

<https://doi.org/10.1038/s40494-025-02210-y>

# Exploring spatial visual characteristics of scenic archetypes through AI multimodal mapping methods in Hangzhou Westlake

Check for updates

Junkai Lan<sup>1</sup>, Mei Liu<sup>2</sup>✉, Eric Luiten<sup>1</sup>, Gregory Bracken<sup>1</sup> & Qian Zhang<sup>3</sup>

Traditional Chinese gardens embody sophisticated spatial design principles often described through abstract terms like “scenic archetypes,” yet systematic methods for analyzing their visual spatial characteristics remain underdeveloped. This study establishes an analytical framework integrating phenomenological theory with AI-enabled multimodal mapping to quantify spatial visual characteristics of four scenic archetypes, including framed, obstructive, porous, and sandwiched scenery, at Hangzhou West Lake. By decomposing scenic compositions and configurations into foreground–middle–background hierarchies characterized through shape, size, position, and texture variables, the framework achieves 94.12% classification accuracy via random forest modeling while revealing each archetype. Statistical analysis identifies archetype-specific spatial strategies: framed scenery employs regular foreground geometry with smooth depth transitions; obstructive scenery utilizes systematic positioning with texture contrasts; porous scenery balances visual permeability with textural variation; sandwiched scenery creates bilateral symmetry with channeling effects. This approach provides replicable methodology for heritage conservation and contemporary landscape design informed by traditional spatial wisdom.

From a design perspective, landscapes are three-dimensional constructions that evolve over time, involving the articulation of abstract notions into physical structures<sup>1</sup>. These structures warrant analysis through visual appreciation, particularly examining the inherent attributes and qualities manifest in the composition and configuration of spatial elements, collectively termed “spatial visual characteristics”<sup>2,3</sup>. These characteristics encompass two fundamental dimensions: first, “what exists” referring to the presence, quantity, distribution, and proportion of spatial elements; and second, “how they are arranged” encompassing the complexity of arrangement, position, orientation, and form<sup>4,5</sup>.

Viewed through this lens, traditional Chinese gardens, a type of landscape architecture, are not as random and loose as they seem; rather, their spatial visual characteristics are carefully orchestrated to create physical structures with highly complex spatial hierarchies from multiple perspectives<sup>6–8</sup>. Building on these characteristics, prominent landscape architects and scholars such as Peng Yigang, Zhou Weiquan and Pan Guxi have identified recurring patterns of spatial composition and configuration, which they termed “scenic archetypes” in traditional Chinese vocabulary and concepts<sup>9–11</sup>. Scenic archetypes translate specific combinations of spatial visual characteristics into culturally embedded organizational principles

that guide the arrangement of space to produce distinctive aesthetic experiences<sup>12</sup>. The influence of these archetypes extends far beyond regional boundaries. Historical evidence demonstrates that the spatial logic they embody has profoundly shaped garden design across cultural contexts, from East Asia to Europe, with particularly notable impacts in Japan, South Korea, and the United Kingdom<sup>13</sup>. This transregional significance is further reflected in the international recognition of several representative gardens, including the Humble Administrator’s Garden, the Linger Garden, and Hangzhou Westlake, all inscribed on the UNESCO World Heritage List in acknowledgment of their enduring design wisdom and innovation<sup>13–15</sup>. In parallel with this international visibility, recent research on landscapes and historic gardens has increasingly emphasized a more precise understanding of spatial organization, particularly as represented by the spatial visual characteristics of scenic archetypes, whose underlying principles have influenced landscape design at broader scales<sup>16</sup>.

However, previous studies on scenic archetypes remains predominantly grounded in experiential descriptions, which limits their potential for systematic knowledge transfer and methodological development in heritage conservation. Early scholars often characterized these archetypes through poetic metaphors such as “scenes changing as steps

<sup>1</sup>Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, The Netherlands. <sup>2</sup>School of Architecture, Harbin Institute of Technology, Shenzhen, China. <sup>3</sup>Department of Engineering, College of Charleston, Charleston, SC, USA. ✉e-mail: [mei.liu@hit.edu.cn](mailto:mei.liu@hit.edu.cn)

move (步移景异)”, “perceiving the vast in the small (小中见大)” and “winding paths leading to secluded spots (曲径通幽)”<sup>9,17</sup>. While such expressions convey the phenomenological essence, they represent personal interpretation rather than spatial attributes that can be objectively described and consistently verified by different observers. More recent studies have sought to document the spatial visual characteristics of scenic archetypes with greater precision: Wang<sup>18</sup> investigated the specific “size” and “scale” of windows and doors in framed scenery; Tong<sup>19</sup> utilized binary oppositions such as “sparse-dense” and “tortuous-straight” to analyze the spatial configurations of obstructive and framed scenery at the Linger Garden’s entrance. Nevertheless, as Liu<sup>3</sup> observes, even these refined vocabularies suffer from terminological inconsistencies and a lack of standardization, with identical terms interpreted variably across studies, thus hindering systematic analysis.

To overcome the limitations of purely descriptive approaches, recent studies have introduced quantitative and computational methods, essentially mapping techniques, to measure spatial visual characteristics with greater precision<sup>1,3,20</sup>. As the fundamental means of capturing, analyzing, and communicating such characteristics, mapping visualizes abstract spatial knowledge and integrates the spatial organization of landscape spaces in both qualitative and quantitative terms<sup>21</sup>, embodying what Corner<sup>22</sup> describes as “ways of seeing” that construct new realities rather than merely recording existing ones. For example, Zhou et al.<sup>23</sup> used angle segment and visibility graph analyses to quantify Daming Temple’s mean depth, connectivity and intelligibility. Chen and Yang<sup>24</sup> linked narratology with VGA and isovist analysis to calculate integration and connectivity in the Humble Administrator’s Garden. Zhang et al.<sup>25</sup> employed DepthmapX for convex, axial and visibility analyses, deriving depth, connectivity, isovist area and integration at West Shu Garden. Chen et al.<sup>26</sup> combined space-syntax indicators (connectivity, step depth, integration) with DBSCAN clustering to extract five indicators: permeability, curvature, visibility, accessibility and differentiation. While these methods offer rigorous tools for measuring individual spatial visual characteristics, they lack the capacity to capture how such characteristics integrate into the coherent spatial gestalts that define scenic archetypes. Addressing this gap calls for approaches capable of linking detailed element-level detection with the holistic perception of design patterns, and one promising direction lies in the use of advanced computer vision mapping. Today’s computer vision technologies represent a transformative expansion of these capabilities: semantic segmentation precisely identifies spatial elements across thousands of images<sup>27,28</sup>, depth estimation reveals three-dimensional relationships from photographs<sup>29,30</sup>, and cutting-edge methods like scale-invariant segmentation<sup>31</sup>, super-resolution mapping<sup>32</sup>, and hyperspectral detection<sup>33</sup> push analytical boundaries even further. Yet a critical limitation still persists: though these technologies excel at element detection, they cannot recognize how elements combine to form scenic archetypes. A wall, for example, may be identified simply as a wall, without understanding its role as the defining “frame” in framed scenery. This disconnect between physical detection and design pattern recognition necessitates developing systematic methods that encode the compositional and configurational logic of scenic archetypes in traditional Chinese gardens.

Above all, understanding the spatial visual characteristics of scenic archetypes is essential for explaining how culturally specific design principles are materialized in physical space and for supporting their conservation and contemporary application. However, existing studies remain largely descriptive or focus on isolated attributes, lacking a systematic framework that can both depict these characteristics holistically and interpret them in a replicable way. This study addresses this gap through two research objectives. First, it establishes an analytical framework that systematically transforms the spatial visual characteristics of scenic archetypes from cultural concepts into measurable attributes enabling holistic representation and verifiable interpretation. Second, it develops an AI-based multimodal mapping methodology that integrates semantic precision with quantitative rigor, ensuring that experiential understanding is preserved while producing measurable, comparable outputs. Using the Hangzhou Westlake (HWL) as a representative case that embodies the full complexity of scenic archetypes,

the study applies the proposed framework to translate tacit design knowledge into explicit, measurable guidelines. The resulting outputs provide a basis for evidence-based conservation strategies and inform contemporary design practices, thereby advancing both the theoretical understanding and the practical implementation of scenic archetype analysis.

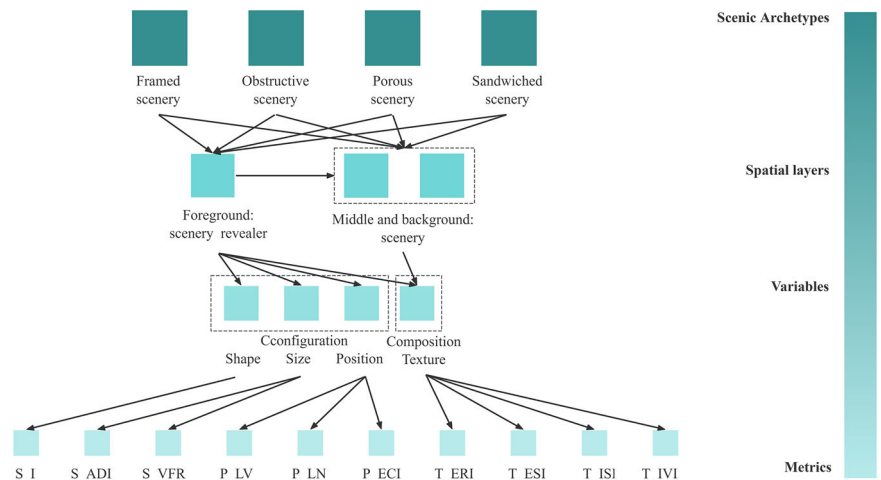
## Methods

### Theoretical foundation and framework overview

To transform scenic archetypes into their constituent spatial visual characteristics, this study first integrates semiotic and phenomenological theories to structure the decomposition process. In semiotics, Peirce’s<sup>34</sup> triadic relationship among sign, object, and interpretant provides a conceptual structure for explaining how meaning is generated and interpreted in relation to physical form<sup>35</sup>. The phenomenological perspective complements this by explaining the cognitive universality of spatial perception. According to Kant, space constitutes “the subjective condition of sensibility” through which “outer intuition is possible for us”<sup>36</sup>. This conceptualization suggests that spatial perception represents an a priori form of human cognition, implying fundamental universality in how individuals perceive and comprehend spatial configurations. Such universality provides crucial justification: if spatial perception follows common cognitive patterns, these patterns can be systematically identified and interpreted. Merleau-Ponty<sup>37</sup> extends this understanding by emphasizing the dialectical relationship between individual exploration and sensorial responses. His work indicates that spatial experience, while transcending pure subjectivity, emerges from observable interactions between humans and their physical environment. Furthermore, several established analytical frameworks have demonstrated the feasibility of translating broad conceptual categories into detailed spatial-visual measures, including Tveit’s<sup>38</sup> landscape visual characterization scheme, Bell’s<sup>2</sup> framework for aesthetic structure, and Liu’s<sup>3</sup> landscape design syntax. These approaches show how abstract design principles can be systematically translated into concrete spatial attributes through multiple interpretive levels, creating a foundation for developing measurable indicators.

Building upon these theoretical and methodological precedents, the proposed framework establishes four hierarchical tiers: the abstract concept level (scenic archetypes) corresponds to the intentional level of meaning; the dimensional level (spatial layers) reflects phenomenological modes of perception; the attribute level (variables) represents concrete spatial manifestations; and the measurable indicator level (metrics) corresponds to identifiable and quantifiable physical attributes (Fig. 1). The first tier establishes scenic archetypes as umbrella concepts that extract recurring design patterns from Jiangnan garden-making techniques, as documented in foundational studies<sup>3,17,39–42</sup>. These archetypes represent the highest level of abstraction, encapsulating centuries of accumulated design wisdom. The second tier introduces spatial layers, which deconstruct each archetype into foreground, middle ground, and background components from the observer’s horizontal perspective<sup>1,3,43</sup>. This tripartite division reveals a critical mechanism: the foreground functions as a mediating element that transforms middle ground and background components from isolated objects into integrated scenic compositions. Significantly, each archetype manifests a distinctive spatial organization pattern that becomes most apparent through foreground characteristics, establishing these as primary indicators for archetype identification. The third tier operationalizes visual characteristics through variables, categorizing perceptual qualities within each spatial layer into four measurable physical attributes. These variables, shape, position, size, and texture were strategically selected from Bell’s<sup>2</sup> comprehensive inventory of eleven variables based on their capacity to capture essential distinguishing characteristics while minimizing subjective observer bias. Each variable serves a specific analytical function: shape delineates archetypal boundaries and defines spatial enclosure; position reveals hierarchical spatial relationships and compositional strategies; size indicates visual prominence and establishes perceptual hierarchy; and texture differentiates surface treatments and material qualities. The fourth and final tier implements metrics through three complementary quantification

**Fig. 1 | Framework of scenic archetype-spatial layer-characteristic-metric.** This diagram illustrates the four-tier hierarchy that transforms abstract scenic archetypes into measurable metrics by decomposing them into spatial layers (foreground, middle ground, and background) and visual variables (shape, size, position, and texture).



methods: distribution trend analysis, absolute value measurement, and relative relationship assessment. These methods collectively transform traditionally experiential interpretations of spatial visual characteristics into verifiable, reproducible data analyses applicable across all spatial layers. This quantitative approach maintains analytical rigor while preserving the phenomenological richness inherent in scenic archetype appreciation.

**Four-tier analytical framework components**

The first tier of the proposed framework encompasses scenic archetypes, which represent fundamental spatial organizational principles in traditional Chinese gardens. Since this study aims to quantitatively assess the spatial visual characteristics of scenic archetypes rather than measuring observers’ subjective responses, we focus on archetypes whose inherent spatial properties enable systematic analysis. Based on the nine scenic archetypes identified by Lu and Liu<sup>44</sup>, scenic archetypes can be fundamentally distinguished by their operational mechanisms: static archetypes that create stable spatial configurations versus dynamic archetypes that unfold through temporal and kinesthetic experiences. This distinction reflects different modes of spatial engagement and perceptual activation. Static archetypes, comprising framed scenery, obstructive scenery, porous scenery, and sandwiched scenery, operate through fixed spatial relationships that maintain consistent visual qualities across viewing positions. These archetypes manifest as stable compositional structures:

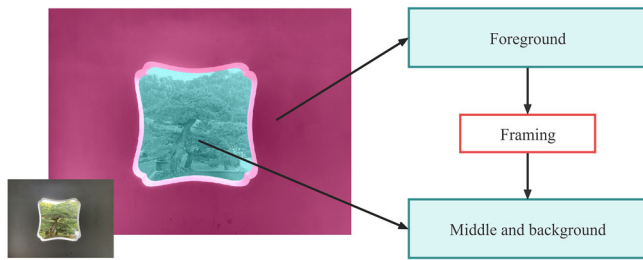
- Framed scenery is composed of four-sided spatial elements, including door frames, window frames, trees, or rocks, which “frame” a specific field of view, producing a “picture frame” effect that highlights the selected scene<sup>9,40,41</sup>.
- Obstructed scenery refers to the partial blockage or interruption of lines of sight through specific spatial elements such as buildings, trees, rocks, or walls, creating a visual effect where the field of view is partially hidden, evoking a sense of mystery<sup>9,17,41,42</sup>.
- Porous scenery refers to the use of spatial elements such as latticed windows, perforated walls, doorways, railings, or bamboo fences that offer partial glimpses of the field of view through openings, resulting in the visual interplay of concealment and exposure<sup>9,17,40</sup>.
- Sandwiched scenery is formed by the placement of spatial elements, such as buildings, trees, rocks, walls, or corridors, on both sides of a field of view. This guides the observer’s sight toward a focal point within the framed scenery, often creating a strong sense of composition and visual direction<sup>9,39,40,42</sup>.

In contrast, dynamic archetypes, including borrowed scenery, hidden scenery, informed scenery, opposite scenery, and segmented scenery, fundamentally rely on temporal unfolding, bodily movement, or cognitive associations that transcend static spatial configurations<sup>9,39,40,44,45</sup>. Du and Ji<sup>46</sup>

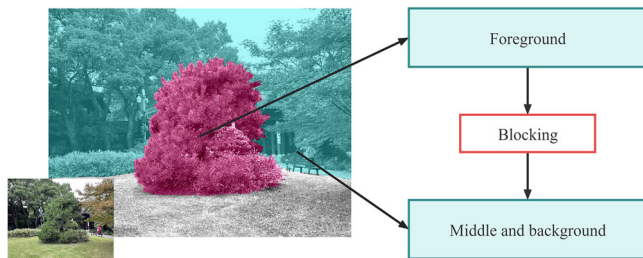
illuminate this distinction through their analysis of “farness” experience in Chinese gardens, where perceived depth fluctuates dramatically with movement: spaces appearing shallow from one position reveal unexpected depth from another. This spatial instability characterizes hidden scenery (藏景) and segmented scenery (隔景), which require kinesthetic exploration to fully manifest. Borrowed scenery (借景) exemplifies a different form of dynamism through its dependence on temporal conditions and intentional cognitive processes. As Lu and Liu<sup>14</sup> demonstrate, this archetype requires distinguishing deliberate visual connections to distant elements from incidental views. This distinction relies on cultural knowledge and atmospheric variability rather than stable spatial relationships. Similarly, opposite scenery (对景) creates reciprocal viewing relationships requiring physical movement between two points to experience the complete spatial dialogue, while informed scenery (点景) operates through metaphorical associations linking physical forms to literary and philosophical concepts<sup>40</sup>.

The selection of the four static archetypes for this study emerges from their inherent potential for systematic analysis. Their stable spatial configurations enable the development of reproducible analytical methods, while contemporary computer vision technologies offer unprecedented capabilities to capture and quantify their compositional logic. This technological potential, combined with the archetypes’ fundamental reliance on measurable spatial relationships, creates opportunities to transform traditionally experiential knowledge into explicit analytical frameworks. Despite these advances, a critical gap persists: no systematic framework currently exists to translate the spatial organizational logic of these static archetypes into quantifiable, reproducible analytical standards.

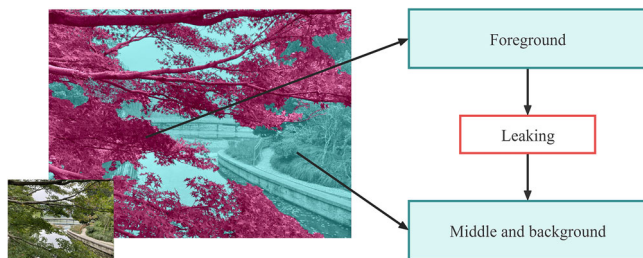
The second tier of our framework comprises spatial layers, which represent the horizontal stratification of scenic archetypes from the observer’s perspective. Each archetype embodies a distinct spatial strategy for directing and modulating visual appreciation through systematic organization of perceptual depth. Spatial layers constitute the visual layout of scenic archetypes as perceived from a horizontal vantage point, specifically from the observer’s eye level during spatial exploration<sup>3,9</sup>. Within this conceptual framework, the observer’s perception of distance and spatial orientation assumes critical importance and conventionally divides into three components: foreground, middle ground, and background<sup>11,43</sup>. Significantly, spatial elements occupying the foreground exert dominant influence in shaping visual perception and mediating the presentation of middle ground and background elements<sup>19,47</sup>. Through this meditative function, the foreground facilitates the transformation of spatial elements in subsequent layers from mere physical objects into integral components of coherent scenery, thereby enabling observers to comprehend and appreciate the compositions and configurations of spatial elements through culturally specific modalities<sup>19</sup>. Consequently, when identifying and differentiating scenic archetypes, their unique spatial visual characteristics derive primarily



**Fig. 2 | Semantic mapping of framed scenery.** The visualization depicts the framing mechanism where the foreground acts as a visual window or boundary, directing the observer's line of sight toward the middle ground focal point.



**Fig. 3 | Semantic mapping of obstructive scenery.** The diagram demonstrates the blocking mechanism where foreground elements (such as trees or walls) create visual barriers that partially conceal the middle and background layers to limit direct viewing.



**Fig. 4 | Semantic mapping of porous scenery.** The visualization illustrates the leaking mechanism where the foreground offers selective views through apertures or gaps, creating focused visual corridors toward the middle ground.



**Fig. 5 | Semantic mapping of sandwiched scenery.** The diagram shows the sandwiching mechanism where bilateral foreground elements channel views in specific directions, making the middle ground the dominant focus through a created visual entrance.

from foreground rather than middle ground and background. The latter function principally as contextual elements that enrich the overall scenic archetypes. Building upon this understanding, this study proposes semantic mappings for each scenic archetype based on the “foreground, and middle ground and background” spatial layers as well as the analysis of framed, obstructive, porous, and sandwiched scenery.

- For framed scenery, the *foreground* functions as a visual window, directing the observers' line of sight through defined boundaries (Fig. 2). Beyond this frame, the *middle ground* forms the primary visual focus through elements such as pavilions or bridges, while the *background* provides contextual support to complete the scenery.

- For obstructed scenery, the *foreground* creates visual barriers that limit direct viewing using elements such as trees or walls (Fig. 3). Both *middle ground* and *background* remain concealed behind these obstacles.

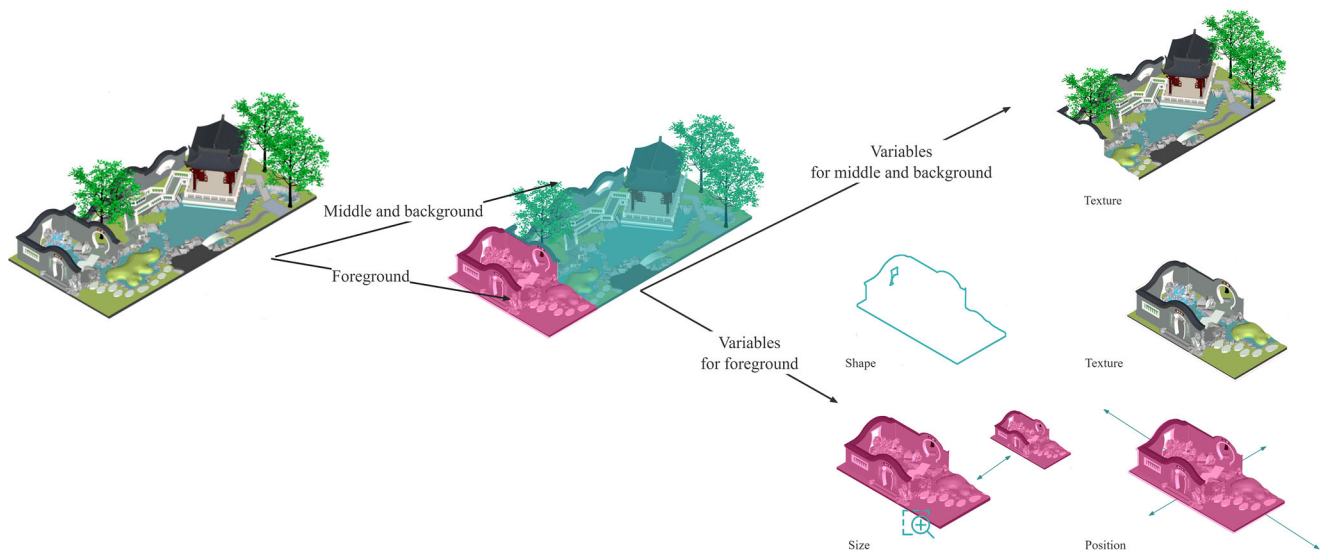
- For porous scenery, the *foreground* offers selective views through gaps, creating focused visual corridors (Fig. 4). The *middle ground* becomes the focal point through these apertures, while the *background* typically includes natural elements like mountains or the sky.

- For sandwiched scenery, the *foreground* uses two major elements to channel views in specific directions (Fig. 5). The *middle ground* becomes the dominant focus through the visual entrance created by these distinct spatial elements, while the open *background* extends the perspective.

The third tier of our framework comprises variables that categorize and measure spatial visual characteristics across the foreground, and middle ground and background of scenic archetypes. Following the establishment of spatial layers for each archetype, the selection of appropriate variables becomes essential for systematic measurement. From Bell's<sup>3</sup> comprehensive inventory of eleven visual variables, this study strategically selects four variables: shape, position, size, and texture (Fig. 6). This selection is grounded in three interconnected rationales that ensure both theoretical rigor and methodological feasibility. First, these four variables demonstrate objectivity and perceptual stability essential for reliable analysis. Visual perception theory establishes shape as the primary characteristic enabling object recognition, with research demonstrating that silhouettes alone suffice for accurate identification<sup>48,49</sup>. Position functions as the foundation of spatial relationships and represents the most accurately perceived dimension according to Gestalt psychology<sup>50</sup>. Size and texture correspond respectively to scale perception and surface characteristic recognition, both fundamental to spatial comprehension. In contrast, other variables proposed by<sup>2</sup>, including color, visual force, and direction, exhibit excessive variability due to lighting conditions, seasonal changes, and viewing angles in garden contexts, thereby lacking the requisite stability for systematic analysis.

Second, these variables align precisely with the spatial organizational principles inherent in traditional Chinese gardens. These gardens achieve specific modes of spatial appreciation through deliberate manipulation of element configuration (shape), dimensional control (size), compositional arrangement (position), and material differentiation (texture)<sup>9</sup>. This alignment manifests distinctly across archetypes: framed scenery delineates visual fields through shape definition; obstructive scenery modulates sight lines through strategic positioning; porous scenery generates perceptual contrast through textural variation; and sandwiched scenery constructs spatial sequences through size relationships. Such correspondence between analytical variables and design principles ensures that the framework captures authentic spatial logic rather than imposing external categories. Third, computer vision technology has achieved sophisticated capabilities in recognizing and quantifying these specific variables. Semantic segmentation accurately extracts element shapes and boundary conditions<sup>27</sup>, while depth estimation reliably determines relative spatial positions<sup>51</sup>. This technological maturity enables systematic data analysis at scales previously unattainable through manual methods. The convergence of theoretical validity and computational feasibility positions these four variables as optimal choices for bridging experiential knowledge and quantitative analysis. Building upon these conceptual foundations, the study operationalizes each variable through specific definitions and measurement protocols.

Shape constitutes the category of spatial visual characteristics generated by element configuration, encompassing the visual appearance of outlines



**Fig. 6 | Pattern mapping of four variables.** The figure details the operationalization of shape, size, position, and texture variables, showing how they are extracted from the foreground, middle ground, and background layers for analysis.

or boundaries that define geometric properties in two-dimensional or three-dimensional space through lines, edges, or surfaces<sup>48,49,52</sup>. Within scenic archetypes, shape specifically denotes the primary contours and boundaries formed by foreground elements, which assume decisive importance in archetype classification and identification<sup>47</sup>. The critical nature of foreground shape reflects the fundamental principle that object identity can be conveyed through basic outline alone<sup>2</sup>.

Size, functioning as a complementary variable to shape, refers to the magnitude, dimension, or scale of spatial elements<sup>53</sup>. Larger forms generate stronger visual impressions and historically convey power or dominance through physical and psychological presence, while smaller elements create subtler visual impacts, particularly when dispersed<sup>2</sup>. In scenic archetypes, the foreground employs size variations to enhance stylistic expression and reinforce spatial hierarchies.

Position represents spatial visual characteristics arising from element configuration, specifically denoting coordinates, orientations, and relational arrangements within three-dimensional space<sup>54</sup>. For scenic archetypes, position describes both the specific locations where elements are composed from particular viewpoints and their relative placements<sup>2,26,42,55,56</sup>. This variable assumes particular significance in Jiangnan gardens, where spatial penetration and hierarchical variation emerge through careful calibration of element separation and connection.

Texture, serving as a complementary variable to position, encompasses spatial visual characteristics generated by element composition, particularly the effects created by the interplay of obstructed and unobstructed visual elements forming recognizable patterns at finer scales<sup>2</sup>. This dual functionality proves crucial for archetype differentiation, as texture influences both the quantitative aspects (presence, quantity, distribution, proportion) and qualitative dimensions of spatial complexity.

The fourth tier operationalizes variables through three complementary metrics that transform experiential interpretations into analyzable data across all spatial layers. Distribution trend metrics employ statistical methods to reveal patterns of spatial variation and compositional dynamics. Absolute value metrics quantify fundamental characteristics including element dimensions, areas, and distances, establishing objective baselines for cross-archetype comparison. Relative relationship metrics utilize ratios and percentages to express proportional analyses, revealing hierarchical relationships and how foreground elements mediate perception of subsequent layers. This tripartite system captures both quantitative measurements and relational logic inherent in scenic archetypes, enabling systematic analysis while preserving the nuanced spatial organizations of traditional garden making. Table 1

presents detailed specifications and mathematical formulations for each metric type.

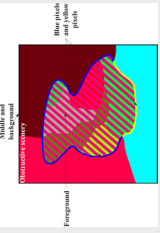
### Case study site and data collection

With the systematically analytical framework established, we now turn to examine this framework through empirical application. HWL was selected as a case study for the application and examination of the proposed framework for the following reasons: (1) HWL constantly interacts with the contemporary environment, representing the harmonious evolution of human activity and nature<sup>15</sup>. This dynamic interaction provides evidence and insights into the applicability of scenic archetypes in contemporary landscape spaces. (2) HWL includes several traditional Chinese gardens<sup>17</sup>, offering a rich data-set of photographic images for scenic archetypes from multiple dimensions, including type, scale, location, and function. (3) As a UNESCO World Heritage Site, information on HWL is publicly available and easily accessible, facilitating data collection and analysis.

To collect data for examination and given the absence of Google Street View coverage on HWL pedestrian paths, we conducted systematic on-site photography using an iPhone 12 Pro, which has a 0.5× zoom that matches human visual perception (120° horizontal field of view) as well as GPS functionality. We capture images at 20-m intervals perpendicular to the path at a height of 1.7 m; three images were taken per location, each facing a different direction (left, front, and right). All photographs were captured during late September under clear daylight conditions, ensuring consistent image quality and optimal visibility. While seasonal variations might affect vegetation density, our semantic segmentation model (trained on ADE20K with diverse seasonal imagery) maintains robust performance across different vegetation states. More importantly, the spatial visual characteristics of the four scenic archetypes are determined by geometric forms (shape), spatial arrangements (position), and proportional relationships (size) remain constant across environmental variations. While texture, particularly of deciduous vegetation, exhibits seasonal variations, these changes occur within predictable ranges that do not alter the archetype’s essential spatial structure. The framework captures the underlying organizational principles rather than ephemeral surface qualities.

From the comprehensive data-set of 1045 photographs, this study selected 168 images representing four scenic archetypes: 54 framed scenery, 20 obstructive scenery, 41 porous scenery, and 53 sandwiched scenery. The remaining photographs were excluded due to their ambiguous classification, as they typically represented transitional views lacking the distinctive spatial configurations that characterize traditional design patterns. The selected images ensure comprehensive coverage of all identifiable scenic archetypes

**Table 1 | Definition of each metric developed in this study**

Metrics		Pattern mapping
<b>S_ERI</b>	Method $\frac{0.25P}{\sqrt{A}}$	
	Description <b>S_ERI</b> is equal to 0.25 times the ratio of $P$ to $\sqrt{A}$ . $P$ represents the pixels on the foreground outline. $A$ denotes the number of pixels contained within the foreground.	
	Unit None	
	Range <b>S_ERI</b> $\geq \frac{\sqrt{A}}{2}$	
	Comments When <b>S_ERI</b> is equal to $\frac{\sqrt{A}}{2}$ , the foreground outline tends to be circular; when <b>S_ERI</b> is equal to 1, the foreground outline tends to be square. As the value of <b>S_ERI</b> increases beyond 1, the foreground outline becomes increasingly irregular. <b>S_ERI</b> is a relative relationship-based metric derived from the isoperimetric inequality $P^2 \geq 4\pi A$ , in order to accurately capture the shape of each foreground outline by comparing its perimeter and area. To normalize the regularity of a square to 1, a scaling factor of 0.25 is applied.	
<b>S_ECI</b>	Method $\frac{P_y}{P_y + P_b}$	
	Description <b>S_ECI</b> is the ratio of $P_y$ to the sum of $P_y$ and $P_b$ . $P_y$ are yellow pixels on the foreground outline. $P_b$ are blue pixels on the foreground outline.	
	Unit None	
	Range $0 \leq \text{S\_ECI} \leq 1$	
	Comments <b>S_ECI</b> of 0 indicates that the element labels of neighboring pixels are perfectly aligned with the pixels on the foreground outline. Conversely, a <b>S_ECI</b> value of 1 indicates that they are completely misaligned. <b>D_ECI</b> is a relative relationship-based metric that describes the differences between the elements forming the foreground and those in the surroundings. For each pixel on the foreground outline, <b>S_ECI</b> checks the element labels of its eight neighboring pixels. If the labels are consistent, the central pixel is marked yellow; otherwise, it is marked blue. The relative proportions of yellow and blue pixels are determined by calculating their ratio.	
<b>S_PSI</b>	Method $\sqrt{\frac{\sum_{i=1}^N \frac{0.25P_i}{\sqrt{A_i}}}{N}}$	
	Description <b>D_PSI</b> is the coefficient of variation for the ratio of $0.25P_i$ to $\sqrt{A_i}$ across all $N$ holes. $P_i$ is the pixel located on the outline of the $i$ -th hole in the foreground. $A_i$ is the pixel contained within the $i$ -th hole in the foreground. $\mu = \frac{0.25P_i}{\sqrt{A_i}}$ represents the average ratio of $P_i$ to $\sqrt{A_i}$ for all holes in the foreground. $N$ is the number of $P_i$ 's in the foreground.	
	Unit None	
	Range <b>S_PSI</b> $\geq 0$	
	Comments <b>S_PSI</b> value of 0 indicates that the shapes of all holes in the foreground are perfectly similar. Conversely, larger <b>S_PSI</b> value suggests increasing dissimilarity in the shapes of the holes. <b>S_PSI</b> is a distributional trend-based metric that employs the coefficient of variation formula to describe the distribution of hole shapes in the foreground. The shape of each hole is calculated using the same method as <b>S_PRI</b> .	
<b>S_API</b>	Method $\sqrt{\frac{\sum_{i=1}^N (P_i - \mu)^2}{N}}$	
	Description <b>S_API</b> is equal to the standard deviation of $P_i$ . $P_i = \frac{P_{i, \text{left}} + P_{i, \text{right}}}{2}$ denotes the midpoint of the line segment connecting the $i$ -th row of pixel pairs on both sides of the foreground outline. $\mu$ is the horizontal coordinate value of the vertical midline of the field of view. $N$ is the number of $P_i$ in the foreground.	
	Unit Pixel	
	Range <b>S_API</b> $\geq 0$	
	Comments <b>S_API</b> of 0 indicates that the vertical distance from each pixel point on both sides of the foreground outline to the midline of the field of view is equal. Conversely, an <b>S_API</b> value approaching $+\infty$ indicates that the vertical distances are completely different.	

**Table 1 (continued) | Definition of each metric developed in this study**

Metrics		Pattern mapping
Comments	<b>S_API</b> is a distributional trend-based metric using the standard deviation formula to describe the tendency of the mean line of the outline on both sides and the dispersion of the line in the field of view.	
<b>S_VFR</b> Method	$A_i/A_j$ <b>S_VFR</b> is the ratio of $A_i$ to $A_j$ .	
Description	$A_i$ represents all pixels in the foreground. $A_j$ represents all pixels in the field of view.	
Unit	None	
Range	$0 \leq \mathbf{S\_VFR} \leq 1$ A value of 0 for <b>S_VFR</b> means that there is no foreground in the field of view. On the contrary, an <b>S_VFR</b> value of 1 means that the field of view consists entirely of foreground.	
Comments	<b>S_VFR</b> is a relative relationship-based metric that describes the visual impact of the foreground on the field of view.	
<b>S_ADI</b> Method	$\frac{SD_j - SD_i}{SD_j + SD_i}$ <b>S_ADI</b> is equal to the ratio of the difference between $SD_i$ and $SD_j$ to $SD_j$ .	
Description	$SD_j = \sqrt{\frac{1}{N} \sum_{i=1}^N (A_i - \mu)^2}$ , is the standard deviation of all $A_i$ in the middle-back-ground. $SD_j = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \sum_{j=1, j \neq i}^N \max(A_i - A_j, 0)^2}$ , is the standard deviation of $A_i$ after removing the largest $A_j$ in the foreground. $A_i$ denotes the number of pixels contained in the $i$ -th hole in the foreground. $\mu_i = \frac{1}{N} \sum_{j=1}^N A_j$ , is the mean of all in the foreground. $\mu_w = \frac{1}{N} \sum_{j=1}^N \sum_{i=1, i \neq j}^N \max(A_i, 0)$ , is the mean of the remaining $A_i$ after removing the one with the largest value. $N$ denotes the number of $A_i$ in the foreground.	
Unit	None	
Range	$0 \leq \mathbf{S\_ADI} \leq 1$ When <b>S_ADI</b> is equal to 0, the size of the largest hole in the foreground is similar to the size of the remaining holes. Conversely, when <b>S_ADI</b> is equal to 1, it indicates that the size of the largest hole is significantly different from the size of the remaining holes.	
Comments	<b>S_ADI</b> is a relative relationship-based metric that characterizes the overall distributional impact of the largest hole in the foreground on all holes.	
<b>P_LV</b> Method	$D_i$ <b>P_LV</b> is equal to the maximum depth value among all pixels in the foreground.	
Description	$D_i$ is a function that identifies the $i$ -th pixel in the foreground with the maximum depth value.	
Unit	Depth	
Range	$\mathbf{P\_LV} \in Z \cap [1, 256]$ $\mathbf{P\_LV}$ equal to 1 means that the starting point of the foreground is furthest from the observer's location. Conversely, $\mathbf{P\_LV}$ equal to 256 means that the starting point of the foreground coincides with the observer's location.	
Comments	<b>P_LV</b> is an absolute value-based metric that directly represents the maximum depth value of the foreground, with its range constrained by the image's grayscale levels.	
<b>P_LN</b> Method	$\sum_{j=1}^N \mathbf{Unique}(D_j)$ <b>P_LN</b> is equal to the total number of $P$ of all pixels in the foreground, which corresponds to the vertical length.	
Description	$N$ is the number of pixels in the foreground. <b>Unique</b> ( $D_j$ ) is a function that identifies the non-repeating depth value of the $i$ -th pixel.	
Unit	Number	
Range	$\mathbf{P\_LN} \in Z \cap [1, 256]$ $\mathbf{P\_LN}$ of 0 indicates that the foreground has no vertical length. Conversely, a $\mathbf{P\_LV}$ of 256 indicates that the foreground has the maximum vertical length.	
Comments	<b>P_LN</b> is an absolute value-based metric that directly represents the vertical length of the foreground, with its range constrained by the image's grayscale levels.	
<b>P_ECI</b> Method	$\frac{\sum_{i=1}^N \sum_{j=1}^N (\sqrt{C_{ij} - \mu_{CV}})^2}{H_{CV}}$ <b>P_ECI</b> is equal to the coefficient of variation of the coefficient of variation of the difference between the depth values of pairs of neighboring pixels in each row and column in the foreground.	
Description	$CV_j = \frac{1}{N} \sum_{i=1}^N \frac{C_{ij} - \mu_j}{\mu_j}$ , is the coefficient of variation of all $C_{ij}$ in the $j$ -th row or column in the foreground. $C_{ij} =  dx_i - dx_{i+1} $ , is the difference in depth values between the $i$ -th pair of neighboring pixels in the $j$ -th row or column. $\mu_j = \frac{1}{N} \sum_{i=1}^N C_{ij}$ , denotes the average of all $C_{ij}$ in the $j$ -th row or column. $N$ is the total number of rows and columns in the foreground.	

**Table 1 (continued) | Definition of each metric developed in this study**

Metrics		Pattern mapping
	<p><math>\mu_{CV} = \frac{1}{N} \sum_{j=1}^N CV_j</math>, is the average of the coefficients of variation <math>CV_j</math> across all rows and columns.</p> <p><math>M_j</math> is the number of <math>C_{ij}</math> in the <math>j</math>-th row or column.</p>	
Unit	None	
Range	$P\_ECI \geq 0$	
Comments	<p><math>P\_ECI</math> of 0 indicates that the depth values in the foreground are continuously consistent. In contrast, as <math>P\_ECI</math> increases, the transition of depth values in the foreground becomes more abrupt, with noticeable jumps or changes.</p> <p><math>P\_ECI</math> is a distributional trend-based metric that uses the coefficient of variation formula to evaluate the continuity of spatial element position changes by calculating the coefficient of variation in depth differences between consecutive location points.</p>	
Method	$\sum_{j=1}^N \text{Unique}(C_j)$	$T\_ERI$ is equal to the number of element classes for all pixels in the foreground, or middle ground and background.
Description	$N$ represents the total number of pixels in a given area.	
Unit	$\text{Unique}(C_j)$ is a function used to identify the unique element category represented by the $i$ th pixel.	
Range	Number	
	$T\_ER \in Z \geq 0$	
Comments	<p><math>T\_ER</math> value of 0 indicates that there are no elements in the foreground, or middle ground and background. As the <math>T\_ER</math> value increases, the number of distinct element categories in the foreground, or middle ground and background increases.</p> <p><math>T\_ER</math> is an absolute value-based metric that directly represents the number of element categories in the foreground, or middle ground and background.</p>	
Method	$\frac{A_g^2}{\sum_{j=1}^N \sum_{i=1}^N A_{ij}^2}$	$T\_ESI$ is equal to the ratio of $A_g^2$ to the sum of all $A_{ij}^2$ .
Description	<p><math>A_g^2</math> is the square of the number of pixels in the foreground, or middle ground and background.</p> <p><math>N</math> represents the number of independently connected regions for elements of the same category.</p> <p><math>A_{ij}^2</math> is the square of the number of pixels in the independently connected region of the <math>i</math>th element under category <math>j</math>.</p>	
Unit	None	
Range	$1 \leq T\_ESI \leq A_g^2$	
Comments	<p>When the <math>T\_ESI</math> value is equal to 1, it indicates that the foreground, or middle ground and background consists of only one independently connected region of elements. Conversely, as the <math>T\_ESI</math> value gradually increases toward <math>A_g^2</math>, each pixel in the given region becomes an independently connected region of elements.</p> <p><math>T\_ESI</math> is a relative relationship-based metric that uses normalized weights to describe the contribution of independently connected regions for each element in the foreground, or middle ground and background, relative to the field of view.</p>	
Method	$\frac{A_o}{A_o + A_e}$	$T\_ISI$ is equal to the ratio of $A_o$ to the sum of $A_o$ and $A_e$ .
Description	<p><math>A_o</math> is the number of pixels in the foreground, or middle ground and background labeled as open attributes.</p> <p><math>A_e</math> is the number of pixels in the foreground, or middle ground and background labeled as enclosed attributes.</p>	
Unit	None	
Range	$0 \leq T\_ISI \leq 1$	
Comments	<p>When the <math>T\_ISI</math> value is equal to 0, it indicates that the foreground, or middle ground and background is completely open. Conversely, when the <math>T\_ISI</math> value is equal to 1, it indicates that the ground is completely closed.</p> <p><math>T\_ISI</math> is a relative relationship-based metric to describe the degree to which open elements in the foreground, or middle ground and background contribute to the field of view.</p>	

**Table 1 (continued) | Definition of each metric developed in this study**

Metrics		Pattern mapping
<b>T_MVI</b>	Method	$\sqrt{\frac{\sum_{i=1}^N (0.25P_i - \mu)^2}{N}}$ T_MVI equals the coefficient of variation of the ratio of 0.25P <sub>i</sub> to $\sqrt{A_i}$ .
	Description	P <sub>i</sub> is the pixel located on the outline of the i <sup>th</sup> open or enclosed connected regions in the foreground, or middle ground and background. A <sub>i</sub> is the pixel contained within the i <sup>th</sup> open or enclosed connected regions in the foreground, or middle ground and background. $\mu = \frac{1}{N} \sum_{i=1}^N \frac{0.25P_i}{\sqrt{A_i}}$ , represents the average of all the ratios of P <sub>i</sub> to $\sqrt{A_i}$ in the foreground, or middle ground and background. N is the number of P <sub>i</sub> in the foreground, or middle ground and background.
	Unit	None
	Range	T_MVI ≥ 0 A value of 0 for T_MVI indicates that the shapes of both open and enclosed independently connected regions in the foreground, or middle ground and background are perfectly similar; conversely, larger values of T_MVI indicate less similarity.
	Comments	T_MVI is a distributional trend-based metric that uses the coefficient of variation formula to describe the distribution of shapes of open or enclosed connected regions in the foreground, or middle ground and background. The shape of each connected region is calculated using the same method as for S_ERI.

P<sub>i</sub>  
A<sub>i</sub>

along West Lake paths while demonstrating clear spatial stratification across foreground, middle ground, and background (Fig. 7). The uneven sample distribution reflects the actual prevalence of scenic archetypes at HWL rather than sampling bias. This distribution, articulately the lower frequency of obstructive scenery, reflects both historical design preferences and contemporary landscape modifications. Each scenic archetype was analyzed independently to reveal its specific spatial visual characteristics. This approach not only aligns with our research goal of characterizing distinct design patterns but also mitigates potential biases arising from the uneven sample distribution, as each archetype’s characteristics are identified without reference to the prevalence of other types.

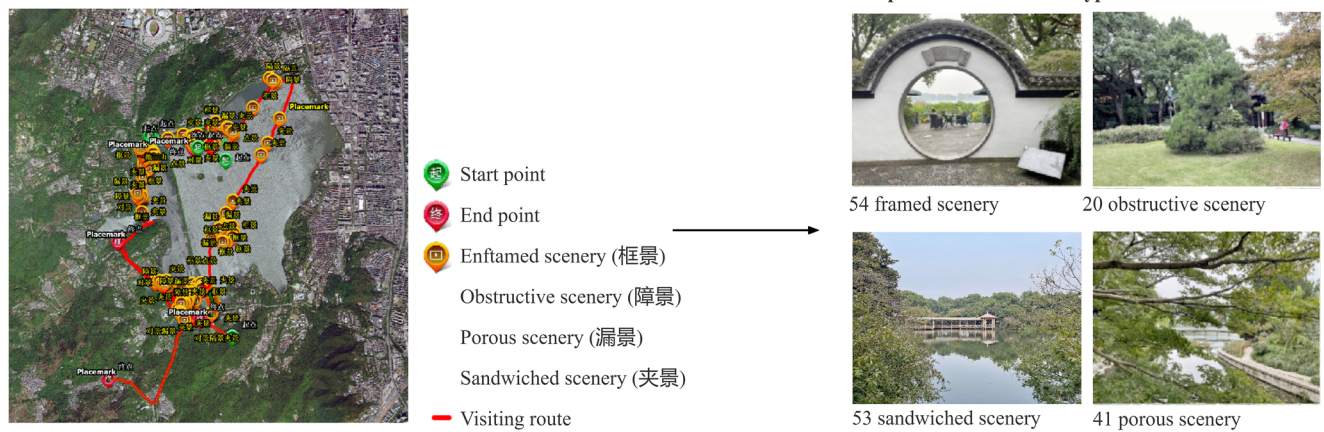
**Image processing and mapping**

Following data collection, image processing operationalizes the analytical framework through computational measurement. The four-tier framework (Fig. 1) establishes what spatial visual characteristics require measurement, specifically the decomposition of scenic archetypes into foreground-middle-background layers characterized by shape, size, position, and texture variables, while computational tools execute these specifications through systematic mapping protocols. This distinction between characteristic specification (framework-determined) and characteristic extraction (tool-executed) positions the study’s contribution in analytical logic transferable across multiple computational implementations, thereby prioritizing replicability over algorithm-specific reproducibility. This methodological stance addresses the distinction between reproducibility (exact numerical replication with identical tools) and replicability (comparable pattern identification with transferable logic), prioritizing the latter as appropriate for design pattern research. Each image underwent processing according to the framework’s hierarchical structure to generate three types of spatial visual characteristic units (Fig. 8): element mapping, openness mapping, and depth mapping. All images were standardized to 600 × 800 pixels at 72 dpi prior to processing.

Element mapping employed PSPNet with ResNet-101 backbone for initial semantic segmentation, pre-trained on ADE20K<sup>57</sup>. This architecture was selected for its demonstrated robustness in complex scene understanding, achieving 41.96% mIoU and 80.64% pixel accuracy on ADE20K validation, representing 4.73% absolute mIoU improvement over ResNet-50 baseline<sup>27</sup>. The pyramid pooling module’s multi-scale context aggregation proves particularly effective for traditional gardens, where elements span multiple scales. However, automated segmentation exhibits systematic limitations: vegetation overlap produces boundary ambiguity, shadow variation reduces contrast discrimination, and irregular organic surfaces challenge recognition algorithms. We therefore implemented interactive boundary refinement using SAM ViT-H variant<sup>58</sup>, selected for its exceptional zero-shot generalization across 16 of 23 evaluation datasets and superior mask quality (7–9/10 ratings in human assessments)<sup>58</sup>. This human-in-the-loop mechanism positions artificial intelligence as supervised augmentation: when PSPNet generated indeterminate boundaries, particularly at vegetation-architecture interfaces or shadow-obscured regions, manual point-click intervention through SAM overrode automated outputs. This refinement additionally incorporated six culturally-specific categories absent from ADE20K’s universal taxonomy (herbaceous plants, aquatic plants, lawn, embankment, architectural inscriptions), yielding 34 semantic classes balancing automated efficiency with supervised accuracy.

Openness mapping incorporated cross-modal validation to ensure data quality. Element maps received binary occlusion labels indicating each component’s contribution to visual permeability, directly supporting texture variable calculation. We implemented cross-modal validation that systematically exploits redundancy across independent data sources to detect processing errors. Element maps, openness maps, and depth maps must maintain logical consistency; violations signal errors undetectable through single-modality analysis. For instance, an element labeled ‘tree’ in semantic segmentation yet appearing transparent in openness mapping indicates classification error. Similarly, depth orderings contradicting

Photography collection in HWL



**Fig. 7 | Data collection for four scenic archetypes.** The left panel displays the visiting route and sampling points mapped at Hangzhou Westlake; the right panel provides representative photographic examples for each of the four collected scenic archetypes.

element positions reveal spatial inconsistencies. When such contradictions emerged, manual inspection and correction were triggered. This validation mechanism transforms multiple mapping outputs from parallel data streams into mutually reinforcing error-detection architecture.

Depth mapping employed ordinal validation to ensure perceptual accuracy. MiDaS v3.1 DPT-Hybrid generated relative depth maps, selected for its robust zero-shot cross-dataset transfer capability (36% relative error reduction versus v3.0 baseline)<sup>51</sup>. Critically, MiDaS performs relative depth estimation, outputting ordinal depth relationships (0=nearest, 1=farthest) appropriate for phenomenological requirements: foreground-middle-background stratification emerges from perceptual depth experience rather than absolute measurements, as demonstrated in visual landscape analysis<sup>43</sup>. The model’s demonstrated generalization across diverse scene types, from indoor spaces to outdoor scenarios to general web imagery, which ensures reliability for varied spatial compositions in traditional gardens<sup>51</sup>. We validated MiDaS outputs against visual inspection to ensure generated depth orderings matched phenomenological perception, with implausible stratification at occlusion boundaries receiving manual resolution. These protocols embody a fundamental methodological principle: artificial intelligence augments analytical efficiency while human oversight ensures interpretive reliability, treating computational models as supervised assistants requiring critical evaluation rather than autonomous systems.

Upon completion of initial mapping procedures, these AI-generated mappings underwent division into discrete mapping units to facilitate calculation of spatial visual characteristics across foreground, middle ground, and background layers. This critical segmentation ensures dedicated input data for each metric while preventing overlaps or interactions between different spatial visual characteristics, thereby maximizing measurement precision. The specific procedural steps encompassed converting color depth maps to grayscale using the standard Inferno color map and delineating preliminary foreground, middle ground, and background masks through the natural breaks classification method in ArcGIS, thus generating area mappings. To refine these preliminary results, we implemented a two-stage correction protocol designed to enhance area mapping accuracy. First, we removed openness elements located in the foreground and enclosed elements in the middle and background that properly belonged to the foreground, based on openness mapping data. Second, we adjusted element boundaries between foreground, and middle ground and background according to element mapping information. In the final processing phase, element, openness, and depth mappings were intersected with area mappings to generate definitive mapping units. Through this systematic approach, three mapping units were produced for both the foreground, and the middle ground and background of each image: element mapping, openness mapping, and depth mapping, with each map corresponding to at least one analytical metric.

**Data measurement and statistical analysis**

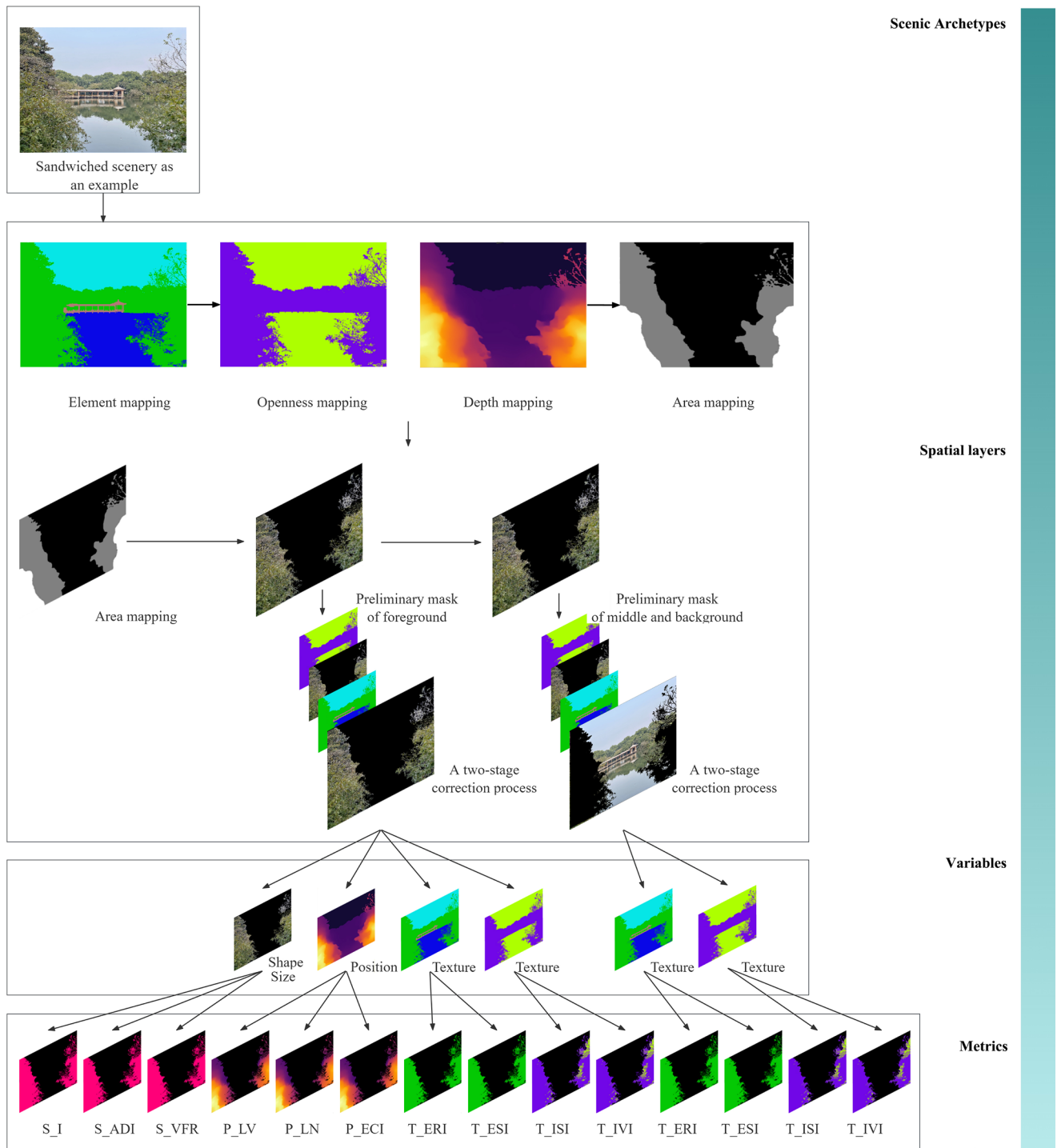
Following image processing and mapping, data analysis proceeded through two distinct phases: measurement and assessment. During the measurement phase, twelve proprietary mapping algorithms developed for this study processed the four variables of shape, size, depth, and texture. Notably, only the texture variable applies to all spatial layers (foreground, middle ground, and background), while the remaining variables were developed exclusively for foreground analysis. The operational architecture of these algorithms comprised four integrated components: mapping unit reading, characteristic extraction, spatial calculation, and result output (Fig. 9). First, the mapping unit reading module converts the pixel-based spatial visual characteristic units into arrays, generating the raw data used for subsequent calculations. Next, the characteristic extraction module extracts the specific values required for metric calculations. Then, the spatial calculation module analyzes the distribution, absolute values, and relative relationships of these values to generate results. Finally, these results are converted into two primary outputs in the result output module: (1) precise and unique values, and (2) output mappings that explain the computational process.

To complete our analytical framework, during the assessment phase, various statistical methods were used to conduct an in-depth analysis of the measurement results, including descriptive statistical analysis, importance analysis, correlation analysis, and element contribution analysis. For basic characterization, descriptive statistics were used to summarize the basic trends and patterns in the numerical representations of the four scenic archetypes, focusing on measures of central tendency, dispersion, and distribution. For more advanced analysis, importance analysis was conducted using a random forest algorithm that transformed the identification of the target scenic archetypes into a binary classification problem<sup>59,60</sup>. This method assessed the importance of each characteristic in improving model performance, identifying the most important characteristics in terms of classifying the different scenic archetypes. Additionally, Spearman’s correlation coefficient, a non-parametric statistical method suitable for analyzing small samples, non-linear relationships, and datasets with outliers, was used to evaluate correlations between composite items in the dataset<sup>61,62</sup>. To complete our comprehensive analysis, element contribution analysis was conducted by comparing the original measurement results with the results obtained after removing the target element, which quantified the contribution of the target element to a specific composite item.

**Results**

**Examining the hierarchical structure of the framework**

The hierarchical structure of the proposed framework, from scenic archetypes to metrics, was examined using descriptive statistical analysis and importance analysis via random forest classification. First, importance analysis was used to evaluate the conversion of scenic archetypes

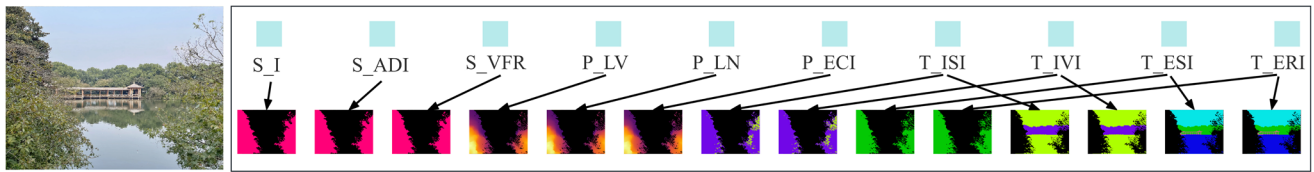


**Fig. 8 | Data pre-processing based on the framework.** This flowchart outlines the image processing pipeline, including the generation of element, openness, and depth mappings, followed by a two-stage correction process to refine foreground and background masks.

into spatial layers; the results showed that the framework exhibited high recognition accuracy for the different scenic archetypes, achieving an overall accuracy of 94.12% and a weighted F1-score of 0.94. This validates the division of scenic archetypes into the foreground, and the middle ground and background. Specifically, random forest models used in the importance analysis exhibited recognition accuracy of 97.06% for framed scenery, 91.18% for obstructed scenery, 97.06% for porous scenery, and 100% for sandwiched scenery, supporting the framework’s theoretical premise that “the foreground plays a dominant role by shaping the observer’s view and mediating the presentation of middle and background

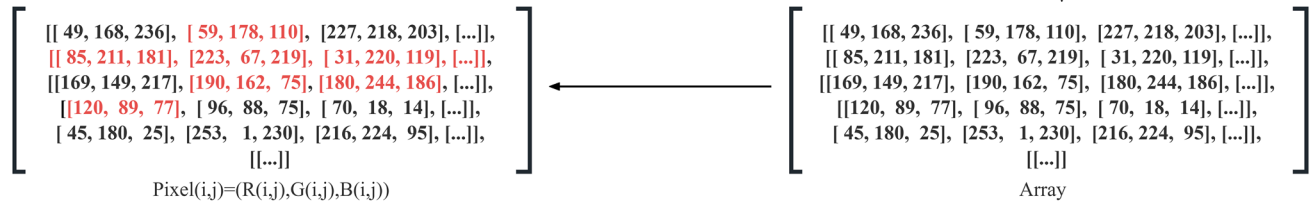
elements.” Second, upon conversion from spatial layers to variables, the random forest models revealed that the most influential spatial visual characteristics were concentrated in the foreground across all four variables: shape (e.g., S\_I contributing  $\approx 31.5\%$ ), size (e.g., S\_ADI contributing  $\approx 12.1\%$  and S\_VFR contributing  $\approx 9.9\%$ ), position (e.g., P\_LV  $\approx 24.3\%$  in obstructed scenery), and texture (e.g., T\_IVI contributing  $\approx 9.3\%$  and T\_ISI  $\approx 8.7\%$ ). These foreground spatial visual characteristics collectively contributed over 70% of total importance, validating the premise that “the foreground facilitates the transformation of spatial elements in the middle ground and background from mere

### Mapping unit reading

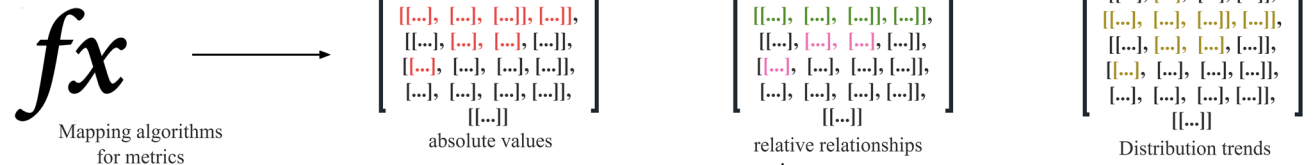


Sandwiched scenery as an example

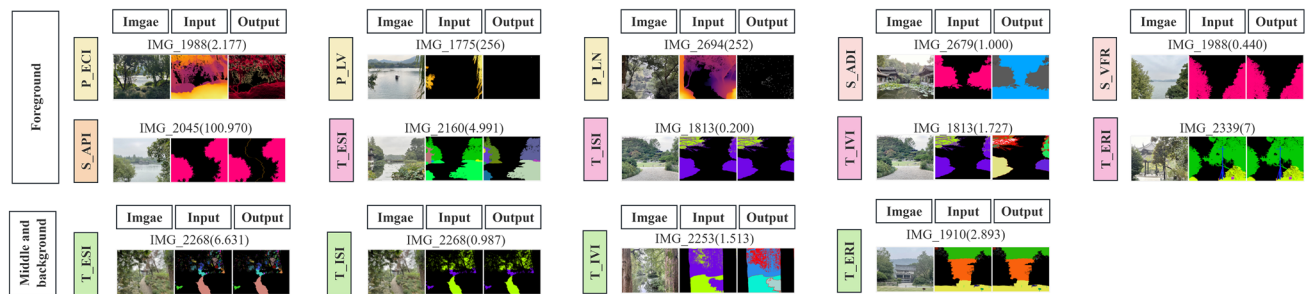
### Feature extraction



### Spatial calculation



### Result output



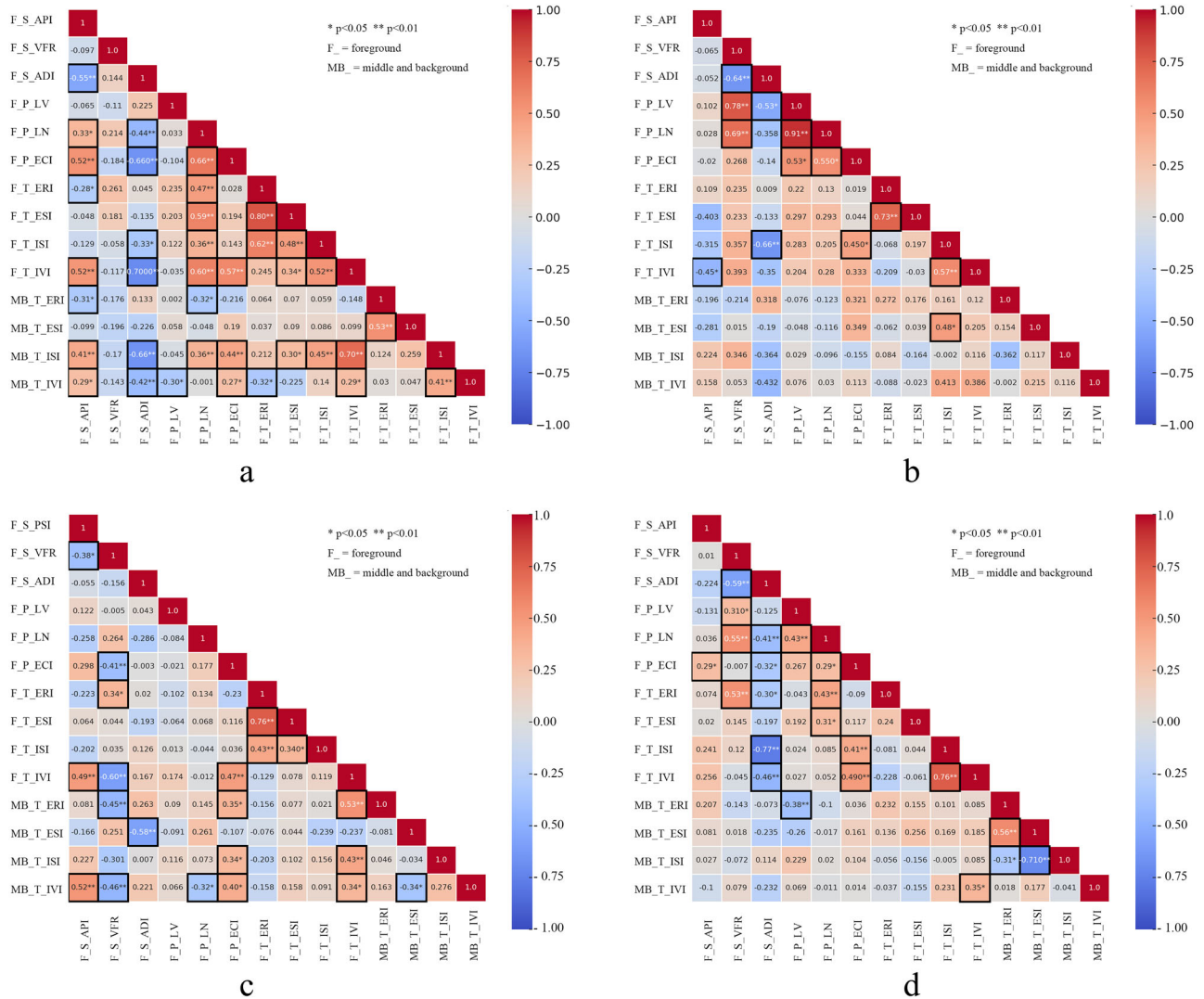
**Fig. 9 | Operational modules of the algorithms used to generate the metrics.** The diagram visualizes the four integrated components of the mapping algorithms: mapping unit reading, characteristic extraction, spatial calculation, and result output.

“objects” into integral parts of the “scenery.”” Finally, when converting from variables to metrics, descriptive statistical analysis with 95% confidence intervals revealed that 55.36% of metrics displayed a near-normal distribution (absolute K-value < 10, absolute Sk < 3; see refs. 63,64), while 33.93% exhibited a positive skew, indicating that metrics developed based on distribution trends, absolute values, and relative relationships can reliably capture the spatial visual characteristics of each scenic archetype. Additionally, significant differences in the mean values (with non-overlapping confidence intervals) of identical metrics across scenic archetypes highlighted the ability of metrics to effectively identify and distinguish different scenic archetypes. Together, these findings demonstrate that the framework’s hierarchical structure is capable of converting abstract concepts into specific, quantifiable metrics.

#### Summarizing and characterizing four scenic archetypes

Descriptive statistical analysis was combined with correlation analysis and element contribution analysis to characterize each scenic archetype based on their position, size, shape, and texture (Figs. 10–11).

Framed scenery is characterized by three interconnected aspects: foreground framing, focal point formation, and spatial hierarchy enhancement. The results of the correlation analysis are detailed in Fig. 10a, and the results of the descriptive statistical analysis are presented in Table 2. The results show that the foreground effectively frames the view through its carefully orchestrated physical attributes. The analysis shows that the foreground maintains regular outlines (S\_ERI) with smooth edge continuity (P\_ECI), creating a stable framing structure. This framing structure is strategically positioned, starting near the observer’s viewpoint (P\_LN) and extending to sufficient depth (P\_LV), with a significant positive correlation between the starting position and depth transition ( $r = 0.656, p < 0.01$ ). The foreground occupies a substantial portion of the field of view (S\_VFR) with an optimal area distribution (S\_ADI). This framing structure effectively directs attention to create a clear visual focal point. This is achieved through deliberate textural contrasts between the foreground and the middle ground and background. The foreground exhibits enclosed, simple characteristics (T\_ESI, T\_ISI, T\_ER), while the middle ground and background exhibit greater openness and diversity (T\_ESI, T\_ISI, T\_ER). The effectiveness of



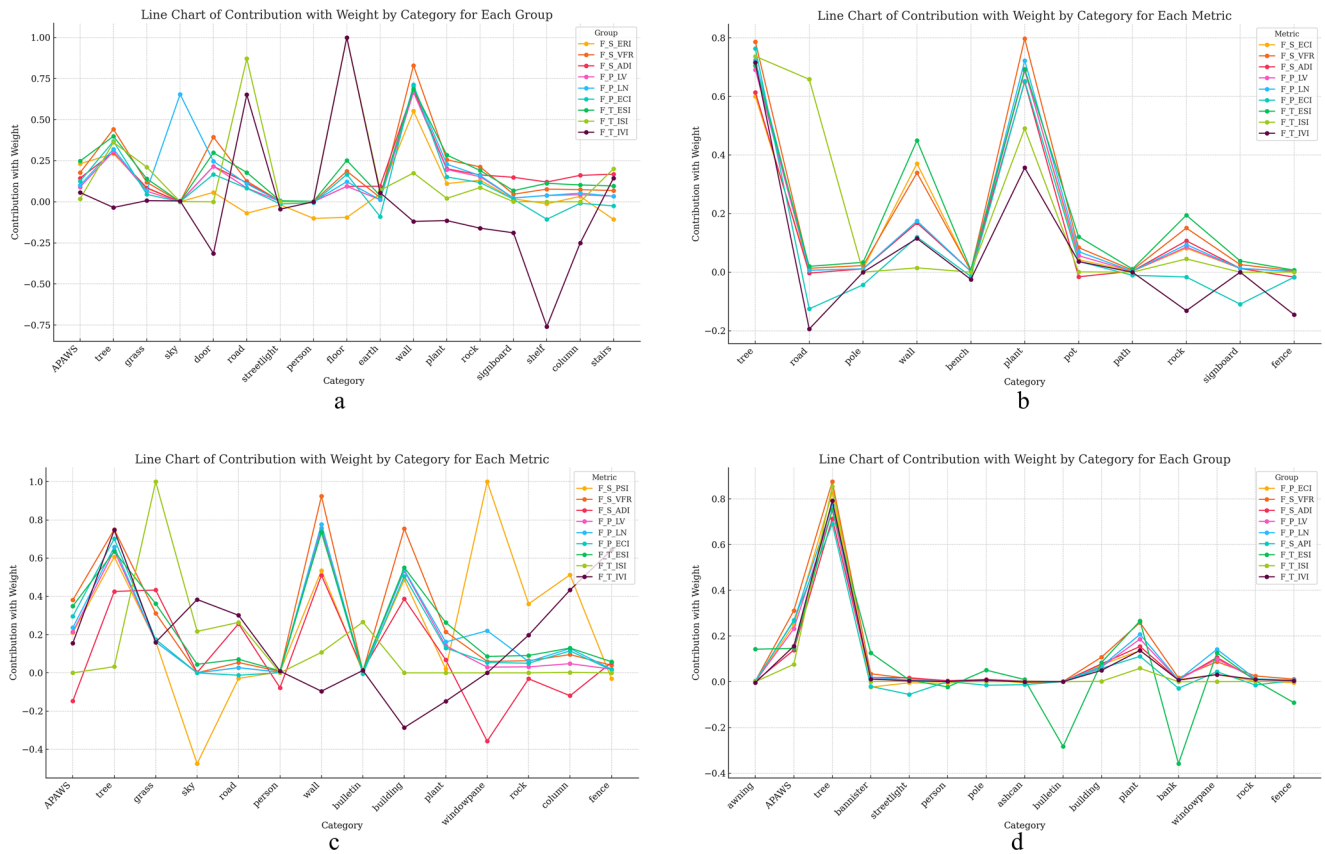
**Fig. 10 | Correlation heatmap of metrics of four scenic archetypes. a** Framed scenery, **b** obstructive scenery, **c** porous scenery, and **d** sandwiched scenery.

focal point creation is further supported by significant correlations between regular edge shapes and positioning ( $r = 0.331, p < 0.05$ ) as well as between texture intervals and area distribution ( $r = -0.326, p < 0.05$ ), indicating coordinated visual guidance. Finally, the spatial hierarchy is enhanced through sophisticated layering mechanisms. The analysis reveals consistent texture variations (T\_IVI autocorrelation = 0.287,  $p < 0.05$ ) that strengthen the hierarchical contrast in space (T\_ESI and P\_LN correlation:  $r = 0.594, p < 0.01$ ). The spatial structure maintains stability despite complex texture variations (T\_ISI and T\_IVI correlation:  $r = 0.699, p < 0.01$ ), while area distribution and edge continuity complement each other in reinforcing spatial depth (S\_ADI and P\_ECI correlation:  $r = -0.66, p < 0.01$ ). This hierarchical organization is materially supported by specific spatial elements, including trees and walls, which exhibit strong and centralized contributions (mean  $r > 0.3$ ) across multiple foreground metrics (Fig. 11a).

These measured correlations align with spatial organization principles documented in traditional Chinese garden treatises and recent empirical validations. The positive correlation between foreground shape regularity (low S\_ERI) and smooth depth transitions (high P\_ECI,  $r = 0.656, p < 0.01$ ) observed at HWL can be interpreted through the lens of principles articulated in Ji<sup>40</sup>, though not in these modern terms. Ji Cheng emphasized that architectural openings such as door and window frames should “collect fine views while excluding mundane sights” (佳境宜收, 俗尘安到). The conceptual foundation of framed scenery as a deliberate compositional device was further developed by Li<sup>41</sup>, who described creating “frameless paintings”

(无心画) by positioning paper borders around window openings to transform architectural apertures into pictorial frames (Fig. 12a). This historical precedent demonstrates that designers understood framing not merely as structural necessity but as a sophisticated tool for controlling visual perception and spatial depth—principles that manifest in contemporary West Lake gardens through systematic foreground geometry (Fig. 12b).

While Ji Cheng did not employ geometric or perceptual psychology terminology, as his philosophy centered on “skillful borrowing and appropriate adaptation” (巧于因借, 精在体宜) and organic flexibility rather than geometric regularity, contemporary research has validated that framing elements do systematically structure visual depth perception. Experimental studies demonstrate that frame positioning, rather than frame geometry alone, determines depth perception in framing contexts<sup>65</sup>. The significant relationship between foreground positioning, denoted as P\_LN, and depth extension, denoted as P\_LV, reflects what traditional texts described as systematic spatial layering<sup>40,41</sup>, where designers deliberately positioned framing elements to create hierarchical views. Contemporary computational analyses confirm that buildings function as primary mechanisms for controlling views and creating framed scenery, with Building Visual Index exerting the strongest influence on visual complexity in garden spaces at  $\beta = 0.683, p < 0.05$ <sup>66</sup>. Thus, the statistical patterns observed in framed scenery at HWL reflect not incidental geometric arrangements but intentional compositional strategies encoded in



**Fig. 11 | Line chart of contribution with weight by category for metrics of four scenic archetypes. a Framed scenery, b obstructive scenery, c porous scenery, and d sandwiched scenery.**

**Table 2 | Descriptive statistical analysis of each metric and their interpretations in the context of framed scenery**

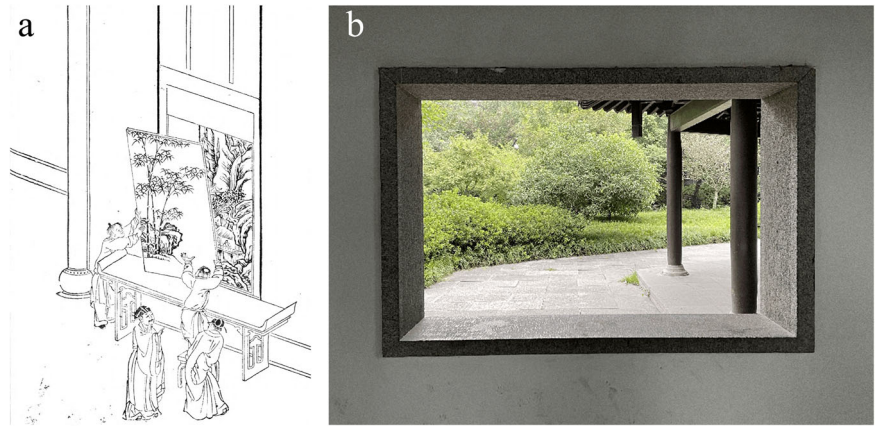
Metrics	Statistical Values	Interpretation	Implication for Design
S_ERI	Mean = 1.634; Mdn = 0.986; 95% CI [1.300, 1.968]	Lower range	Controls regularity in framing outline
P_ECI	Mean = 1.237; Mdn = 1.143; 95% CI [1.074, 1.400]	Lower range concentration	Ensures smooth depth transitions
P_LN	Mean = 50; Mdn = 255; 95% CI [119.441, 152.929]	Near maximum value of 256	Optimal alignment with observer position
P_LV	Mean = 136.185; Mdn = 148.5; 95% CI [250.213, 254.343]	Above median of 128	Extended depth reached in the foreground
S_VFR	Mean = 0.734; 95% CI [0.706, 0.762]	Substantial proportion	Dominant visual presence
S_ADI	Mean = 0.957; 95% CI [0.926, 0.988]; Sk = -3.806; K = 16.2	Near maximum of 1, left-skewed	Effective visual guidance
Foreground T_ESI, T_ISI, T_ER	T_ESI = 2.101, 95% CI [1.727, 2.475]; T_ISI = 0.097, 95% CI [0.072, 0.122]; T_ER = 2.963, 95% CI [2.504, 3.422]	Lower values	Controlled, enclosed textural pattern
Middle ground and background T_ESI, T_ISI, T_ER	T_ESI = 2.963, 95% CI [3.590, 4.648]; T_ISI = 0.396, 95% CI [0.320, 0.472]; T_ER = 6.204, 95% CI [5.644, 6.764]	Higher values	Open and diverse textures

traditional design knowledge, though expressed through cultural vocabularies distinct from modern spatial analysis terminology, and validated through both historical documentation and contemporary empirical research.

Obstructive scenery is characterized by three interrelated aspects: visual barrier creation, sight line guidance, and spatial hierarchy enhancement. The results of the correlation analysis are detailed in Fig. 10b, and the results of the descriptive statistical analysis are presented in Table 3. In obstructive scenery, the foreground effectively establishes a visual barrier through precise control of its physical attributes. Analysis reveals that the foreground maintains distinct outlines (S\_ECI) with smooth depth

transitions (P\_ECI) to create a clear obstruction. This obstruction is strategically configured through optimal area distribution (S\_ADI) while maintaining moderate visual field occupation (S\_VFR). Furthermore, there is a significant negative correlation between area and visual field ratio ( $r = -0.642, p < 0.01$ ) ensuring effective visual concentration. The obstruction is positioned closer to the observer (P\_LN) at a controlled depth (P\_LV), demonstrating a strong position–depth correlation ( $r = 0.912, p < 0.01$ ) for precise obstruction control. This visual barrier systematically guides the observer’s line of sight through deliberate textural manipulation. The foreground exhibits simplified texture characteristics (T\_ISI, T\_ER, T\_ESI); this is in contrast to the richer middle ground and background

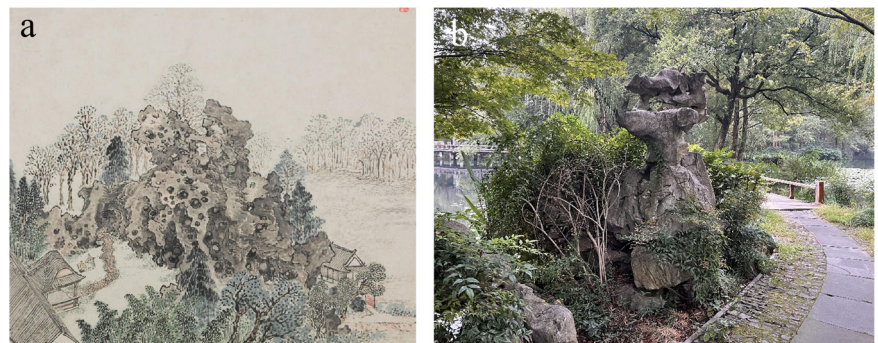
**Fig. 12 | Framed scenery: historical precedent and case manifestation.** **a** Historical illustration from Li<sup>41</sup> depicting the conceptual origin of “frameless painting” (无心画), where architectural apertures transform scenery into composed views. **b** Rectangular stone window frame at HWL demonstrating systematic structuring of depth perception through foreground geometry.



**Table 3 | Descriptive statistical analysis of each metric and their interpretations in the context of obstructive scenery**

Metrics	Statistical Values	Interpretation	Implication for Design
S_ECI	Mean = 0.916; Mdn = 0.968; 95% CI [0.860, 0.972]	Near maximum value of 1	Ensures clear distinction of foreground outlines
P_ECI	Mean = 0.649; Mdn = 0.576; 95% CI [0.451, 0.847]	Moderate range	Enables smooth depth transitions
S_ADI	Mean = 0.864; Mdn = 0.985; 95% CI [0.765, 0.963]	Near maximum value of 1	Strong visual guidance capability
S_VFR	Mean = 0.39; Mdn = 0.359; 95% CI [0.302, 0.478]	Lower proportion	Moderate visual field occupation
P_LN	Mean = 169.65; Mdn = 170; 95% CI [140.335, 198.965]	Above median of 128	Closer positioning to the observer
P_LV	Mean = 99.25; Mdn = 93.5; 95% CI [71.436, 127.064]	Relatively low values	Controlled depth extension
Foreground T_ESI, T_ISI, T_ER	T_ISI = 0.025, 95% CI [0.009, 0.041]; T_ER = 2.05, 95% CI [1.560, 2.540]; T_ESI = 1.324, 95% CI [1.148, 1.500]	Lower values	Simplified textural patterns
Middle and background T_ESI, T_ISI, T_ER	T_ISI = 0.481, 95% CI [0.368, 0.594]; T_ER = 6.25, 95% CI [5.129, 7.371]; T_ESI = 3.543, 95% CI [2.729, 4.357]	Higher values	Rich, complex textures

**Fig. 13 | Obstructive scenery: historical precedent and case manifestation.** **a** Historical illustration from Zhang [67] showing monumental Taihu stone formation as foreground visual barrier in Zhiyuan Garden Album (Ming Dynasty). **b** Taihu stone rockery at HWL maintaining the traditional role of foreground visual barrier.



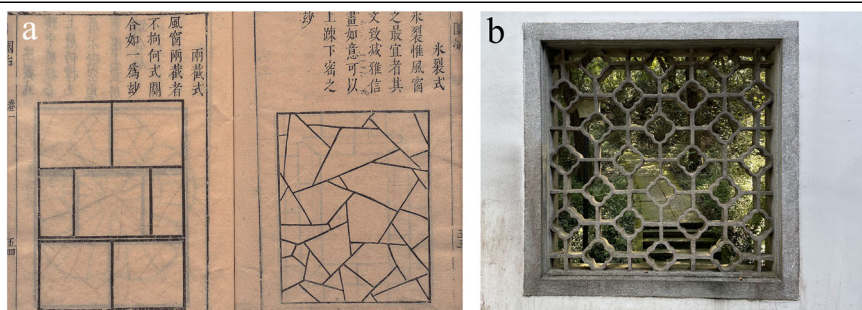
textures (T\_ISI, T\_ER, T\_ESI), which exhibit distinct variation patterns (T\_IVI). This guidance is reinforced through coordinated relationships between textural elements and spacing ( $r = 0.728, p < 0.01$ ), textural variation and intervals ( $r = 0.566, p < 0.01$ ), as well as depth transitions and positioning (P\_LV and P\_LN with P\_ECI:  $r = 0.55$  and  $0.531$  respectively,  $p < 0.05$ ). Finally, spatial hierarchy is enhanced through sophisticated visual and spatial mechanisms: The analysis reveals systematic relationships between visual field proportion and spatial depth ( $r = 0.778, p < 0.01$ ) and positioning ( $r = 0.695, p < 0.01$ ), which are indicative of coordinated spatial progression. Texture intervals are found to significantly influence visual guidance (T\_ISI and S\_ADI correlation:  $r = -0.655, p < 0.01$ ), while precise depth control ensures a smooth spatial transition. This hierarchical organization is materially supported by specific spatial elements, such as trees, walls, and plants (primarily shrubs and tall herbaceous species), which all have a strong and centralized contribution across multiple foreground metrics (Fig. 11b).

These measured spatial patterns, particularly the deliberate textural contrasts between simplified foreground elements and enriched middle-ground/background compositions, correspond to traditional design principles documented in traditional treatises such as Li<sup>41</sup>, which emphasizes the strategic placement of screening elements such as trees, walls, and rockeries to control visual sequences and create progressive spatial revelation. Historical visual evidence illuminates how these principles operated in practice: Zhang<sup>67</sup> systematically depicts rockeries as foreground obstructions, as shown in Fig. 13a, where the monumental Taihu stone formation exemplifies the measured positioning control denoted as P\_LN and depth management indicated by P\_LV identified in our statistical analysis. The rockery’s textural complexity characterized by high T\_ER against simplified surrounding vegetation demonstrates the same foreground-background contrast measured through T\_ESI and T\_ISI indices that our data reveal as characteristic of obstructive scenery. This compositional strategy persists in contemporary practice: at HWL, Taihu stone rockeries continue to function

**Table 4 | Descriptive statistical analysis of each metric and their interpretations in the context of porous scenery**

Metrics	Statistical Values	Interpretation	Implication for Design
S_PSI	Mean = 0.523; Mdn = 0.6; 95% CI [0.464, 0.582]; Sk = -1.424	Moderate value with negative skew	Complex and diverse foreground outlines
P_ECI	Mean = 1.615; 95% CI [1.449, 1.781]; Sk = -0.229	Relatively high with a slight negative skew	Long and varied depth transitions
P_LN	Mean = 252.073; 95% CI [249.750, 254.396]; Sk = -4.333	High value near maximum, strong negative skew	Very close positioning to the observer
S_VFR	Mean = 0.734; Mdn = 0.793; 95% CI [0.675, 0.793]	High proportion	Significant occupation of the visual field
S_ADI	Mean = 0.559; Mdn = 0.61; 95% CI [0.473, 0.645]	Moderate value	Balanced influence on middle ground and background
P_LV	Mean = 168.707; 95% CI [157.056, 180.358]; Sk = -0.571	High value with moderate negative skew	Extended depth reached in the foreground
Foreground T_ESI, T_IVI, T_ISI	T_ESI = 1.835, 95% CI [1.511, 2.159]; T_IVI = 183.349, 95% CI [141.457, 225.241]; T_ISI = 0.02, 95% CI [-0.002, 0.042]	Lower T_ESI, higher T_IVI, very low T_ISI	Uniform texture with varied openness patterns
Middle and background T_ESI, T_IVI, T_ISI	T_ESI = 2.963, 95% CI [9.418, 18.194]; T_IVI = 91.937, 95% CI [80.204, 103.670]; T_ISI = 0.507, 95% CI [0.406, 0.608]	Higher values across metrics	More fragmented and open textures

**Fig. 14 | Porous scenery: historical precedent and case manifestation.** **a** Historical window lattice designs from Ji [40] showing Two-Section Style (两截式) and Ice-Crack Pattern (冰裂纹). **b** Floral lattice window at HWL demonstrating contemporary implementation of the “half-transparent” filtering effect.



as carefully positioned visual barriers, as illustrated in Fig. 13b, their placement near the observer with controlled depth extension creating the “conceal-then-reveal” effect termed (先藏后露) that defines obstructive design.

The strong correlation between foreground positioning and depth control (P\_LN and P\_LV,  $r = 0.912, p < 0.01$ ) observed in our data quantitatively validates what traditional treatises describe qualitatively as “systematic obstruction” (障). HWL examples demonstrate that effective obstructive scenery requires precise calibration: the obstruction must be close enough, reflected in low P\_LN values, to command attention yet allow sufficient depth variation through P\_LV to guide visual exploration around its edges. The textural hierarchy we measured, showing simplified foreground through T\_ISI, T\_ER, and T\_ESI against enriched backgrounds, aligns with the visual strategy evident in both Zhang<sup>57</sup> and WHL, where screening elements heighten anticipation through deliberate contrast. Recent computational analyses confirm that while excessive enclosure reduces dwell duration with  $\beta = -0.789$  and  $p < 0.001$ , strategic screening creates offset-alternating synergies that enhance spatial experience<sup>67</sup>. These convergences between our statistical findings from HWL, Ming Dynasty garden albums and traditional design treatises, demonstrate that the spatial patterns characterizing obstructive scenery reflect centuries-refined compositional strategies rather than incidental arrangements.

Porous scenery is characterized by three interconnected aspects: partial perspective exhibition, balanced spatial enclosure, and spatial hierarchy enhancement (Fig. 10c). The results of the correlation analysis are detailed in Fig. 10c, and the results of the descriptive statistical analysis are presented in Table 4. In porous scenery, the foreground effectively creates partial perspective openings through sophisticated control of its physical attributes. Analysis reveals that the foreground maintains complex and diverse outlines (S\_PSI) with significant depth variations (P\_ECI) that help establish a

sophisticated porous structure. The porous structure is strategically positioned close to the observer (P\_LN) and occupies a substantial portion of the visual field (S\_VFR) while maintaining a balanced influence on the middle ground and background (S\_ADI). This considerable depth extension (P\_LV) further supports the systematic revelation of external scenery through these openings. This design achieves a delicate balance between openness and enclosure through deliberate textural manipulation. The foreground exhibits uniform textural characteristics (T\_ESI) with varied openness patterns (T\_IVI) and relative enclosure (T\_ISI), that contrasts with the more fragmented (T\_ESI), stable (T\_IVI), and open (T\_ISI) middle ground and background sections. This balance is reinforced through significant correlations between foreground segmentation and texture variation ( $r = 0.492, p < 0.01$ ), as well as negative relationships between the visual field ratio and both textural similarity ( $r = -0.598, p < 0.01$ ) and segmentation ( $r = -0.384, p < 0.05$ ), indicating a systematic regulation of visuals through these perspective openings. Spatial hierarchy in this scenic archetype is enhanced through sophisticated layering mechanisms. The analysis reveals continuous textural variations (T\_IVI autocorrelation = 0.341,  $p < 0.05$ ) that are coordinated with depth transitions (T\_IVI and P\_ECI correlation:  $r = 0.472, p < 0.01$ ) to ensure smooth spatial progression. The foreground elements remain relatively simple (T\_ER) compared to the richer middle ground and background sections (T\_ER), creating a clear visual distinction. This hierarchical organization is materially supported by specific landscape elements, such as trees, grass, walls, buildings, and columns, which all have a strong contribution across multiple foreground metrics (Fig. 11c).

The balanced spatial configuration and textural variations characterizing porous scenery correspond to the emptiness-substance (虚实) principle systematically articulated in traditional garden theory. This principle found systematic codification in Ji<sup>40</sup>, which documented canonical window

**Table 5 | Descriptive statistical analysis of each metric and their interpretations in the context of sandwiched scenery**

Metrics	Statistical Values	Interpretation	Implication for Design
S_API	Mean = 43.241; Mdn = 39.368; 95% CI [37.480, 49.002]; Sk = 0.678	Moderate positive skewness	Symmetrical arrangement of foreground elements
P_ECI	Mean = 1.332; 95% CI [1.213, 1.451]; Sk = -0.207	Slight negative skew	Controlled depth transitions in the sandwiching structure
P_LN	Mean = 243.057; 95% CI [235.416, 250.698]; Sk = -2.243	High value with strong negative skew	Very close positioning to the observer
S_ADI	Mean = 0.976; Mdn = 0.994; 95% CI [0.962, 0.990]	Near maximum value	Highly optimized spatial distribution
S_VFR	Mean = 0.451; Mdn = 0.44; 95% CI [0.398, 0.504]	Moderate proportion	Balanced visual field occupation
P_LV	Mean = 165.189; 95% CI [149.898, 180.480]; Sk = -0.64	High value with moderate negative skew	Extended depth reached in the foreground
Foreground T_ISI, T_IVI, T_ESI, T_ER	T_ISI = 0.068, 95% CI [0.053, 0.083]; T_IVI = 114.652, 95% CI [107.875, 121.429]; T_ESI = 2.421, 95% CI [2.183, 2.659]; T_ER = 2.717, 95% CI [2.283, 3.151]	Low T_ISI, moderate T_IVI, T_ESI, T_ER	Enclosed, uniform texture with consistent variations
Middle ground and Background T_ISI, T_IVI, T_ESI, T_ER	T_ISI = 0.767, 95% CI [0.714, 0.820]; T_IVI = 113.588, 95% CI [89.656, 99.028]; T_ESI = 3.005, 95% CI [2.699, 3.311]; T_ER = 5.868, 95% CI [5.425, 6.311]	Higher values across metrics	Open, fragmented textures with rich elements

designs embodying “adjacency to emptiness everywhere, framed views in every direction” (处处邻虚,方方侧景), specifically the strategic placement of porous openings to create “seemingly separated yet not separated” (似隔非隔) spatial relationships (Fig. 14a). For example, the floral lattice window at HWL (Fig. 14b) demonstrates the enduring application of these design principles, where ornamental geometry functions as both aesthetic object and spatial filter to achieve the characteristic “half-transparent” effect. Song et al.<sup>68</sup> document how traditional designers employed “substance within emptiness” and “emptiness within substance” to create complementary pairings: architecture with trees and water bodies, rockeries with vegetation and water, creating variations of density and sparseness that highlight layered spatial structures. The significant negative correlations between visual field ratio and both textural similarity ( $r = -0.598, p < 0.01$ ) and segmentation ( $r = -0.384, p < 0.05$ ) observed in our data reflect this principle’s practical implementation, whereby designers systematically regulated visual access through openings to balance enclosure and revelation. The complex foreground outlines (high S\_PSI) combined with varied depth transitions (high P\_ECI) align with the documented design goal of creating “limitless vision, endless recurrence” through strategic arrangement of sightlines and pathways<sup>69</sup>. Space syntax analyses confirm that traditional designers used doorways and windows to implicitly frame scenic moods, creating clustering centers positioned to maintain attractive scenery while requiring multiple turns for complete spatial perception<sup>70</sup>. The statistical patterns in porous scenery observed at HWL thus validate the traditional design strategy of achieving spatial richness through controlled permeability rather than uniform openness or complete enclosure; this principle has now been empirically confirmed through contemporary computational heritage studies.

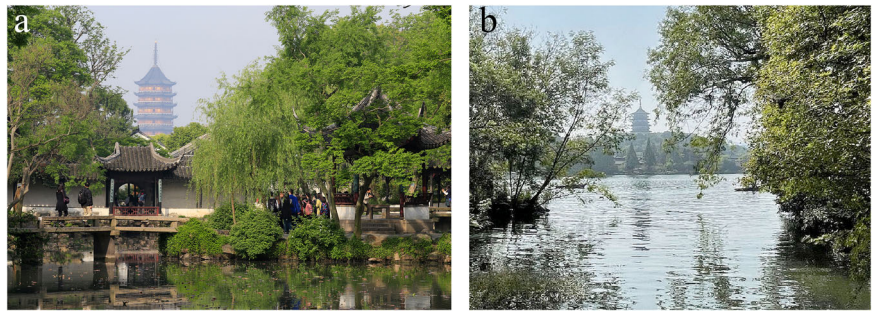
Sandwiched scenery is characterized by three interconnected aspects: structural clamping formation, bilateral visual guidance, and spatial hierarchy enhancement. The results of the correlation analysis are detailed in Fig. 10d, and the results of the descriptive statistical analysis are presented in Table 5. In sandwiched scenery, the foreground effectively establishes clamping structures through the precise control of its physical attributes. The analysis reveals that the foreground maintains highly symmetrical outlines (S\_API) with dynamic depth transitions (P\_ECI) to create a stable sandwiching structure. This sandwiching structure is strategically positioned close to the observer (P\_LN) and demonstrates optimal area distribution (S\_ADI) while occupying a moderate visual field (S\_VFR). The systematic formation of sandwiching structures is further supported by its considerable depth extension (P\_LV). Bilateral elements in the scene effectively guide visual attention through deliberate textural manipulation. The foreground exhibits enclosed characteristics (T\_ISI) with consistent

morphological variations (T\_IVI), simple and uniform textures (T\_ESI), and fewer elements (T\_ER); this is in contrast to the more open (T\_ISI), similarly varied (T\_IVI), fragmented (T\_ESI), and rich (T\_ER) middle ground and background regions. This guidance is reinforced through significant correlations between symmetrical forms and depth variation ( $r = 0.291, p < 0.05$ ), element richness and spatial hierarchy ( $r = 0.434, p < 0.01$ ), and enclosure and visual concentration ( $r = 0.49, p < 0.01$ ). Spatial hierarchy is enhanced through sophisticated organizational mechanisms. Specifically, the analysis reveals systematic relationships between visual field proportion and several variables, including spatial depth ( $r = 0.549, p < 0.05$ ), element richness ( $r = 0.532, p < 0.01$ ), and area distribution ( $r = -0.588, p < 0.01$ ), indicating coordinated spatial progression. Morphological variations exhibit significant consistency (T\_IVI correlation:  $r = 0.764, p < 0.01$ ), with element richness being positively correlated with morphological variation ( $r = 0.532, p < 0.01$ ) and negatively correlated with area distribution ( $r = -0.71, p < 0.01$ ). This hierarchical organization is materially supported by specific spatial elements, including aquatic plants, trees, plants (primarily shrubs and tall herbaceous species), and window-panes; which all have a strong and centralized contribution across multiple foreground metrics (Fig. 11d).

The bilateral symmetry and channeling characteristics defining sandwiched scenery reflect design principles embedded in traditional Chinese garden-making practice. While Ming-dynasty treatises such as Wen<sup>71</sup> emphasized aesthetic principles of spatial rhythm, specifically “positioning with balance between density and sparseness” (位置疏密), and Ji<sup>40</sup> systematically articulated spatial sequencing through borrowed scenery techniques, the specific organizational logic of bilateral framing developed through centuries of iterative practice rather than explicit theoretical formulation. The measured high symmetry in foreground outlines (S\_API mean = 43.241) and its correlation with depth variation ( $r = 0.291, p < 0.05$ ) empirically validates this embedded design intention.

Historical precedents confirm this spatial logic as conscious design knowledge. (Fig. 15a) The Humble Administrator’s Garden, created in the 1510 s, demonstrates the systematic application of sandwiched scenery: curved corridors and elevated walkways lined with bilateral plantings of bamboo and flowering trees create multi-tiered lateral framing that channels sightlines toward the “borrowed” North Temple Pagoda beyond the garden boundary, representing an integration of sandwiched scenery and borrowed scenery documented in Wen’s album<sup>71</sup> depicting the garden’s spatial sequences. This Ming-dynasty built example substantiates that bilateral framing was an intentional compositional strategy for creating “two-sided obstruction, visual guidance, and endpoint emphasis.” HWL exhibits these same organizational principles: trees positioned along both shores generate

**Fig. 15 | Sandwiched scenery: historical precedent and case manifestation.** **a** Curved corridors and bilateral plantings in the Humble Administrator's Garden, Suzhou (Ming dynasty, 1510 s). **b** Trees positioned along both shores at HWL creating visual corridors across the water surface.



visual corridors across the water surface toward distant pagodas, demonstrating the persistence of this spatial logic in contemporary landscape experience (Fig. 15b). The systematic relationships observed between visual field proportion and multiple variables (depth:  $r = 0.549$ , element richness:  $r = 0.532$ , area distribution:  $r = -0.588$ ,  $p < 0.01$ ) reflect the intentional orchestration of bilateral elements to channel visual attention and structure spatial progression. Recent fractal analyses demonstrate that successful Chinese gardens exhibit scale-dependent complexity ranges that correspond to traditional design goals<sup>72</sup>; in these studies, designers did not scale proportionally but applied different compositional strategies at different scales, validating that measured patterns reflect systematic design knowledge. The convergence of our statistical findings with space syntax validations, historical built evidence, and fractal geometry analyses establishes that sandwiched scenery's spatial patterns observed at HWL reflect conscious compositional strategies developed through centuries of practice and transmitted through both built examples and theoretical discourse.

In summary, the full implementation of the proposed framework is capable of capturing, presenting, and measuring the spatial visual characteristics of scenic archetypes. The results are consistent with our hypotheses, demonstrating the applicability and effectiveness of the framework in quantifying scenic archetypes.

## Discussion

This study advances the analysis of scenic archetypes by addressing persistent limitations in both theory-driven and measurement-driven approaches. Existing research on traditional Chinese gardens has evolved along four loosely connected paradigms: (1) the inheritance of abstract design vocabulary, which vividly describes spatial effects yet lacks systematic procedures for extraction and comparison; (2) landscape representation frameworks, which provide hierarchical categorization but do not capture the multi-layered spatial configurations specific to scenic archetypes; (3) quantitative visual indicators, which measure isolated spatial properties without accounting for the configurational relationships that generate overall spatial effect; and (4) computational vision systems, which deliver high detection accuracy but cannot directly translate element recognition into spatial-visual analysis. These paradigms, while valuable in isolation, share a methodological gap: they are unable to connect high-level theoretical concepts with verifiable spatial metrics in a structured and reproducible way. Our framework directly addresses this gap by structuring scenic archetypes into spatial-visual components and linking them to measurable indicators, enabling cross-paradigm integration and application in both research and heritage practice.

### From abstract vocabulary to operational spatial variables

Design traditions in Chinese gardens have long relied on a rich lexicon, such as borrowed scenery, framed scenery, and obstructed scenery, that has transmitted spatial knowledge across generations<sup>9,17</sup>. These poetic descriptions excel at capturing experiential richness and cultural resonance but suffer from interpretive indeterminacy: practitioners interpret “appropriate density” or “depth with layers” through personal experience, creating inconsistencies that impede systematic application and cross-cultural

communication. As Bandarin and van Oers<sup>73</sup> note in their analysis of Historic Urban Landscape approaches, and as Smith<sup>74</sup> argues regarding intangible heritage, the challenge lies not in preserving terminology but in maintaining operational knowledge. This study addresses the issue by translating abstract vocabulary into operational spatial variables without reducing it to technical jargon. For example, when we demonstrate that framed scenery exhibits specific correlation coefficients (0.549 between foreground and background), we reveal the mathematical relationships underlying poetic experience, not replacing metaphor with measurement but uncovering the quantitative structures that enable qualitative experience. This dual preservation of semantic richness and analytical precision enables what recent heritage management discourse terms “value-based indicators”<sup>75</sup> and what contemporary Chinese heritage conservation identifies as urgently needed: transparent communication across different knowledge systems without sacrificing cultural authenticity.

### From static quantities to configurational spatial patterns

Building on the preceding discussion, most existing studies, whether employing conventional landscape metrics or recent AI-based analyses of high-quality imagery, have focused on measuring single spatial-visual attributes in isolation. This tendency is reflected in the emphasis on quantitative measures such as the green view index<sup>76</sup>, sky view factor<sup>77</sup>, and other visual landscape metrics<sup>78</sup>. While these approaches have advanced measurement precision, they often conflate quantification with interpretation when applied to heritage contexts. A 65% green view ratio or 0.75 sky view factor quantifies a state, a static condition at a moment in time, but fails to capture the pattern through which such states generate meaning. As Silva<sup>79</sup> demonstrates in analyzing Historic Urban Landscapes in the Asia-Pacific, and as Veldpaus and Pereira Roders<sup>80</sup> argue in their assessment framework for historic urban landscapes, the essence of designed space emerges from relational configurations rather than aggregate indicators. This study transforms this paradigm by distinguishing between percentage (how much) and pattern (how configured). When identifying framed scenery, we analyze not the quantity of framing elements but the spatial configuration through which framing operates as a phenomenological experience. This distinction proves crucial: two gardens with identical quantitative metrics can produce entirely different spatial experiences because their configurational patterns differ. The shift from state to pattern thus represents more than methodological refinement; it constitutes an epistemological reorientation that aligns measurement with the fundamental nature of spatial experience.

### From universal frameworks to culturally specific spatial logic

Many established landscape representation frameworks, such as Tveit et al.'s<sup>38</sup> nine key concepts for visual landscape character, Bell's<sup>2</sup> elements of visual design, and Liu and Nijhuis's<sup>3</sup> spatial-visual vocabulary, provide valuable hierarchical structures that inspired our approach. Yet these frameworks encounter insurmountable limitations when applied to culturally specific landscapes: they presuppose universal aesthetic principles that transcend cultural boundaries, failing to recognize that scenic archetypes embody not merely visual arrangements but culturally constituted ways of

perceiving and inhabiting space. The Chinese concept of scenery, as Jin<sup>81</sup> elucidates in his exploration of *jing* (scenery) in traditional Chinese garden texts, fundamentally refuses the subject-object dichotomy inherent in western analytical frameworks, the observer does not view the garden from outside but participates in its continuous unfolding. As Lu and Liu<sup>14</sup> demonstrate through spatial-experiential analysis of the Master of Nets Garden, Chinese gardens operate through embodied experience rather than detached observation. Building on this premise, the proposed four-tier hierarchy does not simply append Chinese categories to existing frameworks; it restructures the analytical process to mirror what Sun<sup>82</sup> identifies as the unity of knowledge and practice fundamental to Chinese epistemology. Each tier maintains this unity: scenic archetypes preserve experiential wholeness, spatial layers translate experience into perceptual structures, visual variables extract measurable qualities without fragmenting meaning, and metrics quantify relationships while maintaining semantic integrity. This structural alignment with Chinese spatial thinking explains our 94.12% recognition accuracy success, which derives not from superior algorithms but from epistemological congruence with the phenomena being analyzed.

### From element detection to functional spatial relationships

From a technical perspective, recent computational methods, such as semantic segmentation models and computer vision systems, have achieved remarkable technical precision. YOLOv8's 93.9% element detection accuracy<sup>83</sup> and YOLOv4's 90.20% damage identification rate<sup>84</sup> demonstrate the power of contemporary AI. Yet these tools remain confined to what Liu<sup>3</sup> conceptualizes as "the world of data," unable to reach "the world of concern" where design meaning resides. A semantic segmentation model identifies walls, trees, and rocks with near-perfect accuracy but cannot distinguish whether a wall frames a view, obstructs a sightline, or merely defines a boundary: distinctions fundamental to design practice and heritage value. This study transforms computational capability into design understanding by applying scenic archetype logic as an interpretive layer: element detection provides raw data, but our four-tier structure analyzes spatial relationships to determine design function. This represents not post-processing but fundamental reorientation from asking "what exists?" to asking "what does it mean?"

Building on these four lines of inquiry, a unifying insight emerges: scenic archetypes resist reductive analysis precisely because they function as holistic design patterns where meaning emerges from relational totality rather than component aggregation. The apparent incompatibility between data thinking (emphasizing decomposition and measurement) and design thinking (prioritizing synthesis and experience) reveals not a methodological problem to solve but an ontological reality to acknowledge. Scenic archetypes encode what Ji<sup>40</sup> termed the "living method" of garden creation, principles that exist only through embodied practice, where knowledge is inseparable from action, and patterns dissolve when reduced to isolated elements. Our approach does not bridge this divide through compromise; instead, it recognizes scenic archetypes as translational mechanisms in their own right, capable of converting abstract principles into spatial organizations and, in turn, transforming spatial organizations into visual appreciation.

This recognition has direct implications for heritage conservation, shifting practice from recording only discrete physical changes to monitoring and safeguarding the spatial relationships that sustain scenic integrity. While UNESCO's Guidance and Toolkit for Impact Assessments calls for evaluating impacts on Outstanding Universal Value<sup>85</sup>, and threats to the visual integrity of World Heritage properties are well documented<sup>35,86</sup>, prevailing tools still mainly track building heights, vegetation coverage, and new construction. Recent scholarship underscores the central role of spatial organization in heritage value<sup>87,88</sup>; this research makes that organization measurable and manageable, operationalizing the long-observed fact that deterioration of spatial relationships often precedes physical degradation<sup>89,90</sup>. In impact assessments, the method quantifies how proposed changes affect scenic integrity, for example, reducing the correlation between foreground and background in a framed-scenery view from 0.549

to 0.300 indicates not merely visual intrusion but the loss of the archetype itself, even when structures remain. These analyses provide objective, defensible criteria that go beyond subjective judgments and align with UNESCO's requirements for evidence-based assessment of OUV attributes<sup>84</sup>. Establishing baseline measurements of scenic archetypes further enables proactive management, including early-warning systems that detect gradual erosion of spatial integrity before irreversible damage occurs<sup>91,92</sup>. Operationally, the approach integrates scenic archetypes preservation into conservation planning alongside traditional physical monitoring, consistent with contemporary calls to protect both tangible and spatial dimensions of heritage value while supplying the quantitative tools needed for implementation.

### From quantitative precision to integrated heritage assessment

While the preceding discussion establishes the methodological contribution of quantitative spatial analysis to heritage conservation, responsible scholarship requires explicitly acknowledging the epistemological boundaries of such approaches. What dimensions of heritage value lie beyond the reach of quantification, and how should our framework be positioned relative to these irreducible intangible dimensions? Heritage embody tangible and intangible dimensions<sup>93</sup>, where spatial visual characteristics constitute only one component of a broader constellation of values encompassing historical memory, local identity, community narratives, and intangible cultural practices that give designed spaces profound human meaning<sup>94,95</sup>. As Lian et al.<sup>16</sup> demonstrate in their systematic review of historic garden conservation approaches, effective heritage management necessarily integrates multiple analytical frameworks: landscape mapping identifies physical and spatial attributes; landscape planning addresses conservation strategies; landscape design facilitates development and reuse. Critically, their framework positions spatial analysis within broader landscape context through conceptual "layers" connecting tangible architectural elements with intangible cultural processes, temporal evolution patterns, and community value systems. This layered approach acknowledges that while spatial visual characteristics can be quantified, the cultural significance they embody requires complementary assessment through ethnographic methods, oral history documentation, and participatory evaluation engaging core communities whose lived experiences constitute irreplaceable dimensions of heritage value<sup>96,97</sup>.

This recognition of heritage value's multidimensional nature leads directly to a critical methodological question: how should our quantitative framework be positioned relative to these broader assessment requirements? The present study's analytical framework should therefore be positioned not as comprehensive heritage assessment methodology but as specialized contribution addressing specific evidence gaps in conservation practice. Our quantitative approach excels at particular tasks: providing systematic evidence for heritage inscription processes; detecting gradual spatial changes that might escape qualitative monitoring; enabling comparative analysis across multiple sites; translating design principles into implementable guidelines for contemporary practice. However, these capabilities complement rather than replace methods capturing values our framework cannot measure. A framed scenery opening, for instance, may be precisely characterized through our metrics, yet these measurements cannot convey the cultural associations viewers bring to the scene: literary references accumulated through centuries of poetic tradition, historical events that transformed physical space into commemorative place, personal recollections that link individual memory to collective heritage, or spiritual significance attributed through religious or philosophical practice<sup>98</sup>. These intangible dimensions do not supplement spatial analysis; they constitute parallel and equally valid forms of heritage value requiring distinct methodological approaches<sup>98,99</sup>.

Having established this methodological positioning, we turn to the theoretical frameworks emerging in heritage scholarship that provide conceptual foundation for integrating quantitative and qualitative approaches. Recent theoretical work establishes frameworks for this methodological integration. Mason<sup>99</sup> distinguishes heritage-centered values (historical,

aesthetic, architectural) from societal values (social, economic, environmental), arguing that while heritage-centered values may be more amenable to expert quantification, societal values require participatory assessment engaging diverse stakeholders. Robson<sup>96</sup> demonstrates through case study analysis that different assessment methods—quantitative spatial analysis, qualitative interviews, participatory mapping, photo-elicitation—surface different types of knowledge, with findings sometimes converging but often revealing dissonances requiring negotiation rather than resolution. Waterton and Smith<sup>100</sup> caution that privileging expert-defined quantifiable attributes risks marginalizing community-defined values that resist measurement, potentially creating what they term “authorized heritage discourse” that legitimizes certain forms of knowledge while delegitimizing others. These critiques do not invalidate quantitative approaches but situate them within broader epistemological landscape where multiple ways of knowing heritage coexist, each revealing distinct dimensions of significance<sup>101,102</sup>.

The preceding analysis of epistemological boundaries establishes the theoretical foundation for understanding how quantitative spatial analysis should function within comprehensive heritage assessment. This foundation enables us to articulate the broader implications for heritage science as a discipline, particularly regarding the relationship between advancing computational capabilities and the enduring necessity of humanistic methods. The broader implication for heritage science is that advancing computational analysis capabilities, as this study does through AI-enabled multimodal mapping, does not diminish the importance of qualitative methods but rather increases the imperative for their integration. More powerful quantification tools create greater risk of privileging measurable attributes simply because they are measurable, potentially marginalizing equally important values resisting quantification. The solution lies not in rejecting computational approaches but in designing assessment frameworks where quantitative precision serves rather than supplants qualitative understanding. Our framework provides template for this integration: by translating abstract scenic archetypes into measurable spatial variables, we enable systematic comparison and pattern identification across multiple sites and temporal periods; by situating these measurements within phenomenological understanding of how spatial configurations generate experiential qualities, we maintain connection between quantitative evidence and qualitative meaning; by acknowledging that spatial visual characteristics constitute only one dimension of heritage value, we position our contribution within broader assessment frameworks requiring multiple methodological approaches.

In conclusion, this study establishes an analytical framework that explores the spatial visual characteristics of scenic archetypes with AI-enabled multimodal mapping methods. By deconstructing abstract spatial concepts into measurable variables such as shape, size, position, and texture across foreground, middle ground, and background, the framework bridges traditional design principles with computational analysis, enabling replicable and interpretable mapping of landscape visual logic. While it effectively captures static visual configurations for defined archetypes, it does not yet address the temporal and kinesthetic dimensions central to sequential landscape perception or scene types requiring extended sightlines and dynamic compositional shifts. These methodological boundaries point to clear directions for advancement, including integrating eye-tracking and immersive virtual environments for dynamic visual modeling, applying advanced deep-learning architectures for element recognition in visually complex settings, extending metrics to seasonal, diurnal, and weather-induced variations, and validating the framework across diverse garden traditions to assess transferability. Equally important, future research should explore systematic integration of this spatial analysis framework with ethnographic documentation methods, oral history protocols, and participatory assessment techniques, ensuring that measurable spatial characteristics are interpreted within full cultural context encompassing intangible heritage values that our quantitative approach cannot directly capture<sup>16,103</sup>.

The broader significance of this work extends beyond disciplinary boundaries. As heritage landscapes confront accelerating pressures from

urbanization, tourism, and climate change, analytical frameworks that synthesize traditional wisdom with contemporary technology become not merely useful but essential for cultural survival. By demonstrating that quantification, when properly designed, can reveal rather than obscure complexity, this study shows how computational approaches can illuminate patterns and relationships that purely qualitative methods might overlook. By rendering the implicit explicit, the tacit measurable, and the cultural computational, we enable new modalities for preserving, transmitting, and evolving landscape design traditions. Yet this enabling occurs not through replacing humanistic understanding with technical measurement but through creating complementary forms of evidence that together support more robust conservation decisions. The scenic archetypes of Chinese gardens, refined through centuries of iterative practice, encode spatial wisdom directly relevant to contemporary challenges of place-making in an increasingly mediated world. This framework ensures such wisdom remains not preserved solely as static heritage but actively operational as living knowledge, capable of informing and inspiring future practice while maintaining continuity with its cultural origins. Achieving this aspiration requires recognizing that quantitative spatial analysis and qualitative cultural understanding are not competing paradigms but mutually necessary components of comprehensive heritage stewardship, each revealing dimensions of significance the other cannot access, together enabling conservation approaches that honor both the measurable and the ineffable dimensions of heritage value.

### Data availability

All data generated or analyzed during this study are available from the corresponding author on reasonable request. The Hangzhou Westlake image dataset and metric results have been archived and can be shared for research purposes.

### Code availability

The custom code and algorithms developed for computing spatial visual metrics (shape, size, position, texture) and performing statistical analyses are available from the corresponding author upon reasonable request. This includes scripts for image pre-processing and metric calculation.

Received: 16 February 2025; Accepted: 20 November 2025;

Published online: 16 December 2025

### References

1. Nijhuis, S. GIS-based landscape design research: Stourhead landscape garden as a case study. *A+BE | Archit. Built Environ.* **5**, 1–338 (2015).
2. Bell, S. *Elements of Visual Design in the Landscape* (Routledge, 2004).
3. Liu, M. Mapping landscape spaces: understanding, interpretation, and the use of spatial-visual landscape characteristics in landscape design. *A+BE | Archit. Built Environ.* **10**, 1–248 (2020).
4. Cushman, S. A., Evans, J. S. & McGarigal, K. Landscape ecology: past, present, and future. in *Spatial Complexity, Informatics, and Wildlife Conservation* (eds. Cushman, S. A. & Huettmann, F.) 65–82 (Springer, 2010).
5. Qi, J. et al. Development and application of 3D spatial metrics using point clouds for landscape visual quality assessment. *Landsc. Urban Plan.* **228**, 104585 (2022).
6. Dong, J., Wang, Y. & Yu, R. Application of the semantic network method to sightline compensation analysis of the Humble Administrator’s Garden. *Nexus Netw. J.* **23**, 209–225 (2021).
7. Bullington, J. East-West relational imaginaries: traditional Chinese gardens & self cultivation. *Educ. Philos. Theory* **54**, 1552–1557 (2021).
8. Lin, Y. Spatiotemporal narrative structure of the Lingering Garden based on traditional Chinese conception of time and space. *Landsc. Res.* **48**, 45–63 (2023).

9. Peng, Y. *Analysis of the Traditional Chinese Garden* (China Architectural Industry Press, 1986).
10. Zhou, W. The garden art of the summer residence. *J. Archit.* **6**, 29–32 (1960).
11. Pan, G. Viewing points and routes of Suzhou gardens. *J. Archit.* **6**, 14–18 (1963).
12. Liu, T. Research on Suzhou traditional Landscape Architecture Based on Semantic Network (Master's thesis, Northeast Forestry University, 2020).
13. Yang, H. A treatise on the garden of Jiangnan: a study on the art of Chinese traditional Garden; <https://doi.org/10.1007/978-981-16-6924-8> (Springer Nature, 2022).
14. Lu, L. & Liu, M. Exploring a spatial-experiential structure within the Chinese literati garden: the Master of the Nets Garden as a case study. *Front. Archit. Res.* **12**, 963–977 (2023).
15. UNESCO. West Lake Cultural Landscape of Hangzhou <https://whc.unesco.org/uploads/nominations/1334.pdf> (2011).
16. Lian, J. et al. Conservation and development of the historic garden in a landscape context: a systematic literature review. *Landsc. Urban Plan.* **246**, 105027 (2024).
17. Zhou, W. *History of traditional Chinese Gardens* (Tsinghua University Press, 1999).
18. Wang, J. *The research on Chinese painting frame from the form of traditional of enframed scenery* (Master's thesis, Huazhong Agricultural University, 2013).
19. Tong, M. Towards a view with vista: tectonics of visual culture in the gardens of eastern China. *Time+ Archit.* **5**, 56–66 (2016).
20. Abrams, J. & Hall, P. *Else/where: Mapping New Cartographies of Networks and Territories* (University of Minnesota Press, 2006).
21. Pinzon Cortes, C. *Mapping Urban Form. Morphology Studies in the Contemporary Urban Landscape* (Dissertation, 2009).
22. Corner, J., Corner, J. M. & MacLean, A. S. *Taking Measures Across the American Landscape* (Yale University Press, 1996).
23. Zhou, K., Wu, W., Dai, X. & Li, T. Quantitative estimation of the internal spatio-temporal characteristics of ancient temple heritage space with space syntax models: a case study of Daming Temple. *Buildings* **13**, 1345 (2023).
24. Chen, H. & Yang, L. Analysis of narrative space in the Chinese traditional garden based on narratology and space syntax—taking the Humble Administrator's Garden as an example. *Sustainability* **15**, 12232 (2023).
25. Zhang, C., Lv, Z., Liu, Z. & Sun, Y. A case study based on space syntax theory: West Shu Garden of Qingxi, Dujiangyan scenic area. *Sustainability* **16**, 9459 (2024).
26. Chen, X. H., Yu, H. T., Xiong, R. J. & Ye, Y. Construction of an analytical framework for spatial indicator of Chinese traditional gardens based on space syntax and machine learning. *Landsc. Archit.* **31**, 123–131 (2024).
27. Zhao, H., Shi, J., Qi, X., Wang, X. & Jia, J. Pyramid scene parsing network. In *Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2881–2890 (IEEE, 2017).
28. Zhou, Z., Fan, X., Shi, P. & Xin, Y. R-MSFM: recurrent multi-scale characteristic modulation for monocular depth estimating. In *Proc. 2021 IEEE/CVF International Conference on Computer Vision (ICCV)* 12777–12786; <https://doi.org/10.1109/ICCV48922.2021.01254> (IEEE, 2021).
29. Bochkovskii, A. et al. Depth Pro: sharp monocular metric depth in less than a second; <https://doi.org/10.48550/arXiv.2410.02073> (2024).
30. Li, Z., Bhat, S. F. & Wonka, P. PatchRefiner: leveraging synthetic data for real-domain high-resolution monocular metric depth estimation. In *Proc. 2025 European Conference on Computer Vision (ECCV)* 250–267 (Springer, 2025).
31. Li, C. et al. UrbanSAM: learning invariance-inspired adapters for segment anything models in urban construction; <https://doi.org/10.48550/arXiv.2502.15199> (2025).
32. Deng, J., Hong, D., Li, C. & Yokoya, N. Joint super-resolution and segmentation for 1-m impervious surface area mapping in China's Yangtze River Economic Belt; <https://doi.org/10.48550/arXiv.2505.05367> (2025).
33. Li, C. et al. Learning disentangled priors for hyperspectral anomaly detection: a coupling model-driven and data-driven paradigm. *IEEE Trans. Neural Netw. Learn. Syst.* **36**, 6883–6896 (2025).
34. Peirce, C. S. *Peirce on Signs: Writings on Semiotic* (UNC Press Books, 1991).
35. Liu, M. & Nijhuis, S. Talking about landscape spaces: towards a spatial-visual landscape design vocabulary. *Des. J.* **25**, 263–281 (2022).
36. Hatfield, G. Kant on the perception of space (and time). in *The Cambridge Companion to Kant and Modern Philosophy* (ed. Guyer, P.) 61–93 (Cambridge University Press, 2006).
37. Merleau-Ponty, M. *Phenomenology of Perception* (Routledge, 1962).
38. Tveit, M., Ode, A. & Fry, G. Key concepts in a framework for analysing visual landscape character. *Landsc. Res.* **31**, 229–255 (2006).
39. Chen, C. *On Chinese Gardens* (Better Link Press, 2008).
40. Ji, C. *The Craft of Gardens* (Yale University Press, 1988).
41. Li, Y. *Xianqing Ouyi* (Good Fortune Culture Press, 2020).
42. Liu, D. *Suzhou traditional Gardens* (China Architecture & Building Press, 2005).
43. Higuchi, T. *Visual and Spatial Structure of Landscapes* (MIT Press, 1988).
44. Lu, S. & Liu, S. *Dictionary of Ancient Chinese Architecture* (Beijing Institute of Cultural Relics Press, 1992).
45. Lu, A. Lost in translation: modernist interpretation of the Chinese garden as experiential space and its assumptions. *J. Archit.* **16**, 499–527 (2011).
46. Du, X. T. & Ji, F. Q. Experience of the sense of “farness”: instability of spatial depth in traditional Chinese gardens. *Landsc. Archit.* **30**, 130–136 (2023).
47. Beijing Landscape Architecture School. *Garden Planning and Design* (Beijing Science and Technology Press, 1988).
48. Attneave, F. Some informational aspects of visual perception. *Psychol. Rev.* **61**, 183–193 (1954).
49. Palmer, S. E. *Vision Science: Photons to Phenomenology* (MIT Press, 1999).
50. Bertin, J. *Semiology of Graphics: Diagrams, Networks, Maps* (transl. Berg, W. J.) (University of Wisconsin Press, 1983).
51. Ranftl, R., Lasinger, K., Hafner, D., Schindler, K. & Koltun, V. Towards robust monocular depth estimation: mixing datasets for zero-shot cross-dataset transfer. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 1623–1637 (2022).
52. Arnheim, R. *Art and Visual Perception: A Psychology of the Creative Eye* (University of California Press, 1954).
53. Kosslyn, S. M. Information representation in visual images. *Cognit. Psychol.* **7**, 341–370 (1975).
54. Ching, F. D. K. *Architecture: Form, Space, and Order* (John Wiley & Sons, 2023).
55. Li, H. Chinese architecture in 1962, from style to composition and others. *Time + Architecture* **3**, 20–26 (2021).
56. Li, K. & Jan, W. Phenomenological landscape study: the modernity of Chinese traditional perception of landscape reflected in serial scenes. *Chin. Landsc. Archit.* **25**, 29–33 (2009).
57. Zhou, B. et al. Scene parsing through ADE20K dataset. In *Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 633–641 (IEEE, 2017).
58. Kirillov, A. et al. Segment anything. In *Proc. 2023 IEEE/CVF International Conference on Computer Vision (ICCV)* 4015–4026; <https://doi.org/10.1109/ICDAR.1995.598994> (IEEE, 2023).
59. Ho, T. K. Random decision forests. In *Proc. 3rd International Conference on Document Analysis and Recognition* 1, 278–282 (IEEE, 1995).

60. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
61. Arndt, S., Turvey, C. & Andreasen, N. C. Correlating and predicting psychiatric symptom ratings: Spearman's  $r$  versus Kendall's tau correlation. *J. Psychiatr. Res.* **33**, 97–104 (1999).
62. Hauke, J. & Kossowski, T. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaest. Geogr.* **30**, 87–93 (2011).
63. Drezner, Z., Turel, O. & Zerom, D. A. modified Kolmogorov-Smirnov test for normality. *Commun. Stat. Simul. Comput.* **39**, 693–704 (2010).
64. Kline, R. B. *Principles and Practice of Structural Equation Modeling* (Guilford Publications, 2023).
65. Zhu, H., Kong, Y., Zhang, H., Gu, Z. & Ohno, R. Effects of scenery frame on visual depth perception in traditional Chinese gardens: a case study of the Lvyin Pavilion in Lingerling Garden. *Front. Archit. Res.* **14**, 402–415 (2025).
66. Chen, Y. et al. Unveiling the dynamics of “scenes changing as steps move” in a Chinese traditional garden: a case study of Jingxinzhai Garden. *Herit. Sci.* **12**, 131 (2024).
67. Li, J. & Cahill, J. Paintings of Zhi Garden by Zhang Hong: *Revisiting a Seventeenth-Century Chinese Garden* (Los Angeles County Museum of Art, 1996).
68. Song, Z., Jiang, H., Cui, T. Exploring the correlation of space creation in Suzhou traditional gardens and the Chinese calligraphy Yan Zhenqing's three manuscripts. *J. Asian Archit. Build. Eng.* <https://doi.org/10.1080/13467581.2024.2358202> (2023).
69. Chen, H., Li, Y. & Yang, L. Creating an endless visual space: an isovist analysis of a small traditional Chinese garden. *Environ. Plann. B Urban Anal. City Sci.* <https://doi.org/10.1177/23998083241298739> (2024).
70. Wu, W., Zhou, K., Li, T. & Dai, X. Spatial configuration analysis of a traditional garden in Yangzhou city: a comparative case study of three typical gardens. *J. Asian Archit. Build. Eng.* **24**, 593–604 (2024).
71. Wen, Z. *Zhangwu Zhi [Treatise on Superfluous Things]* (trans. and annot. Li, R.) (Zhonghua Book Company, 2021).
72. Sun, C., Jiang, Z. & Yu, B. How to interpret Jiangnan gardens: a study of the spatial layout of Jiangnan gardens from the perspective of fractal geometry. *Herit. Sci.* **12**, 353 (2024).
73. Bandarin, F. & van Oers, R. *The Historic Urban Landscape: Managing Heritage in an Urban Century* (Wiley-Blackwell, 2012).
74. Smith, L. *Uses of Heritage* (Routledge, 2006).
75. Avrami, E., Macdonald, S., Mason, R. & Myers, D. (eds.) *Values in Heritage Management: Emerging Approaches and Research Directions* (Getty Conservation Institute, 2019).
76. Liu, Y., Pan, X., Liu, Q. & Li, G. Establishing a reliable assessment of the green view index based on image classification techniques, estimation, and a hypothesis testing route. *Land* **12**, 1030 (2023).
77. Gong, F.-Y. et al. Mapping sky, tree, and building view factors of street canyons in a high-density urban environment. *Build. Environ.* **134**, 155–167 (2018).
78. Ode, Å, Hagerhall, C. M. & Sang, N. Analysing visual landscape complexity: theory and application. *Landsc. Res.* **35**, 111–131 (2010).
79. Silva, K. D. (ed.) *The Routledge Handbook on Historic Urban Landscapes in the Asia-Pacific* (Routledge, 2020).
80. Veldpaus, L. & Pereira Roders, A. R. Historic urban landscapes: an assessment framework, part I. In *Proc. 33rd Annu. Conf. Int. Assoc. Impact Assess. (IAIA13)* (International Association for Impact Assessment, 2013).
81. Jin, X. Jing, the concept of scenery in texts on the traditional Chinese garden: an initial exploration. *Stud. Hist. Gard. Des. Landsc.* **18**, 339–365 (1998).
82. Sun, C. H. China's traditional empirical way of thinking and its influence. *Jiangxi Soc. Sci.* **32**, 27–31 (2012). (in Chinese).
83. Gao, C. et al. Cross-cultural insights into traditional Jiangnan gardens of China and Japanese gardens through algorithm-enhanced comparative analysis. *npj Herit. Sci.* **13**, 315 (2025).
84. Yan, L. et al. Application of computer vision technology in surface damage detection and analysis of shed thin tiles in China: a case study of the traditional gardens of Suzhou. *Herit. Sci.* **12**, 72 (2024).
85. Court, S., Jo, E., Mackay, R., Murai, M. & Therivel, R. Guidance and Toolkit for Impact Assessments in a World Heritage Context (UNESCO, ICCROM, ICOMOS & IUCN, Paris, France, 2022). <https://unesdoc.unesco.org/ark:/48223/pf0000382347>
86. Ashrafi, B., Kloos, M. & Neugebauer, C. Heritage impact assessment, beyond an assessment tool: a comparative analysis of urban development impact on visual integrity in four UNESCO World Heritage Properties. *J. Cult. Herit.* **47**, 199–207 (2021).
87. Smith, L. Intangible heritage: a challenge to the authorised heritage discourse?. *Rev. Etnol. Catalunya* **40**, 133–142 (2015).
88. Nic Craith, M. & Kockel, U. (Re-)Building heritage: integrating tangible and intangible. in *A Companion to Heritage Studies* (eds. Logan, W., Nic Craith, M. & Kockel, U.) 426–442 (Blackwell Publishing Ltd, 2015).
89. DeSilvey, C. *Curated Decay: Heritage Beyond Saving* (University of Minnesota Press, 2017).
90. Vandesande, A. & Van Balen, K. Preventive conservation applied to built heritage: a working definition and influencing factors. in *Innovative Built Heritage Models* 63–72 (CRC Press, 2018).
91. Della Torre, S. Italian perspective on the planned preventive conservation of architectural heritage. *Front. Archit. Res.* **10**, 108–116 (2021).
92. Van Balen, K. Preventive conservation of historic buildings. *Int. J. Restor. Build. Monum.* **21**, 99–104 (2015).
93. Azzopardi, E., Mason, K., Strlic, M. & Dillon, C. The social value of built heritage: an interdisciplinary discourse. *Built Herit* **9**, 3 (2025).
94. Tang, C. et al. Heritage perspectives on cultural memory and spatial identity in Yuan River Basin, Hunan, China. *npj Herit. Sci.* **2**, 12 (2025).
95. Zhang, S. et al. Spatial distribution and pedigree age of intangible cultural heritage along the Grand Canal of China. *npj Herit. Sci.* **12**, 198 (2024).
96. Robson, E. Assessing the social values of built heritage: participatory methods as ways of knowing. *Architecture* **3**, 489–510 (2023).
97. Li, J. et al. Community participation in cultural heritage management: a systematic literature review comparing Chinese and international practices. *Cities* **96**, 102476 (2020).
98. Taylor, K. & Lennon, J. Cultural landscapes: a bridge between culture and nature?. *Int. J. Herit. Stud.* **17**, 537–554 (2011).
99. Mason, R. Assessing values in conservation planning: methodological issues and choices. in *Assessing the Values of Cultural Heritage* (ed. de la Torre, M.) 5–30 (Getty Conservation Institute, 2002).
100. Waterton, E. & Smith, L. The recognition and misrecognition of community heritage. *Int. J. Herit. Stud.* **16**, 4–15 (2010).
101. de la Torre, M. (ed.) *Assessing the Values of Cultural Heritage* (Getty Conservation Institute, 2002).
102. Pereira Roders, A. & van Oers, R. Outstanding universal value, world heritage cities and sustainability: mapping assessment processes. In *Heritage 2010: Heritage and Sustainable Development* (eds. Amoêda, R., Lira, S. & Pinheiro, C.) 1419–1428 (Green Lines Institute, 2010).
103. Brennan-Horley, C. & Gibson, C. GIS, ethnography, and cultural research: putting maps back into ethnographic mapping. *Inf. Soc.* **26**, 92–103 (2010).

## Acknowledgements

The authors are grateful for the support provided by their institutions during this research. We thank our colleagues for assistance with field photography and data annotation at Hangzhou Westlake. This work was supported by the Basic and Applied Basic Research Foundation of Guangdong Province [grant number 2023A1515011360]; the China Sponsorship Council [grant

number 202208420034]; and the Young Scientists Fund of the National Natural Science Foundation of China [grant number 52208054].

### Author contributions

The first author, Junkai Lan, is responsible for the conception, methodology, data collection, analysis, writing of the manuscript, and preparing all figures. Mei Liu contributed to the development of the research idea, provided writing suggestions, reviewed the manuscript, and supported the project with funding. Eric Luiten and Gregory Bracken offered valuable feedback and suggested revisions to the manuscript. Qian Zhang provided guidance on the data analysis.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to Mei Liu.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025