

Visual Semantics and Multimodal Perception of Digitally Reconstructed Garden Scenes of Jiucheng Palace

Jingru Gan

Northwest University

Chao Wen

wenchao@nwu.edu.cn

Northwest University

Yun Xiao

Northwest University

Jing Liu

Northwest University

Jiani Zhang

Northwest University

Article

Keywords:

Posted Date: September 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10714462/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Visual Semantics and Multimodal Perception of Digitally

Reconstructed Garden Scenes of Jiucheng Palace

Jingru Gan^a, Chao Wen^{a, c*}, Yun Xiao^b, Jing Liu^{a*}, Jiani Zhang^a

a. School of Art, Northwest University, Xi'an, Shaanxi 710127, China,

b. School of Computer Science, Northwest University, Xi'an, Shaanxi 710127, China

c. National–Local Joint Engineering Research Center for Digitalization of Cultural Heritage, Northwest University, Xi'an, Shaanxi 710127, China

***Corresponding author:**

Chao Wen, E-mail: wenchao@nwu.edu.cn

Jing Liu, E-mail: 20200023@nwu.edu.cn

Abstract

Jiucheng Palace is an important Sui–Tang imperial detached palace garden site, and digital reconstruction is key to understanding its lost spatial form. Existing evaluations emphasize historical evidence and morphological accuracy but lack empirical validation of how reconstructed scenes convey spatial relationships and experiential characteristics. This study established an integrated evaluation framework based on multimodal perception for reconstructed garden scenes. Basing the study on Jiucheng Palace reconstructions, we selected architectural, Mountain–water, and vegetation scenes, quantified features using deep learning–based semantic segmentation, integrated eye-tracking, electroencephalography, and semantic differential data, and conducted immersive validation. Results showed significant differences in multiple subjective ratings and Alpha power, but not eye-tracking measures; Spearman analysis revealed associations between visual semantic features, subjective evaluations, and perceptual responses. A comprehensive evaluation matrix, Perceived Garden Experience, and Perceptual Effectiveness Index extended assessment from morphological accuracy to spatial expression and experience conveyance, providing quantitative support for reconstruction comparison and presentation optimization.

Introduction

Chinese classical gardens constitute an important expression of traditional Chinese culture and an invaluable part of humanity's cultural heritage. They embody the integration of ancient Chinese philosophy, religious thought, culture, and the arts, while reflecting the ideal of the “unity of heaven and humanity” shaped by Confucian, Buddhist, and Taoist traditions^[1]. The fundamental characteristic that distinguishes Chinese classical gardens from other garden traditions lies not in the pursuit of a particular outward form, but in the creation of *yijing* (artistic conception), whereby finite spatial forms evoke seemingly boundless emotions and reflections in the viewer^[2]. Within the Chinese aesthetic system, *yijing* is regarded as a key category that links objective scenery with subjective emotion and embodies an aesthetic ideal extending beyond visual beauty. Its essence resides not in any specific physical space or individual landscape element, but in the holistic experience jointly produced by the scenery and the viewer's emotions and associations during appreciation^[3]. As a distinctive category of Chinese classical gardens, imperial detached palace gardens differed from the relatively enclosed palace gardens located within imperial capitals. They were typically set amid mountains and water and formed integrated complexes for recreation,

residence, and political activities^[4]. This study focuses on Jiucheng Palace, a renowned imperial detached palace of the Sui and Tang dynasties. However, only remnants of building platforms, column bases, and localized topographic features survive at the site, making its historical spatial organization and aesthetic intent difficult to discern. Consequently, its historical landscape configuration and intended *yijing* can be understood only through textual research and reconstruction-based inference.

Advances in technologies such as terrestrial laser scanning^[5], unmanned aerial vehicle (UAV) photogrammetry^[6], and digital modeling^[7] have provided important technical support for the accurate documentation of cultural heritage spaces, the representation of historical scenes, and three-dimensional reconstruction. Relevant studies have shown that multisource spatial data and digital modeling technologies can not only preserve detailed morphological information about heritage objects but also provide a reliable basis for representing historical scenes digitally^[8-9]. These technologies have also been applied to research on Chinese classical gardens, and existing studies have focused mainly on two areas. The first focuses on the digital documentation and reconstruction of garden spaces, architectural remains, and landscape elements using terrestrial laser scanning, UAV photogrammetry, image fusion, and virtual modeling^[10-11]; the second examines visual attention, spatial perception, landscape appreciation, and emotional responses in existing gardens or real-world landscapes using computer vision, spatial analysis, eye tracking, and related methods^[12-15]. These studies have laid an important foundation for the digital reconstruction of Chinese classical gardens and the study of how they are perceived. However, research on garden perception has focused mainly on existing gardens and real-world landscapes, while the relationship between spatial representation and perceptual experience in digitally reconstructed scenes of historic gardens has not yet been sufficiently validated. For historical gardens such as Jiucheng Palace, where only incomplete remains survive and the original landscape can no longer be experienced directly, digital reconstruction must do more than recreate spatial form and historical scenes. It must also accurately convey spatial relationships and landscape imagery, making it essential to evaluate the perceptual effectiveness of the reconstructed scenes.

The experience of *yijing* in gardens arises from viewers' integrated perception of spatial relationships and garden imagery. This experience is shaped by the spatial organization, relative proportions, and visual representation of landscape elements. However, these visual characteristics are difficult to convert directly into measurable indicators. For digitally reconstructed scenes, objectively identifying different landscape elements and quantifying their composition within images are fundamental steps in analyzing perceptual experience. Manual interpretation can be influenced by researchers' experience and differences in classification criteria, making it difficult to maintain consistent identification across scenes. With advances in computer vision and deep learning, semantic segmentation has become an important method for image-based scene parsing. It enables different objects and regions within an image to be classified automatically and provides a technical basis for identifying and quantifying complex visual information^[16-17]. In landscape research, semantic segmentation can identify landscape elements in images and extract their proportions and spatial distributions^[18-19]. This process converts the visual composition of scenes into comparable quantitative data, improves consistency across scenes, and reduces the influence of subjective judgment. Research has also been extended to historical landscapes and Chinese classical gardens. Ding et al. (2026)^[20] used SegFormer-B5 to identify landscape elements in historical photographs of water towns and combined semantic segmentation with depth estimation to analyze

spatial perception. Similarly, Chai et al. (2026)^[21] integrated DeepLab V3+ semantic segmentation with spatial grid analysis to identify differences in the visual preferences of Chinese and Western visitors in Chinese classical gardens. To meet the analytical needs of the digitally reconstructed Jiucheng Palace scenes, this study uses semantic segmentation to extract the proportions and spatial distributions of different landscape elements, thereby providing a quantitative basis for the objective comparison of their visual composition.

Yijing experience in gardens is formed through a dynamic perceptual process. Semantic features can reveal the composition of a scene, but they cannot explain how its landscape elements are viewed and experienced. To obtain feedback on viewers' experiences, studies of spatial experience commonly use questionnaires to assess viewers' subjective perceptions and emotional evaluations of scenes^[22-23]. However, questionnaires mainly capture overall judgments made after viewing and cannot directly record changes in visual attention or neurophysiological responses during viewing. With advances in perceptual measurement technologies, increasing attention has been paid to individuals' immediate responses during information processing^[24-25]. Multimodal perceptual measurement methods, including eye tracking and electroencephalography (EEG), have been widely used in studies of spatial experience in architectural, landscape, and cultural heritage settings to examine the relationships among visual attention, emotional responses, and environmental experience^[26-28]. Under controlled stimulus-presentation conditions, eye-tracking metrics can capture the distribution and concentration of visual attention across different landscape scenes, providing a quantitative basis for analyzing scene-specific patterns of visual attention^[29]. In contrast, EEG signals provide complementary neurophysiological evidence of emotional arousal and cognitive processing during viewing^[30]. The semantic differential (SD) questionnaire can be used to quantify viewers' subjective evaluations of different scenes^[31]. By integrating these three approaches, multimodal perceptual measurement provides complementary evidence across three dimensions: overt visual attention, neurophysiological responses, and subjective evaluations. This integration combines subjective judgments obtained through questionnaires with objective responses recorded during viewing, thereby supporting a more comprehensive assessment of reconstructed scenes and the analysis of the relationship between the semantic features of garden spaces and viewers' experiences.

Accordingly, this study focuses on the digitally reconstructed scenes of Jiucheng Palace and constructs a multimodal perception-based evaluation framework for digitally reconstructed garden scenes. The framework uses semantic segmentation to quantify the visual semantic features of the scenes and integrates subjective and objective data obtained through eye tracking, EEG, and SD questionnaires. The integrated data are then used to analyze differences in visual attention, neural responses, and subjective experience across reconstructed scene types and to examine the relationships between visual semantic features and multimodal perceptual responses. The framework thereby provides a quantitative approach to evaluating and optimizing the presentation of digitally reconstructed garden heritage scenes.

Based on this framework, the study addresses the following three research questions:

RQ1: What differences in multimodal perception emerge among different types of digitally reconstructed garden scenes?

RQ2: What relationships exist between the visual semantic features of reconstructed scenes and multimodal perceptual responses?

RQ3: How can visual semantic features and multimodal perceptual responses be combined to construct an integrated evaluation framework for digitally reconstructed garden scenes?

Methods

Study area

Jiucheng Palace of the Tang dynasty, known as Renshou Palace during the Sui dynasty, is located in Jiuchenggong Town, Linyou County, Baoji, Shaanxi Province, China, and is a representative imperial detached palace of the Sui and Tang periods^[32]. Built approximately 1,400 years ago, the site was included in the fourth batch of Major Historical and Cultural Sites Protected at the National Level in 1996. Historical records indicate that Jiucheng Palace held a prominent position and was regarded as the foremost imperial detached palace. For a time, it also served as a center for political, cultural, military, diplomatic, and religious activities during the Sui and Tang dynasties. As recorded in the *Treatise on Geography* in the *New Book of Tang*^[33], “The perimeter wall extended 1,800 bu, and an imperial park, government storehouses, official offices, and other facilities were established within it”, it suggests that Jiucheng Palace was enclosed by a clearly defined perimeter wall and combined garden and administrative functions, reflecting the integrated spatial organization typical of an imperial detached palace. The site consisted of two main components: the Imperial Precinct and the Imperial Park. The palace buildings were concentrated in the valley, whereas the Imperial Park was enclosed by walls that followed the mountain ridges, forming an extensive garden area characterized by nine surrounding mountains and seven winding waterways (**Fig. 1**). The overall garden layout followed the design principles of adapting to the mountain terrain and creating scenery through water. The mountain landforms, spring-fed watercourses, and building platforms together created a coherent sequence of views and a distinct rhythm of spatial experience^[34]. This spatial organization shows how the architecture was closely integrated with the surrounding mountains, terrain, and natural water system, reflecting the principles of working with nature and adapting the design to local conditions. This dual quality, involving both spatial and psychological dimensions, not only provides a basis for the perception of *yijing* but also makes Jiucheng Palace a suitable case for landscape semantic analysis and perceptual experiments.

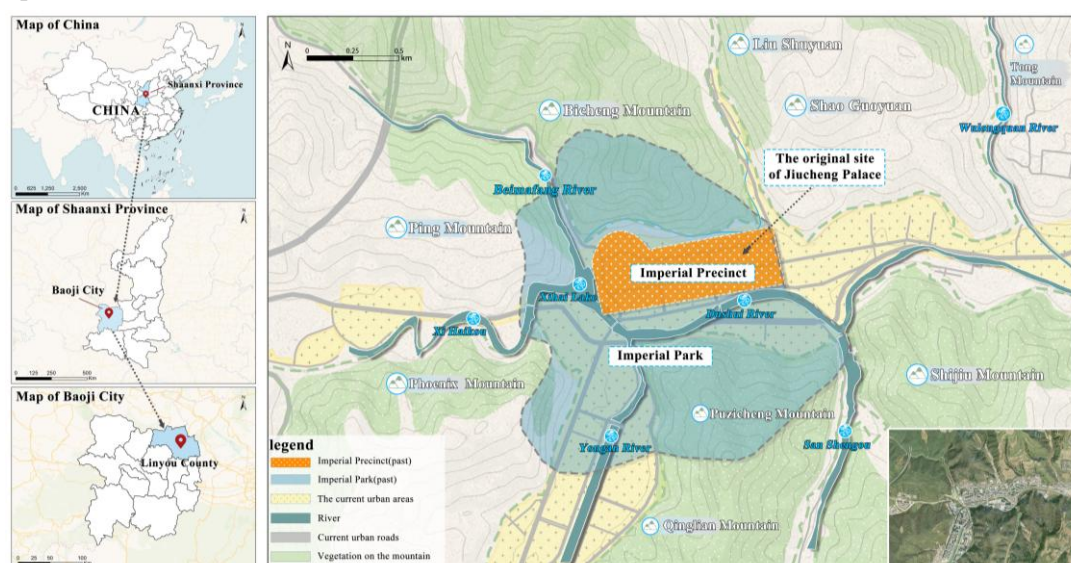


Fig.1 Location and spatial context of the Jiucheng Palace site.

Source: The three location maps on the left were drawn by the authors using the CARTO Voyager basemap (no labels; © OpenStreetMap contributors and © CARTO); satellite imagery in the lower-right corner: Esri World Imagery.

Research framework

The overall research framework consists of four main stages (**Fig. 2**). First, a three-dimensional digital reconstruction of Jiucheng Palace was created using historical documents and archaeological evidence. Three representative landscape scene types were then selected through expert evaluation, and standardized visual stimuli were generated for the subsequent experiments. Second, a deep-learning-based semantic segmentation model was used to identify visual elements in the stimulus images and extract their proportions and spatial distributions, thereby revealing the visual semantic structure of each scene. Third, under controlled conditions, eye-tracking and EEG experiments were conducted, and SD questionnaires were administered, to collect data on participants' visual attention, neurophysiological responses, and subjective evaluations. Interviews were also conducted to further examine their perceptual experiences. The main findings from the static-image experiment were then validated through an immersive virtual reality experiment. Finally, visual semantic features and multimodal perceptual responses were integrated to analyze the relationships between the semantic features of the reconstructed scenes and the public's perceptual responses. A comprehensive multimodal evaluation matrix oriented toward public perception was then constructed to provide a basis for evaluating digitally reconstructed scenes and optimizing their presentation.

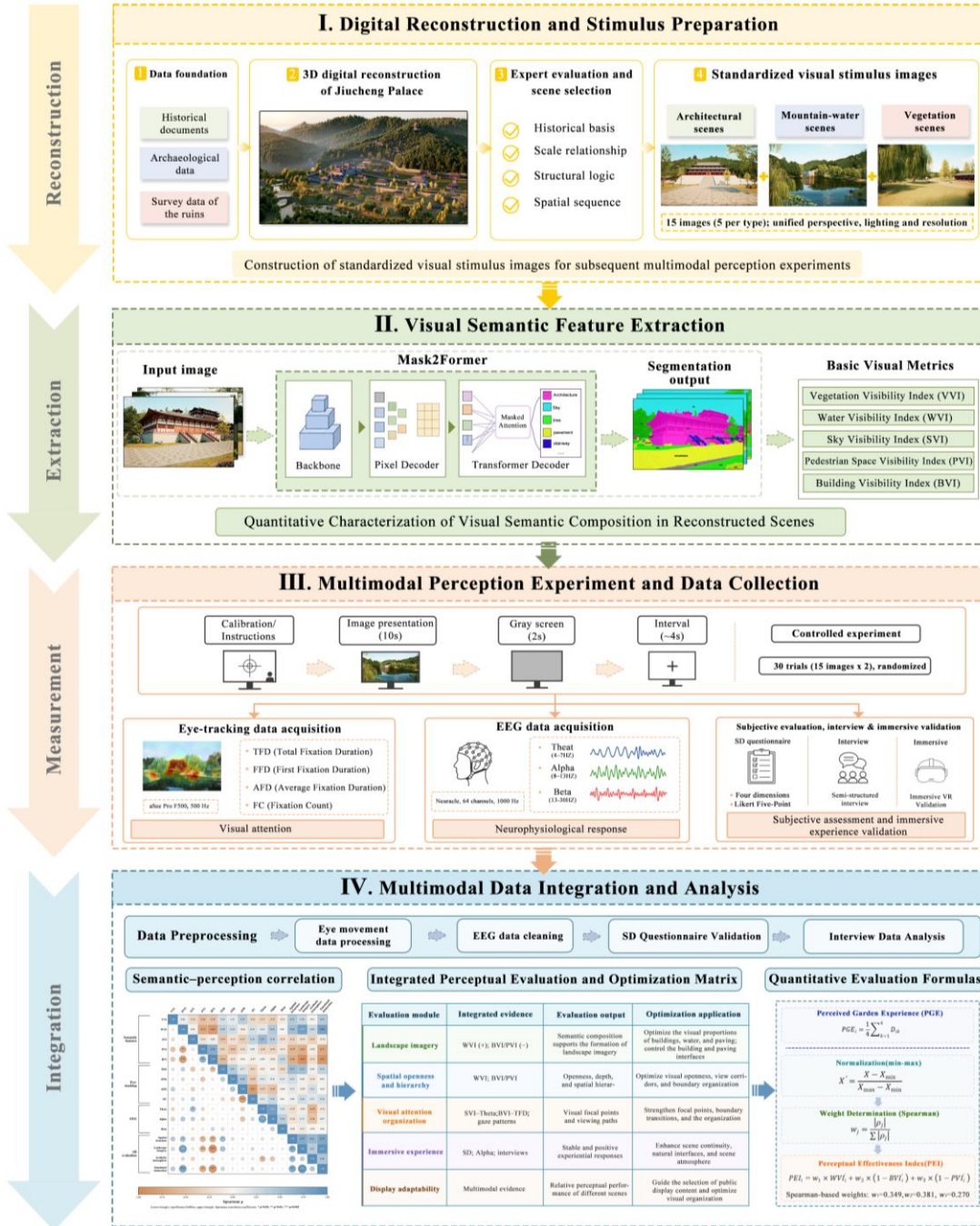


Fig.2 Research framework for the multimodal perceptual evaluation of digitally reconstructed garden scenes.

Selection and Preparation of Experimental Stimuli

The experimental materials used in this study were derived from the comprehensive digital reconstruction of the Jiucheng Palace site previously completed by our research team. The reconstruction integrated historical documents, archaeological records, site survey data, and terrain models. It covered 37 identified locations within the Jiucheng Palace site and their surrounding environments, including individual buildings, the palace and garden layout, terrain and water systems, and the wider landscape setting. The reconstruction covered not only the architectural forms, spatial positions, and scale relationships of the individual sites, but also the spatial connections among them and their relationship with the surrounding mountain and water landscape.

The reconstruction process began with SketchUp, which was used to create preliminary massing models of the main buildings and quickly check the overall spatial layout and scale relationships. The models were then imported into 3ds Max for further refinement and construction of the overall environment, including terrain reconstruction, detailed building modeling, the placement of landscape elements, and the creation of materials and textures. Specifically, terrain contours were adjusted using existing DEM data and archaeological survey records. The forms, proportions, and structural features of the buildings were inferred from archaeological reports, other documentary sources, and relevant previous reconstruction studies. After the preliminary reconstruction was completed, experts in traditional Chinese architectural history and garden history were invited to evaluate the model based on historical evidence, scale relationships, construction logic, and spatial sequence. Based on their feedback, targeted adjustments were made to key elements, including the dimensions of certain buildings, terrain gradients, and watercourse alignments. These revisions resulted in a three-dimensional reconstructed environment with a relatively high level of historical credibility.

Based on the digital reconstruction, the reconstructed environment was classified into landscape types by considering historical descriptions, the overall mountain-water pattern of Jiucheng Palace, and the distribution of buildings. These landscape types were then used in the subsequent semantic segmentation and multimodal perception experiments. Drawing on the landscape classification principles proposed by Xiao and Zhong (1998)^[35] and considering the characteristics of Chinese classical gardens, three representative landscape scene types were identified within the reconstructed environment. 1) Architectural scenes: spaces dominated by architectural elements, such as palace halls and covered corridors. 2) Mountain-water scenes: composite scenes characterized by the spatial combination of undulating mountains, water bodies, and natural landforms. 3) Vegetation scenes: landscape spaces in which trees, other vegetation, and natural ground cover form the main visual elements. This classification was based on the dominant visual elements in each scene and was intended to produce experimental samples with clear and representative visual differences.

To ensure consistency throughout the experiment and subsequent analyses, images representing the three selected types of reconstructed scenes were generated in D5 Render. All images were rendered using fixed camera viewpoints, focal lengths, and heights, with consistent lighting, time-of-day, environmental, material, and rendering settings to minimize visual interference from variables unrelated to the experiment. The output images were standardized to the same dimensions to meet the presentation requirements of the subsequent semantic segmentation analysis and multimodal perception experiments. In total, 15 static images were produced across three landscape scene types: architectural scenes, Mountain-water scenes, and vegetation scenes. The procedure used to determine the experimental materials is shown in **Fig. 3**.

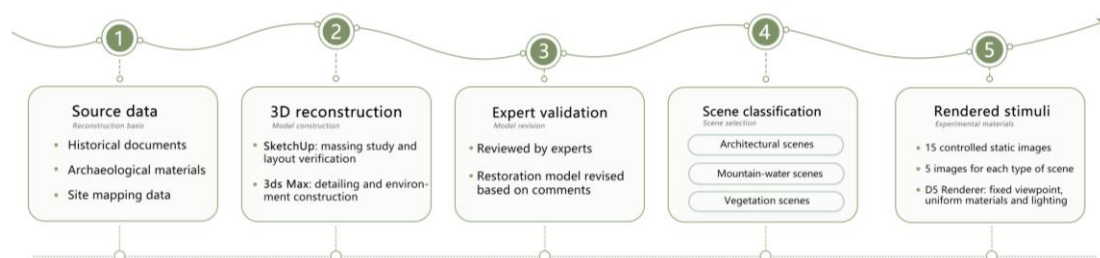


Fig.3 Workflow for selecting and preparing the tested materials

Semantic Segmentation and Construction of Objective Visual Indicators

Semantic segmentation is a fundamental method in computer vision for analyzing landscape scenes and provides quantitative data for subsequent visual indicator extraction and multimodal perception analysis. In this study, the Mask2Former model trained on the ADE20K dataset was used to segment the stimulus images and extract the pixel proportions of different spatial elements in each scene^[36]. The model identified over 20 semantic categories, including buildings, sky, trees, grass, water, roads, rocks, mountains, and waterfalls (**Fig. 4**). To improve the accuracy of the segmentation results, over-segmented regions were merged. The segmentation results showed that the architectural scenes consisted mainly of sky, buildings, and paved pedestrian spaces, followed by trees. The Mountain-water scenes were dominated by water and also included elements such as sky and trees. The vegetation scenes were dominated by trees and also included sky, grass, and small areas of ground. In Chinese classical garden scenes, mountains generally contribute to the landscape as integrated forms covered by vegetation and are visually perceived mainly through changes in elevation and background contours beneath the vegetation cover. Therefore, mountain areas were represented mainly by vegetation-related categories in the semantic segmentation results. Merging these categories improved the stability of subsequent indicator calculations and made the results easier to interpret.

Previous studies have shown that the proportions and spatial arrangement of natural and artificial elements influence viewers' overall perception of landscape spaces^[37]. Based on these findings and the spatial characteristics of Chinese classical gardens, a set of objective visual indicators was developed to describe the visual composition of landscape scenes^[38]. The system includes five basic visual indicators (**Table 1**): the Vegetation Visibility Index (VVI), Water Visibility Index (WVI), Sky Visibility Index (SVI), Pedestrian Space Visibility Index (PVI), and Building Visibility Index (BVI). Among these indicators, VVI, WVI, and SVI primarily reflect the visual presence of natural elements in a scene, whereas PVI and BVI reflect the visual prominence of built elements in heritage environments.

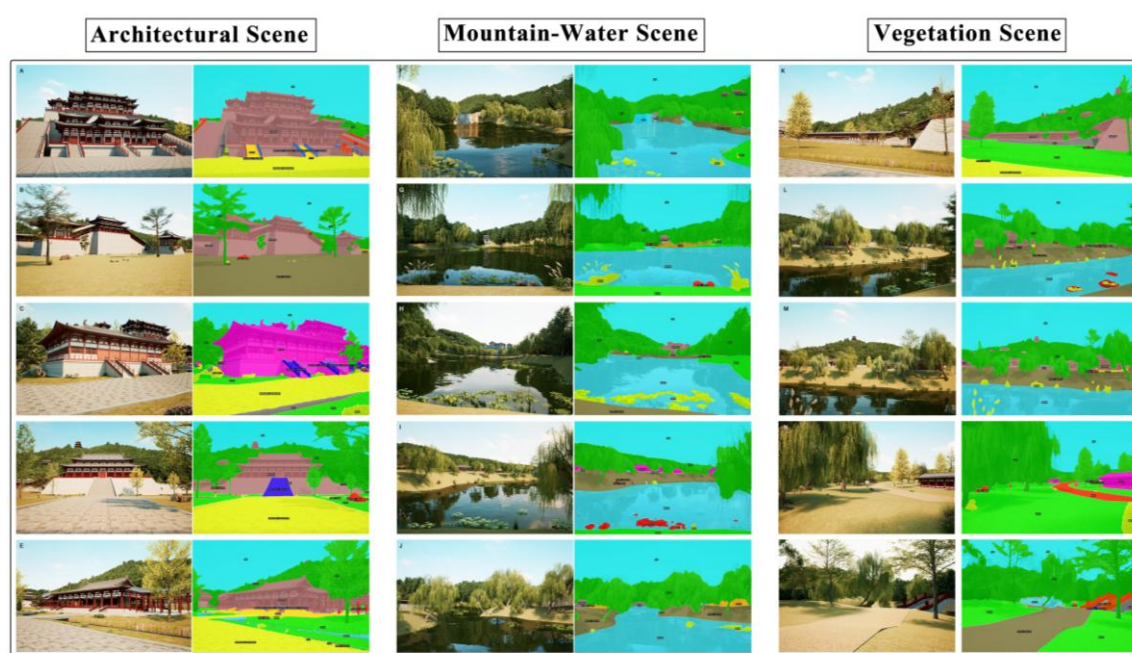


Fig.4 Reconstructed images and semantic segmentation results.

Table 1. Definitions of the objective visual indices

Index	Description
Vegetation Visibility Index (VVI)	The proportion of vegetation elements, including trees, grass, plants, flowers, and ground, visible within the field of view.
Water Visibility Index (WVI)	The proportion of water elements, including water and waterfalls, visible within the field of view.
Sky Visibility Index (SVI)	The proportion of sky visible within the field of view.
Pedestrian Space Visibility Index (PVI)	The proportion of pedestrian space, including floors, sidewalks, paths, stairs, bridges, and roads, visible within the field of view.
Building Visibility Index (BVI)	The proportion of building related elements, including walls, buildings, houses, sculptures, and banisters, visible within the field of view.

Experimental procedure

A total of 48 participants were recruited for this study, including 16 men and 32 women. All participants had normal or corrected-to-normal vision and no neurological or visual impairments. Based on their academic backgrounds, the participants were divided into two groups for between-group comparisons: 24 participants with backgrounds in garden studies or related fields, including landscape architecture, architecture, and art and design, and 24 participants without relevant academic backgrounds.

The multimodal perception experiment was conducted at the VR Joint Laboratory of Northwest University in June 2026. The experimental equipment included a Neuracle EEG system, an aSee Pro F500 desktop eye tracker, and a Dell P2222H monitor, which were used to record participants' neurophysiological and eye-movement responses while they viewed the stimulus images. Participants were seated approximately 50–60 cm from the monitor, and the EEG equipment was fitted before the eye tracker was calibrated. Participants were informed that they could stop the experiment and withdraw at any time if they experienced discomfort. Before the formal experiment, two practice images were presented to help participants become familiar with the task. During the formal experiment, 15 stimulus images were presented. Each image was presented twice, resulting in a total of 30 trials and improving the stability of the measurements. To reduce order and fatigue effects, the 30 trials were divided into three blocks of 10 trials, and the presentation order was randomized. Each image was presented for 10 s, with a 2-s gray screen displayed before and after each presentation as a baseline period. This resulted in an approximately 4-s gray-screen interval between consecutive images. After each block, participants rested for approximately 2 min, after which the eye tracker was recalibrated.

After the experiment, participants completed a survey on their subjective perceptions of the garden spaces in a quiet environment. The survey included an SD-based aesthetic questionnaire and a semi-structured interview. Considering the characteristics of Chinese classical gardens, the questionnaire consisted of 16 pairs of bipolar semantic items covering four dimensions: perception of spatial structure, landscape imagery, *yijing* experience, and emotional and immersive experiences. It was used to quantify participants' subjective evaluations of the garden spaces^[39]. The interviews focused on participants' descriptions of the overall *yijing* of the garden spaces, key visual elements, and emotional experiences. The questionnaire and interviews were used together to interpret the subjective results of the multimodal experiment and provide a more detailed understanding of participants' perceptions of the garden spaces and their *yijing* experiences.

Data collection and preprocessing

During the experiment, eye-tracking and EEG data were recorded simultaneously while participants viewed the digitally reconstructed scenes. The eye-tracking data were used to analyze

patterns of visual attention, whereas the EEG data were used to analyze changes in neural activity. After the experiment, the eye-tracking and EEG data were preprocessed separately. The main procedures included the removal of invalid data, data segmentation, indicator extraction, and data export. Data preprocessing and statistical analysis were conducted mainly in MATLAB R2023a and with custom Python scripts.

Participants' eye movements were recorded using an aSee Pro F500 desktop eye tracker operating in binocular pupil-tracking mode at 500 Hz. Eye-tracking data were collected and analyzed using aSee Studio 2. To analyze patterns of visual attention while participants viewed different types of digitally reconstructed scenes, four principal eye-tracking indicators were selected: total fixation duration (TFD), first fixation duration (FFD), average fixation duration (AFD), and fixation count (FC). Together, these indicators were used to characterize participants' visual attention during scene viewing^[40]. The definitions of the four indicators are provided in **Table 2**.

Eye-tracking data were preprocessed using the viewing period for each image as the basic unit of analysis. A total of 30 trials were obtained for each participant. To ensure the reliability of the eye-tracking metrics, the proportion of valid gaze samples in each trial was first calculated based on the validity of the gaze coordinates. Trials with a valid gaze-sample proportion below 0.60, or for which the system detected no valid data, were excluded. Trials with a valid gaze-sample proportion between 0.60 and 0.75 were flagged for further review and temporarily retained. To prevent fixation duration from being counted more than once when the same fixation event was recorded repeatedly in the raw data, fixation events were grouped by Fixation Index, and only the portion of each fixation that fell within the 10-s image presentation window was included in the calculation. After metric extraction, the eye-tracking metrics were averaged for each participant within each of the three scene types to compare visual attention across the scene types.

Table 2. Definitions and interpretations of eye-tracking metrics

Metric	Definition	Interpretation
Total Fixation Duration (TFD)	The total duration of all fixation events during image presentation.	Reflects the overall level of visual attention that participants devote to the digitally reconstructed scene images.
First Fixation Duration (FFD)	The duration of the first fixation event after image onset.	Reflects the duration of the initial fixation response after image onset.
Average Fixation Duration (AFD)	The average duration of all fixation events during image presentation.	Reflects the average duration of individual fixations.
Fixation Count (FC)	The number of fixation events during image presentation.	Reflects the frequency of fixation activity and the intensity of visual search.

For EEG data, a Neuracle EEG system was used to simultaneously record participants' EEG signals at a sampling rate of 1,000 Hz while they viewed different reconstructed scenes. EEG data were preprocessed in MATLAB R2023a using the EEGLAB toolbox. The main preprocessing steps included event correction, channel inspection, filtering, artifact removal, data segmentation, and frequency-band power extraction. The theta (4–7 Hz), alpha (8–13 Hz), and beta (13–30 Hz) bands were selected for analysis. These bands are commonly used in studies of visual stimulus processing, environmental perception, and emotional arousal to reflect attentional states, cognitive load, and emotional activation ^[41–43]. The definitions and research interpretations of the EEG indicators are presented in **Table 3**.

Table 3. Principal EEG frequency-band indicators and their interpretations in this study

Indicator	Frequency range (Hz)	General interpretation	Interpretation in this study
Alpha band	8–13	Commonly associated with relaxation, attentional regulation, and relatively low levels of external activation.	Used to reflect differences in relative levels of relaxation, attentional regulation, and visual processing elicited by different reconstructed scenes during viewing.
Beta band	13–30	Commonly associated with attentional engagement, cognitive processing, and task-related activation.	Used to reflect differences in attentional engagement and cognitive processing across different reconstructed scenes during viewing.
Theta band	4–7	Commonly associated with cognitive load, internally directed attention, and memory processing.	Used to provide supplementary evidence of differences in visual information processing and cognitive engagement across different reconstructed scenes.

After the raw EEG data were imported, the image-stimulus events were corrected based on the experimental Mark file, and the onset of each image was defined as 0 ms for data segmentation and feature extraction. During preprocessing, the continuous EEG signals were first inspected for data quality. Auxiliary channels, including ECG and EOG, and EEG channels showing clear abnormalities or poor signal quality were then removed. The data were then band-pass filtered from 1 to 30 Hz and re-referenced to the common average. Independent component analysis (ICA) was subsequently used to identify and remove artifact components associated with eye movements, blinks, and muscle activity. Bad channels identified during preprocessing were interpolated from neighboring electrodes after artifact correction to ensure consistency in the number and spatial configuration of channels across participants. A total of 59 EEG channels were retained for each participant for subsequent analysis.

After preprocessing, the continuous EEG data were segmented based on the image-stimulus events using a time window from −2 to 12 s. The period from 0 to 10 s after image onset was defined as the valid viewing period. Frequency-domain analysis was then performed on the EEG data from each trial. Power in the theta (4–7 Hz), alpha (8–13 Hz), and beta (13–30 Hz) bands was extracted and averaged across all valid channels to obtain trial-level power features for each frequency band. To reduce the influence of residual artifacts, robust z-scores were calculated from the log-transformed power values. A trial was classified as an outlier and excluded from subsequent statistical analysis if the absolute robust z-score in any frequency band exceeded 3.5.

Result

Spatial semantic composition of reconstructed scenes

Based on the semantic segmentation results, the mean values and standard errors of VVI, WVI, SVI, PVI, and BVI were calculated for each scene type. The results showed that the visual characteristics of the three scene types were consistent with the patterns observed in the semantic segmentation results. The architectural scenes had a mean VVI of 0.280 ± 0.080 . In this study, grass was included in the calculation of VVI to reflect the overall visual presence of vegetation. Compared with the other two scene types, the architectural scenes had a relatively high PVI (0.196 ± 0.053) but a very low WVI (0.003 ± 0.003), reflecting the visual prominence of built elements. The Mountain-water scenes were dominated by natural elements, with mean VVI, WVI, and SVI values of 0.389 ± 0.010 , 0.324 ± 0.016 , and 0.274 ± 0.020 , respectively. PVI and BVI were both very low ($PVI < 0.001$; $BVI = 0.007 \pm 0.002$), indicating the visual dominance of natural elements and relatively similar proportions among the three main natural indicators. The vegetation scenes had relatively high VVI and SVI values ($VVI = 0.548 \pm 0.090$; $SVI = 0.274 \pm 0.046$). Although the indicator values showed some variation, natural elements remained visually dominant. The error bars represent the standard error (SE) and show the variability among samples within each scene

type (**Fig. 5**). Overall, the three scene types exhibited distinct visual compositions: the architectural scenes were dominated by built elements, the Mountain-water scenes were characterized by a combination of water and vegetation, and the vegetation scenes were dominated by vegetation.

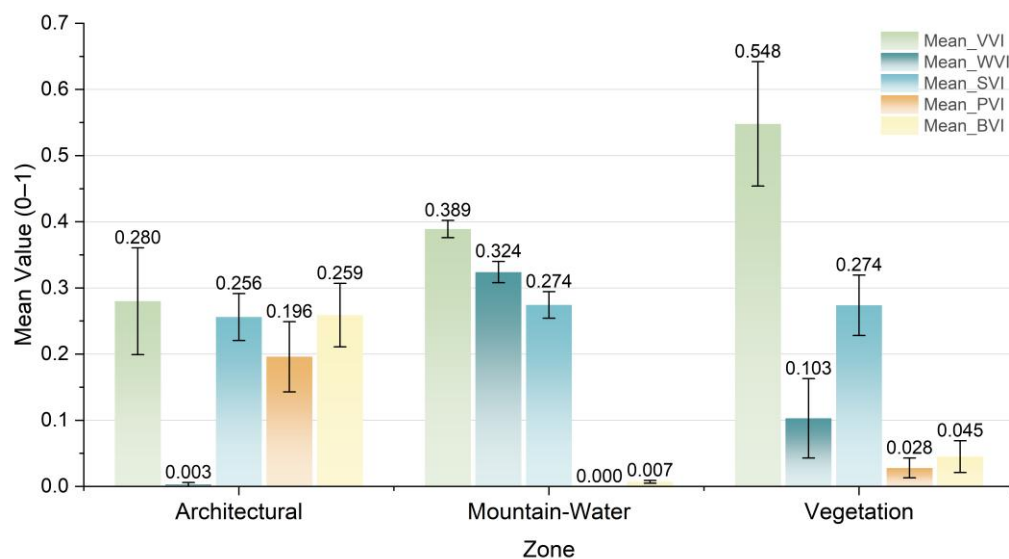


Fig.5 Mean Visual Indicator Values ($\pm$ SE) Across Landscape Zones (Normalized 0 - 1)

Objective perceptual responses across scene types

Based on the semantic segmentation results, which revealed differences in visual composition among the three types of digitally reconstructed scenes, this study further examined whether the different scene types elicited measurable differences in perceptual responses using eye-tracking and EEG indicators. The two sets of objective indicators were compared across scene types to determine whether differences in visual composition were accompanied by differences in participants' perceptual responses.

Eye-tracking data from all 48 participants were extracted within the predefined time window and subjected to quality screening. The professional and general groups each comprised 24 participants. Each participant completed 30 trials, yielding a total of 1,440 trials. After quality control, 36 trials were excluded because their proportion of valid eye-tracking samples was below 0.60 or because they contained no valid samples. A total of 1,404 trials were retained for subsequent analysis, representing a trial retention rate of 97.5%. Among the retained trials, the mean proportion of valid eye-tracking samples was 92.3%. Of these, 466, 471, and 467 trials corresponded to the architectural, Mountain-water, and vegetation scenes, respectively.

Descriptive statistics showed that the mean FC values for the architectural, Mountain-water, and vegetation scenes were 14.73 ± 1.90 , 14.78 ± 1.83 , and 14.86 ± 2.16 , respectively. The values were similar across the three scene types, indicating only small differences in overall fixation frequency. The mean TFD was slightly higher for the Mountain-water scenes (5.94 ± 0.88 s) than for the architectural scenes (5.80 ± 0.90 s) and vegetation scenes (5.79 ± 1.03 s). The mean FFD was also higher for the Mountain-water scenes (0.45 ± 0.26 s) than for the architectural scenes (0.43 ± 0.29 s) and vegetation scenes (0.40 ± 0.20 s). The mean AFD values were 0.49 ± 0.21 s for the Mountain-water scenes, 0.48 ± 0.24 s for the vegetation scenes, and 0.47 ± 0.22 s for the architectural scenes. Overall, FC varied little among the three scene types, whereas the Mountain-water scenes showed slightly higher TFD, FFD, and AFD values. Nevertheless, the differences in

all four eye-tracking indicators were small (**Fig. 6**).

A 2×3 mixed-design analysis of variance (ANOVA) was further conducted to examine differences in the eye-tracking metrics. The results showed that FC, TFD, FFD, and AFD had no significant main effects of Group or SceneType and no significant Group $\times$ SceneType interactions ($p > 0.05$). Among the four eye-tracking metrics, the main effect of SceneType on TFD had the largest effect size, although it did not reach statistical significance, $F(2, 92) = 1.94, p = 0.149, \eta_p^2 = 0.041$. Together with the descriptive results, these findings indicated that the three types of digitally reconstructed scenes differed only slightly in TFD, FFD, AFD, and FC, with none of these differences reaching statistical significance.

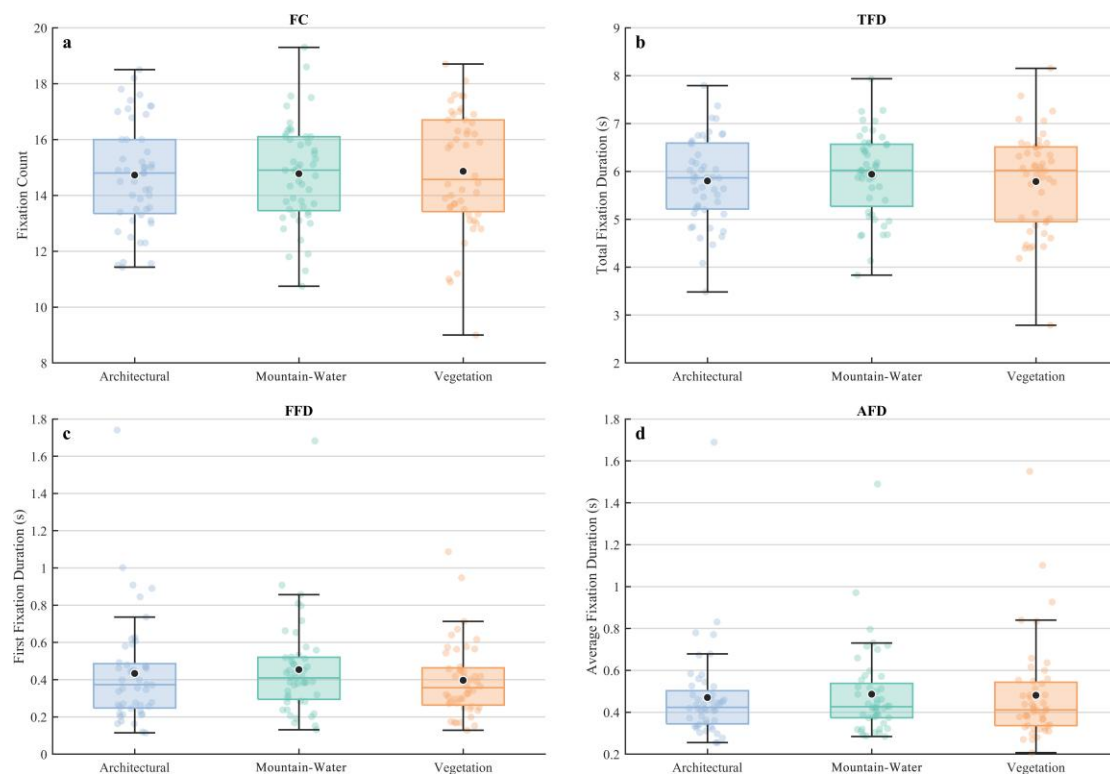


Fig.6 Eye-tracking responses across reconstructed scene types.

Following the overall analysis of the eye-tracking metrics, fixation heat maps were generated for the 15 reconstructed scene images using valid fixation data to examine the spatial distribution of visual attention across the three scene types (**Fig. 7**). The heat maps showed that fixation hotspots were mainly concentrated in areas with clear changes in direction, boundary transitions, and spatial layering. This result was consistent with previous studies on visual attention, which have shown that fixation locations are generally shaped by a combination of local saliency, object features, and scene meaning rather than by a single visual element^[44-45]. In the architectural scenes, hotspots were mainly distributed around clearly defined architectural forms, including the main building structures, doors, and windows, indicating attention to recognizable structures and cues of spatial order. Although the Mountain-water scenes contained large areas of water, hotspots were more often located along waterfront boundaries, distant buildings, and vegetation edges. This pattern suggests that continuous water surfaces do not necessarily attract dense fixations, whereas boundary transitions and landscape nodes are more likely to become visual focal points. In the vegetation scenes, hotspots were relatively dispersed and occurred mainly along vegetation edges, around distant buildings, and in locally enclosed or open spaces. This distribution reflected attention to local

visual differences and changes in spatial layering. The heat-map results showed that, although the three reconstructed scene types differed little in the overall eye-tracking metrics, their fixation hotspots were unevenly distributed and tended to concentrate in areas with clear boundaries, distinct spatial structures, and pronounced changes in spatial layering.

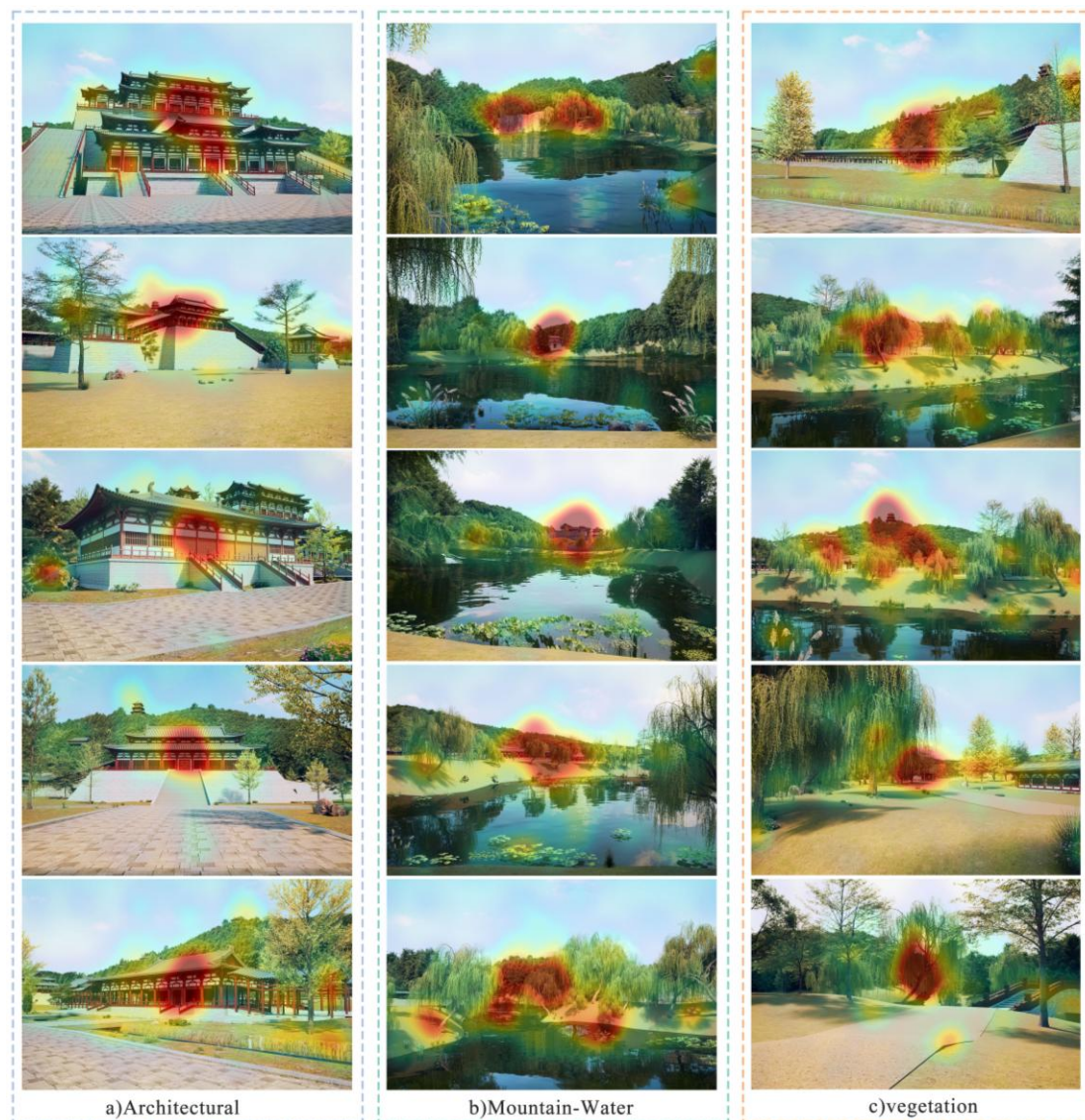


Fig.7 Fixation heatmaps of the 15 reconstructed scene images

Following the eye-tracking analysis, EEG data were analyzed to examine whether neurophysiological responses during viewing differed among the three types of digitally reconstructed scenes. After preprocessing and the exclusion of outlier trials, a total of 1,321 valid trials from 48 participants were retained. The EEG trials were grouped by scene type into architectural, Mountain-water, and vegetation scenes, and power in the theta, alpha, and beta bands was extracted separately for the image-viewing period. To reduce the influence of trial-level fluctuations, mean power in each frequency band was calculated for each participant and scene type. A 2×3 mixed-design ANOVA was then conducted to examine the effects of Group, SceneType, and their interaction on EEG band power.

The results showed that neither the main effect of Group nor the Group $\times$ SceneType interaction was significant in any frequency band. This finding indicated that the professional and general

groups did not differ significantly in overall band-power levels or in their response patterns across scene types. The main effect of SceneType (**Table 4**) was significant in the alpha band, approached significance in the theta band, and was not significant in the beta band. Bonferroni-corrected post hoc comparisons showed that alpha-band power was significantly higher in the Mountain-water scenes than in the vegetation scenes, whereas the architectural scenes did not differ significantly from either of the other two scene types. Across the three frequency bands, the effect of scene type was most evident in the alpha band, specifically in the difference between the Mountain-water and vegetation scenes. **Fig. 8** presents the distributions of log-transformed theta-, alpha-, and beta-band power for the two participant groups across the three types of digitally reconstructed scenes.

Regarding the interpretation of the EEG indicators, alpha-band power has generally been associated with the regulation of visual attention, neural inhibition, and relatively relaxed perceptual states. In this study, the higher alpha-band power observed in the Mountain-water scenes suggested a more stable and relatively relaxed neural state during viewing. In relation to scene composition, water bodies, rocks, variations in spatial enclosure and openness, and layered near-to-distant views in the Mountain-water scenes jointly provided relatively coherent spatial organization cues. In contrast, the vegetation scenes were dominated by continuous natural textures and contained less distinct spatial layering. The EEG results showed differences between the two natural scene types at the neurophysiological level. Among the three reconstructed scene types, the Mountain-water scenes elicited a more stable viewing state and a stronger immersive experience.

Table 4. SceneType effects on EEG spectral power across reconstructed scene types

Frequency band	SceneType effect	p	η_p^2	Bonferroni post-hoc comparison
Theta	$F(2, 92) = 2.48$	0.089	0.051	n.s.
Alpha	$F(2, 92) = 3.97$	0.022	0.080	Mountain-Water > Vegetation, $t(47) = 3.26$, $p = 0.006$, Cohen's $d_z = 0.47$
Beta	$F(2, 92) = 1.36$	0.261	0.029	n.s.

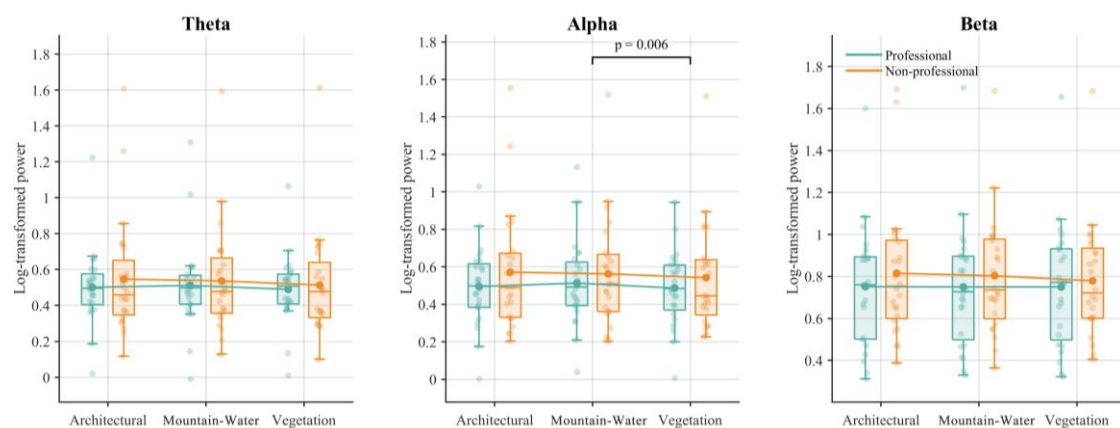


Fig.8 EEG spectral power across reconstructed scene types and participant groups.

Subjective evaluation based on SD questionnaire and interviews

This study further combined the SD questionnaire with semi-structured interviews to examine differences in subjective evaluations across the three reconstructed scene types. Together, these methods provided a more comprehensive understanding of participants' perceptual and cognitive responses and strengthened the evidence base for the multimodal perceptual evaluation. Based on the SD questionnaire responses, Cronbach's α coefficients were calculated for each of four SD dimensions across the 15 stimulus images: perception of spatial structure, perception of landscape imagery, *yijing* experience, and emotional and immersive experience. The mean Cronbach's α

values for the four dimensions were 0.827, 0.846, 0.783, and 0.861, respectively, indicating good internal consistency across all four dimensions. Mean scores were then calculated for each stimulus image in each dimension and aggregated by scene type for comparison among the architectural, Mountain-water, and vegetation scenes (**Fig. 9**). The distribution of mean scores revealed clear differences in participants' subjective evaluations across the three reconstructed scene types. Participants gave the most positive overall evaluations to the Mountain-water scenes, particularly for perception of landscape imagery ($M = 4.068$, $SD = 0.612$) and emotional and immersive experience ($M = 4.064$, $SD = 0.575$). The vegetation scenes received intermediate overall ratings, with relatively high scores for perception of spatial structure ($M = 3.820$, $SD = 0.615$) and emotional and immersive experience ($M = 3.747$, $SD = 0.669$). The architectural scenes received a moderate rating for perception of spatial structure ($M = 3.643$, $SD = 0.543$) but the lowest rating for perception of landscape imagery ($M = 2.997$, $SD = 0.602$).

To examine differences among the three reconstructed scene types, repeated-measures ANOVAs were conducted for each of the four dimensions. The results showed significant differences among the scene types in perception of spatial structure, perception of landscape imagery, and emotional and immersive experience. Among these dimensions, perception of landscape imagery showed the largest effect size, suggesting that this dimension was most strongly affected by scene type. No significant difference was found in *yijing* experience, suggesting relatively similar evaluations across the three scene types in this dimension. Bonferroni-adjusted pairwise comparisons showed that Mountain-water scenes scored significantly higher than both vegetation and architectural scenes for perception of landscape imagery and emotional and immersive experience. They also scored significantly higher than the architectural scenes for perception of spatial structure (**Table 5**).

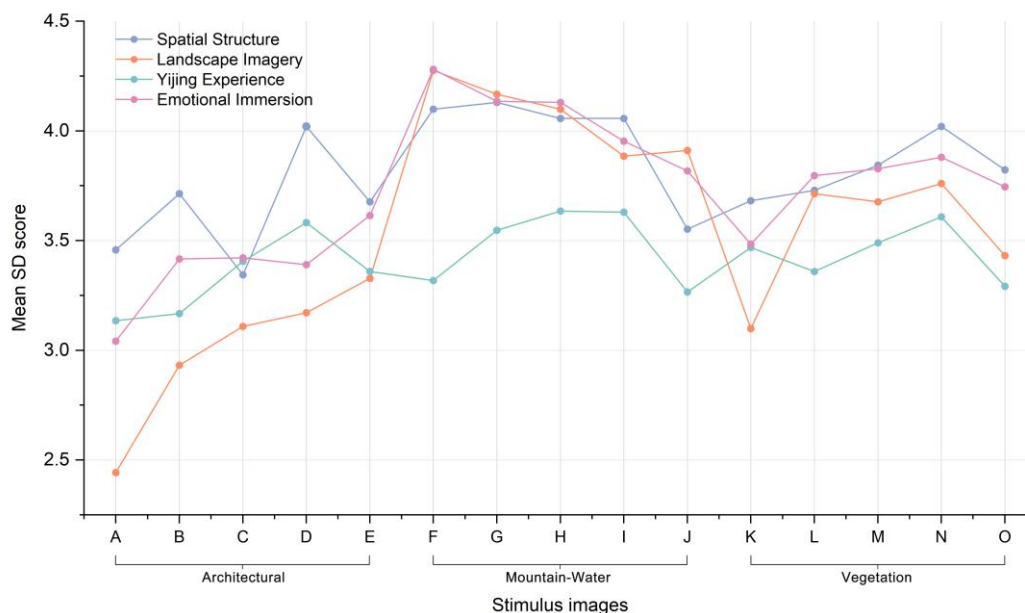


Fig.9 Subjective evaluation scores of the 15 reconstructed scene images across four SD dimensions.

Table 5. Repeated-measures ANOVA results of SD scores across scene types.

Dimension	$F(df)$	p	η_p^2	Pairwise comparison
Spatial Structure	$F(1.68, 79.06) = 7.131$	0.003	0.132	Mountain-Water > Architectural
Landscape imagery	$F(1.78, 83.82) = 69.335$	<0.001	0.596	Mountain-Water > Vegetation > Architectural
<i>Yijing</i> Experience	$F(1.56, 73.21) = 1.736$	0.189	0.036	n.s.

Dimension	$F(df)$	p	η_p^2	Pairwise comparison
Emotional immersion	$F(1.66, 77.85) = 28.829$	<0.001	0.380	Mountain-Water > Vegetation > Architectural

To further explain the differences in subjective evaluations revealed by the SD questionnaire, semi-structured interviews were conducted. The interviews focused on participants' overall impressions of the space, the landscape elements that attracted the most attention, the historical atmosphere and *yijing* of imperial gardens, and their emotions and associations during viewing. Compared with the quantitative scores, the interview responses provided more direct insight into the specific aspects considered by participants when evaluating the reconstructed scenes. The results showed that participants' responses focused mainly on architectural forms, mountain-water and vegetation elements, a classical atmosphere, and a sense of involvement during viewing (Table 6). The Mountain-water scenes more readily evoked associations with garden imagery and a sense of involvement during viewing. The architectural scenes were associated mainly with spatial order and the recognition of architectural forms, whereas the vegetation scenes were more strongly associated with a sense of natural comfort. The interview results indicated that participants' subjective evaluations were shaped primarily by perceived spatial form, landscape composition, and viewing experience.

Table 6. Main perceptual cues extracted from semi-structured interviews.

Interview focus	Main perceptual cues	Representative keywords
Spatial impression	Architectural order	Open space; regularity; orderliness; grandeur; solemnity
Landscape elements	Landscape and natural features	Water; reflections; vegetation; trees; grass
Historical atmosphere	Classical and cultural association	Architectural forms; colors; Mountain-water garden elements; imperial dignity; classical charm
Emotion and association	Viewing experience	Relaxation and comfort; calmness and tranquility; touring and leisure; associations with ancient life

Associations between spatial semantic features and multimodal perceptual responses

To analyze the relationships between the visual semantic features of the reconstructed scenes and multimodal perceptual responses, Spearman's rank correlation analysis was conducted at the image level, with the 15 reconstructed images serving as the units of analysis (Fig. 10). The visual semantic indicators for each image were obtained directly from the semantic segmentation results, whereas image-level means for the eye-tracking, EEG, and SD indicators were calculated from valid data collected from the 48 participants. This analysis was used to identify visual semantic indicators that were significantly associated with multimodal perceptual responses.

The correlation analysis showed that, among the SD dimensions, WVI was significantly positively correlated with perception of landscape imagery ($\rho = 0.84, p < 0.001$), emotional and immersive experience ($\rho = 0.80, p < 0.001$), and perception of spatial structure ($\rho = 0.54, p = 0.036$). These results indicated that images with higher proportions of water generally received more positive evaluations across all three dimensions. BVI showed significant negative correlations with perception of spatial structure ($\rho = -0.66, p = 0.007$), perception of landscape imagery ($\rho = -0.92, p < 0.001$), and emotional and immersive experience ($\rho = -0.88, p < 0.001$). PVI also showed significant negative correlations with the same three dimensions, with coefficients of $\rho = -0.59$ ($p = 0.021$), $\rho = -0.65$ ($p = 0.009$), and $\rho = -0.71$ ($p = 0.003$), respectively. Together, these results indicated that images with higher proportions of buildings and pedestrian spaces received lower evaluations of spatial structure, landscape imagery, and emotional and immersive experience. Overall, the visual semantic indicators were clearly associated with the SD evaluations. These

findings support the subjective evaluation results presented above and show that, in digitally reconstructed garden scenes, spatial semantic compositions characterized by water and natural landscape elements are more conducive to enhancing landscape imagery and immersive experience.

Compared with the SD indicators, the eye-tracking metrics showed only one significant correlation: BVI showed a significant negative correlation with TFD ($\rho = -0.54, p = 0.038$). Apart from this relationship, TFD, FFD, AFD, and FC did not vary consistently as the area proportion of any single semantic element increased, indicating that viewers' overall fixation behavior was not directly determined by the proportion of a single element. When these findings were considered together with the heat-map results presented above, fixation hotspots were more concentrated along building boundaries, waterfront interfaces, distant landscape nodes, vegetation edges, and locations with pronounced changes in spatial hierarchy. This pattern indicated that visual attention was jointly influenced by local boundary relationships, spatial organization, and visual points of interest. The EEG results showed that SVI was significantly positively correlated with theta-band power ($\rho = 0.57, p = 0.027$), indicating that sky visibility and the spatial openness it reflects were important semantic features associated with spatial processing states during viewing. Alpha- and beta-band power showed no significant correlations with any spatial semantic indicators, indicating that no stable correspondence existed between power in these two frequency bands and the area proportions of the visual elements.

Overall, WVI, BVI, and PVI were primarily associated with subjective evaluations of garden experience, whereas the EEG results showed that SVI was mainly associated with theta-band power. Among the correlations between the visual semantic indicators and eye-tracking metrics, only the correlation between BVI and TFD was significant. The different modalities provided complementary evidence at the levels of subjective evaluation, visual attention, and neural responses. These findings indicated that the perceptual experience of digitally reconstructed scenes cannot be determined by a single semantic element alone but should instead be comprehensively assessed by integrating spatial semantic composition with multimodal perceptual outcomes. Therefore, identifying key visual elements associated with evaluations of digitally reconstructed garden experiences from an image-semantic perspective can provide a quantitative basis for the subsequent optimization of the visual organization of reconstructed scenes.

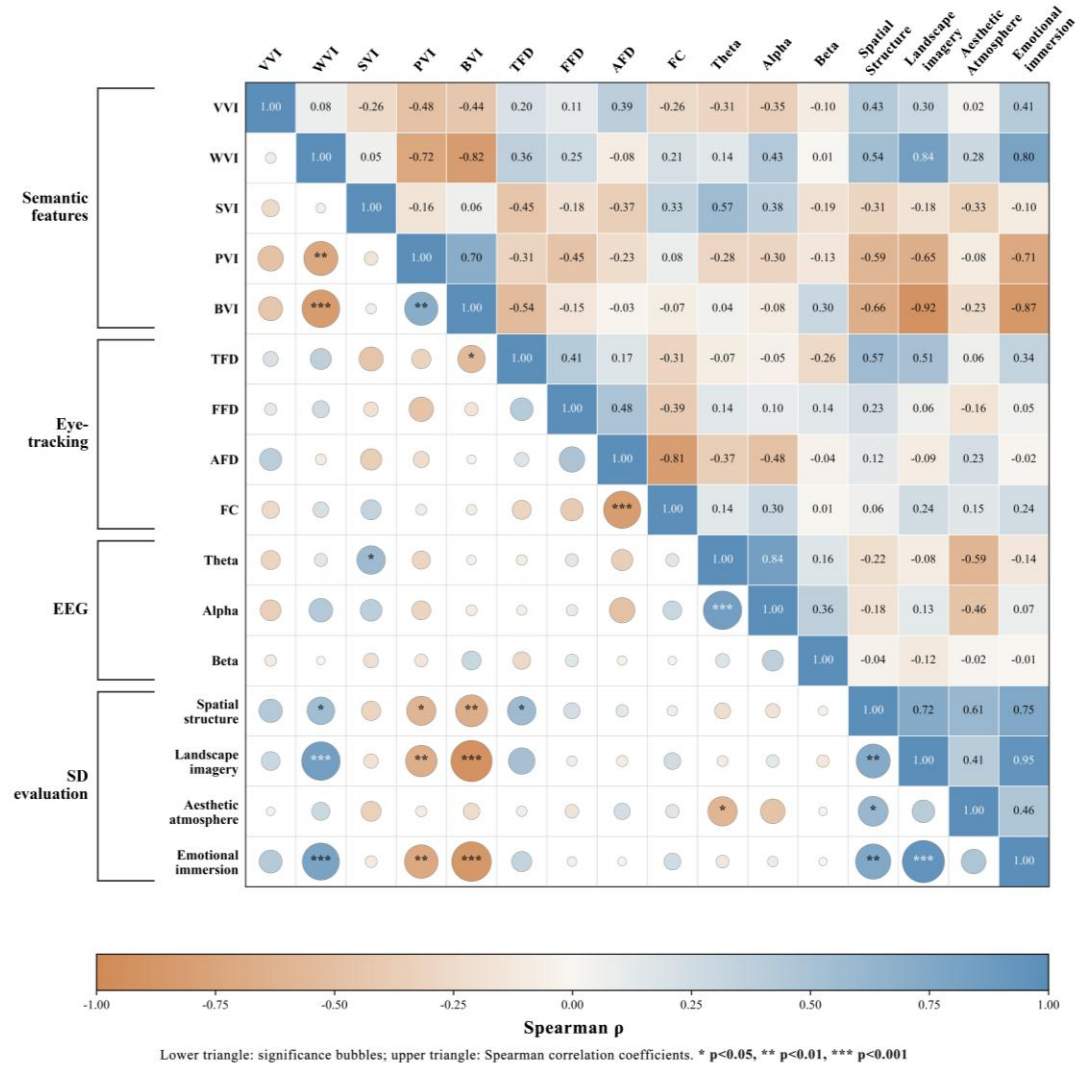


Fig.10 Spearman correlation matrix of spatial semantic features and multimodal perceptual indicators.

Immersive validation of the reconstructed spatial experience

In this study, an immersive virtual reality experiment was conducted using a Meta Quest 3 headset (Fig. 11). Three representative nodes corresponding to the viewpoints used in the static images were selected from each of the architectural, Mountain-water, and vegetation scene types, resulting in nine fixed viewpoints in total. A total of 12 participants took part in the experiment, including six who had participated in the preceding multimodal perception experiment and six who had not taken part in the main experiment. After viewing the scenes, participants completed an immersive experience questionnaire adapted from relevant dimensions of the Igroup Presence Questionnaire (IPQ) [46] to evaluate the reconstructed scenes in terms of presence, spatial perception, touring and viewing experience, and *yijing* experience.

The questionnaire results showed that the 12 participants generally evaluated the digitally reconstructed panoramic scenes of Jiucheng Palace positively. On the five-point scale, the mean scores were 4.83 ± 0.31 for immersive presence and spatial presence, 4.78 ± 0.30 for perception of the reconstructed space and touring and viewing experience, and 4.78 ± 0.26 for aesthetic and *yijing* experience, with an overall mean of 4.80 ± 0.24 . Of all responses to the questionnaire items, 97.8%

were rated as “agree” or “strongly agree,” indicating that the immersive scenes received positive evaluations in terms of presence, spatial perception, and aesthetic experience.

The open-ended responses showed that the scenes that made the strongest impressions on participants were mainly those featuring water bodies, lakeside buildings, and large-scale architectural nodes. Most participants reported that the immersive experiment was more conducive to developing a sense of spatial immersion and scale and gaining a better understanding of the spatial relationships among different landscape nodes. Participants also made suggestions for optimizing dynamic effects, daylight intensity, material authenticity, vegetation layering, and path continuity. These findings indicate that the current reconstructed scenes performed well in terms of spatial immersion, although the level of detail and the continuity of the touring and viewing experience require further improvement.

To evaluate the immersive scenes from a professional perspective, three experts in cultural heritage digitization were invited to assess the scenes in terms of spatial configuration, landscape organization, and the representation of historical atmosphere. The experts considered that the current scenes reflected the overall characteristics of the reconstructed space of Jiucheng Palace and presented relatively clear organizational relationships among architectural, mountain–water, and vegetation elements. They also recommended further improvements to architectural details, path continuity, vegetation layering, and material representation. The immersive virtual reality experiment further supported the conclusion that the digitally reconstructed scenes of Jiucheng Palace performed well in terms of spatial perception, landscape imagery, and immersive experience.



Fig.11 Immersive virtual reality experiment

Integrated perceptual evaluation framework for reconstructed garden scenes

Based on the empirical results presented above, the perceptual characteristics reflected by the different modalities were summarized as four core modules: support for landscape imagery, spatial openness and hierarchy, organization of visual attention, and immersive experience performance. With presentation suitability serving as the integrated output, an evaluation pathway was established, moving from the integration of perceptual evidence to judgments regarding the optimization of the reconstructed scenes. Each module corresponds to different types of experimental indicators and can provide a basis for selecting presentation viewpoints, organizing spatial hierarchy and sightlines, developing image compositions, and optimizing landscape nodes in the reconstructed scenes. The

integrated perception evaluation and optimization matrix is presented in **Table 7**.

Table 7. Integrated perceptual evaluation and optimization matrix for reconstructed garden scenes.

Evaluation module	Basis for evidence integration	Evaluation output	Applications to reconstruction optimization
Landscape imagery support	Positive association of WVI with landscape imagery scores; negative associations of BVI and PVI with these scores	Extent to which semantic composition supports the perception of landscape imagery	Selection of presentation viewpoints; adjustment of the water visibility proportion; control of building and paved-space interfaces
Spatial openness and hierarchy	Positive correlation between WVI and perception of spatial structure; negative correlations of BVI and PVI with perception of spatial structure; positive correlation between SVI and theta-band power	Openness, sense of depth, and spatial hierarchy	Openness of the visual field, sightline corridors, and boundary organization
Organization of visual attention	Heat maps and eye-tracking metrics; single significant correlation between BVI and TFD; boundary and node characteristics	Visual focal points and viewing paths	Focal-point enhancement, boundary transitions, node presentation, and image composition
Immersive experience performance	Ratings of emotional and immersive experience; differences in alpha-band power; interview results and immersive validation	Stable and positive experiential responses	Landscape continuity, natural elements, and scene atmosphere
Presentation suitability	Semantic indicators, SD scores, eye-tracking metrics, EEG indicators, and interview results	Relative perceptual performance of different scenes	Selection of images for public presentation and optimization of visual organization

Based on the evaluation matrix described above, this study further developed two quantitative measures: (1) the Perceived Garden Experience (PGE), which characterizes viewers' overall subjective judgments of each reconstructed scene; and (2) the semantic-level Perceptual Effectiveness Index (PEI), which reflects the extent to which the spatial semantic composition supports the perception of landscape imagery.

The PGE was first calculated from the four evaluation dimensions of the SD questionnaire. Because all four dimensions used the same rating scale, their scoring directions were first standardized, and the mean score across the four dimensions was then used as the PGE value for each reconstructed scene:

$$PGE_i = \frac{1}{4} \sum_{k=1}^4 D_{ik} \quad (1)$$

Here, PGE_i denotes the overall Perceived Garden Experience value for reconstructed scene i , and D_{ik} denotes the mean score for scene i on SD evaluation dimension k . A higher PGE value indicates a more positive overall evaluation of the garden experience associated with that reconstructed scene.

On this basis, the PEI was constructed to further explain the relationship between the perception of landscape imagery and spatial semantic composition. The preceding analysis showed that WVI was positively correlated with landscape imagery scores, whereas BVI and PVI were negatively correlated with these scores. This result indicated relatively clear correspondences between semantic elements such as water, buildings, and paved spaces and the perception of landscape imagery. Therefore, the PEI incorporates WVI, BVI, and PVI as semantic-level indicators to characterize the extent to which visual semantic composition supports the perception of landscape imagery.

Before the PEI was calculated, WVI, BVI, and PVI were normalized using the min–max method:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (2)$$

Here, X' denotes the normalized indicator value, X denotes the original indicator value, and $X_{\min}$ and $X_{\max}$ denote the minimum and maximum values of the indicator within the sample, respectively. The weights were then determined based on the absolute values of the Spearman's rank correlation coefficients between the semantic indicators and the landscape imagery scores:

$$w_j = \frac{|\rho_j|}{\sum |\rho_j|} \quad (3)$$

Here, w_j denotes the weight of semantic indicator j , and ρ_j denotes the Spearman correlation coefficient between that indicator and the landscape imagery score. Given that WVI was positively correlated with landscape imagery scores, whereas BVI and PVI were negatively correlated with these scores, the PEI is expressed as follows:

$$PEI_i = w_1 \times WVI'_i + w_2 \times (1 - BVI'_i) + w_3 \times (1 - PVI'_i) \quad (4)$$

Based on the strength of the correlation of each of WVI, BVI, and PVI with the landscape imagery scores in the study sample, the weights assigned to the three indicators were 0.349, 0.381, and 0.270, respectively. The formula can therefore be further written as follows:

$$PEI_i = 0.349 \times WVI'_i + 0.381 \times (1 - BVI'_i) + 0.270 \times (1 - PVI'_i) \quad (5)$$

Because the PEI was calculated using min – max normalization within the study sample, its value reflects the relative performance of the reconstructed scenes. Based on the method described above, the PEI values for the three reconstructed scene types were, in descending order, 0.960 for the Mountain–water scenes, 0.685 for the vegetation scenes, and 0.237 for the architectural scenes. The higher PEI value for the Mountain-water scenes was mainly associated with their higher water visibility and lower proportions of buildings and pedestrian spaces, indicating that, within the study sample, this type of semantic composition provided stronger support for the perception of landscape imagery. PGE and PEI characterize, respectively, viewers' overall garden experience in the reconstructed scenes and the extent to which semantic composition supports the perception of landscape imagery. Together with the evaluation matrix, these two measures constitute an integrated perception evaluation framework. As indicated by the evaluation matrix, the perceptual evaluation of digitally reconstructed garden scenes requires consideration not only of visual semantic composition but also of multiple aspects, including spatial hierarchy, visual attention, and immersive experience. This framework can provide a quantitative reference for selecting presentation viewpoints, optimizing image composition, and evaluating the perceptual effects of reconstructed scenes.

Discussion

Focusing on the digitally reconstructed scenes of Jiucheng Palace, this study established an evaluation framework for the digital reconstruction of historic garden sites that integrates visual semantics with multimodal perception. The comprehensive evaluation matrix, PGE, and PEI were used to compare the perceptual performance of different reconstructed scenes. The results showed that different types of digitally reconstructed scenes of Jiucheng Palace exhibited different patterns of variation in spatial structure recognizability, perception of landscape imagery, emotional experience, and immersive experience, and that visual semantic features were differentially associated with viewers' subjective evaluations and objective perceptual indicators. These findings

indicate that the perceptual effects of digitally reconstructed scenes are related not only to spatial form but also to the organization of landscape elements and their visual expression.

Among the three scene types, the Mountain-water scenes elicited relatively positive perceptual responses at both the subjective and neurophysiological levels. Specifically, Mountain-water scenes received higher ratings for the perception of landscape imagery and emotional and immersive experience, and the alpha-band power elicited by these scenes was also significantly higher than that elicited by the vegetation scenes. These results indicated differences between the two types of natural scenes in both subjective experience and neural responses. Research on blue spaces in environmental psychology provides a relevant explanation for the higher subjective evaluations of the Mountain-water scenes. By comparing images of natural and built environments both with and without water, White et al. (2010)^[47] found that the presence of water significantly enhanced environmental preference, positive affect, and perceived restorativeness.

Traditional garden-making theory likewise emphasizes the integral relationship between water and site organization. In his Ming-dynasty treatise *Yuanye*, Ji Cheng stated: “In selecting a site for construction, proximity to water is highly valued; before laying the foundations, one should first investigate the source, trace the course of the water, and ascertain its origin”^[48]. This passage identifies the water source, water flow, and their relationship with the site as important considerations in garden site selection and spatial layout, reflecting the fundamental role of water in the spatial creation of traditional gardens. It also provides a historical context for understanding the related perceptual results observed in the digitally reconstructed scenes of Jiucheng Palace. Compared with the vegetation scenes, which were dominated by continuous vegetation textures, the Mountain-water scenes contained water, rocks, and vegetation, together with more pronounced transitions between enclosure and openness and between near and distant views. This indicates that the contribution of water to garden imagery lies not merely in the addition of another natural element but also in the organization of relationships among natural scenery, built elements, and visual layers within the image. This composite configuration of blue-green elements and spatial hierarchy enabled viewers to more readily recognize the scenes as garden spaces characterized by mountain-water relationships and qualities of touring and viewing. Such recognition corresponded to the higher subjective evaluations and alpha-band power observed for the Mountain-water scenes.

Further analysis showed relatively clear associations between visual semantic features and subjective evaluations. By contrast, significant associations involving the eye-tracking and EEG indicators were largely limited to the association between BVI and TFD and that between SVI and theta-band power. The limited differentiation in objective perceptual responses can be explained from two perspectives: the similarity of the stimulus materials and the way in which the indicators were aggregated. All three stimulus types were derived from the same digital reconstruction system of Jiucheng Palace and shared consistent scene styles, presentation methods, and viewing tasks; consequently, the overall viewing patterns across the different scenes were relatively similar. Moreover, the aggregation of the indicators incorporated local or short-term variations into overall mean values, thereby reducing the differences among the scene types. The eye-tracking heat maps showed that visual attention was concentrated mainly around boundaries, nodes, and spatial transitions, indicating that differences among the scenes were reflected more strongly in fixation locations and local spatial organization than in the overall level of fixation activity. Previous research likewise found no corresponding relationships between landscape preference and either fixation count or average fixation duration, indicating that overall eye-tracking metrics cannot

directly represent viewers' preference judgments^[49]. The EEG results further showed a significant positive correlation between SVI and theta-band power, indicating that images with higher sky visibility generally corresponded to higher theta-band power. This finding suggests that spatial enclosure and openness, together with overall visual organization, may be important factors influencing neural responses. Meanwhile, subjective evaluations and objective perceptual indicators showed different patterns of variation, indicating that the perception of digitally reconstructed scenes involves multiple levels, including visual attention, neural responses, and overall experience. Eye tracking and EEG recorded immediate responses during viewing, whereas the SD questionnaire reflected the integrated judgments formed after viewing. Therefore, scene characteristics that elicit differences in subjective evaluations do not necessarily produce corresponding changes in overall fixation measures or across multiple EEG frequency bands. Although the modalities differed in their sensitivity to scene differences, they provided complementary evidence regarding visual organization, neural responses, and subjective experience, thereby supporting an integrated assessment of the perceptual effects of digitally reconstructed garden scenes.

These differences further reveal two interrelated but distinct levels in the evaluation of digital reconstructions of historic garden sites: the recognizability of spatial structure and the effectiveness of experience conveyance. Although the architectural scenes conveyed basic spatial order relatively well, the perception of landscape imagery in these scenes was relatively weak, indicating that clear architectural forms and spatial boundaries did not translate into a corresponding garden experience. By contrast, the spatial hierarchy and visual cues formed by water, rocks, vegetation, and open views in the Mountain-water scenes were more conducive to the formation of landscape imagery and immersive experience. Therefore, beyond historical evidence and morphological accuracy, digital reconstruction also needs to examine whether reconstructed scenes can effectively convey their spatial relationships and experiential characteristics.

The comprehensive evaluation matrix, together with the PGE and PEI, can be used to compare the overall perceptual performance of different scenes, analyze the relationship between spatial semantics and experiential evaluations, and translate the multimodal analysis results into feedback for adjusting the presentation and expression of reconstruction outcomes. On this basis, the presentation of the reconstructed scenes can be optimized by focusing on the proportional relationship between natural landscape elements and built elements, the organization of visual boundaries and nodes, and variations in spatial enclosure and openness and the hierarchy between near and distant views. Without altering the historical basis or reconstructed content, adjustments to viewpoints and image composition can improve the legibility of key landscape elements and spatial relationships.

In summary, the evaluation of digital reconstruction can follow an iterative pathway of "evidence-based historical modeling - expert validation - integrated perception evaluation - revalidation and presentation optimization." Evidence-based historical modeling and expert validation ensure the reliability of the reconstructed content; integrated perception evaluation and revalidation provide feedback for subsequent presentation optimization, enabling digital reconstruction outcomes to account for both historical plausibility and perceptual effectiveness. By introducing multimodal perceptual evidence on the basis of historical research and expert validation, this framework extends the scope of evaluation from morphological accuracy and model precision to spatial expression and experience conveyance, forming a practical and reproducible evaluation method that provides a methodological basis for assessing reconstruction outcomes, optimizing their

presentation, and further validating and applying the framework to other historic garden sites.

This study has some limitations regarding the experimental presentation method and sample scope. The main experiment used static reconstructed images presented from fixed viewpoints, which facilitated control over the stimulus conditions but could not fully represent the continuous movement, viewpoint transitions, and route choices involved in actual touring and viewing. Although the immersive experiment supported the results of the static-image experiment to some extent, the number of participants and test nodes remained limited. In addition, PGE and PEI were used for integrated comparisons among the reconstructed scenes examined in this study, and the PEI weights were calculated based on the current set of 15 images and the participant data. The evaluation framework can be extended; however, the stability of its specific parameters requires further examination using larger samples and cross-case validation. Future research could incorporate dynamic walkthroughs, VR interaction, and different historic garden sites to further calibrate the multimodal evaluation method and its associated parameters.

Human Ethics and Consent to Participate

The study was conducted in accordance with the ethical principles of the Declaration of Helsinki and was approved by the Medical Ethics Committee of Northwest University. All participants in this study were adults. Before the experiment, the researchers explained the study objectives, experimental procedures, and the data to be collected to the participants, who voluntarily participated after providing verbal informed consent. All data collected were de-identified. The results are presented only in aggregate form and contain no information that could identify individual participants. The data were used solely for this study and related academic publications.

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgment

This study was supported by the National Natural Science Foundation of China (Grant Nos. 62572394 and 62372371). The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Author Contributions

J.G.: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Software, Visualization, Writing - original draft; C.W.: Conceptualization, Methodology, Supervision, Writing -review & editing; Y.X.: Resources, Methodology, Investigation; J.L.: Supervision, Writing - review & editing; J.Z.: Investigation, Data curation.

Competing interests

The authors declare no competing financial or non-financial interests.

References

- [1] Keswick, M. *The Chinese Garden: History, Art and Architecture*. (Harvard University Press, 2003).
- [2] Wu, J.Y., Xiao, Y. Viewing Chinese gardens from the perspective of traditional Chinese culture. *Chinese Landscape Architecture* **3**, 85–87 (2001) (in Chinese).
- [3] Gu, G.C. On the artistic conception of Chinese classical gardens. *Chinese Landscape Architecture* **30**, 17–21 (2014)(in Chinese).

- [4] Yang, X.Z. A study on gardening characteristics of imperial gardens of Sui and Tang dynasties. Master's thesis, Xi'an University of Architecture and Technology. <https://doi.org/10.27393/d.cnki.gxazu.2019.001705> (2019) (in Chinese).
- [5] Liu, S., Bin Mamat, M.J. Application of 3D laser scanning technology for mapping and accuracy assessment of the point cloud model for the Great Achievement Palace heritage building. *Herit. Sci.* **12**, 153. <https://doi.org/10.1186/s40494-024-01252-y> (2024).
- [6] Xu, Z. et al. Tridimensional reconstruction applied to cultural heritage with the use of camera-equipped UAV and terrestrial laser scanner. *Remote Sens.* **6**, 10413–10434. <https://doi.org/10.3390/rs6110413> (2014).
- [7] Chen, Q., Chen, T.-N., Ouyang, Z.-Y., Wang, X.-Y. Virtual reconstruction of Wanfoge in Beijing Zhihua Temple. *J. Cult. Herit.* **73**, 382–392. <https://doi.org/10.1016/j.culher.2025.04.016> (2025)
- [8] Bent, G.R., Pfaff, D., Brooks, M., Radpour, R., Delaney, J. A practical workflow for the 3D reconstruction of complex historic sites and their decorative interiors: Florence As It Was and the church of Orsanmichele. *Herit. Sci.* **10**, 118. <https://doi.org/10.1186/s40494-022-00750-1> (2022).
- [9] Milosz, M., Kęsik, J., Abdullaev, U. 3D scanning and modeling of highly detailed and geometrically complex historical architectural objects: The example of the Juma Mosque in Khiva, Uzbekistan. *Herit. Sci.* **12**, 99. <https://doi.org/10.1186/s40494-024-01207-3> (2024).
- [10] Liang, H. et al. The integration of terrestrial laser scanning and terrestrial and unmanned aerial vehicle digital photogrammetry for the documentation of Chinese classical gardens: A case study of Huanxiu Shanzhuang, Suzhou, China. *J. Cult. Herit.* **33**, 222–230. <https://doi.org/10.1016/j.culher.2018.03.004> (2018).
- [11] Zhang, C., Deng, K., Yan, D., Mao, J., Yang, X. Research on multi-source image fusion technology in the digital reconstruction of classical garden and ancient buildings. *Int. Rev. Spat. Plan. Sustain. Dev.* **11**, 116–131. https://doi.org/10.14246/irspsd.11.3_116 (2023).
- [12] Huang, X. et al. Modeling visual perception of Chinese classical private gardens with image parsing and interpretable machine learning. *npj Herit. Sci.* **14**, 105. <https://doi.org/10.1038/s40494-026-02372-3> (2026).
- [13] Shen, C., Yu, C. The virtual-real measurement of Chinese garden impression: A quantitative analysis of cognitive experience of Jiangnan gardens with virtual reality experiments. *Front. Archit. Res.* **13**, 895–911. <https://doi.org/10.1016/j.foar.2024.02.007> (2024).
- [14] Zhou, L., Li, R., Liu, D., Pang, T. An integrated framework to quantify tour-view experience in classical gardens using computer vision and spatial analysis. *npj Herit. Sci.* **14**, 119. <https://doi.org/10.1038/s40494-026-02387-w> (2026).
- [15] Zhu, H., Kong, Y., Zhang, H., Gu, Z., Ohno, R. Effects of scenery frame on visual depth perception in classical Chinese gardens: A case study of the Lvyin Pavilion in Lingering Garden. *Front. Archit. Res.* **14**, 402–415. <https://doi.org/10.1016/j.foar.2024.08.009> (2025).
- [16] Garcia-Garcia, A. et al. A survey on deep learning techniques for image and video semantic segmentation. *Applied Soft Computing*, **70**, 41 – 65. <https://doi.org/10.1016/j.asoc.2018.05.018> (2018).
- [17] Sultana, F., Sufian, A., Dutta, P. Evolution of image segmentation using deep convolutional neural network: A survey. *Knowledge-Based Systems* 201 – 202, 106062. <https://doi.org/10.1016/j.knosys.2020.106062>. (2020).
- [18] Hong, S., Kwak, S., Han, B. Weakly supervised learning with deep convolutional neural networks for semantic segmentation: Understanding semantic layout of images with minimum human supervision. *IEEE Signal Process. Mag.* **34**, 39–49. <https://doi.org/10.1109/MSP.2017.2742558> (2017).
- [19] Long, J., Shelhamer, E., Darrell, T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, pp. 3431–3440. doi: 10.1109/CVPR.2015.7298965. (2015).
- [20] Ding, J. et al. Rethinking the landscape perception of traditional built environment from the perspective of depth. *Front. Archit. Res.* **15**, 825–844. <https://doi.org/10.1016/j.foar.2025.08.004> (2026).

- [21] Chai, H. et al. Semantic segmentation and spatial grid analysis of Chinese heritage landscape photographic compositions with cross-cultural perspectives. *npj Herit. Sci.* **14**, 176. <https://doi.org/10.1038/s40494-026-02439-1> (2026).
- [22] Sektani, H.H.J., Khayat, M., Mohammadi, M., Roders, A.P. Erbil city built heritage and wellbeing: An assessment of local perceptions using the semantic differential scale. *Sustainability* **13** (7), 3763. <https://doi.org/10.3390/su13073763> (2021).
- [23] Li, W., Liu, Y. The influence of visual and auditory environments in parks on visitors' landscape preference, emotional state, and perceived restorativeness. *Humanities and Social Sciences Communications* **11** (1), 1491. <https://doi.org/10.1057/s41599-024-04064-4> (2024).
- [24] Chen, X., Ibrahim, Z. A comprehensive study of emotional responses in AI-enhanced interactive installation art. *Sustainability* **15**, 15830. <https://doi.org/10.3390/su152215830> (2023).
- [25] Li, Z., Wang, K., Hai, M., Cai, P., Zhang, Y. Preliminary study on gender differences in EEG-based emotional responses in virtual architectural environments. *Buildings* **14**, 2884. <https://doi.org/10.3390/buildings14092884> (2024).
- [26] Castiblanco Jimenez, I.A. et al. Exploring user engagement in a museum scenario with EEG: A case study in the MAV Craftsmanship Museum in the Valle d'Aosta Region, Italy. *Electronics* **12**, 3810. <https://doi.org/10.3390/electronics12183810> (2023).
- [27] Chen, Z., Takatori, C. Is green alone enough? A mixed-design virtual reality study on landscape elements and restorative effects across different occupations. *Landsc. Urban Plan.* **261**, 105392. <https://doi.org/10.1016/j.landurbplan.2025.105392> (2025).
- [28] Peng, H. et al. Depression or recovery? A study of the influencing elements of urban street environments to alleviate mental stress. *Front. Archit. Res.* **14**, 846–862. <https://doi.org/10.1016/j.foar.2024.11.006> (2025).
- [29] Dorr, M., Martinetz, T., Gegenfurtner, K.R., Barth, E. Variability of eye movements when viewing dynamic natural scenes. *J. Vis.* **10**, 28. <https://doi.org/10.1167/10.10.28> (2010).
- [30] Koelstra, S. et al. DEAP: A database for emotion analysis using physiological signals. *IEEE Trans. Affect. Comput.* **3**, 18–31. <https://doi.org/10.1109/T-AFFC.2011.15> (2012).
- [31] Li, Y., Cao, L., Gao, W. Perceptual dimensions of heritage canal landscapes and their impact on conservation value cognition. *Ecological Indicators* **176**, 113626. <https://doi.org/10.1016/j.ecolind.2025.113626> (2025).
- [32] Institute of Archaeology, Chinese Academy of Social Sciences, 2008. *Sui Renshou Palace and Tang Jiucheng Palace: Archaeological Excavation Report* (Science Press, 2008) (in Chinese).
- [33] Ouyang, X., Song, Q. *Xin Tang Shu [New Book of Tang]*, vol. 37, Di li zhi [Treatise on Geography]: Fengxiang Fu [Fengxiang Prefecture]. (Zhonghua Book Company, 1975) (in Chinese)
- [34] Zhang, Y., Liu, H. Research on construction techniques of “landscape of advantage landform” based on Chinese landscape culture. *Landscape Architecture* **26**, 80–84. <https://doi.org/10.14085/j.fjyl.2019.11.0080.05> (2019) (in Chinese).
- [35] Xiao, D.N., Zhong, L.S. Ecological principles of landscape classification and assessment. *Chinese Journal of Applied Ecology* **9**, 217–221. <https://doi.org/10.13287/j.1001-9332.1998.0047> (1998) (in Chinese).
- [36] Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., Girdhar, R. Masked-attention mask transformer for universal image segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1290–1299. <https://doi.org/10.1109/CVPR52688.2022.00135> (2022).
- [37] Jahani, A., Saffariha, M. Aesthetic preference and mental restoration prediction in urban parks: An application of an environmental modeling approach. *Urban For. Urban Green.* **54**, 126775. <https://doi.org/10.1016/j.ufug.2020.126775> (2020).
- [38] Wang, Y. et al. 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.* **12**, 36. <https://doi.org/10.1186/s40494-024-01145-0> (2024)
- [39] Ren, J.Y. Landscape visual evaluation and place attachment in historical and cultural districts: A study based on semantic differential scale and eye-tracking experimental methods. *Multimed. Syst.* **30**, 306. <https://doi.org/10.1007/s00530-024-01514-6> (2024).

- [40] Chen, W., Ruan, R., Deng, W., Gao, J. The effect of visual attention process and thinking styles on environmental aesthetic preference: An eye-tracking study. *Front. Psychol.* **13**, 1027742. <https://doi.org/10.3389/fpsyg.2022.1027742> (2023).
- [41] Foxe, J.J., Snyder, A.C. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Front. Psychol.* **2**, 154. <https://doi.org/10.3389/fpsyg.2011.00154> (2011)
- [42] Spitzer, B., Haegens, S. Beyond the status quo: A role for beta oscillations in endogenous content reactivation. *eNeuro* **4**, ENEURO.0170-17.2017. <https://doi.org/10.1523/ENEURO.0170-17.2017> (2017).
- [43] Suhaimi, N.S., Mountstephens, J., Teo, J. EEG-based emotion recognition: A state-of-the-art review of current trends and opportunities. *Comput. Intell. Neurosci.* **2020**, 8875426. <https://doi.org/10.1155/2020/8875426> (2020).
- [44] Henderson, J.M., Hayes, T.R. Meaning guides attention in real-world scene images: Evidence from eye movements and meaning maps. *J. Vis.* **18**, 10. <https://doi.org/10.1167/18.6.10> (2018).
- [45] Itti, L., Koch, C., Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 1254–1259. <https://doi.org/10.1109/34.730558> (1998).
- [46] Schubert, T., Friedmann, F., Regenbrecht, H. The experience of presence: Factor analytic insights. *Presence Teleoperators Virtual Environ.* **10**, 266–281. <https://doi.org/10.1162/105474601300343603> (2001).
- [47] White, M.et al. Blue space: The importance of water for preference, affect, and restorativeness ratings of natural and built scenes. *J. Environ. Psychol.* **30**, 482–493. <https://doi.org/10.1016/j.jenvp.2010.04.004> (2010).
- [48] Ji, C. *Yuan Ye [The Craft of Gardens]*. (Zhonghua Book Company, 2017) (in Chinese).
- [49] Liu, Q.et al. The impact of landscape complexity on preference ratings and eye fixation of various urban green space settings. *Urban For. Urban Green.* **66**, 127411. <https://doi.org/10.1016/j.ufug.2021.127411> (2021).