Research Square

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

Tourists' Spatial Perception of Traditional Villages Based on UGC Data: A Case Study of Heshun Ancient Town, China

Yun Zhang
Dalian University of Technology
Xixuan Fan
fxx@dlut.edu.cn

Dalian University of Technology
Dongqiang Zhang
Baoshan University
Xiaodong Lu
Dalian University of Technology

Article

Keywords: Traditional villages, Spatial perception, Multimodal analysis framework, User-generated content, Heshun Ancient Town

Posted Date: June 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9720828/v1

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

Additional Declarations: No competing interests reported.

Tourists' Spatial Perception of Traditional Villages Based on UGC Data: A Case Study of Heshun Ancient Town, China

Abstract

As a core component of China's cultural heritage, traditional Chinese villages (TCVs) have become a key driver of regional economic revitalization. However, the rapid expansion of the tourism industry poses a serious challenge to their spatial patterns and cultural ecosystems. The complexity of tourists' spatial experiences makes it difficult to accurately quantify their perceptions. This paper constructs a quantitative analysis framework based on User-generated content (UGC) multimodal data, matching tourists' emotional tendencies and visual preferences to specific points of interest (POI), and combines this with spatial kernel density estimation (KDE) to generate heatmaps, revealing the spatial distribution patterns of tourist hotspots. The method proposed in this paper lays an empirical foundation for targeted micro-updates and the prevention of excessive commercialization in digital heritage management.

Keywords Traditional villages, Spatial perception, Multimodal analysis framework, User-generated content, Heshun Ancient Town

1. Introduction

TCVs are rural settlements that have developed unique regional cultural characteristics through long-term historical evolution. They possess a wealth of tangible and intangible cultural heritage resources, encompassing a wide range of heritage elements such as traditional architecture, historic streets and alleys, folk arts and crafts, and local beliefs, and constitute an important component of the cultural heritage system (Chen et al., 2024; Wu et al., 2020). With the rapid development of the digital economy, tourism in traditional villages has become one of the core drivers of regional economic revitalization (Bao et al., 2023). However, the protection and development of traditional villages face severe challenges stemming from the rapid expansion of the tourism industry. Excessive commercialization, homogenization of resources, the loss of cultural significance, and the destruction of traditional spaces have weakened the core competitiveness of traditional villages and caused irreversible damage to the local ecological environment and sociocultural fabric (Yang et al., 2022). Exploration of tourists' perceptions of traditional villages and their spatial preferences is a key pathway to optimizing the spatial allocation of tourism resources and achieving the sustainable utilization of cultural heritage (Ren et al., 2025). Developing a perception evaluation scale is an effective method for exploring the relationship between cultural landscape quality, architectural character, and tourist satisfaction (Shen et al., 2025). Participatory observation and qualitative content analysis can also provide in-depth insights into the sense of place and perceived value that tourists develop in historic and cultural towns (Lu et al., 2025; Zhang et al., 2025). Furthermore, comprehensive evaluation systems for the

tourism image of traditional villages are widely used to investigate the driving effects of natural environments and historical and cultural elements on tourist behavior (Jia et al., 2025; Yanan et al., 2024). However, existing evaluation systems for traditional villages often focus on macro-level heritage conservation policies or qualitative and scale-based analyses, making it difficult to accurately capture tourists' actual on-site emotional fluctuations and attention distribution. A quantitative analysis framework that integrates multi-source data can accurately capture tourists' authentic perceptions, thereby reducing the risk of planning strategies becoming disconnected from real-world needs.

Research into tourist perceptions and spatial preferences urgently requires support from objective, up-to-date data sources. Traditional research methods based on questionnaires and in-depth interviews are costly to implement, have limited sample sizes, and are susceptible to cognitive biases among respondents. Furthermore, such static data lacks geospatial coordinates, making it difficult to track the spatiotemporal behavior of large-scale groups and rendering it unable to keep pace with the rapid evolution of the tourism industry (Hu et al., 2024; Li et al., 2025; Sánchez-Franco et al., 2024). In contrast, UGC is decentralized and records tourists' authentic experiences across multiple stages—before, during, and after travel—making it the most valuable data source in tourism research (Hu et al., 2024). UGC not only contains rich textual sentiment but also carries a large volume of visual images and location tags, significantly enhancing the credibility of tourist perceptions (Sánchez-Franco et al., 2024). At the same time, UGC has evolved from single-text data into a multimodal data system integrating text, images, and geotags, enabling a multidimensional portrayal of tourists’ cognition, emotions, and spatial behavior (liu et al., 2026 ; Ai et al., 2025).

Early research on UGC was largely confined to monomodal text mining or image classification, an approach that tends to sever the intrinsic connection between subjective emotions and objective physical contexts (Hu et al., 2024). With the application of artificial intelligence technologies, multimodal data fusion has become a key analytical method in current tourism perception research (Xu et al., 2023). In the field of natural language processing, analytical methods have evolved from discrete word frequency statistics and LDA topic modeling (Guo et al., 2017; Li et al., 2018) to deep context awareness, addressing the challenges traditional methods face in capturing long-range semantic dependencies and complex contextual information. Pre-trained language models based on the Transformer architecture, SnowNLP, and bidirectional long-short-term memory (LSTM) networks combined with attention mechanisms can dynamically adjust the contextual weights of text and are therefore used to analyze tourists’ emotional tendencies in complex scenarios (Yang et al., 2025 ; Tian et al., 2025). In the field of computer vision, lightweight convolutional neural networks (Zhang et al., 2021), object detection algorithms (Deng et al., 2018), and large-scale visual models (Li et al., 2025) have replaced manual image coding, enabling the efficient extraction of features from large-scale scenes, architectural components, and natural elements. These technologies transform pixel-level image features into high-level semantic information, establishing a mapping pathway from physical space to cognitive space, which is then used to parse visual preferences within images (He et al., 2022). Furthermore, by structurally aligning the textual sentiment polarity of tourists with the spatial entity features in images, it is possible to achieve cross-validation

between textual sentiment mining and visual feature analysis. This establishes a technical framework for objectively and dynamically quantifying tourists' authentic multidimensional perceptions (Liu et al., 2025; Wang et al., 2025).

Although multimodal data provides rich dimensions for quantifying tourist perception, existing research still faces limitations in establishing a coupling relationship between digital features and physical spaces. tourists' spatial perception is tied to specific geographic entities and spatial elements. However, current studies primarily focus on the fusion of multimodal data features and rarely address the mapping of data to physical spaces. This makes it difficult to translate evaluation results into spatial optimization strategies for specific streets, alleys, or buildings (Salas-Olmedo et al., 2018). POI, as digital representations of geographic entities, contain spatial attributes, business types, and structural elements, and thus provide a data foundation for establishing mapping relationships (Gao et al., 2017). Using Geographic Information Systems (GIS), sentiment scores extracted by natural language processing and computer vision algorithms can be matched to specific POI nodes (Li et al., 2018). Through spatial statistical methods such as Kernel Density Estimation (KDE), standard deviation ellipses, and hotspot analysis, discrete location data can be transformed into continuous spatial density surfaces, thereby visualizing the spatial patterns, diffusion trends, and evolution of tourist hotspots and emotional tendencies (Zheng et al., 2017; García-Palomares et al., 2015). Research integrating multimodal data quantification methods with spatial geographic analysis provides a methodological framework for identifying spatial resource misallocations, delineating tourist hotspots and cold spots, and addressing issues of excessive commercialization (Gao et al., 2017; Liu et al., 2022).

Using Heishun Ancient Town as a case study, this study constructs a multimodal quantification framework that integrates text mining, visual analysis, and spatial mapping. First, we collected multimodal UGC and POI data and constructed a location-based dictionary through semantic merging to establish a cognitive baseline. Second, we employed the SnowNLP algorithm to quantify text sentiment scores and capture emotional tendencies. Next, we utilized the cross-modal pre-trained large model CLIP for scene classification and label analysis to extract tourists’ objective visual preferences. Finally, we linked the sentiment and visual analysis results to POI nodes and rendered them as spatial heatmaps using GIS-based KDE, revealing the heterogeneous distribution patterns of tourist perceptions. This paper proposes a UGC-based analytical framework that addresses the urgent need for objective spatial assessment amid homogenized competition in heritage tourism. It provides a scientific basis for heritage conservation, tourist management, itinerary optimization, and spatial resource reallocation in TCVs.

2. Materials and Methods

2.1 Study Area

This paper selects Heishun Ancient Town in Tengchong City, Yunnan Province, China, as a case study for research on perceptions of traditional villages and spatial preferences (Fig. 1). The study area covers a total of 17.4 km² and preserves a relatively intact traditional village, rich in local folk culture, clan culture, and the culture of the Ancient Tea-

Horse Road. Core heritage elements include ancient streets and alleys, ancestral halls and academies, ancient bridges and wells, feng shui forests, and water inlet systems. Due to the high level of tourism activity, the town has accumulated a large volume of multimodal UGC data on social media. In recent years, with the intensification of tourism development, it has faced increasingly complex challenges in balancing heritage conservation, commercial expansion, and the diverse experiential needs of tourists. Using this site as a case study for spatial perception analysis can objectively reflect tourists' authentic experiences of this composite heritage space and can serve as a representative model for exploring the refined governance and revitalization of traditional villages.

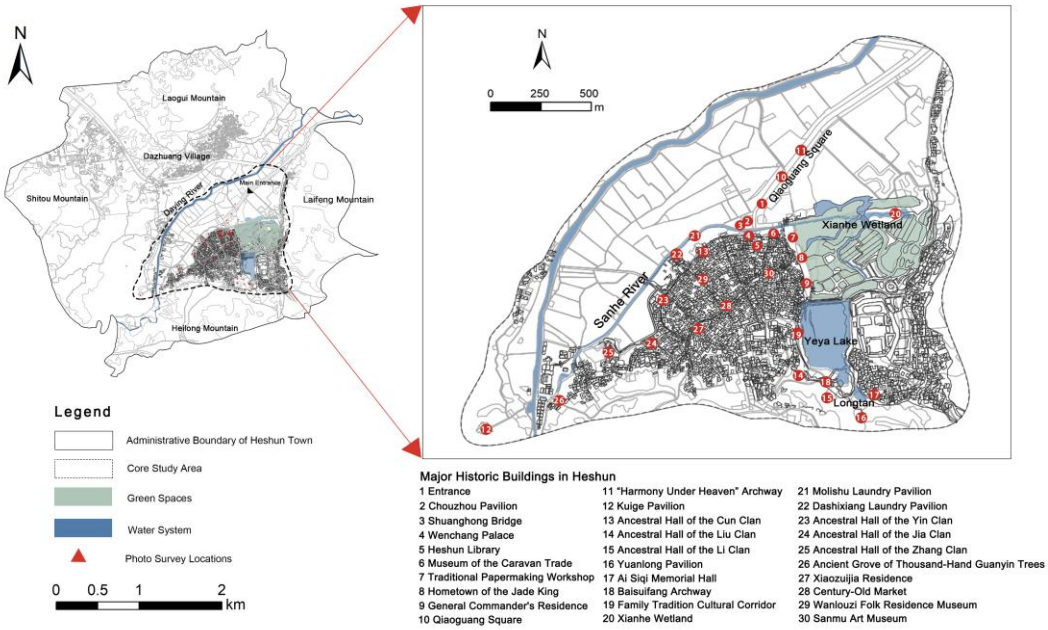


Fig. 1 Spatial Layout Map of Heshun Ancient Town

2.2 Data Sources

2.2.1 Multimodal UGC Data

Weibo and Xiaohongshu are China's two leading social media platforms. Weibo focuses on real-time updates and emotional expression, providing a rich source of textual data. Xiaohongshu is a visually oriented lifestyle community that has accumulated a vast collection of high-quality travel photos and records of spatial experiences. Combining these two platforms provides representative multimodal data for analyzing tourists' emotional characteristics and spatial preferences. This paper used Weibo and Xiaohongshu as data sources and searched using the keyword "Heshun Ancient Town." The time window was set from January 1, 2021, to December 31, 2025. Using a Python crawler, we collected 12,806 valid UGC text-image samples, containing 38,154 images. After filtering, we ultimately retained 17,684 valid images. The resulting multimodal dataset preserves key metadata, including user ID, publication time, title, body text, and IP location.

2.2.2 Points of Interest

POI data records the coordinates and attributes of various elements in physical space

and can be used for spatial analysis. This paper uses Amap as the data source (Peng et al., 2023) and employs a Python program to call its web service API. Centering on the ancient town of Heshun, a search radius of 3,000 meters was set, and data were selectively extracted from three dimensions: historical features, natural landscapes, and tourism services. Global deduplication and data cleansing were performed based on Amap's unique IDs and field verification, ultimately yielding 1,496 POI spatial nodes (Wu et al., 2025). This provides the spatial data foundation for subsequent analyses of spatial density and the spatial overlay of objective and subjective factors (Zhang et al., 2024).

2.3 Research methods

This paper integrates multimodal UGC data with POI spatial features to construct a research framework comprising "text mining, visual analysis, and spatial mapping." This framework enables the study of tourists' perceptual characteristics and behavioral preferences from multiple dimensions. The research process consists of the following four steps (Fig. 2):

(1) Data Acquisition and Lexicon Construction: A Python crawler was used to collect multimodal UGC data and POI data from traditional villages. By extracting high-frequency words from UGC text, we preliminarily identified tourists' overall perceptions of and focal points of interest regarding Heishun Ancient Town. We further performed semantic merging on these high-frequency words to generate a "location information word set."

(2) Sentiment Quantification: UGC text containing the "location information word set" was extracted, and sentiment analysis algorithms were applied to calculate sentiment scores, thereby analyzing tourists' emotional tendencies.

(3) Visual Preference Analysis: Scenes were categorized based on the location information word set. An image recognition model was introduced to classify and decode UGC image data, thereby quantifying tourists' photographic preferences for each scene. Key visual elements within each scene were further identified to explore tourists' visual attention patterns.

(4) Spatial Mapping and Visualization: Map UGC image data to corresponding POI nodes and calculate the frequency of image records for each POI. Generate spatial heatmaps using a GIS platform in conjunction with the KDE method to reveal the spatial distribution patterns of tourist preferences.

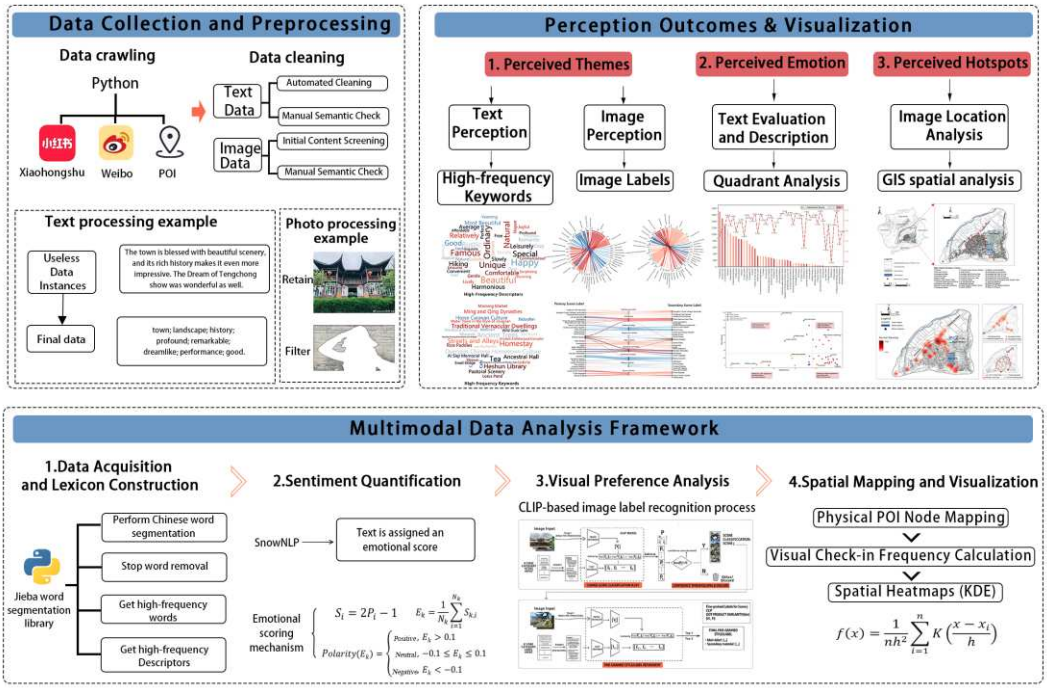


Fig. 2. Research framework

2.3.1 Text Segmentation and Lexicon Construction

Text segmentation, as a natural language processing technique, can divide continuous unstructured text into discrete lexical units. This study utilizes the Python third-party library jieba to mine UGC text data, following these specific steps[36]:

(1) Preprocess the UGC text data by removing text noise, punctuation, and redundant information, retaining only valid Chinese characters.

(2) Calculate the frequency of words after segmentation and extract keywords describing cultural heritage, natural landscapes, tourism services, and intangible culture. Filter out high-frequency words that appear more than 40 times.

(3) Dedupe and semantically merge the selected high-frequency words to generate a "location information vocabulary" comprising 30 core elements.

2.3.2 Text Sentiment Analysis Based on the Location Lexicon

Text sentiment analysis can identify subjective text and quantify its sentiment polarity. This paper uses the SnowNLP model as a sentiment measurement tool. This model is an algorithm library specifically designed for Chinese text mining, and it demonstrates high accuracy and robustness when processing short Chinese texts from social media. The cleaned and deduplicated full-sample UGC texts are input into the SnowNLP model to calculate the initial sentiment probability, P_i. This probability is then linearly mapped to the [-1, 1] interval to obtain the sentiment score, S_i, for each individual text:

$$S_i = 2P_i - 1$$

where $P_i \in [0,1]$ is the original sentiment probability of the i-th text. An S_i value approaching 1 indicates an extremely positive emotion, while a value approaching -1 indicates an extremely negative emotion.

Subsequently, targeted searches were conducted within the corpus based on the

location-based lexicon to extract text sets associated with specific spatial nodes. Using a single spatial element as the analysis unit, the mean sentiment score of the corresponding text set was calculated:

$$E_k = \frac{1}{N_k} \sum_{i=1}^{N_k} S_{k,i}$$

where E_k is the overall mean sentiment score of the k-th spatial node; N_k is the total number of associated texts for the k-th spatial node; and $S_{k,i}$ is the sentiment score of the i-th text within the associated text set of the k-th spatial node.

To quantify tourists' emotional attitudes toward a location, an emotional score greater than 0.1 is defined as "positive," less than -0.1 as "negative," and values between the two as "neutral." Based on this classification criterion, the overall mean sentiment for each factor and the proportion of samples in each polarity are calculated, thereby revealing differences in tourists' emotional attitudes and perceptions toward different locations. The classification function is defined as follows:

$$Polarity(E_k) = \begin{cases} Positive, & E_k > 0.1 \\ Neutral, & -0.1 \leq E_k \leq 0.1 \\ Negative, & E_k < -0.1 \end{cases}$$

2.3.3 Image Visual Decoding Based on the CLIP Model

To extract structured visual features from massive image data, we constructed an image perception framework utilizing CLIP (Fig. 3). Proposed by OpenAI, CLIP is a cross-modal large model pre-trained on massive image-text pairs via contrastive learning. This architecture circumvents the traditional deep learning paradigm that relies heavily on domain-specific annotated data, thus possessing robust multi-scene generalization and zero-shot reasoning capabilities.

(1) Scene Classification Decoding and Shooting Preference Quantification

Based on the spatial characteristics of the ancient town, the visual corpus was classified into nine mutually exclusive scene categories: Architecture, Streets and Alleys, Bridges and Pavilions, Water Systems, Mountains and Forests, Commercial Facilities, Catering Facilities, Service Facilities, and Festivals and Events. To ensure mutual exclusivity, 'Commercial Facilities' were strictly defined as venues primarily for retail or lodging (e.g., guesthouses, souvenir shops), whereas 'Catering Facilities' were designated as venues for on-site food and beverage service.

Using Prompt Engineering, the K categories of scenes were converted into text prompts with contextual semantics. The CLIP model mapped the i-th input image and the j-th scene prompt text into a shared multi-modal feature space, extracting feature vectors $\mathbf{v}_i$ and $\mathbf{t}_j$. to calculate their cosine similarity:

$$sim(\boldsymbol{v}_i, \boldsymbol{t}_j) = \frac{\boldsymbol{v}_i \cdot \boldsymbol{t}_j}{\| \boldsymbol{v}_i \| \| \boldsymbol{t}_j \|}$$

To quantify shooting preferences, the Softmax function was applied to normalize the cosine similarity, yielding the probability distribution of the image belonging to each scene:

$$P_{i,j} = \frac{\exp\left(sim(\boldsymbol{v}_i, \boldsymbol{t}_j)\right)}{\sum_{k=1}^{K} \exp\left(sim(\boldsymbol{v}_i, \boldsymbol{t}_k)\right)}$$

where $P_{i,j}$ represents the predicted probability that the i-th image belongs to the j-th scene category; and K is the total number of extracted scene categories.

To filter out high-noise images, we set a confidence threshold, θ. We confirmed the scene attribution only when the highest predicted probability $P_{i,j} > \theta$. Otherwise, we uniformly classified the image into an irrelevant category.

(2) Detailed Feature Analysis of Multi-scene Images

While CLIP possesses robust generic zero-shot capabilities, applying it directly to TCVs posed semantic challenges due to the domain shift between general internet pre-training and highly localized vernacular heritage features. To mitigate this gap, we curated a localized prompt dictionary (detailed in Table A2). For images with a confirmed scene attribution j, this study further matched visual labels. For example, for the "architecture" scene, we mapped a set $E_j = \{e_1, e_2, \ldots, e_M\}$ containing M labels. Examples include "white walls and black tiles" and "wooden components." In this stage, the CLIP model performed multi-label feature reasoning. It calculated the similarity score, $S_{i,m}$, between the image feature vector $\mathbf{v}_i$ and each label text feature vector $\mathbf{t}_{e_m}$. This explored the visual focus of tourists in that specific scene:

$$S_{i,m} = sim(\boldsymbol{v}_i, \boldsymbol{t}_{e_m})$$

We sorted all candidate labels in descending order based on their scores. We then extracted the top two items with the highest probabilities. These form the "primary label" and "secondary label" set, $Tags_i$, for that image. This process transforms unstructured image data into a structured feature matrix:

$$Tags_i = \text{argmax}^{\text{top-2}}_{\text{m}\in\{1,2,\ldots,\text{M}\}}(S_{i,m})$$

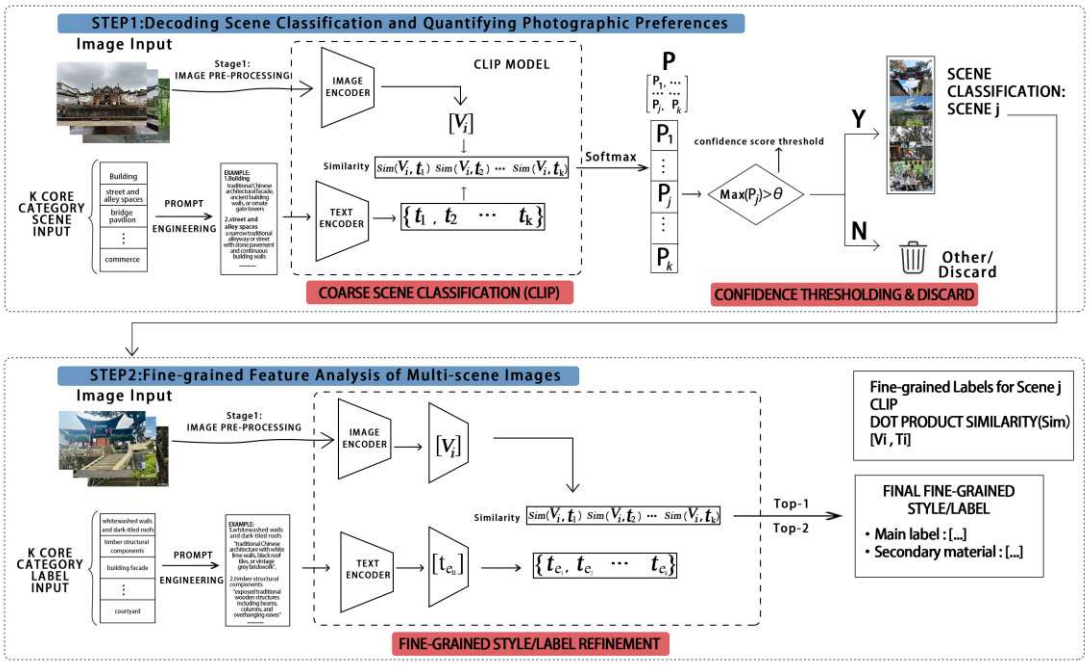


Fig. 3. CLIP-based image label recognition process

2.3.4 GIS-Based Quantification of Spatial Distribution Patterns

To investigate the spatial distribution patterns of tourists' visual perceptions and

photography preferences, this section maps UGC images to POI nodes. Since most social media images lack high-precision GPS metadata, this section implements image mapping based on the CLIP model. Representative images for each POI were manually curated to establish a reference feature set. The CLIP model was then used to perform similarity matching across the entire UGC image dataset. Ambiguous samples undergo a secondary manual verification to ensure mapping accuracy. By aggregating the frequency of photos taken at each POI, the unstructured image corpus is successfully transformed into a structured geospatial dataset containing physical coordinates and perceived frequency. Subsequently, this dataset is imported into the ArcGIS platform for spatial clustering and feature analysis.

Furthermore, this section introduces kernel density estimation to reveal the spatial distribution structure of tourist hotspots. Using the ArcGIS 10.8 Kernel Density Analysis tool, a spatial heatmap was generated based on a field weighted by the capture frequency of POI nodes, thereby quantifying the spatial distribution patterns of tourist perceptions regarding different geographic features. The kernel density estimation method centers on spatial points and uses a kernel function to calculate the density contribution of the estimated point within its surrounding neighborhood, reflecting the degree of spatial clustering of geographic features.

$$f(x) = \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

where $f(x)$ is the estimated kernel density value at spatial location x, n is the total sample size of POIs, h is the bandwidth, and its value determines the KDE result. $K(\cdot)$ is the kernel function for spatial weighting, and $x - x_i$ represents the straight-line spatial distance between the estimated point x and the i-th sample point x_i.

2.3.5 Validation of Model Applicability

To ensure the methodological rigor and domain adaptability of the pre-trained model used in this paper, a reliability validation was performed on the model prior to the formal analysis. From a theoretical perspective, the CLIP model is pre-trained on contrastive learning using massive amounts of image-text pairs from the internet and possesses zero-shot generalization capabilities, enabling it to recognize previously unseen architectural and natural landscape photography scenes. To evaluate its performance in the context of traditional Chinese villages, this section randomly selected 200 image samples from the dataset for manual annotation and cross-validation. The validation results show that CLIP achieved an accuracy of 87% in scene classification. These metrics confirm that, even without domain-specific fine-tuning, this model can still provide a highly reliable quantitative benchmark for subsequent spatial perception analysis.

3. Results

3.1 Identification and Summarization of Perception Themes

3.1.1 Textual Perception Themes

Based on tourist review texts, word segmentation, word frequency analysis, clustering analysis, and word cloud visualization (Fig. 4) were employed to systematically extract and

summarize high-frequency terms and evaluation perspectives related to perceptions of cultural heritage (Table A1). This analysis investigated the key areas of tourist interest and overall evaluation trends.

The results of the high-frequency word analysis showed that terms such as “vernacular dwellings,” “temples,” “commemorative archways,” “Ming and Qing dynasties,” and “ancient trees” appeared with high frequency, reflecting the core attractions of Heshun Ancient Town for tourists. This indicated that tourists’ focus was primarily concentrated on three aspects, including traditional architectural heritage with distinct local characteristics, natural landscape elements represented by ancient and famous trees and the historic water network, and the cultural heritage of the village rooted in the history of the Ming and Qing dynasties.

Based on the analysis of evaluative descriptors, positive evaluations primarily centered on terms such as “unique,” “authentic,” and “rich,” reflecting tourists’ high appreciation for the historic atmosphere and cultural diversity of the ancient town. Negative evaluations were mostly expressed through phrases like “ordinary” and “not worthwhile,” indicating that some tourists experienced a discrepancy between their expectations and their actual experiences. Overall, tourists held a positive attitude toward the historic built environment and cultural heritage of Heshun Ancient Town, particularly appreciating its pleasant atmosphere and rich historical resources. However, the reviews also revealed several factors contributing to negative experiences. The spatial scale of the town was relatively large, resulting in long walking distances. Frequent motorcycle traffic introduced disturbances, while the concentration of sloped pathways further reduced walking comfort. The combination of these physical environmental and management-related constraints, to some extent, diminished overall tourist satisfaction and constituted the primary source of negative feedback.



Fig. 4. Word cloud of evaluative keywords

3.1.2 Visual Perception Themes

Based on the CLIP model, this section performs scene classification and label generation on 17,684 images uploaded by tourists. Based on the spatial characteristics of the ancient town, the images were categorized into nine scene types: Architecture, Streets and Alleys, Bridges and Pavilions, Water Systems, Mountains and Forests, Commercial

Facilities, Catering Facilities, Service Facilities, and Festivals and Events. Representative tags were extracted based on the visual features of each scene type, and classification rules were formulated accordingly to construct a systematic image perception and analysis framework (Table A2).

(1) Distribution of Scene Types

Overall scene statistics (Fig. 5) showed that architectural scenes accounted for the highest proportion (8,579 images, 48.51%), forming the core of tourists' visual perception. These were followed by mountains and forests (2,742 images, 15.51%) and festive events (2,238 images, 12.65%). Bridges and pavilions (1,234 images, 6.98%) and dining facilities (1,151 images, 6.51%) accounted for a moderate proportion, while streets and alleys (265 images, 1.50%) and commercial facilities (144 images, 0.81%) accounted for the lowest proportions. Overall, heritage spatial elements characterized by traditional architectural features dominated tourists' visual documentation, while visual representations of consumption and service scenarios were relatively limited.

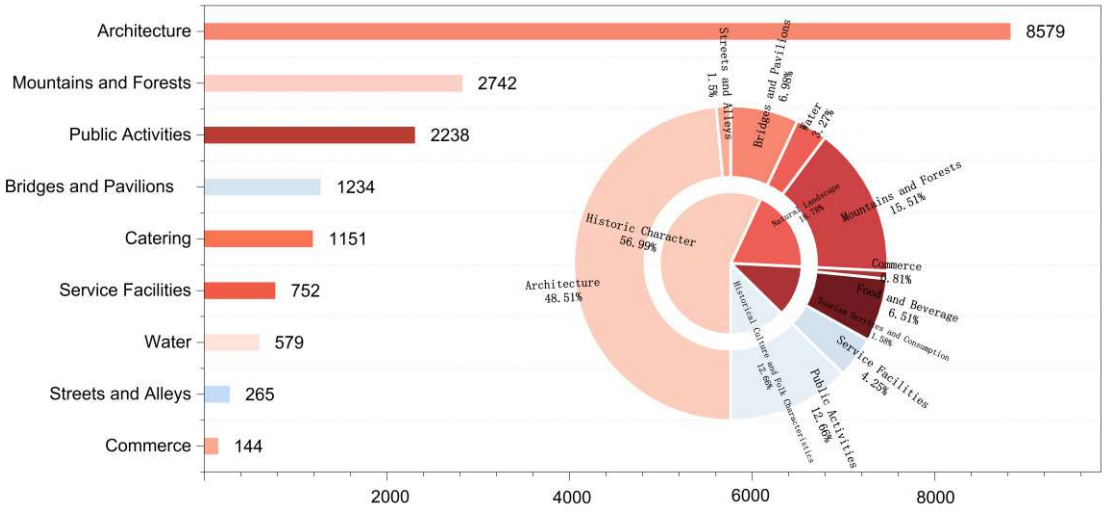


Fig. 5. Scene classification and recognition proportions

(2) Characteristics of Primary and Secondary Labels Across Scenes

The primary and secondary tag chord diagrams (Fig. 6) reveal differences in visual composition within each scene. In architectural scenes, the primary tags (Fig. 6-a) focused on overall morphological elements such as building facades, white walls and black-tiled roofs, and wooden structural components, while the secondary tags (Fig. 6-b) extended further to detailed features such as carved window lattices and courtyards, indicating that visitors' visual attention to traditional architecture encompassed both macro-level features and material details. In mountain and forest scenes, both primary and secondary tags primarily featured dense vegetation and mountain contours, often appearing as an overall ecological backdrop. In water-related scenes, primary tags focused on water surfaces and riverbanks, while the frequency of reflections and boats in secondary tags increased, reflecting visitors' heightened attention to the details of aquatic landscapes. In public event scenes, primary tags were dominated by traditional festivals and song-and-dance performances, while the proportion of intangible cultural heritage crafts among secondary

tags increased. This indicated that the depiction of cultural experiences in visitors' images was gradually expanding from event-based depictions to detailed documentation of artisanal techniques.

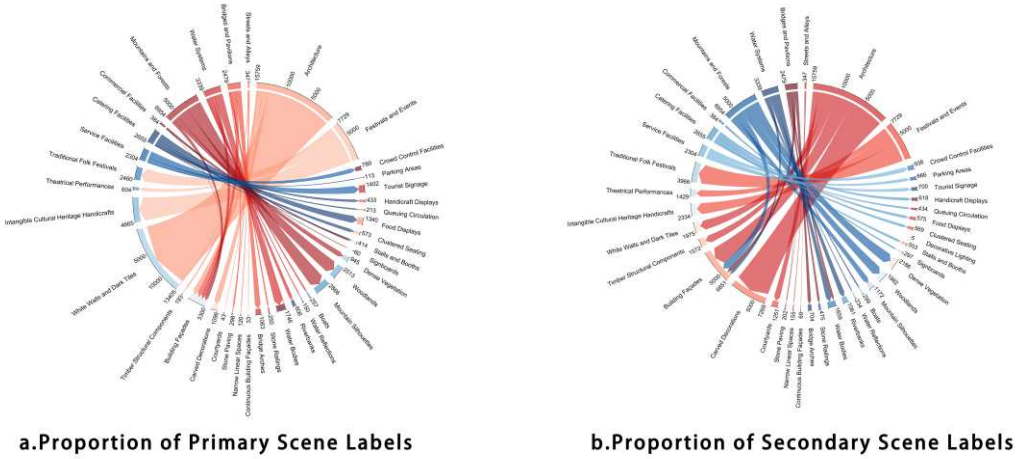


Fig. 6. Chord diagrams of primary and secondary scene tags

(3) Distributional Logic of Scene Labels

The Sankey diagram illustrating the relationship between primary and secondary labels (Fig. 7) further clarified the flow and cross-distribution of visual elements across different scenes. Architecture-related labels frequently emerged as secondary labels within the Streets and Alleys and Bridges and Pavilions scenes, indicating that architectural interfaces persisted across various categories within the tourists' field of vision. Similarly, labels associated with Mountains and Forests appeared extensively as secondary labels within the Architecture and Bridges and Pavilions scenes, corroborating the supportive role of the natural ecological background in shaping the spatial imagery. Conversely, labels pertaining to Service Facilities and Commercial Facilities functioned primarily as secondary elements embedded within the Architecture and Streets and Alleys scenes, constituting an everyday experiential backdrop for heritage tourism.

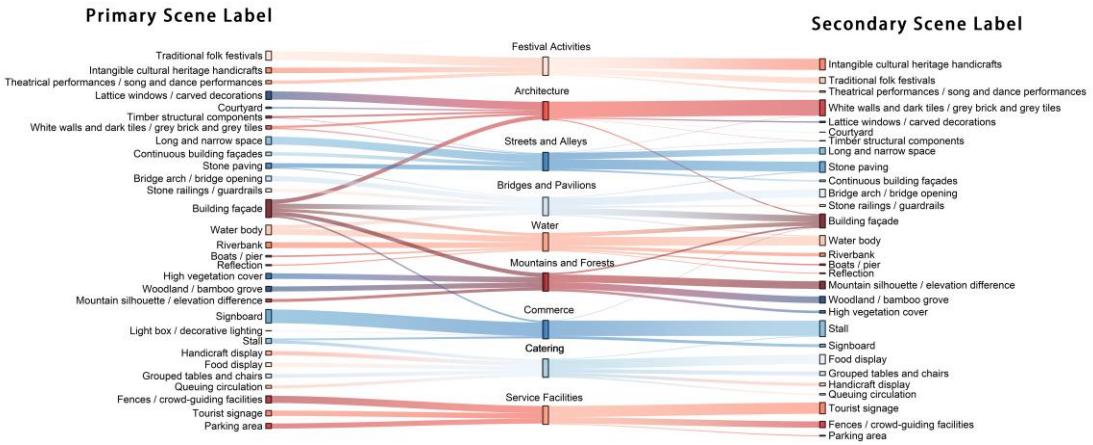


Fig. 7. Comparison of primary and secondary scene tag classifications

In summary, the perception of Heshun Ancient Town exhibited a composite visual structure. Traditional architectural character served as the visual core, supported by natural

landscapes, and augmented by cultural events and lived experiences. Furthermore, the cross-scene correlations between primary and secondary labels indicated that these perceptual elements do not exist in isolation， rather, they synergistically formed a multilayered cognitive understanding of the ancient town's overall spatial imagery.

3.1.3 Synthesis of Textual and Visual Perception Themes

Based on the classification results of high-frequency text descriptors and visual scenes, this study summarized tourist perception into four primary themes according to the semantic attributes and spatial types of heritage elements: Historic Character, Natural Landscape, Tourism Services and Consumption, and Historical and Folk Culture. Accordingly, we constructed a comprehensive heritage element framework (Table 1).

Historic Character was the most concentrated theme in tourist perception. Within the textual data, high-frequency descriptors such as "vernacular dwellings," "ancestral halls," "temples," "memorial archways," and "historic alleyways" directly pointed to architectural morphology. Correspondingly, in the visual data, Architecture scenes ranked first (48.51%), with labels primarily focused on elements such as building facades, white walls and dark tiles, and timber structural components. The high cross-modal consistency between these two data sources consolidated the dominant position of Historic Character as the core perceptual theme.

Natural Landscape served as a vital environmental backdrop across both data modalities. Textual descriptors such as "ancient trees," "rural farmland," "lotus ponds," and "streams" reflected significant tourist attention toward the natural environment. This observation was also corroborated visually, wherein the Mountains and Forests (15.51%) and Water Systems scenes were prominent. Specific visual labels—namely mountain silhouettes, dense vegetation, riverbanks, and water reflections—collectively constituted the site context of the ancient town.

The theme of Tourism Services and Consumption exhibited a structural divergence in perceptual intensity between textual and visual data. Textual descriptors such as "guesthouses," "coffee shops," and "morning markets" appeared with high frequency, indicating that accommodation and consumption experiences constituted vital components of tourists' overall perception. However, in the visual data, the proportions of Catering Facilities and Commercial Facilities scenes remained relatively low. This disparity suggested that tourists' attention toward service and consumption spaces in their written expressions significantly exceeded their actual visual recording behavior, revealing a notable decoupling between experiential perception and visual documentation.

The theme of Historical and Folk Culture maintained consistent visibility across both data modalities. Textual descriptors such as "handicrafts," "cultural and creative products," and "festival activities" indicated that tourists' comprehension of cultural value depended heavily on participatory activities and in-situ displays. In the visual data, the Festivals and Events scenes (12.65%) were primarily characterized by traditional folk festivals and stage performances, which further corroborated the event-oriented presentation of living heritage.

On the basis of the above analysis, this study classified the heritage elements into two major fields: Tangible Elements and Intangible Elements, and further subdivided them into four secondary categories. The classification was combined with the visual scene recognition data to build a statistical framework for the frequencies of heritage elements

(Table 1). The cross-modal frequencies in each category together reflected the perception intensity of the corresponding heritage elements. The higher the aggregated values, the more concentrated the perception of tourists was.

Table 1. Summarization of textual and visual perception themes

Primary Category	Secondary Category	Photographic Scene	Freq.	Textual Elements	Freq.
Tangible Elements	Historic Character	Architecture	8579	6. Residential dwellings, 9. Museum, 11. Heshun Library, 12. Ancestral hall, 17. Temple, 25. Integration of Chinese and Western Styles, 26. White walls and dark tiles, 27. Huizhou-style architecture, 30. Courtyard, 31. Volcanic stone, 45. Former residence	4403
		Bridges and Pavilions	1234	16. Commemorative archway(*pailou*)， 20. Arched stone bridge， 22. Laundry pavilion， 34. Stone stele， 35. Moon platform， 36. screen wall(*zhaobi*)， 37. Garden pavilion， 44. Neighborhood gateway(*lǚmén*)	1082
		Streets and Alleys	265	8. Historic alleyway， 29. Heritage promenade， 38. Family moral culture corridor	1001
	Natural Landscape	Water Systems	579	13. stream， 18. Lotus pond， 19. Jiangnan water-town landscape， 24. Mountain-water integrated setting， 33. Sunken river wetland	1521
		Mountains and Forests	2742	2. Ancient trees, 5. Rural farmland	2505
	Tourism Services and Consumption	Catering Facilities	144	7. Coffee shop, 21. Restaurant, 39. Teahouse	1210
		Commercial Facilities	1151	1. Guesthouse, 15. Morning market	2025
		Service Facilities	752	32. Parking lot, 40. Visitor center, 42. Restroom, 46. Directional signage	142
Intangible Elements	Historical and Folk Culture	Festivals and Events	2238	4. Culture, 14. Ming and Qing dynasty	1731
				3. Local specialties, 10. Local cuisine, 23. Handicraft, 28. Cultural and Creative Products, 41. Festival activities, 43. Opera stage	2522

3.2 Perceptual Evaluation and Sentiment Structure

Combining textual sentiment analysis with visual frequency metrics enabled a comprehensive evaluation of tourists' spatial perception. As indicated by high-frequency descriptor statistics, the overall evaluation of the ancient town was positive. Sentiment scores were primarily concentrated between 0.6 and 1.0 (Fig. 8). Furthermore, we constructed a quadrant map of various elements based on the calculation results. This map used the median values of frequency and sentiment scores as coordinate split points. It displayed the perceptual characteristics of each category. Quadrants I, II, III, and IV were classified as the Improvement Zone, Advantage Zone, Marginal Zone, and Observation Zone, respectively (Fig. 9).

As shown in the figure, Quadrant I represented areas with high frequency but low

satisfaction. Tourists had a high level of perception in this zone but reported low satisfaction. The main elements included tourism service and consumption spaces such as "guesthouses" and "coffee shops". Quadrant II was the area where both frequency and satisfaction were high. This indicated that tourists had both high perception and high satisfaction. These elements primarily consisted of Historic Character spaces such as "residential dwellings" and the "Heshun Library". Quadrant III was the area where both frequency and satisfaction were low. The main elements in this zone were basic infrastructure facilities like "parking lots" and "public restrooms". This phenomenon suggested that the authentic historical environment of the ancient town helped foster positive place attachment. However, the overall tourism experience remained constrained by the functional limitations of basic infrastructure. Quadrant IV was the area with low frequency but high satisfaction. In this zone, tourists had high satisfaction but low perception. The distributed elements included parts of Historical and Folk Culture and Historic Character.

Consequently, the tourist perceptual evaluation system was driven positively by natural and historical assets. Yet, it was negatively constrained by commercial services and physical infrastructure. Compared to unimodal evaluations, this quadrant-based approach provided a more comprehensive understanding of tourist cognition. It laid an objective analytical foundation for target-oriented heritage space optimization.

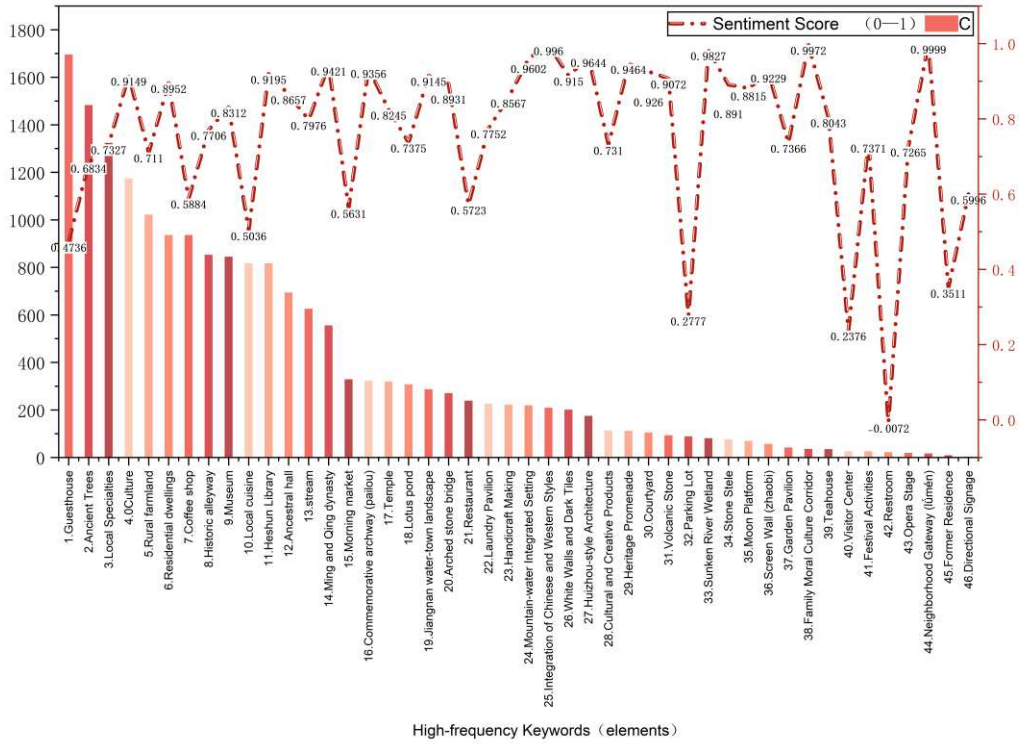


Fig. 8 Sentiment scores of high-frequency textual keywords

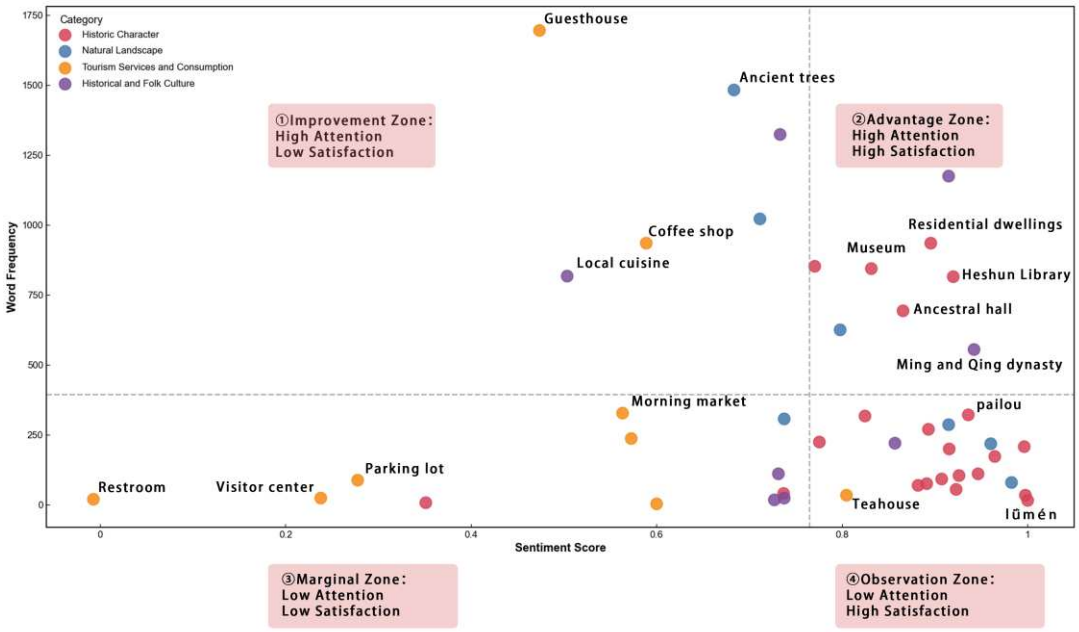


Fig. 9 Quadrant distribution of word frequency and sentiment score for heritage elements

3.3 Spatial Distribution Patterns of Perception

Based on the KDE spatial distribution analysis of photo location data (Fig. 10), the perception of tourists in Heshun Ancient Town showed a significant concentration in space. The overall distribution pattern could be summarized as "centralized clustering, expansion along corridors, and attenuation at the periphery".

KDE distribution results showed that there were differences in the spatial aggregation trends of different perceived themes. The perception density of Historic Character was the highest, mainly concentrated in the core sightseeing areas in the north-central part of the ancient town, especially along the Sanhe River, around Qiaoguang Square, and near historical nodes such as stone arch bridges and archways. The area was dense with traditional buildings, complete ancestral hall patterns, and concentrated public activities, which constituted the absolute core of the perception distribution. The perception density of Natural Landscape was secondary. The water bodies and mountain and forest nodes in the southeast formed obvious secondary hotspots, including Xianhe Wetland, Yeya Lake and Longtan. Tourism Services and Consumption perception was mainly attached to the commercial service gathering areas around Qiaoguang Square, such as guesthouses and morning market, but the photo shooting density in these areas was relatively low, which spatially confirmed the disconnect between the high-frequency text expression and the actual visual record mentioned above. Historical and Folk Culture perception was distributed in dots, relying on specific nodes such as traditional paper workshops and family cultural corridors, and the perception coverage was relatively narrow. At the edge of the ancient town, the outskirts of the mountain and the areas far from the main tour routes, the intensity of various perceptions decreased significantly, forming a number of "perceptual blind zones".

Combined with the analysis of perception path and flow distribution, perception distribution presented the spatial organization logic of "single core, multiple nodes, and corridor connections". The core scenic spots dominated by Historic Character elements constituted the main core of perception, and the water system nodes, historical bridge pavilions, lakeside wetlands and ancestral halls along the Sanhe River formed secondary perception nodes with multiple themes. The internal streets and alleys and the village loops connected these nodes in series to form a continuous perception corridor. Tourists mainly entered from the entrance of Qiaoguang Square in the north and spread south and southeast along the tour route. Therefore, the perception of Historic Character and Natural Landscape unfolded in turn along these corridors, and gradually decayed to the periphery after the core area was highly concentrated.

Overall, the spatial analysis intuitively quantified the perceived distribution pattern of Heshun Ancient Town. The spatial structure was characterized by the dominance of Historic Character, the support of Natural Landscape, the partial adhesion of consumption culture, and the lack of peripheral perception, which verified the heterogeneous distribution of perception intensity in physical space.



Fig.10. Spatial kernel density and organizational structure of perception

4. Discussion

4.1 Structural Divergence in Spatial Perception

Previous research on the perception of tourism-oriented TCVs has largely been constrained by retrospective scales, making it difficult to accurately map spatial experiences in response to real-world environmental stimuli. The multimodal perception framework established in this study translates holistic subjective evaluations into quantitative results. These results are grounded in specific spatial elements, which encompass physical morphology, natural settings, and cultural characteristics. Empirical results reveal a structural divergence in the spatial perception mapping of traditional villages. This divergence manifests not only as a discrepancy in the coupling relationship

between visual attention and emotional feedback. It also appears as a significant separation of agglomeration centers within the geographic spatial distribution.

4.1.1 Emotional Mapping and Experiential Divergence of Spatial Morphology

The data indicates that physical architectural clusters centered on "traditional dwellings and ancestral halls" and natural settings characterized by "lotus ponds and ancient trees" exhibit the highest frequency of visual documentation and positive emotional mapping within the multimodal perception framework. This confirms from a morphological perspective that a built environment maintaining its original spatial scale, stylistic features, and historical fabric constitutes the core foundation of the ancient town's spatial attractiveness. In the context of homogenized competition, such authentic physical spaces demonstrate robust morphological resilience, effectively anchoring public place attachment and deep spatial identity.

In contrast to the high positive feedback associated with authentic spaces, symbolic commercial nodes exhibit significant perceptual asymmetry. Specifically, their textual attention is exceptionally high, yet their visual recording rate and SnowNLP sentiment scores are significantly below the sample average. Exploring the underlying spatial mechanisms reveals a severe marginalization of commercial nodes in visual records. This phenomenon indicates that such modern consumption interfaces have failed to achieve an effective, non-intrusive integration into the original historical fabric of the ancient town. This disconnect in visual perception disrupts the overall continuity of the traditional streets' spatial morphology. Although such consumption interfaces—characterized by their "social media hotspot" attributes—can temporarily capture visual attention, they fragment the holistic context of the historical environment. This phenomenon suggests that the authentic "placeness" of traditional villages is gradually being dissolved by homogenized modern commercial aesthetics and consumption interfaces.

4.1.2 Imbalance in Perceptual Distribution and the Generation of Perceptual Blind Zones

Beyond the divergence of morphological attributes, the spatial KDE analysis further reveals the spatial structural characteristics of the circulation routes. Perceptual trajectories are highly concentrated in the north-central core area, presenting a hierarchical "core-corridor-node" distribution that adheres to the primary circulation axes. Such spatial characteristics are not solely the result of tourists' spontaneous behaviors; rather, they are significantly constrained by the ancient town's single main entrance configuration, the scarcity of guided tour routes, and the physical boundaries of its infrastructure. The study finds that core historical elements and secondary natural landscapes (e.g., the water system hotspots in the southeast) can unfold sequentially along the main routes. However, the rigid primary circulation acts as an invisible barrier. This barrier restricts permeability into secondary spaces, resulting in the marginalization of numerous intangible cultural nodes scattered outside the main thoroughfares, thereby generating "perceptual blind zones." This indicates that the current spatial organization pattern, to a certain extent, restricts the comprehensive realization of peripheral heritage values.

4.2 Implications for Spatial Intervention and Heritage Management

The multimodal perception model constructed in this study provides referential spatial

coordinates for optimizing the built environment of traditional villages. In response to the imbalance in perceptual distribution revealed previously, spatial interventions in heritage sites should shift from homogenized development toward differentiated and precise strategies.

First, regarding the spatial micro-regeneration of low-sentiment nodes, interventions should not stop at programmatic adjustments to commercial formats; rather, they must delve into the integration of physical spatial morphology. For nodes with high frequency but low satisfaction, it is necessary to control building scales and facade materials to restrict the disorderly sprawl of standardized commercial interfaces. Conversely, for low-frequency and low-satisfaction infrastructure, while their functional limitations must be addressed, localized materials and concealed design should be adopted. This approach enables them to be anchored into the historical environment through non-intrusive means, thereby restoring the morphological coherence of the traditional street fabric.

Second, Route Restructuring and Blind Zone Activation. Targeting nodes with high satisfaction but low perceptual frequency, as well as the peripheral areas identified by the KDE analysis, interventions should focus on breaking the insularity of the existing circulation structure. Specific measures include adding multi-themed secondary trails along water systems, mountains and forests, or intangible cultural heritage workshops. This will guide tourists away from the high-density core area in the north-central sector toward the water bodies, wetlands, and scattered historical remnants in the southeast. This strategy is not merely about adding tour paths; rather, it aims to redistribute perceptual density by reshaping the circulation network. By integrating peripheral spaces—which were previously obscured by the main tour structure—into the holistic spatial vitality network, a systemic restoration of the built environment's perceptual pattern can be achieved.

4.3 Study Limitations and Future Research

Although combining multimodal UGC with deep learning models offers advantages in exploring public spatial perception of traditional villages, the current study exhibits certain limitations that should be addressed in future research.

First, user-generated data cannot fully represent all demographic groups, particularly the elderly and children, who are less active on social media platforms. Groups with varying social attributes (e.g., age, social roles) exhibit distinct demands for the built environment. Furthermore, due to data availability constraints, this study was unable to obtain complete user identity information or effectively distinguish between local residents and out-of-town tourists in the comments. It is worth noting that, based on general understanding, users who post comments online are mostly out-of-town tourists, while local residents are relatively less vocal. Therefore, the findings of this study may primarily reflect the perceptions of out-of-town tourists. In terms of behavioral differences, local residents tend to be more concerned with the long-term maintenance of village infrastructure and the convenience of living environments, whereas out-of-town tourists focus more on landscape characteristics and visual experiences, and are relatively less sensitive to management issues. Moreover, different age groups also diverge in their concerns: younger people generally prefer vibrant and interactive spaces, while older people place greater value on tranquility and accessibility. These unaddressed demographic characteristics may

influence the outcomes of the perception analysis.

Second, there are limitations regarding the generalizability of the research sample and spatial typologies. As this study focuses solely on Heshun Ancient Town as a single case, it does not encompass or differentiate among various types of traditional villages. Consequently, the perceptual patterns revealed may lack generalizability across a broader range of contexts. Different typologies of traditional villages (e.g., commercially-oriented versus residential-oriented; plain versus mountainous) exhibit substantial differences in spatial scale, programmatic functions, and morphological design. These variations not only alter tourists' visual foci and environmental preferences but also influence their sensitivity and tolerance toward specific spatial features. Given that varying demographic groups hold distinct expectations for different village typologies, future research should conduct more granular typological comparisons to delve into the variations in spatial perception patterns across different types of traditional villages. Such endeavors will provide more targeted strategic guidance for the refined management of cultural heritage.

5. Conclusions

This study expands the research on the spatial perception of traditional villages into a multimodal, spatialized quantitative analytical pathway by constructing a quantitative evaluation framework based on massive multimodal UGC. The findings demonstrate that cross-modal data can effectively capture spatial perceptual characteristics. These characteristics encompass the physical built environment, natural settings, and intangible culture, thereby revealing the deep coupling mechanisms between public perceptual intensity and physical spatial distribution.

The contribution of this study lies in quantitatively defining the patterns of structural divergence in spatial perception. Empirical results corroborate that authentic spaces exhibit robust morphological resilience. These spaces preserve their original morphological features and historical fabric, serving as a critical factor in sustaining place attachment and high levels of satisfaction. Although excessive symbolic commercial erosion may capture instantaneous visual attention, the homogenization of its architectural interfaces and its decoupling from cultural cognition have emerged as a "perceptual trap," constraining the enhancement of spatial quality in traditional villages.

Furthermore, the perceptual model constructed in this study successfully identifies the "perceptual blind zones" obscured by existing circulation routes and infrastructural boundaries. By translating abstract public spatial perception into spatially explicit quantitative indicators, heritage managers can implement highly precise optimization strategies, thereby achieving a sustainable balance between heritage value conservation and spatial vitality for traditional villages in the digital era.

Competing interests

The authors declare no competing interests.

Data availability

The datasets and code generated and analysed during the current study have been

deposited in the Mendeley Data public repository. The repository contains: (1) raw and anonymised user-generated content (images and text) from social media platforms; (2) Cleaned analysis datasets: Versions processed through text filtering, image deduplication, and other procedures; (3) Analysis code and scripts: Python scripts covering data pre-processing, CLIP model invocation, SnowNLP sentiment analysis, and more; (4) Supplementary documentation: Including a data dictionary and shapefile files. These materials can be accessed at: https://doi.org/10.5281/zenodo.20489445.

Ethical approval

All procedures performed in this study adhered to the ethical standards outlined in the Declaration of Helsinki and its subsequent amendments. Ethical approval was granted on June 1, 2026, by the Biomedical and Medical Ethics Committee of Dalian University of Technology. All research activities complied with applicable guidelines and regulations for studies involving human participants. The approved project was titled "Tourists' Spatial Perception of Traditional Villages Based on UGC Data: A Case Study of Heshun Ancient Town, China." The reviewed document included the Ethics Review Application Form (which incorporated the study protocol). The approval ID is DUTSAFA260601-02.

Informed consent

Informed consent was not required as the study did not involve human participants .

Figures legends

Fig. 1 Spatial Layout Map of Heshun Ancient Town
Fig. 2 Research framework
Fig. 3 CLIP-based image label recognition process
Fig. 4 Word cloud of evaluative keywords
Fig. 5 Scene classification and recognition proportions
Fig. 6 Chord diagrams of primary and secondary scene tags
Fig. 7 Comparison of primary and secondary scene tag classifications
Fig. 8 Sentiment scores of high-frequency textual keywords
Fig. 9 Quadrant distribution of word frequency and sentiment score for heritage elements
Fig.10 Spatial kernel density and organizational structure of perception

References

Chen, Z. *et al.* Exploring the association between the built environment and positive sentiments of tourists in traditional villages in Fuzhou, china. *Ecol. Inf.* 80, 102465 (2024).

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

Bao, Y., Jiang, H., Ma, E., Sun, Z. & Xu, L. A longitudinal spatial-temporal analysis of ancient village tourism development in zhejiang, china. *Sustainability* 15, 143 (2023).

Yang, W., Fan, B., Tan, J., Lin, J. & Shao, T. The Spatial Perception and Spatial Feature of Rural Cultural Landscape in the Context of Rural Tourism. *Sustainability* 14, 4370 (2022).

Ren, X., Zhang, D. & Qi, Y. Analysis of tourists' cultural perception in cultural tourism villages based on online review data: a case study of dangjia village, shaanxi province, china. *Buildings* 15, 3891 (2025).

Shen, H. *et al.* Perceptions of cultural landscapes: exploring tourist satisfaction in traditional villages. *AIMS Geosci.* 11, 7–26 (2025).

Lu X , Tan D , Zhou Y ,et al. Traditional village perception and protection behavior: the mediating role of local identity and the impact of different population differences. *Journal of Asian Architecture & Building Engineering*, 24(6) , (2025).

Zhang, Y., Guo, W., Sheng, Y. & Li, S. Word-of-Mouth Evaluation of Ancient Towns in Southern China Using Web Comments. *Tour. Hosp.* 6, 25 (2025).

Jia, M. *et al.* Coupling Relationship Between Tourists' Space Perception and Tourism Image in Nanxun Ancient Town Based on Social Media Data Visualization. *Buildings* 15, 1465 (2025).

Yanan, L., Ismail, M. A. & Aminuddin, A. How has rural tourism influenced the sustainable development of traditional villages? A systematic literature review. *Heliyon* 10, e25627 (2024).

Hu, T. & Geng, J. Research on the perception of the terrain image of the tourism destination based on multimodal user-generated content data. *PeerJ Comput. Sci.* 10, e1801 (2024).

Li, K., Zou, Y., Wang, H. & Chen, S. Exploring the relationship between the tourist behavior and the spatial characteristics for rural tourism. *Sci. Rep.* 15, 9229 (2025).

Sánchez-Franco, M. J. & Rey-Tienda, S. The role of user-generated content in tourism decision-making: an exemplary study of andalusia, spain. *Manag. Decis.* 62, 2292–2328 (2024).

Liu, S., Wang, L., Yang, X. & Peng, Y. Posting the urban tourism experience: motivations behind multimodal UGC sharing. *Urban Sci.* 10, 88 (2026).

Ai, D., Kuang, D., Tao, Y. & Zeng, F. Integrating image recognition, sentiment analysis, and UWB tracking for urban heritage tourism: a multimodal case study in macau. *Sustainability* 17, 7573 (2025).

Xu, H. *et al.* Understanding the influence of user-generated content on tourist loyalty behavior in a cultural world heritage site. *Tour. Recreat. Res.* 48, 173–187 (2023).

Guo, Y., Barnes, S. J. & Jia, Q. Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation. *Tour. Manag.* 59, 467–483 (2017).

Li, J., Xu, L., Tang, L., Wang, S. & Li, L. Big data in tourism research: a literature review. *Tour. Manag.* 68, 301–323 (2018).

Yang, T. & Hsu, C. H. C. An AI-driven framework for continuous tourist sentiment scoring using longitudinal and group-level insights with pre-trained language models (RoBERTa-CSS). *Tour. Rev.* 81, 167–187 (2025).

Tian, Z., Wang, X., Wang, Y. & Lee, J. H. Enhancing museum visitor forecasting using deep learning and sentiment analysis: a transformer-based approach for sustainable management. *PLOS One* 20, e0335623 (2025).

Zhang, K., Lin, Z. & Zhang, J. Tourist gaze through computer vision: differences between asian, north American, and European tourists. *Ann. Touris. Res.* 88, 103039 (2021).

Deng, N. & Li, X. (Robert). Feeling a destination through the "right" photos: a machine learning model for DMOs' photo selection. *Tour. Manag.* 65, 267–278 (2018).

Li, Y., Meg Lee, H. H. & Blasco-Arcas, L. Computer vision in branding: a conceptual framework and future research agenda. *J. Bus. Res.* 193, 115329 (2025).

He, Z., Deng, N., Li, X. (Robert) & Gu, H. How to "read" a destination from images? Machine learning and network methods for DMOs' image projection and photo evaluation. *J. Travel Res.* 61, 597–619 (2022).

Liu, W., An, G., Xue, Y. & Hong, C. Multimodal analysis of the destination image of traditional villages based on tourist comments and photos: the case of hongcun, china. *J. Hosp. Tour. Technol.* 17, 561–581 (2025).

Wang, X., Zhang, C., Xu, Z. & Yin, L. Measuring tourism destination image from multimodal online reviews: a novel framework using deep learning approach. *Inf. Technol. Tour.* 28, (2025).

Salas-Olmedo, M. H., Moya-Gómez, B., García-Palomares, J. C. & Gutiérrez, J. Tourists' digital footprint in cities: comparing big data sources. *Tour. Manag.* 66, 13–25 (2018).

Gao, S., Janowicz, K. & Couclelis, H. Extracting urban functional regions from points of interest and human activities on location-based social networks. *Trans. GIS.* 21, 446–467 (2017).

Zheng, W., Huang, X. & Li, Y. Understanding the tourist mobility using GPS: where is the next place? *Tour. Manag.* 59, 267–280 (2017).

García-Palomares, J. C., Gutiérrez, J. & Mínguez, C. Identification of tourist hot spots based on social networks: a comparative analysis of European metropolises using photo-sharing services and GIS. *Appl. Geogr.* 63, 408–417 (2015).

Gao, J. & Wu, B. Revitalizing traditional villages through rural tourism: a case study of yuanjia village, shaanxi province, china. *Tour. Manag.* 63, 223–233 (2017).

Liu, S. *et al.* Toward classification-based sustainable revitalization: assessing the vitality of traditional villages. *Land Use Policy* 116, 106060 (2022).

Peng, R. & Gao, W. Spatial distribution pattern and driving mechanism of tourist attractions in Gansu Province based on POI data. *PLOS ONE* 18, e0292165 (2023).

Liang, W., Ahmad, Y. & Mohidin, H. H. B. Spatial pattern and influencing factors of tourism based on POI data in Chengdu, China. *Environ Dev Sustain* 1–17 (2023).

Wu J , Zhang Q , Li Z .Spatial Coupling Characteristics Between Tourism Point of Interest (POI) and Nighttime Light Data of the Changsha–Zhuzhou–Xiangtan Metropolitan Area, China.*Sustainability*, 17(6) ,(2025).

Zhang, M., Di, Q. & Liu, Y. Perceived differences in coastal tourism image under tourist experience-IPA analysis based on UGC data of 12 coastal cities. *PLoS One* 19, e0299431 (2024).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix.docx