public Perception and Concerns. *Türkiye İnsansız Hava Araçları Derg.* **2023**, *5*, 50–58. [CrossRef]
15. Muhmad Kamarulzaman, A.M.; Wan Mohd Jaafar, W.S.; Mohd Said, M.N.; Saad, S.N.; Mohan, M. UAV Implementations in Urban Planning and Related Sectors of Rapidly Developing Nations: A Review and Future Perspectives for Malaysia. *Remote Sens.* **2023**, *15*, 2845. [CrossRef]
16. Jiang, F.; Ma, J.; Webster, C.J.; Alain, C.; Zhou, Y.; Zhao, Z.; Zhang, X. Generative urban design: A systematic review on problem formulation, design generation, and decision-making. *Prog. Plan.* **2024**, *180*, 100795. [CrossRef]
17. Kim, B.C.; Kim, S.N.; Joo, Y. Visual impact control of urban waterfront development on the background mountain view: Examining its justifiability through two types of immersive virtual reality experiments. *Environ. Impact Assess. Rev.* **2024**, *106*, 107500. [CrossRef]
18. Puspitasari, A.W.; Kwon, J. Analysis of the visual quality of riverfront skyline through the feature of height and spatial arrangement of tall building. *Archit. Res.* **2019**, *21*, 91–98.
19. Chen, C.; Li, H.; Luo, W.; Xie, J.; Yao, J.; Wu, L.; Xia, Y. Predicting the effect of street environment on residents' mood states in large urban areas using machine learning and street view images. *Sci. Total Environ.* **2022**, *816*, 151605. [CrossRef]
20. Xu, Z.; Jiang, T.; Zheng, N. Developing and analyzing eco-driving strategies for on-road emission reduction in urban transport systems-A VR-enabled digital-twin approach. *Chemosphere* **2022**, *305*, 135372. [CrossRef]
21. Tabrizian, P.; Baran, P.K.; Van Berkel, D.; Mitasova, H.; Meentemeyer, R. Modeling restorative potential of urban environments by coupling viewscape analysis of lidar data with experiments in immersive virtual environments. *Landsc. Urban Plan.* **2020**, *195*, 103704. [CrossRef]
22. Lyu, R.; Pang, J.; Tian, X.; Zhao, W.; Zhang, J. How to optimize the 2D/3D urban thermal environment: Insights derived from UAV LiDAR/multispectral data and multi-source remote sensing data. *Sustain. Cities Soc.* **2023**, *88*, 104287. [CrossRef]
23. D'Urso, F.; Santoro, C.; Santoro, F.F. An integrated framework for the realistic simulation of multi-UAV applications. *Comput. Electr. Eng.* **2019**, *74*, 196–209. [CrossRef]
24. Martínez-Carricondo, P.; Carvajal-Ramírez, F.; Yero-Paneque, L.; Agüera-Vega, F. Combination of nadiral and oblique UAV photogrammetry and HBIM for the virtual reconstruction of cultural heritage. Case study of Cortijo del Fraile in Níjar, Almería (Spain). *Build. Res. Inf.* **2020**, *48*, 140–159. [CrossRef]
25. Ardiny, H.; Beigzadeh, A.; Mahani, H. Applications of unmanned aerial vehicles in radiological monitoring: A review. *Nucl. Eng. Des.* **2024**, *422*, 113110. [CrossRef]
26. Langhammer, J.; Janský, B.; Kocum, J.; Minařík, R. 3-D reconstruction of an abandoned montane reservoir using UAV photogrammetry, aerial LiDAR and field survey. *Appl. Geogr.* **2018**, *98*, 9–21. [CrossRef]
27. Vacca, G.; Dessì, A.; Sacco, A. The use of nadir and oblique UAV images for building knowledge. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 393. [CrossRef]
28. Zhang, X.; Zhao, P.; Hu, Q.; Ai, M.; Hu, D.; Li, J. A UAV-based panoramic oblique photogrammetry (POP) approach using spherical projection. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 198–219. [CrossRef]
29. Emilien, A.V.; Thomas, C.; Thomas, H. UAV & satellite synergies for optical remote sensing applications: A literature review. *Sci. Remote Sens.* **2021**, *3*, 100019.
30. Ding, L.; Zhou, J.; Meng, L.; Long, Z. A practical cross-view image matching method between UAV and Satellite for UAV-based geo-localization. *Remote Sens.* **2020**, *13*, 47. [CrossRef]
31. Fagin, T.D.; Vadjunec, J.M.; Boardman, A.L.; Hinsdale, L.M. Use of Participatory sUAS in Resilient Socioecological Systems (SES) Research: A Review and Case Study from the Southern Great Plains, USA. *Drones* **2024**, *8*, 223. [CrossRef]
32. Watkins, S.; Burry, J.; Mohamed, A.; Marino, M.; Prudden, S.; Fisher, A.; Kloet, N.; Jakobi, T.; Clothier, R. Ten questions concerning the use of drones in urban environments. *Build. Environ.* **2020**, *167*, 106458. [CrossRef]
33. Wiltenburg, R.; Mendoza, F.R.; Hurst, W.; Tekinerdogan, B. Virtual Reality for Spatial Planning and Emergency Situations: Challenges and Solution Directions. *Appl. Sci.* **2024**, *14*, 3595. [CrossRef]
34. Meng, L.; Li, S.; Zhang, X. Assessing biodiversity's impact on stress and affect from urban to conservation areas: A virtual reality study. *Ecol. Indic.* **2024**, *158*, 111532. [CrossRef]
35. Wang, Y.; Wang, S.; Pan, Y.; Li, C.; Wang, J. Immersive virtual reality and computer vision for heritage: Visual evaluation and perception of the industrial heritage sites along the Yunnan–Vietnam railway (Yunnan section). *Herit. Sci.* **2024**, *12*, 36. [CrossRef]
36. Papadopoulou, E.E.; Papakonstantinou, A. Combining Drone LiDAR and Virtual Reality Geovisualizations towards a Cartographic Approach to Visualize Flooding Scenarios. *Drones* **2024**, *8*, 398. [CrossRef]
37. Balestra, M.; Choudhury, A.M.; Pierdicca, R.; Chiappini, S.; Marcheggiani, E. UAV-Spherical Data Fusion Approach to Estimate Individual Tree Carbon Stock for Urban Green Planning and Management. *Remote Sens.* **2024**, *16*, 2110. [CrossRef]
38. Liao, B.; van den Berg, P.E.W.; van Wesemael, P.J.V.; Arentze, T.A. Individuals' perception of walkability: Results of a conjoint experiment using videos of virtual environments. *Cities* **2022**, *125*, 103650. [CrossRef]
39. Keil, J.; Weißmann, M.; Korte, A.; Edler, D.; Dickmann, F. Measuring physiological responses to visualizations of urban planning scenarios in immersive virtual reality. *KN-J. Cartogr. Geogr. Inf.* **2023**, *73*, 117–126. [CrossRef]
40. Zhou, K.; Wang, Z.; Ni, Y.Q.; Zhang, Y.; Tang, J. Unmanned aerial vehicle-based computer vision for structural vibration measurement and condition assessment: A concise survey. *J. Infrastruct. Intell. Resil.* **2023**, *2*, 100031. [CrossRef]

41. Mouratidis, K.; Hassan, R. Contemporary versus traditional styles in architecture and public space: A virtual reality study with 360-degree videos. *Cities* **2020**, *97*, 102499. [CrossRef]
42. Hu, C.B.; Zhang, F.; Gong, F.Y.; Ratti, C.; Li, X. Classification and mapping of urban canyon geometry using Google Street View images and deep multitask learning. *Build. Environ.* **2020**, *167*, 106424. [CrossRef]
43. Xu, F.; Wong, M.S.; Zhu, R.; Heo, J.; Shi, G. Semantic segmentation of urban building surface materials using multi-scale contextual attention network. *ISPRS J. Photogramm. Remote Sens.* **2023**, *202*, 158–168. [CrossRef]
44. Manakitsa, N.; Maraslidis, G.S.; Moysis, L.; Fragulis, G.F. A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision. *Technologies* **2024**, *12*, 15. [CrossRef]
45. Marasinghe, R.; Yigitcanlar, T.; Mayere, S.; Washington, T.; Limb, M. Computer vision applications for urban planning: A systematic review of opportunities and constraints. *Sustain. Cities Soc.* **2023**, *2023*, 105047. [CrossRef]
46. Lyu, Y.; Vosselman, G.; Xia, G.S.; Yilmaz, A.; Yang, M.Y. UAVid: A semantic segmentation dataset for UAV imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *165*, 108–119. [CrossRef]
47. DePasquale, A.; Hogan, J.D.; Guadamuz Araya, C.; Dominy, N.J.; Melin, A.D. Aeroscapes and the sensory ecology of olfaction in a tropical dry forest. *Front. Ecol. Evol.* **2022**, *10*, 849281. [CrossRef]
48. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]
49. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 12077–12090.
50. Mishra, A.; Agnihotri, A.K.; Pipil, S.; Gaur, S.; Ohri, A. Surveying techniques for urban areas. In *Earth Observation in Urban Monitoring*; Elsevier: Amsterdam, The Netherlands, 2024; pp. 69–91.
51. Wang, Z.; Ito, K.; Biljecki, F. Assessing the equity and evolution of urban visual perceptual quality with time series street view imagery. *Cities* **2024**, *145*, 104704. [CrossRef]
52. Ogawa, Y.; Oki, T.; Zhao, C.; Sekimoto, Y.; Shimizu, C. Evaluating the subjective perceptions of streetscapes using street-view images. *Landsc. Urban Plan.* **2024**, *247*, 105073. [CrossRef]
53. Widdowson, C.; Yoon, H.J.; Hovakimyan, N.; Wang, R.F. A Novel Measure of Human Safety Perception in Response to Flight Characteristics of Collocated UAVs in Virtual Reality. *IEEE Trans. Hum.-Mach. Syst.* **2023**, *54*, 1–10. [CrossRef]
54. Zhang, X.; Qi, J.; Lin, E.S.; Tan, P.Y.; Ho, R.; Sia, A.; Song, X.P.; Waykool, R.; Olszewska-Guizzo, A. Towards healthy cities: Modeling restorative potential of urban environments by coupling LiDAR-derived 3D metrics with panorama-based online survey. *Environ. Impact Assess. Rev.* **2024**, *106*, 107497. [CrossRef]
55. Li, X.; Huang, K.; Zhang, R.; Chen, Y.; Dong, Y. Visual Perception Optimization of Residential Landscape Spaces in Cold Regions Using Virtual Reality and Machine Learning. *Land* **2024**, *13*, 367. [CrossRef]
56. Clay, G.R.; Daniel, T.C. Scenic landscape assessment: The effects of land management jurisdiction on public perception of scenic beauty. *Landsc. Urban Plan.* **2000**, *49*, 1–13. [CrossRef]
57. Wu, C.; Ye, Y.; Gao, F.; Ye, X. Using street view images to examine the association between human perceptions of locale and urban vitality in Shenzhen, China. *Sustain. Cities Soc.* **2023**, *88*, 104291. [CrossRef]
58. Ma, Z. Deep exploration of street view features for identifying urban vitality: A case study of Qingdao city. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *123*, 103476. [CrossRef]
59. Mirri, S.; Prandi, C.; Roccetti, M.; Salomoni, P. Walking under a different sky: Urban colored routes for creative engagement and pleasure. *Int. J. Hum.–Comput. Interact.* **2017**, *33*, 1010–1021. [CrossRef]
60. Ma, B.; Hauer, R.J.; Xu, C.; Li, W. Visualizing evaluation model of human perceptions and characteristic indicators of landscape visual quality in urban green spaces by using nomograms. *Urban For. Urban Green.* **2021**, *65*, 127314. [CrossRef]
61. Seremetakis, C.N. The Memory of the Senses Historical Perception, Commensal Exchange, and Modernity. In *Visualizing Theory*; Routledge: Vienna, Austria, 2014; pp. 214–229.
62. Lee, C.; Park, A.H.; Lee, H.; Bratman, G.N.; Hankey, S.; Li, D. Measuring urban nature for pedestrian health: Systematic review and expert survey. *Landsc. Urban Plan.* **2024**, *250*, 105129. [CrossRef]
63. Abd-Alhamid, F.; Kent, M.; Wu, Y. Assessment of window size and layout impact on a view quality perception in a virtual reality environment. *Leukos* **2024**, *20*, 239–260. [CrossRef]
64. Pöhlmann, K.; Wilson, G.; Mcgill, M.; Medeiros, D.P.S.; Li, G.; Brewster, S. *Towards Cross-Platform and Multi-Transport Real Time XR-Experiences*; ACM: New York, NY, USA, 2022.
65. Luo, J.; Liu, P.; Xu, W.; Zhao, T.; Biljecki, F. A perception-powered urban digital twin to support human-centered urban planning and sustainable city development. *Cities* **2025**, *156*, 105473. [CrossRef]
66. Chum, L.; Subramanian, A.; Balasubramanian, V.N.; Jawahar, C.V. Beyond supervised learning: A computer vision perspective. *J. Indian Inst. Sci.* **2019**, *99*, 177–199. [CrossRef]
67. Liu, D.; Jiang, Y.; Wang, R.; Lu, Y. Establishing a citywide street tree inventory with street view images and computer vision techniques. *Comput. Environ. Urban Syst.* **2023**, *100*, 101924. [CrossRef]
68. Mo, Y.; Wu, Y.; Yang, X.; Liu, F.; Liao, Y. Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing* **2022**, *493*, 626–646. [CrossRef]
69. Zhao, L. Image semantic segmentation method based on GAN network and FCN model. *J. Eng.* **2022**, *2022*, 1–9. [CrossRef]

70. Wang, J.; Liu, W.; Gou, A. Numerical characteristics and spatial distribution of panoramic Street Green View index based on SegNet semantic segmentation in Savannah. *Urban For. Urban Green.* **2022**, *69*, 127488. [CrossRef]
71. Zhao, T.; Liang, X.; Tu, W.; Huang, Z.; Biljecki, F. Sensing urban soundscapes from street view imagery. *Comput. Environ. Urban Syst.* **2023**, *99*, 101915. [CrossRef]
72. Linling, L. Dividing Ecological units of Rural Landscape Based on Clustering Algorithm. In Proceedings of the 2024 International Conference on Integrated Circuits and Communication Systems (ICICACS), Raichur, India, 23–24 February 2024; pp. 1–5.
73. Heydari, H.; Momeni, M.; Nadi, S. Innovative data clustering method improves drought prediction in heterogeneous landscapes using GEE-derived remote sensing indices. *Remote Sens. Appl. Soc. Environ.* **2024**, *33*, 101112. [CrossRef]
74. Zhang, H.; Peng, Y. Image clustering: An unsupervised approach to categorize visual data in social science research. *Sociol. Methods Res.* **2024**, *53*, 1534–1587. [CrossRef]
75. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0, fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef] [PubMed]
76. Zhou, W.; Cao, W.; Wu, T.; Zhang, T. The win-win interaction between integrated blue and green space on urban cooling. *Sci. Total Environ.* **2023**, *863*, 160712. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.