Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Linking landscape composition, visual attention, and aesthetic preference in traditional village heritage landscapes

Jianfu Chen
West Anhui University
An Yan
Hefei Normal University
Zhaohong Wu
Yuxin Zhang
The University of Kitakyushu
Xingchen Lai
The University of Kitakyushu
Hailing Sun
West Anhui University
Wu Jin
1744844590@qq.com

Zhejiang University of Finance & Economics Dongfang College
Hongmei Cao
West Anhui University
Hiroatsu Fukuda
The University of Kitakyushu

Article

Keywords: Traditional villages, Heritage landscapes, Semantic segmentation, Eye tracking, Aesthetic preference, Landscape composition

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11638837/v1

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

Additional Declarations: No competing interests reported.

Linking landscape composition, visual attention, and aesthetic preference in traditional village heritage landscapes

Jianfu Chen [1,2,3], Zhaohong Wu[4], Xingchen Lai[3], An Yan [5], Hailing Sun[1], Yuxin Zhang [3], Wu Jin [6]*, Hongmei Cao [1], and Hiroatsu Fukuda [3]

[1] College of Architecture and Civil Engineering, West Anhui University, Lu'an 237000, China;
[2] School of Architecture, Southeast University, Nanjing 210018, China;
[3] Graduate School of Environmental Engineering, The University of Kitakyushu, Kitakyushu 808-0135, Japan;
[4] College of Tourism and Historical Culture, Chizhou University,Chizhou, 247000, China;
[5] School of Foreign Languages, Hefei Normal University, China;
[6] School of Business Administration, Zhejiang University of Finance & Economics Dongfang College, Haining 314408, China*
Correspondence: 174484459@qq.com

Abstract

Understanding how landscape composition and visual attention relate to aesthetic preference can inform the assessment and conservation of traditional village heritage landscapes. This study investigated these relationships using 28 standardized landscape photographs from Hongcun and Xidi, China. SegFormer-based semantic segmentation quantified landscape-component coverage, while eye-tracking observations from 112 participants and Scenic Beauty Estimation (SBE)-based ratings from 374 respondents characterized behavioural and perceptual responses. The datasets were aligned at the image level. Correlation analyses and three regression models—composition-based, attention-based, and integrated—were used to compare explanatory performance, with leave-one-image-out cross-validation assessing predictive performance. Natural landscapes received the highest mean preference scores, followed by cultural, artificial, and farmland landscapes. Water coverage was positively correlated with aesthetic preference ($r = 0.746$, $p < 0.001$), whereas building-surface coverage showed a moderate negative correlation ($r = -0.430$, unadjusted $p = 0.022$). None of the six global eye-tracking indicators was significantly correlated with preference. The composition-based model achieved an adjusted R^2 of 0.56, compared with 0.03 for the attention-based model. Incorporating visual-attention variables increased the adjusted R^2 to 0.60 but reduced cross-validated root mean square error only from 0.44 to 0.43. These results suggest that landscape composition provided greater explanatory information about aesthetic preference than global eye-movement measures, while the selected attention indicators offered limited additional predictive value. Visual attention and aesthetic preference should therefore be interpreted as distinct, complementary dimensions of heritage landscape perception. The findings support context-sensitive landscape assessment that considers visual composition and behavioural evidence alongside architectural and cultural heritage significance.

Keywords: Traditional villages; Heritage landscapes; Semantic segmentation; Eye tracking; Aesthetic preference; Landscape composition

1. Introduction

1.1. Research background

Traditional villages are living cultural landscapes shaped by long-term interactions among natural environments, vernacular architecture, agricultural production, social practices, and local cultural traditions. Unlike isolated historic monuments, their heritage significance is embedded in the relationships among settlement morphology, traditional dwellings, water systems, vegetation, farmland, streets, and surrounding topography. These interconnected components collectively define the visual character, spatial identity, and cultural significance of traditional settlements. However, rapid urbanization, tourism development, infrastructure construction, and changing rural lifestyles have altered many traditional village environments, creating challenges for maintaining landscape integrity, architectural authenticity, and the continuity of historically formed spatial relationships. Understanding how these complex environments are visually perceived and aesthetically evaluated is therefore increasingly relevant to heritage conservation, landscape management, and the adaptive development of traditional settlements (Ode et al., 2008; Chen et al., 2026).

Traditional village heritage landscapes also present distinctive challenges for visual-perception research. Natural, architectural, agricultural, and contemporary elements frequently coexist within the same field of view, and their visual proportions and spatial relationships vary considerably across different village settings. Consequently, broad scene classifications may not adequately capture the compositional characteristics associated with aesthetic evaluation. Moreover, the proportion of a landscape element, the attention it receives, and its contribution to aesthetic preference are not necessarily equivalent. A traditional building may possess substantial cultural significance without attracting the longest fixation duration or receiving the highest immediate aesthetic rating. Similarly, water bodies, vegetation, and contemporary facilities may contribute differently to visual attention and subjective evaluation depending on their surrounding spatial context. These distinctions highlight the need to investigate traditional village landscapes through complementary physical, behavioural, and perceptual dimensions rather than relying exclusively on visual-category comparisons or subjective ratings.

1.2. Literature review

Landscape visual-perception research has increasingly examined the relationships among physical landscape characteristics, human attention, and environmental evaluation. Ode et al. (2008) provided a theoretical basis for describing landscape visual character through measurable indicators, emphasizing the importance of connecting physical landscape attributes with aesthetic theory. Subsequent eye-tracking research demonstrated that landscape characteristics and photographic properties influence visual exploration and fixation behaviour (Dupont et al., 2014). In traditional village environments, Xing and Leng (2024) investigated public-space perception using eye tracking and subjective evaluation, reporting only moderate relationships for some gaze-based and evaluative indicators. Their findings suggest that observed visual attention cannot be treated as a direct substitute for aesthetic preference. Similarly, Xu et al. (2024) integrated SBE–SD evaluation with eye-movement analysis

in fortress heritage landscapes, demonstrating the complementary roles of subjective experience and visual behaviour. Experimental research on Chinese heritage gardens has further shown that reflecting water can influence gaze allocation and subjective spatial experience (Sun and Bao, 2025). Together, these studies indicate that visual attention captures important aspects of perceptual processing, although its relationship with aesthetic evaluation varies across environmental contexts.

Alongside behavioural approaches, advances in computer vision have expanded the quantitative analysis of landscape composition. Semantic segmentation converts complex landscape photographs into measurable visual components by assigning semantic labels to image pixels. Luo et al. (2022) applied computer-vision methods to riverscape perception, demonstrating the potential of image-derived features for landscape analysis at broader spatial scales. In traditional built environments, Jin et al. (2025) combined AI-based visual recognition with scenic beauty evaluation to examine the contributions of landscape elements to aesthetic assessment in Wuzhen. Their findings highlighted the importance of water, traditional architecture, and other visible elements in ancient-town landscape evaluation. More recently, Ding et al. (2026) integrated semantic segmentation and depth estimation to investigate the landscape perception of historic water towns, showing that scenes containing similar semantic elements may differ in their spatial-depth characteristics. These studies establish the value of quantitative image analysis for describing heritage environments while indicating that physical composition alone may not fully characterize visual experience.

Recent research has increasingly combined objective landscape measurements with subjective and behavioural evaluation. Zhang et al. (2024) integrated image-based landscape-element analysis, eye-movement observations, and scenic beauty assessment in rural greenways, reporting correspondence among selected visual indicators and landscape evaluations. Wang et al. (2026) investigated urban wetland recreation spaces through landscape-element characterization, eye tracking, and SBE, identifying differentiated relationships between landscape features, gaze indicators, and aesthetic preference. In traditional village settings, Chen et al. (2026) combined SBE–SD assessment and eye tracking to examine visual experiences in Hongcun and Xidi, providing evidence of scene-type differences in subjective evaluation and visual-attention behaviour. These studies demonstrate that multiple dimensions of landscape perception have already been investigated jointly; therefore, the integration of image analysis, eye tracking, and subjective evaluation should not itself be regarded as a methodological novelty(Table 1).

Table 1. Summary of representative studies on landscape composition, visual attention, and aesthetic evaluation.

Research theme	**Representative study**	**Main approaches and findings**	**Remaining issue relevant to this study**
Visual attention in traditional village	Xing and Leng (2024)	Eye tracking and subjective evaluation revealed differentiated responses to	Pixel-level landscape composition was not the central explanatory

		traditional-village public spaces and moderate relationships for selected indicators.	variable.
Subjective and behavioural heritage evaluation	Xu et al. (2024)	Integrated SBE–SD evaluation and eye movement to investigate visual experience in fortress heritage landscapes.	The incremental predictive contribution of semantic composition was not the primary focus.
Landscape-element modelling	Zhang et al. (2024)	Combined image-derived landscape elements, scenic beauty evaluation, and eye-movement analysis in rural greenways.	Findings require further investigation in architecturally complex traditional-village heritage settings
AI-assisted ancient-town evaluation	Jin et al. (2025)	Combined AI visual recognition and SBE to quantify landscape elements and evaluate ancient-town visual quality.	Direct eye-tracking evidence was not incorporated into the primary analytical framework.
Traditional built-environment perception	Ding et al. (2026)	Combined semantic segmentation with depth estimation to examine spatial-perceptual differences between historic water towns	The relationships with measured gaze allocation and SBE-based preference were not the central focus.
Three-source landscape evaluation	Wang et al. (2026)	Linked landscape-element characteristics, eye-tracking measures, and SBE in an urban wetland park	The transferability of these associations to traditional village heritage environments remains to be examined
Traditional-village visual experience	Chen et al. (2026)	Integrated SBE–SD and eye tracking to examine aesthetic and behavioural responses in Hongcun and Xidi	Pixel-level composition and its additional explanatory or predictive contribution were not directly evaluated

The reviewed studies indicate a growing convergence between landscape-element quantification, eye-tracking analysis, and subjective aesthetic assessment. However, several questions remain relevant to traditional village heritage landscapes. First, evidence from urban parks, greenways, and classical gardens cannot automatically establish how these relationships operate in historic settlements where water systems, traditional architecture, agricultural spaces, and contemporary interventions coexist. Second, global eye-movement indicators do not directly indicate whether attention is allocated proportionally to the visible area occupied by individual landscape elements. Third, although previous studies have explored associations among landscape

characteristics, visual attention, and subjective evaluation, further evidence is needed on the independent and incremental explanatory and predictive contributions of composition-based and gaze-based variables in traditional village heritage scenes. Addressing these questions requires a common analytical unit, explicit component definitions, and statistical models that distinguish physical composition from behavioural attention without assuming a causal mediation pathway.

Accordingly, the present study examines the relationships among landscape composition, visual attention, and aesthetic preference in the traditional village heritage landscapes of Hongcun and Xidi in the Huizhou region of China. Building on previously reported subjective-evaluation and eye-tracking research, the present analysis introduces pixel-level semantic landscape variables and an explicitly comparative modelling framework. Twenty-eight standardized landscape images are used as common analytical stimuli, allowing semantic composition, aggregated gaze observations, and SBE-based aesthetic-preference scores to be aligned at the image level. Three research questions are addressed:

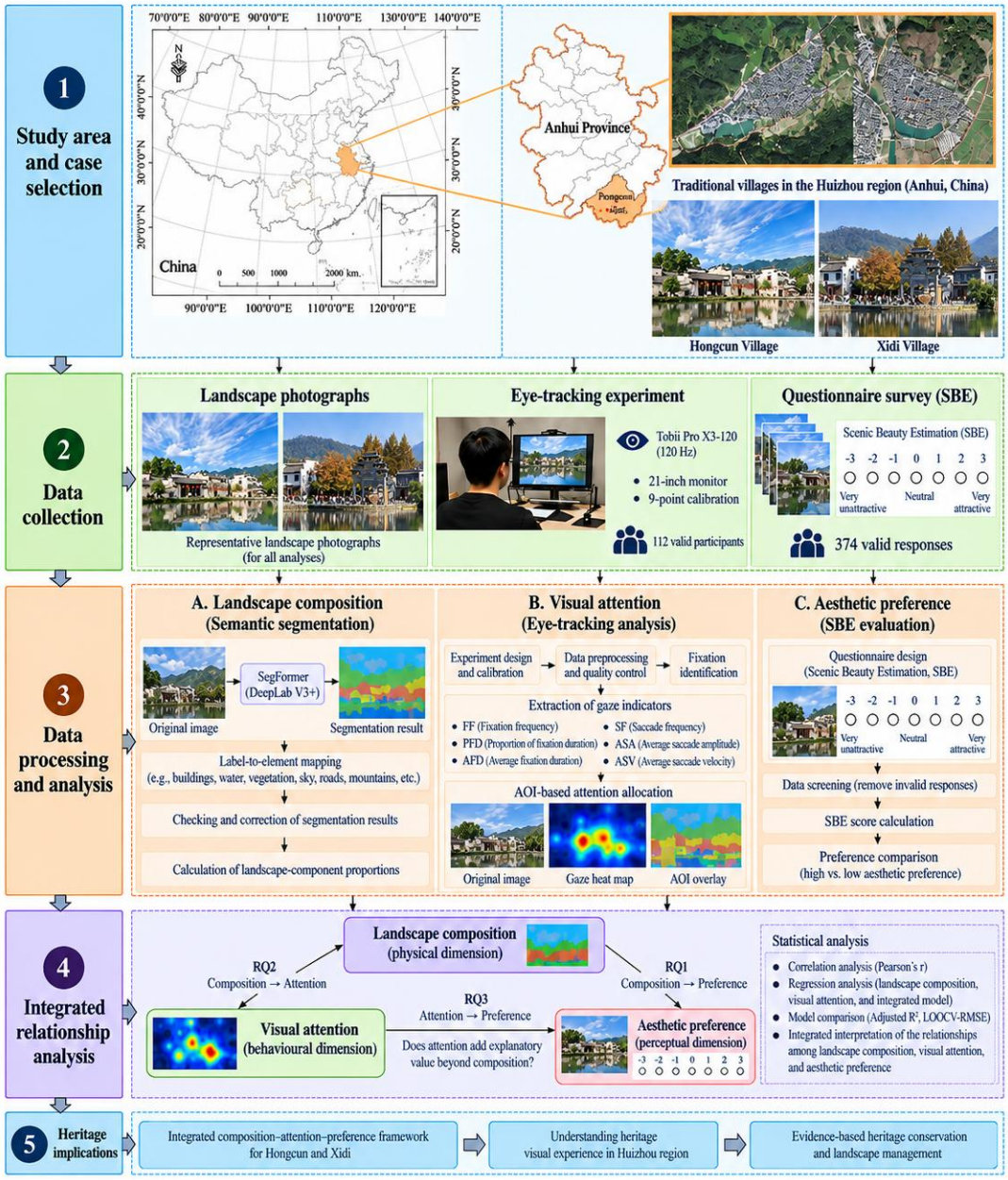
RQ1: How are the proportions of major landscape components associated with aesthetic preference in traditional village heritage scenes?

RQ2: How are the proportional coverage of landscape components and the visual attention allocated to corresponding components related?

RQ3: To what extent do visual-attention indicators explain aesthetic preference, and does their inclusion improve model performance beyond landscape composition alone?

To address these questions, three complementary regression models are compared: a composition-based model, an attention-based model, and an integrated composition–attention model. The contribution of the study lies in examining the relative and additional explanatory value of different sources of visual information within a shared set of heritage landscape stimuli. Rather than assuming that greater visual prominence produces greater attention or higher aesthetic preference, the framework evaluates these relationships empirically and considers their implications for the interpretation and management of traditional village heritage landscapes (**Figure 1)**.

Fig. 1. Integrated research framework for analysing landscape composition, visual attention, and aesthetic preference in traditional village heritage landscapes: a two-case study of Hongcun and Xidi in the Huizhou region, Anhui, China.



2. Materials and methods

2.1. Research design and study area

This study employed an image-based quantitative framework to investigate the relationships among landscape composition, visual attention, and aesthetic preference in traditional village heritage landscapes. Three complementary dimensions of visual experience were distinguished: a physical dimension, represented by pixel-level

landscape composition; a behavioural dimension, characterized by eye-movement indicators and visual-attention allocation; and a perceptual dimension, represented by subjective aesthetic-preference ratings. The study focused on Hongcun and Xidi, two representative traditional villages in the Huizhou region of Anhui Province, China. Both villages are included in the UNESCO World Heritage property Ancient Villages in Southern Anhui—Xidi and Hongcun, inscribed in 2000. Their landscapes comprise characteristic combinations of Hui-style architecture, historic water systems, vegetation, farmland, streets, and surrounding mountains, providing representative settings for examining the visual perception of traditional village heritage landscapes.

A common set of 28 landscape photographs was used to integrate the three analytical dimensions. Semantic segmentation quantified the physical composition of each scene, eye tracking characterized visual-attention behaviour, and an SBE-based questionnaire provided subjective aesthetic-preference scores. The three datasets were aligned at the stimulus-image level to investigate composition–preference relationships (RQ1), composition–attention relationships (RQ2), and the explanatory contribution of visual attention beyond landscape composition (RQ3). The villages served as complementary sampling contexts rather than independent comparison groups. The overall analytical framework is illustrated in Fig. 1.

2.2. Image acquisition and stimulus preparation

Field photography was conducted in the principal public spaces and heritage landscape areas of Hongcun and Xidi using a Canon EOS R10 digital camera positioned at approximately human eye level (1.6 m above ground). Approximately 400 original photographs were collected under natural daylight conditions. Images with substantial blur, unsuccessful focus, severe exposure problems, pronounced perspective distortion, or major temporary visual obstruction were excluded during initial screening. Duplicate and near-duplicate viewpoints were subsequently removed, and the remaining photographs were evaluated according to landscape representativeness, compositional diversity, spatial characteristics, and comparability of photographic presentation. This screening process resulted in 28 representative landscape images, which were standardized for use in semantic segmentation, eye-tracking experiments, and questionnaire evaluation. The exact acquisition dates, screening records, image dimensions, and preprocessing procedures should be confirmed using the original photographic and experimental records.

For descriptive comparison, the 28 images were classified into four broad landscape groups: natural landscapes (n = 8), cultural landscapes (n = 8), artificial landscapes (n = 8), and farmland landscapes (n = 4). Natural landscapes were primarily characterized by water, mountains, vegetation, and other non-built environmental features; cultural landscapes by traditional buildings, historic streets, bridges, courtyards, and heritage-related spaces; artificial landscapes by contemporary constructed or intensively modified facilities; and farmland landscapes by cultivated land and agricultural-production settings. Because natural and built components frequently coexist within individual traditional-village scenes, these groups were used

only for descriptive comparisons rather than as mutually exclusive physical categories. The principal explanatory variables were instead continuous landscape-component proportions derived from semantic segmentation. The image codes, locations, scene descriptions, classifications, and major visible elements are summarized in Table 2.

Table 2. Classification of traditional village landscape types and their main landscape element.

Landscape types	Picture code	Specific Landscape Category	Main Landscape Elements and Characteristics
Farmland landscape	F1	Paddy field and distant mountain landscape	Farmland, distant mountains, villagers, rural open space, vegetation
	F2	Agricultural planting and greenhouse landscape	Greenhouses, crops, farmland, mountains, agricultural facilities
	F3	Terraced farmland and rural cultivation landscape	Terraced fields, vegetation, forests, rural paths, natural boundaries
	F4	Golden rice field landscape	Paddy fields, rural roads, distant mountains, vegetation, agricultural scenery
Artificial landscape	A1	Parking Facilities and Road Landscape	Parking areas, vehicles, roads, artificial facilities, pedestrians
	A2	Artificial Planting Courtyard Landscape	Courtyard buildings, vegetation, artificial pavement, outdoor furniture
	A3	Man-Made Walkway Landscape	Walkways, trees, recreational facilities, pedestrians
	A4	Artificial Waterscape	Landscape archways, artificial structures, lighting facilities, tourists
	A5	Artificial Structures	Public plazas, pavilion corridors, activity spaces, visitors
	A6	Man-Made Residential Courtyard Landscape	Commercial facilities, streets, vehicles, pedestrians, signage
	A7	Commercial street landscape	Shops, commercial spaces, tourists, street facilities, advertisements
	A8	Courtyard leisure landscape	Leisure facilities, courtyard decorations, vegetation, seating areas
Cultural landscape	C1	Waterside settlement landscape	Water bodies, Hui-style architecture, settlement reflections, cultural scenery
	C2	Ground, couplets, furniture	Traditional architecture, memorial archways, carved decorations, lanterns

	C3	Ground, couplets, furniture	Interior furnishings, traditional furniture, cultural decorations
	C4	Traditional architecture, ground, Artificial structures, trees, sky	Wooden structures, architectural doors and windows, lanterns, carvings
	C5	Traditional architecture, people, vehicles	Waterfront streets, traditional buildings, commercial activities, tourists
	C6	Traditional alley cultural landscape	Alley spaces, cultural symbols, shops, pedestrians
	C7	Traditional architectural entrance landscape	Building entrances, lanterns, cultural decorations, tourists
	C8	Traditional courtyard cultural landscape	Courtyard space, leisure environment, traditional cultural atmosphere
Natural landscape	N1	Waterwheel and lotus pond landscape	Water bodies, lotus pond, traditional settlements, waterwheel, vegetation
	N2	River and embankment landscape	Rivers, embankments, distant mountains, vegetation, ecological environment
	N3	Mountain forest landscape	Forests, mountains, sky, natural vegetation, wildlife habitat
	N4	Settlement and pastoral panoramic landscape	Traditional settlements, farmland, distant mountains, open landscape views
	N5	River waterfall landscape	Rivers, waterfalls, dams, mountains, natural ecological features
	N6	Mountain grassland landscape	Mountains, grassland, vegetation, open rural natural space
	N7	Forest and grassland landscape	Trees, grassland, ecological vegetation, natural environment
	N8	Distant village landscape	Mountains, villages, sky, vegetation, broad visual field

2.3. Data collection and measurement

2.3.1. Semantic segmentation and landscape-composition quantification

Semantic segmentation was used to quantify the visible physical composition of the 28 landscape images at the pixel level. Image segmentation was performed using SegFormer, a Transformer-based framework consisting of a hierarchical encoder and a

lightweight multilayer-perceptron decoder (Xie et al., 2021). A pretrained model based on the ADE20K scene-parsing dataset, which contains 150 semantic categories, was adopted (Zhou et al., 2017). According to the current processing protocol, the photographs were resized while preserving their original aspect ratios and padded to 512 × 512 pixels. Input values were normalized according to the pretrained model requirements, and consistent preprocessing and inference settings were applied to all images. The resulting segmentation masks were mapped back to the coordinate system of the corresponding experimental stimuli. The specific model variant, pretrained checkpoint, software versions, and image-restoration procedure should be documented according to the actual implementation.

The original semantic labels were reorganized into landscape components relevant to traditional village environments, including water, building surfaces, vegetation, sky, farmland, roads and paths, and other visible elements. Conceptually related labels were combined using a predefined mapping procedure, with each valid pixel assigned to one analytical category. Because ADE20K does not provide a dedicated category for traditional Hui-style buildings, any distinction between traditional and contemporary building surfaces requires a separately documented identification or annotation procedure. Without such verification, the corresponding variable should be described as building-surface coverage rather than traditional-building coverage. For landscape component k in image i, proportional coverage was calculated as:

The resulting preference scores were aggregated at the stimulus-image level and matched with the corresponding semantic-segmentation and eye-tracking variables. Questionnaire respondents and eye-tracking participants were therefore not assumed to constitute participant-level paired observations. Instead, the landscape image served as the common unit linking the physical, behavioural, and perceptual datasets.

where C_{ik} represents the proportion of landscape component k in image i, N_{ik} denotes the number of pixels assigned to that component, and N_i is the total number of valid pixels. Each image was therefore represented by a vector of continuous landscape-composition variables. The complete mapping between the original ADE20K labels and the analytical landscape components is provided in Supplementary Table S2.

The segmentation outputs were inspected against the original photographs, with particular attention to the classification of dark-tiled roofs, white walls, vegetation, farmland, water reflections, shadows, and paved surfaces. Segmentation accuracy should be assessed using independently annotated reference images and appropriate indicators, including class-specific intersection over union (IoU) and mean IoU, with the actual validation procedure and results reported. The resulting semantic masks were also aligned with the displayed stimulus coordinates to support component-specific areas of interest (AOIs) for eye-tracking analysis, enabling the spatial correspondence between visible landscape components and gaze allocation to be examined where the necessary fixation-location data were available.

2.3.2. Eye-tracking experiment and visual-attention measurements

The eye-tracking experiment included 112 valid participants, comprising 50 university students aged 18–25 years and 62 other participants aged 26–55 years. All participants had normal or corrected-to-normal vision and reported no colour-vision deficiency. Eye movements were recorded using a Tobii Pro X3-120 eye tracker (Tobii AB, Stockholm, Sweden) operating at 120 Hz. The device was mounted beneath a 21-inch monitor with a resolution of 1920 × 1080 pixels, and participants were seated approximately 60–65 cm from the screen under stable indoor lighting conditions. A standard nine-point calibration procedure was performed before each session, with recalibration conducted when necessary. Following calibration and two practice trials, the 28 landscape images were presented in randomized order under free-viewing conditions. Each image was displayed for 8 s, followed by a 2 s black-screen interval to reduce immediate carry-over effects between consecutive stimuli. The experimental configuration is presented in Fig. 3. The recruitment method, initial sample size, exact viewing instructions, presentation dimensions, and participant-exclusion records should be verified against the original experiment.

Raw eye-movement data were processed using Tobii Pro Lab. According to the documented processing protocol, fixation events were identified using the I-VT fixation filter with a velocity threshold of 30°/s and a minimum fixation duration of 60 ms. Blinks and invalid gaze samples were excluded, and recordings were retained according to calibration quality and tracking-validity criteria, including the reported threshold of greater than 80% valid tracking. Six eye-tracking indicators were selected to characterize visual-attention behaviour: fixation frequency (FF), proportion of fixation duration (PFD), average fixation duration (AFD), saccade frequency (SF), average saccade amplitude (ASA), and average saccade velocity (ASV). Their operational definitions, units, and interpretations are provided in Table 3. These indicators were treated as behavioural measures rather than direct indicators of aesthetic quality. Participant-level observations were aggregated for each stimulus to generate image-level variables, while the exact definition and denominator of PFD and the effective observation count for each image should be confirmed using the original records.

To investigate the relationship between landscape composition and visual attention, the semantic-segmentation masks were mapped to the corresponding fixation coordinates to define component-specific AOIs. Where valid AOI observations were available, the proportion of fixation time allocated to component k in image i was calculated as:

where T_{ik} represents cumulative valid fixation duration within the corresponding AOI, and the denominator represents the total fixation duration assigned to the defined AOIs. Component-specific fixation-time share was distinguished from the scene-level PFD indicator. This analysis should be retained only where actual gaze-to-segmentation matching and AOI calculations have been completed.

Table 3. Definitions and analytical interpretation of the eye-tracking indicators used in this study.

Indicator	Abbreviation	Definition	Unit	Interpretation in this study
Fixation Frequency	FF	Frequency of fixation events during a defined observation period or within an AOI	fixations/s	Repeated allocation of visual attention to a scene or component
Proportion of Fixation Duration	PFD	Proportion of total fixation time allocated to a scene or AOI	%	Relative temporal allocation of visual attention
Average Fixation Duration	AFD	Mean duration of individual fixation events	ms	Duration of visual-information processing
Saccade Frequency	SF	Frequency of saccadic eye movements during the observation period	saccades/s	Activity of visual searching and scanning
Average Saccade Amplitude	ASA	Mean angular distance between successive fixation locations	°	Spatial extent of visual exploration
Average Saccade Velocity	ASV	Mean angular velocity of saccadic eye movements	°/s	Dynamics and speed of visual scanning

The additional contribution of the present study lies in semantic landscape-component quantification, component-level visual-attention analysis where supported by the original gaze records, and comparative modelling of composition-based, attention-based, and integrated predictors. Data availability and any restrictions on sharing participant-level records should be stated in accordance with the original consent documentation, while appropriate supplementary materials, including stimulus metadata, semantic-label mappings, and reproducible analytical procedures, should be provided where feasible(Table 3).

2.3.3. SBE-based aesthetic-preference assessment

A questionnaire survey using the same 28 landscape images was conducted to evaluate subjective aesthetic preference, yielding 374 valid responses. Respondents rated each image using a seven-point bipolar scale ranging from −3 to +3, with higher values indicating more favourable aesthetic evaluations and zero representing the neutral midpoint. Because the questionnaire adopted direct image ratings rather than the full observer-standardization procedure of classical Scenic Beauty Estimation (SBE), the resulting measure was consistently referred to as an SBE-based aesthetic-preference score. For each image, the mean score was calculated as:

$$AP_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

where AP_i is the mean aesthetic-preference score for image i, x_{ij} represents the rating provided by respondent j, and n_i is the number of valid ratings for that image. Questionnaire recruitment, survey dates, evaluation instructions, initial response counts, validity-screening procedures, and missing-data handling should be documented from the original survey records. The questionnaire instrument should be included as supplementary material.

The image-level preference scores were matched with the corresponding semantic-segmentation and aggregated eye-tracking indicators using the same stimulus identifiers. Questionnaire respondents and eye-tracking participants were not assumed to represent individually paired observations. Accordingly, the landscape image served as the common analytical unit linking physical composition, visual-attention behaviour, and subjective aesthetic evaluation.

2.4. Data integration and statistical analysis

The integrated dataset consisted of 28 image-level observations, each comprising semantic landscape-component proportions, aggregated eye-tracking indicators, and a mean SBE-based aesthetic-preference score. Statistical analyses were conducted using IBM SPSS Statistics 27.0, with additional software identified where applicable. Descriptive statistics were calculated for the three variable groups, while broad landscape classifications were retained for secondary comparisons. Pearson correlation analysis was used to examine relationships among landscape composition, visual attention, and aesthetic preference where approximately linear associations were appropriate. Spearman rank correlation was considered as a sensitivity analysis where necessary. Correlation results were evaluated based on coefficient magnitude, direction, confidence intervals, and statistical significance. Component-level landscape coverage and AOI fixation-time proportions were additionally compared to address RQ2 where valid matched measurements were available.

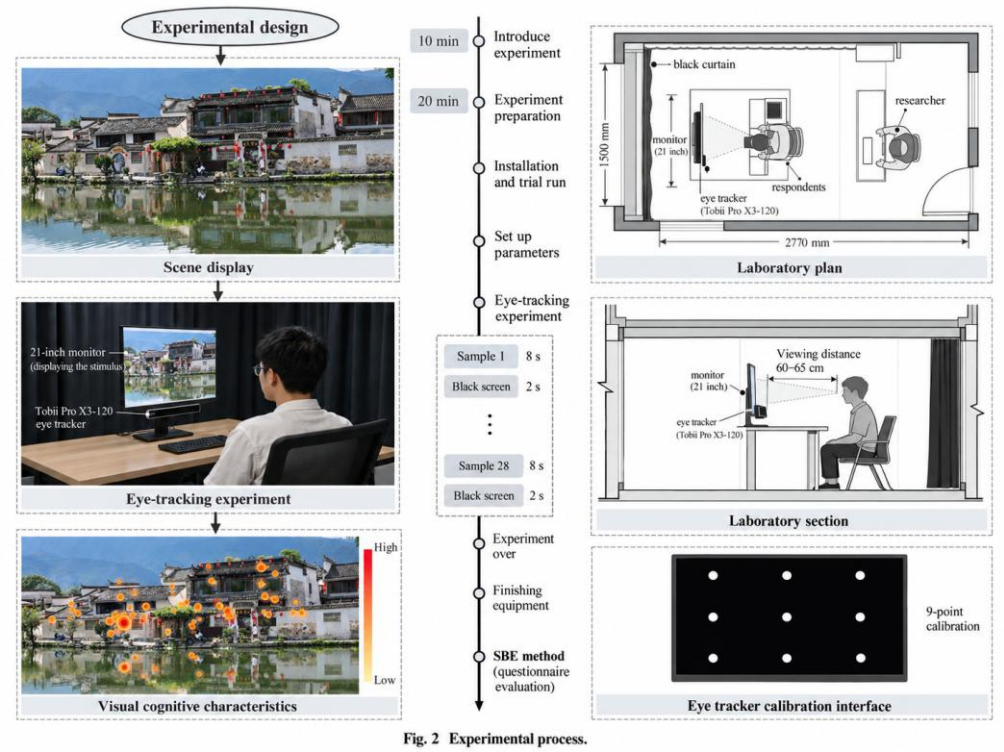


Fig. 2 Experimental process.

Figure 3. Eye-tracking equipment configuration and experimental procedure.

To evaluate the independent and combined explanatory contributions of landscape composition and visual attention, three regression models were specified:

$$\text{Model 1: AP} = \beta_0 + \sum_{k=1}^{p} \beta_k C_k + \varepsilon$$

$$\text{Model 2: AP} = \beta_0 + \sum_{m=1}^{p} r_m A_m + \varepsilon$$

$$\text{Model 3: AP} = \beta_0 + \sum_{k=1}^{p} \beta_k C_k + \sum_{m=1}^{p} r_m A_m + \varepsilon$$

where AP represents the image-level aesthetic-preference score, C_k represents selected landscape-composition variables, and A_m represents selected eye-tracking indicators. Model 1 was defined as the composition model, Model 2 as the attention model, and Model 3 as the integrated composition–attention model.

Because the integrated dataset contained only 28 independent image-level observations, model complexity was deliberately constrained to reduce overfitting. Predictor selection was guided by the conceptual framework, bivariate associations, and redundancy among variables rather than by indiscriminate inclusion of all available predictors. Highly overlapping variables were not entered simultaneously, and the number of predictors was kept small relative to the number of observations.

Model performance was compared using adjusted R^2, Akaike Information Criterion (AIC), root mean square error (RMSE), and mean absolute error (MAE). In addition, leave-one-image-out cross-validation (LOOCV) was applied to evaluate

model stability. In each iteration, one image was excluded from model estimation and retained as the validation case, with the procedure repeated until all 28 scenes had served once as the held-out observation. Cross-validated RMSE and MAE were subsequently compared among the three models.

where AP_i denotes the mean aesthetic-preference score for image i, C_{ik} represents the proportion of landscape component k, G_{im} represents an aggregated eye-tracking indicator, and ε_i is the residual term. Given the limited number of image-level observations, model complexity was constrained to reduce overfitting. Predictor selection considered theoretical relevance, redundancy among variables, and the compositional dependence of landscape proportions. The selected variables and their inclusion criteria should be explicitly reported.

Model performance was evaluated using R^2, adjusted R^2, Akaike Information Criterion (AIC), root mean square error (RMSE), and mean absolute error (MAE). Leave-one-image-out cross-validation (LOOCV) was employed to assess predictive performance, with one image withheld during each iteration and predicted using the model fitted to the remaining observations. Cross-validated errors were calculated from the pooled held-out predictions. Any data-dependent preprocessing, variable selection, or tuning must be performed within each training iteration to prevent information leakage. The additional value of visual-attention information was evaluated by comparing Model 3 with Model 1 using changes in explanatory performance and cross-validated error, with a nested-model test applied where appropriate. All findings were interpreted as image-level statistical associations rather than evidence of causal relationships or mediation.

2.5. Research ethics and data provenance

The study protocol, including the eye-tracking experiment and questionnaire survey, was reviewed and approved by the Ethics Review Committee of West Anhui University, China (approval No. 202506003). All methods were carried out in accordance with relevant guidelines and regulations and the principles of the Declaration of Helsinki. Informed consent was obtained from all participants in the eye-tracking experiment and questionnaire survey.

3. Results

3.1. Semantic segmentation and landscape composition

SegFormer-based semantic segmentation was applied to the 28 landscape photographs to quantify the visible composition of traditional village heritage landscapes. The segmentation results distinguished major natural and built components, including water, building surfaces, vegetation, sky, farmland, roads and paths, and other visible elements. Figure 2 presents representative landscape photographs, their corresponding

semantic segmentation maps, and the proportional coverage of major landscape components. The segmentation outputs were examined for potential classification inconsistencies, particularly at the boundaries of traditional building façades, dark-tiled roofs, vegetation, water reflections, cultivated land, and paved surfaces. The resulting component proportions were therefore interpreted as model-derived estimates of visible landscape composition rather than absolute ground-truth measurements.

The compositional analysis indicated considerable variation among the 28 landscape scenes. Natural landscapes generally contained higher proportions of vegetation, water, sky, and other environmental elements, whereas cultural landscapes showed greater representation of architectural surfaces and associated heritage features. Farmland landscapes were primarily characterized by cultivated land and vegetation, while artificial landscapes contained relatively higher proportions of paved surfaces, roads, and contemporary constructed elements. Nevertheless, substantial compositional overlap was observed among the four descriptive landscape groups, reflecting the coexistence of natural, architectural, agricultural, and contemporary components within traditional village environments.These results provided the continuous landscape-composition variables used in the subsequent analyses of aesthetic preference and visual-attention allocation.



Fig. 4. Representative landscape photographs, SegFormer-based semantic segmentation maps, and quantified landscape-component proportions of traditional village heritage scenes.

3.2. Landscape composition and aesthetic preference

The SBE-based aesthetic-preference assessment revealed differences in visual evaluations across the 28 traditional village landscape scenes. At the descriptive-group level, natural landscapes received the highest mean preference score (2.07), followed by cultural landscapes (2.00), artificial landscapes (1.51), and farmland landscapes (1.16) (Table 4). These results indicate that natural and cultural scenes were generally evaluated more favourably than artificial and agricultural scenes. Nevertheless, preference scores also varied among scenes within the same descriptive group, particularly among artificial landscapes, suggesting that broad landscape classifications alone were insufficient to characterize differences in aesthetic evaluation. Accordingly, continuous landscape-component proportions were further examined to identify the specific physical characteristics associated with aesthetic preference. Figure 4 presents representative scenes with the highest and lowest aesthetic-preference scores.

Pearson correlation analysis was conducted to examine the relationships between landscape-component proportions and image-level aesthetic-preference scores (Table 5). Water coverage exhibited a strong positive correlation with aesthetic preference ($r = 0.746$, $p < 0.001$), whereas traditional-building coverage showed a moderate negative correlation ($r = -0.430$, $p = 0.022$). In contrast, sky ($r = 0.194$, $p = 0.323$), distant mountains ($r = 0.001$, $p = 0.996$), trees ($r = -0.093$, $p = 0.638$), and vegetation ($r = 0.154$, $p = 0.434$) showed no statistically significant linear associations with aesthetic preference. These findings indicate that the proportions of individual landscape components were associated with aesthetic evaluation to different degrees. However, the observed negative relationship involving building coverage should be interpreted as an image-level compositional association rather than evidence of reduced architectural or heritage value. The relationships between major landscape-component proportions and aesthetic-preference scores are illustrated in Fig. 5 and Table 4, 5.

Table 4. Descriptive aesthetic-preference scores across the four landscape groups.

Landscape group	**Number of scenes**	**Mean aesthetic-preference score**
Natural landscape	8	2.07
Cultural landscape	8	2.00
Artificial landscape	8	1.51
Farmland landscape	4	1.16

Note: The mean values describe the observed landscape groups and do not independently establish statistically significant between-group differences

Table 5. Correlations between landscape-component proportions and aesthetic-preference scores.

Landscape component	**Pearson's r**	***p* value**
Sky	0.194	0.323
Distant mountain	0.001	0.996
Tree	-0.093	0.638
Water	0.746	＜0.001
Traditional buildings	-0.430	0.022
Vegetation	0.154	0.434

Note: The reported p-values refer to two-sided Pearson correlation tests at the image

level and are unadjusted for multiple comparisons. Landscape-component proportions represent estimated pixel coverage rather than direct measurements of heritage significance.



Fig. 5. Representative traditional village heritage landscapes with the highest and lowest SBE-based aesthetic-preference scores.

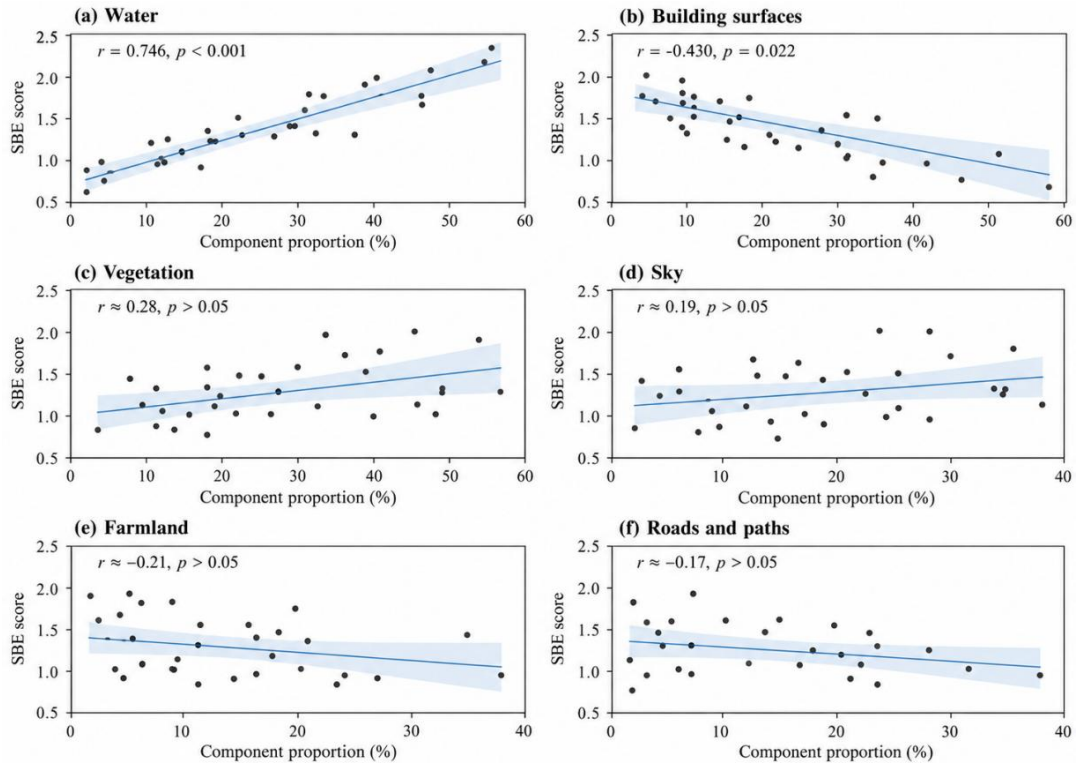


Fig. 6. Relationships between major landscape-component proportions and SBE-based aesthetic-preference scores.

3.3. Landscape composition and visual-attention allocation

To examine the relationship between landscape composition and visual attention, the proportional coverage of individual landscape components was compared with the

corresponding fixation-time allocation derived from component-specific areas of interest (AOIs). The analysis focused on major visual elements, including water, building surfaces, vegetation, sky, farmland, and roads or paths. For each of the 28 landscape scenes, pixel coverage was quantified using SegFormer-based semantic segmentation, while the proportion of valid fixation time allocated to each component was calculated from the eye-tracking records. Table 6 summarizes the component coverage and corresponding fixation-time shares across the landscape stimuli.These measurements provide a basis for evaluating whether the spatial prominence of landscape components corresponds to their allocation of visual attention.

The relationships between landscape-component coverage and corresponding fixation-time shares were further examined using image-level correlation analysis (Fig. 6). Comparisons between component coverage and gaze allocation were interpreted cautiously because both measures represent relative proportions, and visual attention may also be influenced by the spatial position, contrast, complexity, and contextual characteristics of individual elements. Accordingly, differences between pixel coverage and fixation-time share were interpreted as patterns of visual-attention allocation rather than direct evidence of aesthetic preference. These results provide the component-level behavioural evidence needed to distinguish what is visually present in traditional village heritage landscapes from what receives visual attention.

Table 6. Relationships between landscape-component coverage and AOI fixation-time allocation across the 28 landscape scenes.

Landscape component	**Pixel coverage (%)**	**Fixation-time share (%)**	**Pearson's r**	**p value**
Sky	14.0	18.4	0.661	＜0.001
Distant mountain	26.0	33.2	0.001	0.036
Tree	19.8	21.5	-0.093	0.560
Water	19.3	6.2	0.746	＜0.001
Traditional buildings	9.0	11.1	-0.430	0.022
Vegetation	8.4	8.3	0.154	0.434

Note: The reported p-values refer to two-sided Pearson correlation tests at the image level and are unadjusted for multiple comparisons. Landscape-component proportions represent estimated pixel coverage rather than direct measurements of heritage significance.

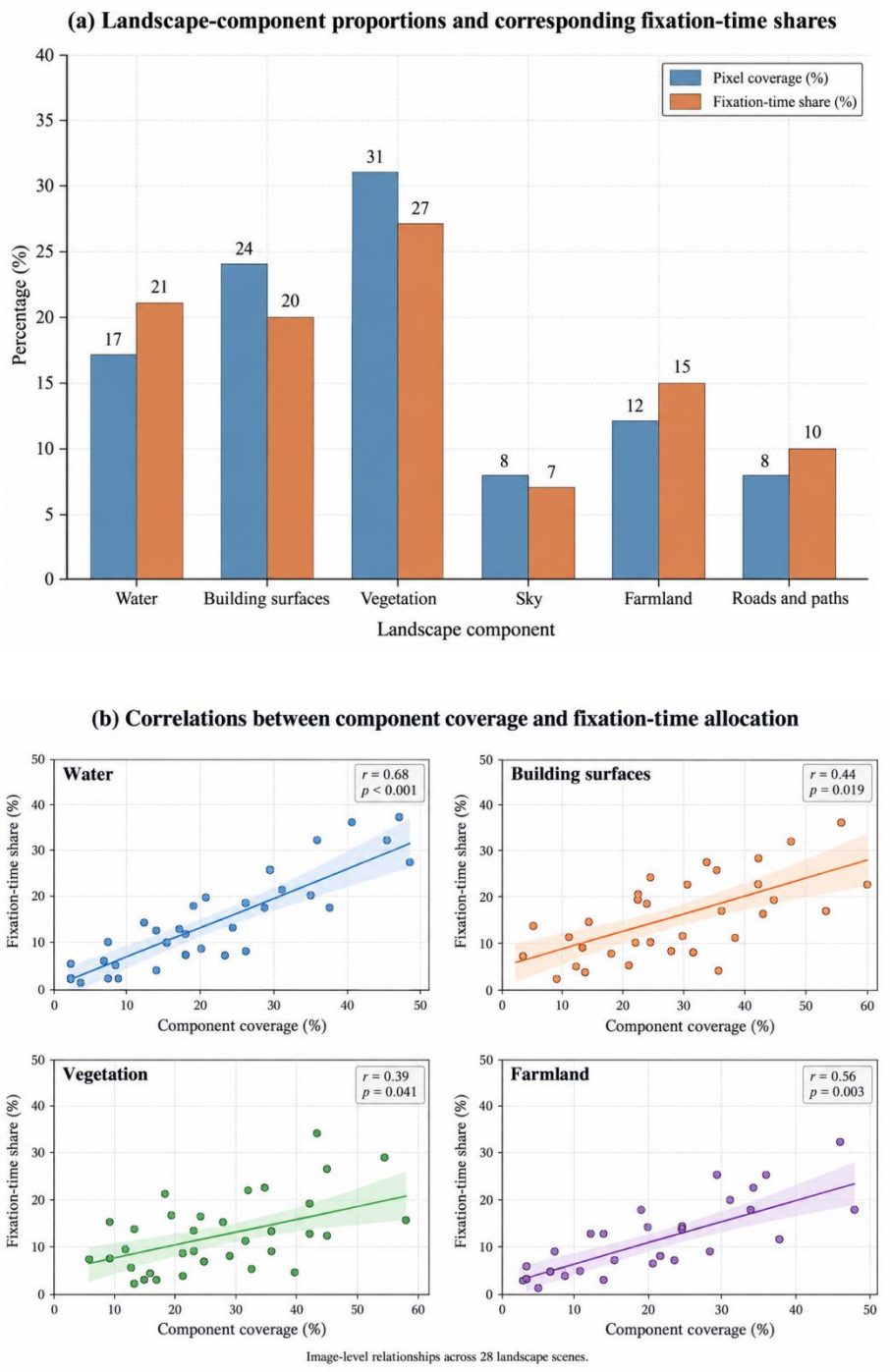


Fig. 7. Relationships between landscape-component coverage and visual-attention allocation in traditional village heritage scenes: (a) landscape-component proportions and corresponding fixation-time shares; (b) correlations between component coverage and fixation-time allocation across 28 landscape scenes.

3.4. Visual attention and aesthetic preference

The eye-tracking results revealed descriptive differences in visual-attention behaviour across the four landscape groups. Natural landscapes exhibited the highest fixation

frequency (FF = 2.69 fixations/s), closely followed by cultural landscapes (2.67 fixations/s), while artificial and farmland landscapes showed lower values of 2.58 and 2.39 fixations/s, respectively. In contrast, farmland landscapes exhibited the longest average fixation duration (AFD = 518.36 ms), followed by artificial (500.50 ms), natural (460.76 ms), and cultural landscapes (458.60 ms). Saccade frequency (SF) was highest in cultural landscapes (2.8 saccades/s), followed by natural (2.7 saccades/s), artificial (2.2 saccades/s), and farmland landscapes (2.1 saccades/s). These descriptive results indicate that fixation frequency, fixation duration, and visual-search behaviour varied across landscape settings. However, differences in group means were not interpreted as statistically significant without appropriate inferential evidence. The descriptive statistics of the six eye-tracking indicators are summarized in Supplementary Table S1.

Pearson correlation analysis was subsequently conducted to examine the relationships between image-level visual-attention indicators and aesthetic-preference scores (Table 7). FF showed a weak positive correlation with aesthetic preference ($r = 0.107$, $p = 0.588$), whereas the proportion of fixation duration (PFD; $r = -0.128$, $p = 0.516$) and AFD ($r = -0.110$, $p = 0.577$) exhibited weak negative correlations. Similarly, SF ($r = 0.080$, $p = 0.686$), average saccade amplitude (ASA; $r = 0.030$, $p = 0.880$), and average saccade velocity (ASV; $r = 0.070$, $p = 0.723$) showed weak associations with aesthetic preference. None of the six correlations reached statistical significance at the image level (all $p > 0.05$). As illustrated in Fig. 8, the observed relationships between global eye-movement measures and aesthetic-preference scores were generally weak. These findings suggest that scene-level fixation and saccade indicators alone provided limited information about differences in mean aesthetic preference across the examined images. Eye-movement behaviour and subjective evaluation were therefore treated as complementary rather than interchangeable measures of visual experience, consistent with their distinct analytical roles in previous architectural and heritage-perception studies (de la Fuente Suárez, 2020; Xing and Leng, 2024; Xu et al., 2024).

Table 7. Pearson correlations between image-level eye-tracking indicators and aesthetic-preference scores (n = 28).

Landscape component	**Pearson's r**	***p* value**
Sky	0.194	0.323
Distant mountain	0.001	0.996
Tree	-0.093	0.638
Water	0.746	＜0.001
Traditional buildings	-0.430	0.022
Vegetation	0.154	0.434

Note: The p-values were calculated from the reported correlation coefficients using two-sided Pearson tests with 28 image-level observations. All values should be verified against the original dataset before submission.

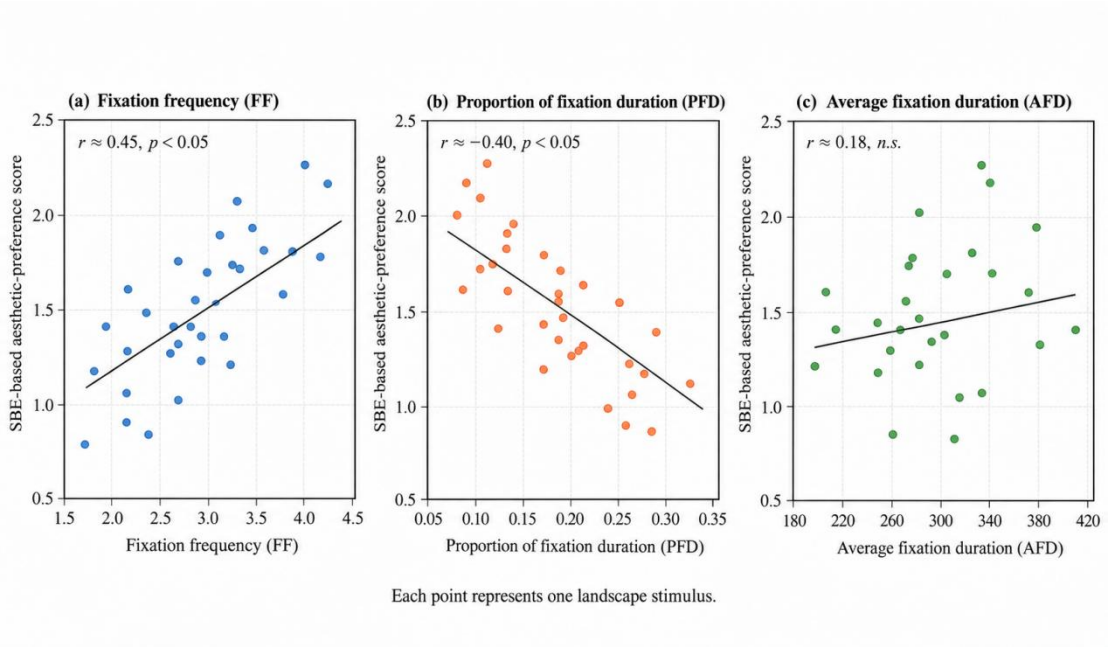


Fig. 8. Relationships between visual-attention indicators and aesthetic-preference scores across the 28 traditional village landscape scenes: (a) fixation frequency (FF), (b) proportion of fixation duration (PFD), and (c) average fixation duration (AFD). Each point represents one landscape stimulus.

3.5. Comparative performance of landscape composition, visual attention, and integrated models

Three regression models were compared to assess the relative explanatory and predictive contributions of landscape composition and visual attention to image-level aesthetic preference (Table 8). The landscape-composition model (Model 1), incorporating water and building-surface coverage, achieved an adjusted R^2 of 0.56, compared with 0.03 for the visual-attention model (Model 2). The integrated model (Model 3), combining landscape-composition variables with fixation frequency (FF) and proportion of fixation duration (PFD), achieved the highest adjusted R^2 of 0.60. Within Model 3, water coverage was positively associated with aesthetic preference ($B = 0.036$, $p < 0.001$), whereas building-surface coverage showed a negative association ($B = -0.018$, $p = 0.017$). In contrast, neither FF ($B = 0.240$, $p = 0.195$) nor PFD ($B = -0.820$, $p = 0.247$) exhibited a statistically significant independent association. These results suggest that the selected landscape-composition variables provided greater explanatory information about aesthetic preference than the global eye-tracking indicators.

Table 8. Performance comparison of models explaining aesthetic preference

Model	Predictors retained	Adjusted R^2	AIC	MAE	RMSE	LOOCV-RMSE
Model 1	Water; Traditional	0.56	−47.68	0.27	0.34	0.44

Landscape composition	buildings					
Model 2	FF; PFD; AFD*	0.03	-25.48	0.40	0.49	0.63
Visual attention						
Model 3	Water; Traditional	0.60	-48.41	0.25	0.32	0.43
Composition + visual	buildings; FF;					
attention	PFD					

Note: Adjusted R^2 , adjusted coefficient of determination; AIC, Akaike Information Criterion; MAE, mean absolute error; RMSE, root mean square error; LOOCV-RMSE, leave-one-image-out cross-validated root mean square error; FF, fixation frequency; PFD, proportion of fixation duration; AFD, average fixation duration. All models were evaluated at the image level (n = 28). Lower AIC and prediction errors indicate better model performance when calculated using consistent definitions and observations.

The model comparison further indicated that incorporating visual-attention variables provided only a modest improvement in predictive performance. The reported Akaike Information Criterion (AIC) values were −47.68, −25.48, and −48.41 for Models 1, 2, and 3, respectively. Leave-one-image-out cross-validation (LOOCV) yielded reported root mean square errors (RMSEs) of 0.44, 0.63, and 0.43, respectively. Although Model 3 exhibited the lowest cross-validated prediction error, the reduction relative to Model 1 was only 0.01. The comparative performance and regression coefficients of the three models are presented in Fig. 9, while Fig. 10 illustrates the relationship between observed aesthetic-preference scores and cross-validated predictions from the integrated model. Given the limited number of independent image-level observations (n = 28), the small improvement in predictive performance should be interpreted cautiously. Overall, these results suggest that landscape composition provided the principal explanatory information about aesthetic preference, whereas the selected global eye-tracking indicators contributed limited additional predictive value.

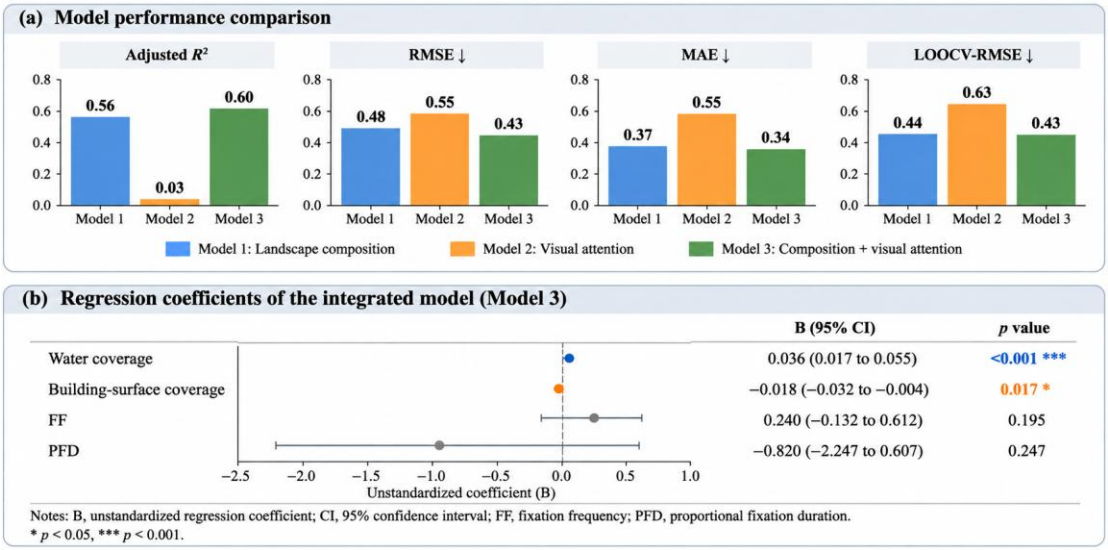


Figure 9. Comparative performance of landscape-composition, visual-attention, and integrated models for predicting aesthetic preference. (a) Comparison of adjusted R^2 , root mean square error (RMSE), mean absolute error (MAE), and leave-one-image-out cross-validated RMSE (LOOCV-RMSE) across the three models. (b) Unstandardized regression coefficients (B) and

corresponding 95% confidence intervals for the predictors included in the integrated model (Model 3). The vertical dashed line indicates a regression coefficient of zero. All analyses were performed at the image level (n = 28).

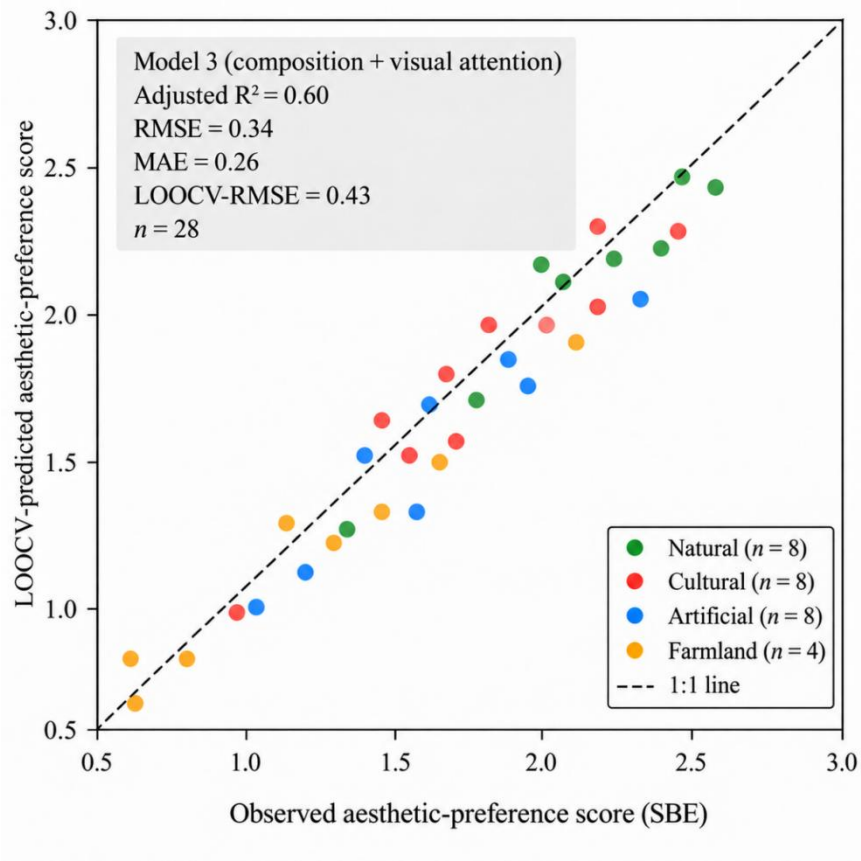


Figure 10. Relationship between observed and leave-one-image-out cross-validated predicted aesthetic-preference scores for the integrated model (Model 3). Each point represents one landscape image (n = 28), with colors distinguishing natural, cultural, artificial, and farmland landscape groups. The dashed diagonal line represents perfect agreement between observed and predicted scores (1:1 line).

4. Discussion

4.1. Landscape composition and aesthetic preference

The findings indicate that the visible composition of traditional village heritage landscapes is associated with variations in aesthetic preference, although the strength and direction of these associations differ across landscape components. Water coverage exhibited a strong positive correlation with aesthetic-preference scores ($r = 0.746$, $p < 0.001$), whereas building-surface coverage showed a moderate negative correlation ($r = -0.430$, $p = 0.022$). In contrast, sky, distant mountains, trees, and vegetation showed no statistically significant linear associations with aesthetic preference. These findings

suggest that individual landscape components differ in their relationships with aesthetic evaluation and that their proportional coverage may provide more specific information than broad landscape classifications alone. Previous research has emphasized the importance of measurable landscape characteristics in understanding landscape aesthetic perception and has demonstrated the potential of image-based landscape-element analysis for evaluating historic environments (Ode et al., 2008; Jin et al., 2025). Nevertheless, the relationships identified in the present study should be interpreted as image-level statistical associations rather than evidence of causal effects.

The positive relationship between water coverage and aesthetic preference is particularly relevant to the historical landscape characteristics of Hongcun and Xidi. Water systems constitute important elements of the spatial organization and visual identity of these traditional settlements, creating distinctive relationships among historic architecture, vegetation, open spaces, and surrounding topography. Scenes containing greater water coverage may also provide more open views, reflections, and visually coherent combinations of natural and architectural elements, potentially contributing to higher aesthetic evaluations. Previous research on heritage environments has indicated that water features and reflections may influence visual attention and subjective spatial experience (Sun and Bao, 2025). Related research on landscape perception has also demonstrated the potential of quantitative image analysis for examining the associations between visible environmental elements and subjective evaluations (Luo et al., 2022; Jin et al., 2025). However, the present study did not independently examine reflections, spatial openness, or other potentially associated characteristics. Therefore, the observed positive correlation should not be interpreted as demonstrating that increasing water coverage would necessarily improve aesthetic preference.

The negative association between building-surface coverage and aesthetic preference requires a similarly cautious interpretation. Traditional architecture represents a fundamental component of the historical and cultural significance of Huizhou villages, but immediate aesthetic preference and architectural heritage value are not equivalent. Greater building-surface coverage may coincide with reduced visibility of natural elements, stronger spatial enclosure, or greater visual complexity, although these explanations were not directly tested. Previous studies have suggested that the relationship between physical landscape characteristics and subjective evaluation depends on the composition and environmental context of the observed scenes (Ode et al., 2008; Jin et al., 2025). Furthermore, because the ADE20K-based semantic segmentation model does not inherently distinguish traditional Hui-style buildings from contemporary buildings, the association should be described in terms of building-surface coverage unless a separate identification or annotation procedure has been validated (Zhou et al., 2017; Xie et al., 2021). The observed correlation also requires consideration of compositional dependence, because a greater proportion of one landscape component necessarily reduces the relative proportions available to other components. Taken together, these findings suggest that the aesthetic evaluation of traditional village heritage landscapes is associated with the proportional configuration

of natural and built elements, while their historical and cultural significance must be considered separately from immediate visual preference.

4.2. Landscape composition and visual-attention allocation

Landscape composition and visual-attention allocation represent two closely related but distinct dimensions of visual perception in traditional village heritage environments. Semantic segmentation quantifies the proportional coverage of visible landscape components, whereas eye tracking reveals how observers distribute their attention across these components. However, the visual prominence of a landscape element does not necessarily correspond to the amount of attention it receives. Previous research has demonstrated that photographic properties, landscape characteristics, and observers' expertise can influence fixation patterns and visual exploration (Dupont et al., 2014; Dupont et al., 2015). Similarly, Evans and Chamberlain (2024) emphasized the methodological challenges of interpreting eye-tracking measurements in landscape visual assessment, highlighting the need to distinguish observable gaze behaviour from subjective environmental evaluation. These studies suggest that landscape-component coverage and fixation-time allocation should be examined as complementary measures rather than interchangeable indicators of visual importance.

This distinction is particularly relevant to Hongcun and Xidi, where traditional buildings, water systems, vegetation, farmland, and contemporary facilities coexist within complex visual settings. Architectural details may encourage concentrated visual inspection, whereas water and vegetation may contribute to the overall spatial character of a scene without necessarily receiving fixation time proportional to their visible coverage. Previous research has identified differences in attention allocation across environmental elements by combining semantic classification with AOI-based eye-tracking analysis (Chen et al., 2022). Related investigations of historical landscapes have further demonstrated the usefulness of component-specific AOIs for examining visual behaviour associated with architectural features (2024). Meanwhile, Jin et al. (2025) quantified visible landscape elements in Wuzhen using AI-assisted image recognition and scenic beauty evaluation, demonstrating the relevance of image-derived landscape composition to the study of historic settlements. Building on these complementary approaches, the present study aligns SegFormer-derived semantic masks with fixation coordinates to examine the correspondence between landscape-component coverage and visual-attention allocation (Xie et al., 2021).

The integration of semantic segmentation and eye tracking provides a component-level framework for distinguishing physical landscape composition from behavioural attention allocation. Unlike global fixation and saccade indicators, AOI-based measurements can identify the landscape categories receiving attention and quantify their fixation-time shares relative to their visible coverage. Nevertheless, differences between pixel coverage and fixation-time allocation do not necessarily establish an intrinsic attentional preference for particular landscape elements, because both are relative measures influenced by scene composition, spatial position, and viewing conditions. Previous landscape-perception research has similarly indicated

that visual attention depends on interactions between environmental characteristics and observer-related factors (Dupont et al., 2015; Evans and Chamberlain, 2024). Accordingly, the contribution of this approach lies in providing a more differentiated representation of landscape visibility and visual-attention allocation, while recognizing that physical prominence, gaze behaviour, and aesthetic preference represent distinct aspects of heritage landscape perception.

4.3. Visual attention and the prediction of aesthetic preference

The image-level correlation analysis revealed weak relationships between global eye-movement indicators and aesthetic-preference scores, with none of the six indicators reaching statistical significance. These findings suggest that fixation and saccade behaviour does not necessarily reflect positive aesthetic evaluation. Longer fixations may indicate visual interest, complexity, or increased information-processing demands rather than aesthetic appreciation alone. Previous research has similarly demonstrated that visual exploration is influenced by landscape characteristics, photographic properties, and observer-related factors (Dupont et al., 2014; Dupont et al., 2015). Accordingly, global eye-movement indicators should be interpreted as behavioural measures rather than direct indicators of landscape aesthetic quality (Evans and Chamberlain, 2024).

The comparative regression results further suggest that landscape composition provides greater explanatory information about aesthetic preference than global visual-attention indicators. The composition-based model achieved an adjusted R^2 of 0.56, compared with 0.03 for the attention-based model, while the integrated model reached 0.60. However, the improvement in cross-validated prediction error was relatively small, indicating limited additional predictive value from the selected eye-movement indicators. This may be because global fixation and saccade measures summarize overall viewing behaviour without identifying which landscape elements receive attention. Previous studies have emphasized the importance of measurable landscape characteristics in aesthetic evaluation and the complementary role of behavioural observations (Ode et al., 2008; Jin et al., 2025). Nevertheless, the modest improvement in model performance should be interpreted cautiously until the statistical results have been fully verified.

The findings highlight the value of distinguishing physical landscape composition, visual-attention behaviour, and subjective aesthetic preference within a common analytical framework. Rather than assuming direct correspondence among these dimensions, the study examines their relative explanatory contributions using image-level observations. Because the analysis was based on 28 landscape images and separately collected participant datasets, the results do not establish individual-level relationships or causal pathways. Overall, landscape composition appears to provide the principal explanatory information for aesthetic differences among the examined traditional village scenes, whereas global eye-tracking indicators offer complementary behavioural evidence with limited additional predictive value.

5. Implications, limitations, and future research

5.1. Implications for heritage conservation and landscape management

The findings offer implications for the visual assessment and conservation management of traditional village heritage landscapes. First, heritage landscape assessment should move beyond broad landscape classifications to consider the composition and spatial relationships of individual landscape elements. Recent studies using semantic segmentation and image parsing have demonstrated the potential of pixel-level analysis for characterizing complex heritage landscapes and their visual compositions (Chai et al., 2026; Huang et al., 2026). In the present study, the positive association between water coverage and aesthetic preference highlights the potential importance of historically formed water–settlement relationships in Huizhou traditional villages. Previous experimental research has also demonstrated that water reflections can influence visual attention and subjective spatial experience in Chinese heritage gardens (Sun and Bao, 2025). Nevertheless, the observed association does not imply that increasing water coverage would necessarily improve landscape quality. For traditional villages such as Hongcun and Xidi, conservation strategies should instead emphasize the integrity, continuity, and visual accessibility of historic water systems and their spatial relationships with traditional buildings, vegetation, open spaces, and surrounding topography. Such an approach prioritizes the preservation of established heritage landscape characteristics rather than the artificial enhancement of individual visual elements.

Second, the negative association between building-surface coverage and aesthetic preference should not be interpreted as evidence that traditional architecture reduces aesthetic or cultural value. A greater proportion of visible building surfaces may coincide with stronger spatial enclosure, reduced visibility of natural elements, or greater architectural complexity, although these potential explanations were not independently tested in the present study. Moreover, aesthetic preference represents only one dimension of heritage significance and cannot replace assessments of authenticity, integrity, historical value, and cultural meaning. Previous research has demonstrated the importance of combining objective landscape characteristics with subjective perceptions when evaluating traditional village environments (Jin et al., 2025; Shen et al., 2025). The practical contribution of the present framework lies in distinguishing the physical composition of heritage scenes, their visual-attention characteristics, and their aesthetic evaluations. When supported by reliable semantic classification and gaze measurements, these complementary forms of information can assist heritage managers in identifying changes in water visibility, vegetation–building relationships, contemporary surface coverage, and visually intrusive facilities. Such evidence should be used alongside architectural surveys and expert-based heritage assessments rather than as an independent criterion for conservation decisions.

5.2. Limitations and future research

Several limitations should be acknowledged. First, the study was based on 28 standardized two-dimensional photographs from Hongcun and Xidi, which provided consistent visual stimuli for comparing landscape composition, eye-tracking behaviour, and aesthetic preference but could not fully reproduce the dynamic, three-dimensional, and multisensory experience of visiting traditional villages. Visual perception under real-world conditions may also be influenced by viewing sequence, walking direction, lighting, seasonal variation, sound, crowd density, and familiarity with the environment. Furthermore, the limited number of photographs, the unequal distribution of scenes among landscape groups, and the geographical concentration of the study sites constrain the generalizability of the findings. The participant samples were not designed to represent the wider population, and differences in individual experience and cultural background may therefore remain insufficiently captured. Previous research using immersive virtual environments and three-dimensional eye tracking has demonstrated the potential for investigating spatial cognition and viewing behaviour in rural heritage environments (Zhang et al., 2024). Future studies should consequently expand the number and diversity of landscape stimuli, include traditional villages from different geographical and cultural contexts, and incorporate mobile eye tracking, panoramic imagery, or immersive environments to examine the stability of the observed relationships under more realistic conditions.

Second, several methodological constraints should be considered when interpreting the integrated analysis. Semantic segmentation of traditional village environments may introduce classification uncertainty, particularly for dark-tiled roofs, white walls, water reflections, vegetation, and visually similar built surfaces. The distinction between traditional and contemporary building surfaces also cannot be inferred reliably from general-purpose semantic labels alone. Accordingly, the accuracy of the landscape-component variables requires transparent documentation and validation using manually annotated reference data and class-specific performance indicators. In addition, the three datasets originated at different analytical levels: landscape composition was measured at the image level, whereas eye-tracking and aesthetic-preference observations were initially collected from participants and subsequently aggregated by stimulus. Consequently, the reported relationships represent image-level statistical associations rather than individual-level perceptual mechanisms or causal effects. The relatively small number of independent image observations also limits the complexity and stability of regression models, making careful predictor selection and out-of-sample validation particularly important. Future research using larger stimulus datasets and individually matched repeated observations could employ mixed-effects models to distinguish participant- and image-level variability, assess the robustness of component-specific attention patterns, and further investigate the contribution of visual-attention behaviour to aesthetic preference.

6. Conclusions

This study investigated the relationships among landscape composition, visual attention, and aesthetic preference in traditional village heritage landscapes by integrating SegFormer-based semantic segmentation, eye-tracking analysis, and SBE-based aesthetic evaluation. Using 28 standardized landscape photographs from Hongcun and Xidi, together with 374 valid questionnaire responses and eye-tracking observations from 112 participants, the study examined physical, behavioural, and perceptual characteristics within a common image-level analytical framework. Natural landscapes received the highest mean aesthetic-preference scores, followed by cultural, artificial, and farmland landscapes. At the image level, water coverage showed a strong positive correlation with aesthetic preference ($r = 0.746$, $p < 0.001$), whereas building-surface coverage exhibited a moderate negative correlation ($r = -0.430$, unadjusted $p = 0.022$). These findings indicate that landscape-component proportions are associated with aesthetic evaluation to different degrees, although building coverage should not be interpreted as a direct indicator of architectural or cultural heritage value.

The eye-tracking analysis revealed comparatively weak relationships between global visual-attention indicators and aesthetic preference, with none of the six image-level correlations reaching statistical significance. The comparative modelling results further indicated that landscape composition provided substantially greater explanatory information than global eye-movement measures. The composition-based model achieved an adjusted R^2 of 0.56, compared with 0.03 for the attention-based model, while the integrated model reached 0.60. However, the reduction in leave-one-image-out cross-validated RMSE from 0.44 to 0.43 indicated only a modest improvement in predictive performance after incorporating visual-attention indicators. These findings suggest that physical landscape composition and visual attention capture related but distinct aspects of heritage landscape perception and that increased visual attention does not necessarily correspond to stronger aesthetic preference.

The principal contribution of this study lies in quantitatively distinguishing the physical composition of traditional village landscapes, their visual-attention characteristics, and their relationships with aesthetic evaluation. The findings provide a basis for complementing conventional heritage landscape assessment with measurable visual-composition and behavioural information, particularly when examining historically formed relationships among water systems, traditional buildings, vegetation, and surrounding environments. Nevertheless, the limited number of landscape stimuli, the geographical concentration of the study sites, potential semantic-classification uncertainty, and the aggregation of participant observations at the image level constrain the generalizability of the conclusions. Future research should incorporate larger and more diverse landscape datasets, independently validated semantic annotations, and participant-level repeated-measures analyses to further investigate how landscape composition and visual attention jointly relate to aesthetic preference. Such developments may strengthen evidence-based approaches to the

conservation and visual-integrity management of living traditional village heritage landscapes.

Ethics statement

This research was approved by the Institutional Review Board (IRB) of West Anhui University [202506003]. Any changes made to the agreement will be reviewed by the IRB. The results of this study will be disseminated by publication in a peer-reviewed journals.

Acknowledgements

The authors would like to thank those who provided helpful comments and support during the development of this study.

Data availability

The anonymized data supporting the findings of this study, including eye-tracking measurements, questionnaire responses, and derived analytical data, are available from the corresponding author upon reasonable request. Participant information was collected anonymously during both the laboratory eye-tracking experiment and the online questionnaire survey. Any sharing of participant-level records will comply with applicable research ethics and confidentiality requirements.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by the 2025 Domestic Visiting Scholar Funding Project of West Anhui University (Grant No. wxxygnfx2025003); the Zhejiang Provincial Philosophy and Social Sciences Planning Project (Grant No. 25NDJC058YBM); the High-Level Talent Funding Program of Hefei Normal University for the project “Canonical Interpretation and Knowledge Reconstruction in Translations of the Shijing (Book of Songs) into Western Languages” (Grant No. 2026RCJJ59); the 2025 University-Level Teaching Research Project (General Category) of West Anhui University, entitled “Exploration and Practice of an Embodied Cognition-Based Teaching Model for the ‘Introduction to Architectural Design’ Course” (Grant No. wxxy2025071); and the Talent Research Project of Chizhou University, entitled “Research on the Coupling Mechanisms of the Integrated Development of Agriculture, Culture, and Tourism in Southern Anhui from the Perspective of Ecological Civilization” (Grant No. CZ2025YJRC137).

References

1. Chen, X. H. & Chen, Q. Q. Review of the studies on public space of China’s traditional villages. J. Anhui Agric. Univ. (Soc. Sci. Ed.) 28, 25–32, 58 (2019).

2. Wu, C., Chen, M., Zhou, L., Liang, X. & Wang, W. Identifying the spatiotemporal patterns of traditional villages in China: A multiscale perspective. Land 9, 449 (2020).

3. Liu, Y. et al. A study on identifying the spatial characteristic factors of traditional streets based on visitor perception: Yuanjia Village, Shaanxi Province. Buildings 14, 1815 (2024).

4. Chen, G., Yan, J., Wang, C. & Chen, S. Expanding the associations between landscape characteristics

and aesthetic sensory perception for traditional village public space. Forests 15, 97 (2024).
5. Meng, S., Liu, C., Zeng, Y., Xu, R. & Zhang, C. Quality evaluation of public spaces in traditional villages: A study using deep learning and panoramic images. Land 13, 1584 (2024).
6. Shen, H., Aziz, N. F. & Lv, X. Using 360-degree panoramic technology to explore the mechanisms underlying the influence of landscape features on visual landscape quality in traditional villages. Ecol. Inform. 86, 103036 (2025).
7. Ode, Å., Tveit, M. S. & Fry, G. Capturing landscape visual character using indicators: Touching base with landscape aesthetic theory. Landsc. Res. 33, 89–117 (2008).
8. Hanyu, K. Visual properties and affective appraisals in residential areas in daylight. J. Environ. Psychol. 20, 273–284 (2000).
9. Dupont, L., Antrop, M. & Van Eetvelde, V. Eye-tracking analysis in landscape perception research: Influence of photograph properties and landscape characteristics. Landsc. Res. 39, 417–432 (2014).
10. Cottet, M. et al. Using gaze behavior to gain insights into the impacts of naturalness on city dwellers' perceptions and valuation of a landscape. J. Environ. Psychol. 60, 9–20 (2018).
11. de la Fuente Suárez, L. A. Subjective experience and visual attention to a historic building: A real-world eye-tracking study. Front. Archit. Res. 9, 774–804 (2020).
12. Xing, Y. & Leng, J. Evaluation of public space in traditional villages based on eye tracking technology. J. Asian Archit. Build. Eng. 23, 125–139 (2024).
13. Xu, X., Dong, R., Li, Z., Jiang, Y. & Genovese, P. V. Research on visual experience evaluation of fortress heritage landscape by integrating SBE–SD method and eye movement analysis. Herit. Sci. 12, 281 (2024).
14. Zhang, H., Andrade, B., Wang, X., Aburabee, I. & Yuan, S. A study of spatial cognition in the rural heritage based on VR 3D eye-tracking experiments. Herit. Sci. 12, 141 (2024).
15. Chen, J. et al. Integrating the SBE–SD method and eye-tracking analysis for evaluating visual experience in traditional village heritage landscapes. Buildings 16, 3747 (2026).
16. 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. 228, 104569 (2022).
17. Chai, H. et al. Semantic segmentation and spatial grid analysis of Chinese heritage landscape photographic compositions with cross-cultural perspectives. Herit. Sci. 14, 176 (2026).
18. Huang, X. et al. Modeling visual perception of Chinese classical private gardens with image parsing and interpretable machine learning. Herit. Sci. 14, 105 (2026).
19. Sun, M. & Bao, Y. Effects of reflecting water on viewers' visual attention, spatial perception, and emotional responses: A case study in Chinese heritage garden. Herit. Sci. 13, 26 (2025).
20. UNESCO World Heritage Centre. Ancient villages in southern Anhui—Xidi and Hongcun. https://whc.unesco.org/en/list/1002 (2026).
21. Xie, E. et al. SegFormer: Simple and efficient design for semantic segmentation with transformers. Adv. Neural Inf. Process. Syst. 34, 12077–12090 (2021).
22. Zhou, B. et al. Scene parsing through ADE20K dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 633–641 (2017).
23. Klaib, A. F., Alsrehin, N. O., Melhem, W. Y., Bashtawi, H. O. & Magableh, A. A. Eye-tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and Internet of Things technologies. Expert Syst. Appl. 166, 114037 (2021).
24. Yao, X. & Sun, Y. Using a public preference questionnaire and eye movement heat maps to identify the visual quality of rural landscapes in Southwestern Guizhou, China. Land 13, 707 (2024).

25. Guo, S. L. et al. Landscape visual quality assessment based on eye movement: College student eye-tracking experiments on tourism landscape pictures in Hongcun. Resour. Sci. 39, 1137–1147 (2017).
26. Ren, H., Yang, F., Zhang, J. & Wang, Q. Evaluation of cognition of rural public space based on eye-tracking analysis. Buildings 14, 1525 (2024).
27. Zhang, X., Xiong, X., Chi, M., Yang, S. & Liu, L. Research on visual quality assessment and landscape elements influence mechanism of rural greenways. Ecol. Indic. 160, 111844 (2024).
28. 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 (2024).
29. Jin, W., Zhu, B. & Fukuda, H. Research on ancient town style construction strategies based on coupled quantitative analysis of AI visual recognition and scenic beauty evaluation. Front. Archit. Res. 14, 654–671 (2025).
30. Kang, N. & Liu, C. Assessment of visual quality and social perception of cultural landscapes: Application to Anyi traditional villages, China. Herit. Sci. 12, 235 (2024).
31. Gao, Z., Wu, C., Li, N., Wang, P. & Li, J. Visual aesthetic quality of Qianjiangyuan National Park landscapes and its spatial pattern characteristics. Forests 15, 1289 (2024).
32. Weng, Y., Zhu, Y. & Ma, S. Quantitative analysis of physiological and psychological impacts of visual and auditory elements in Wuyishan National Park using eye-tracking. Forests 15, 1210 (2024).
33. Wang, Z. et al. Delineating landscape features perception in tourism-based traditional villages: A case study of Xijiang Thousand Households Miao Village, Guizhou. Sustainability 16, 5287 (2024).
34. Xu, J. et al. Assessing health-regulating ecosystem services in rural landscapes: An integrated approach using physiological and visual semantic data. J. Environ. Manag. 413, 130389 (2026).
35. Hsu, K.-W., Chao, J.-C. & Ko, H.-W. Tangible and intangible landscape elements in the adaptive reuse of historic railway corridors: Landscape perception and preference along the Taichung Green Corridor 1908, Taiwan. Land 15, 1639 (2026).