

Research on the Correlation Between Tourists' Facial Emotion Perception and Street Spatial Elements Through Deep Learning: A Case Study of Traditional Tourism-Oriented Villages in China

Yixin Liu¹, Zhimin Li², Yixin Tian³, Hao Wang⁴, Ruqin Wang⁵

¹ School of Architecture at Xi'an University of Architecture and Technology, liu_yixin@xauat.edu.cn

² School of Architecture, Xi'an University of Architecture and Technology, lizhimin@xauat.edu.cn

³ School of Architecture, Xi'an University of Architecture and Technology, tian_yixin@xauat.edu.cn

⁴ School of Architecture at Xi'an University of Architecture and Technology, 1719187202@qq.com

⁵ Hokkaido University, ruqin.wang,c6@elms.hukuai.ac.jp

Abstract

The tourism-driven revitalisation of traditional villages has become a significant pathway to achieving historical and cultural heritage preservation and promoting rural revitalisation in the new era. Among these efforts, constructing the regional characteristics of the built environment in villages is a critical aspect of revitalisation. Tourists' emotional perceptions during their spatial experiences in tourism villages serve as an important evaluation metric for the localization characteristics of the space. This study investigates the correlation mechanisms between 13 spatial elements—such as spatial organization, morphology, and interface—and tourists' emotional types and intensity preferences based on the distribution characteristics of tourists' emotional perceptions in the street spaces of tourism villages. Facial expression images of tourists in different village streets were collected via smartphone photography, and a deep learning model, YOLOv5, was trained using a manually labeled dataset to recognize and quantify four types of emotions: smile, surprise, calm, and negative emotions. To enhance recognition accuracy, the model's performance was improved using techniques such as RGB value adjustment, flipping, and augmentation. Subsequently, a multiple regression analysis was conducted to identify the street space elements influencing tourists' emotional perceptions. The results indicate that the proposed model optimization methods effectively identified tourists' emotions under various scenarios with high accuracy. Regression analysis revealed that spatial segmentation factors of streets had a regression coefficient of 0.184, making it a key factor affecting tourists' perceptions. The findings identified five typical street space elements that positively influence tourists' emotions. Based on these results, recommendations for optimizing street space to enhance tourists' perceptions were proposed, offering support for improving street space tourism experiences and promoting the sustainable development of traditional villages.

Keywords: Facial expression perception; Deep learning; Alley space; Tourist Emotions; Yuan Family Village

1. Introduction

With the rapid development of tourism, the revitalisation of traditional village tourism has become an important way to preserve historical and cultural heritage and promote rural revitalisation. In this process, the construction of village built environment with regional characteristics is an important link. The emotional perception of tourists in the physical space experience of tourism villages has received increasing attention from researchers as an important indicator for evaluating spatial regionalisation characteristics. Since the 1960s, with a rethinking of modern urban planning dominated by functional zoning, there has been a growing emphasis on diverse and vibrant urban spaces [1]. Vibrant cities and streets enhance the happiness of residents and social cohesion [2]. Urban street vitality has become an important indicator for assessing the attractiveness and potential of urban neighbourhoods for sustainable development [3]. Many design theories have been proposed to create a vibrant urban street space and a growing body of methodological and empirical research has been conducted on the role of relevant physical environment variables, such as street interface continuity, greening ratio, ground floor interface transparency, and commercial density, in fostering street activities and vitality [4]. It is postulated that AI may also be employed in the monitoring of human activities in urban settings. For instance, Yamada et al. employed Deep Learning to assess the impression of a city, while Liu et al. utilised machine learning to construct buildings dedicated to the evaluation of the quality of maintenance management. Furthermore, technologies for the monitoring of shop-goers using cameras installed in shops are becoming increasingly feasible [5]. Existing research has mostly focused on the economic and socio-cultural impacts of tourism, with less attention paid to tourists' emotional responses in specific spaces. However, emotional perception, as a core component of tourists' experience, not only affects tourists' overall impression of a tourist village, but also directly relates to their satisfaction and willingness to return. Therefore, it is of great theoretical and practical significance to explore the association mechanisms between the spatial organisation, morphology, and interface of streets and lanes and tourists' emotional types and intensity preferences.

2. Materials and Methods

2.1 Materials

Yuanjia Village, situated in Liquan County, Xianyang City, Shaanxi Province, is one of the second batch of traditional villages included in the list of traditional villages in China. It is a natural village with a long history in the Guanzhong area of Shaanxi Province (Figure 1). The location and layout of Yuanjia Village are centred on the base of the village, with a landscape pattern that reflects the traditional Chinese landscape pattern of "Facing the sunny side and backing the shady side, backed by mountains and facing water". This was a pattern that is typical of the Guanzhong area, and it was one that Yuanjiacun has better preserved and continued. The village was surrounded by mountains and by blue water. Secondly, Yuanjia Village was a popular tourist destination. As of 2023, the number of tourists visiting Yuanjia Village annually exceeded 8 million, with the village's annual tourism output value exceeding RMB 1.2 billion. [6] making it one of the villages with the highest number of annual visitors in the Shaanxi region. Based on the annual increase in tourist visitation and tourism revenue, Yuanjia Village was one of the traditionally successful villages in tourism transformation. The substantial number of tourists provided universality and data credibility for sampling to simulate and test tourist perceptions.

Therefore, the research focused on six typical streets within the core scenic area of Yuanjia Village (as shown in Figure 1): namely, Ancestral Hall Street (CT=248m), Muslim Quarter Street (HM=156m), Snack Street (XC=428m), Sanqin Ancient Street (SQ=704m), Academy Street (SY=413m), and Kanqzhuang Old Street (KZ=383m).

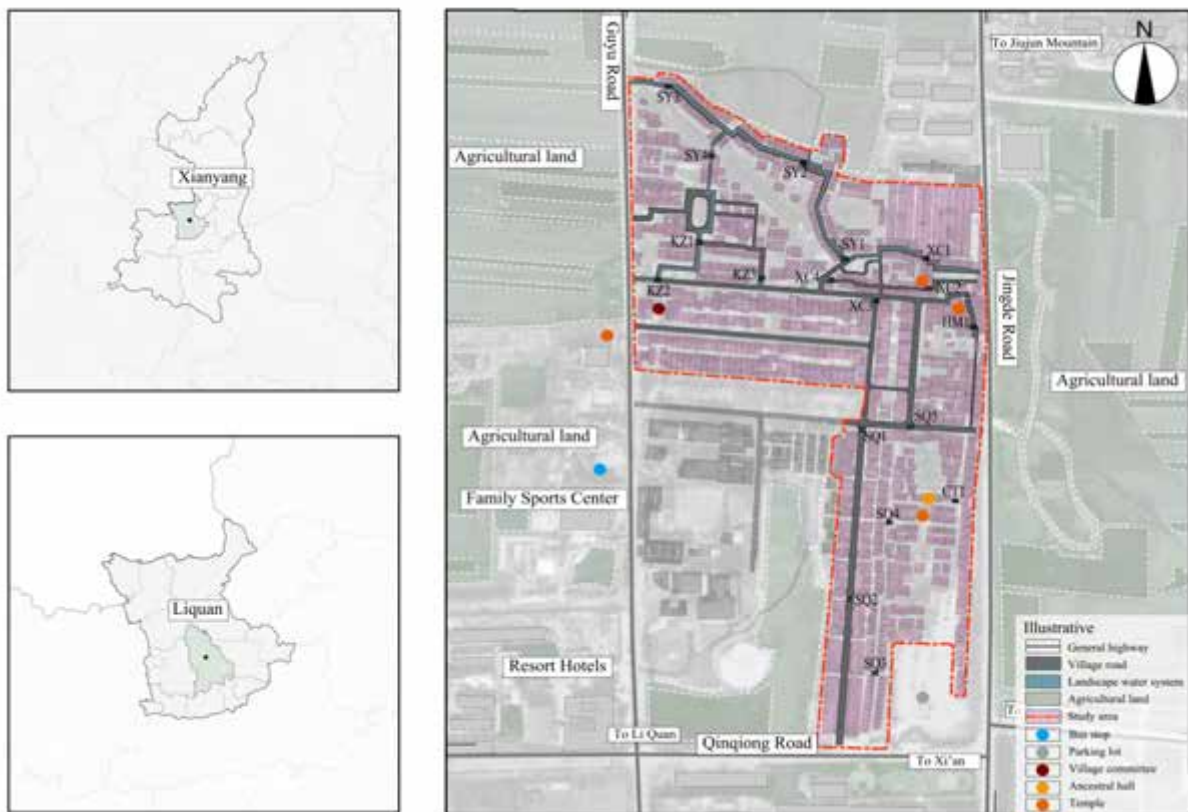


Figure 1 Location and Layout Plan of Yuanjia Village

2.2 Methods

Building upon previous studies, this paper proposes an improved YOLOv5 recognition model and pathway designed to more accurately and comprehensively capture facial information from photographs. Furthermore, it correlates optimised subjective perception values of tourist emotions with built environment contexts. Through multiple regression analysis, the study aims to explore the impact of the optimised approach on facial expressions and emotions of tourists in spatial environments.

2.2.1 Facial expression recognition based on YOLOv5

YOLO, as a mature algorithm in the field of deep learning, has demonstrated superior performance and practical feasibility in numerous domains and real-world applications. The facial expression analysis component of this study drew inspiration from the research of Shungo Noji[7], Liu[8], leveraging deep learning algorithms. By training deep learning models with real-world photos of different time periods and age groups, the study aimed to enhance the accuracy of facial expression recognition for specific populations in a designated area (where the images were collected). Furthermore, in subsequent experiments, the study hypothesized smile rates among different populations at various locations. In this study, the YOLOv5 model was employed (The YOLO-v5 model, which is based on a real-time object detection algorithm). The YOLO network was optimised to minimise the loss function as the training objective to achieve object detection. The YOLO network consisted of convolutional layers specifically designed for object detection tasks and included three key modules as illustrated in Figure 1: 1. Backbone network: This module was used to extract generic feature representations and served as a high-performance classifier network. 2. Neck network. The three convolutional networks collaborated to produce the final output of object detection results. In this training session, we employed common data processing techniques such as Shungo Noji. This included adjustments to the data such as angle rotation, flipping, and slight changes in colour and brightness to achieve more robust results. The model was pretrained using the COCO dataset and further fine-tuned and trained using a proprietary dataset. The proprietary dataset used in this experiment consisted of 1006 images collected in the autumn of 2023, with a resolution of 3840 x 2160 pixels. 85% of these images were used as the training dataset, and 15% were used as the test dataset to validate the efficiency and accuracy of the model.

In this experiment, the processed dataset underwent various data augmentation techniques, including colour variations, image flipping, and image overlays. Mild colour adjustments were applied, encompassing changes in brightness, contrast, and saturation to simulate different lighting conditions, thereby enhancing the model's adaptability to various lighting environments. Additionally, horizontal and vertical flipping of images was performed to help the model better handle variations in object orientations and improve its recognition capability. Through image overlays, we merged two or more images to create more complex scenes. This method helped the model learn how to accurately identify and locate objects amidst complex backgrounds and multiple object occlusions. These data augmentation techniques effectively increased the diversity and complexity of the training data, significantly boosting the model's robustness and generalization ability. This enhancement allows the model to more accurately detect and classify target objects in practical application scenarios of this experiment.

Emotions including Smile, Lugh, Neutral, and Negative Emotion were identified and quantified to determine the facial expression recognition across various street locations. This process culminated in calculating the subjective smile rate (E) for different streets by integrating the total number of recognized expressions, as detailed in the formula.(1):

$$E = (\sum \text{Smile} + \sum \text{Lugh}) / (\sum \text{Smile} + \sum \text{Lugh} + \sum \text{Neutral} + \sum \text{Negative Emotion})$$

2.2.2 Selection and calculation methods of influencing factors

This study drew inspiration from previous research on measurement systems for street and lane indices[9][10], and integrated with the spatial characteristics of Yuanjiacun's streets and lanes, established an index system influencing tourists' emotional perception. These influencing factors were categorized into three aspects: spatial form, spatial function, and spatial interface (Figure 2).

Primary indicator	Secondary indicators	Description	Labeled Name	Assignment Description
Street scale and form	Scale of streets	Width	KD	/
		Ratio of street height to width	GKB	Average building height on both sides/street width
	Street of morphology	Relationship of Streets to Landforms	JX	Responding to the terrain
		Ground Level Elevation	DMTS	Whether the height difference between the transition space and the street space is significant
	Transition space	Space Separation	KJFG	Significant height difference between transitional space and street space
Street floor characteristics	Façade Setbacks/Extensions	Longitudinal Subgrade	LMHT/WY	Separation of spaces from the street through parapets, low walls, planters
			ZSDM	
	Permeability	Longitudinal cross roads	ZXJCDL	Vertical alleyway entrance at ground-floor
		Straight bottom elevation	PZDCLM	Vertical projection of ground-floor elevation is straight
	Curvature	Curvature Curvilinear ground-floor	QZDCLM	ground-floor façade is clearly angled, projecting or recessed
Lane space function	Type of business	Six business types: hotel, food, shopping, attractions, entertainment, and public services	F	Assign values 1-6
	Business density	Density of commercial POI points per 10m street buffer zone	FD	$FD = ? POI_num / Road_length$
	Business mixing degree	Street mix (diversity) is the mix of vitality-related POIs in the Street buffer.	FH	$FH(X) = -\sum P(x_i) \log P(x_i) \quad (i=1,2,...,n)$

Figure 2 Classification and Assignment of Street Space Element Data

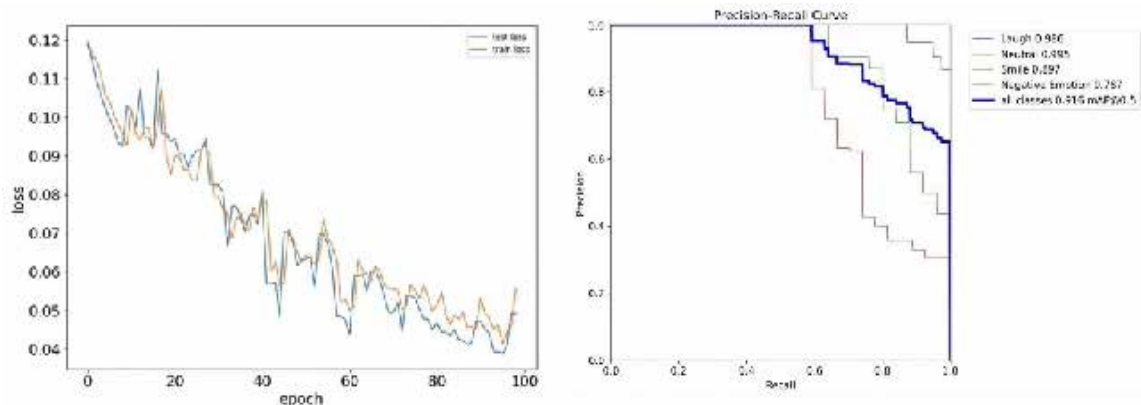
3. Discussion

3.1 A more efficient method for facial expression recognition

This study utilized Anaconda3 to construct the environment, ensuring stability and independence during model training and execution. Additionally, labelling Python scripts were employed for manual annotation, and annotations were cross-checked by fellow research lab members to guarantee dataset accuracy. The implementation was based on the PyTorch framework, and training and testing were conducted on a single NVIDIA RTX A5000 GPU. To optimize the network, we utilized stochastic gradient descent (SGD) optimizer with a learning rate of 0.01 and momentum of 0.9.

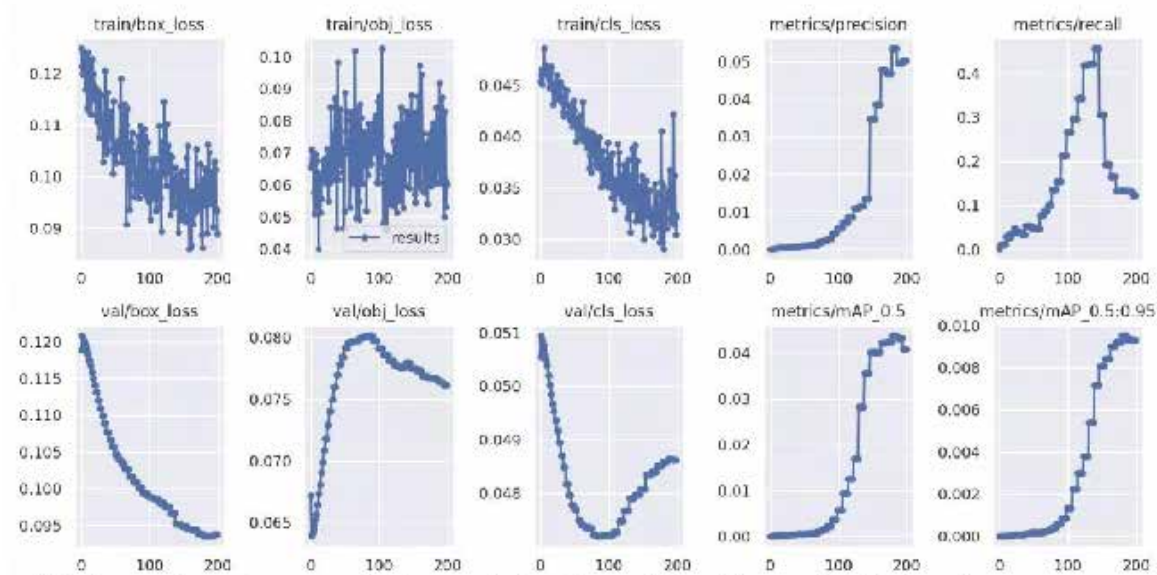
To overcome challenges such as a small private dataset, significant outdoor lighting variations, and complex backgrounds, data augmentation techniques were employed. These included deformation, adjusting RGB values, flipping, and overlaying methods to enhance model robustness by augmenting the dataset. In terms of model performance metrics analysis, we achieved relatively high average precision (AP) values, indicating stable performance on the test data. The deep blue line in Figure 1 represents the average of the four classification results. Through multiple validations, AP values consistently exceeded 0.9, meeting the experiment's precision requirements (Figure 3-1). Additionally, Figure 3-2 illustrates the loss values during the training process. The experiment utilized test loss X and Y axes to monitor performance changes, aiming for minimal validation loss. As shown in the graph, with increasing epochs, the loss gradually decreased, indicating effective training.

In conclusion, graphical representations of AP values and loss values demonstrated expected progress in model training. Significant improvements in classification accuracy and continuous reduction in loss values validate the effectiveness of our approach.



3-1: The Precision-Recall curves for the different mood categories

3-2: Trends of model training and test losses with training rounds



3-3: Changes in performance indicators during the model training and validation phases

Figure 3 Detection of Tourist Facial Expressions Using Recognition Models

The graph depicted various performance metrics of our model during the training and validation phases (Figure 3-3). The horizontal axis represented training epochs, while the vertical axes indicated loss values and performance metrics. The graph was divided into two rows: the upper row illustrated performance on the training data, and the lower row showed performance on the validation data. Throughout the training process, both train/box_loss and train/cls_loss steadily decreased as training epochs progressed, indicating ongoing improvements in bounding box prediction and classification accuracy. Despite fluctuations, train/obj_loss exhibited an overall decreasing trend, signifying a reduction in object detection loss. Precision (metrics/precision) and recall (metrics/recall) on the training data demonstrated noticeable upward trends, underscoring enhanced precision and comprehensiveness in detection and classification tasks. During validation, significant declines were observed in both val/box_loss and val/cls_loss, further confirming the model's efficacy during training. However, val/obj_loss initially showed a minor increase followed by a decline, suggesting potential early-stage overfitting. Average precision on the validation set (metrics/mAP_0.5 and metrics/mAP_0.5:0.95) progressively improved, particularly within stricter threshold ranges (0.5 to 0.95), indicating consistent performance across diverse evaluation criteria and effective detection of various data types. In summary, these findings from the graphical representations underscored the anticipated advancements in model training, with enhancements observed across both training and validation metrics over successive epochs, validating the efficacy and resilience of our methodology.

Although there remained scope to enhance accuracy in recognizing negative emotions, this was attributed to the limited number of data samples for negative emotions and the imbalance in quantities across different data types. Overall, these metric fluctuations reflected effective training of our model and a steady improvement in performance. Despite fluctuations during training, the model consistently demonstrated strong learning capability and adaptability on both training and validation data, suggesting its potential for achieving favorable outcomes in practical applications ()



Figure 4 Target Recognition Performance of the YOLO-v5 Model in Different Scenarios

3.2 Spatial Distribution of Tourists' Emotions in the Streets and Alleys of Yuanjia Village

Based on the results of emotional recognition: In the SY area, the smile rate among tourists peaked at 35% in the afternoon, while negative emotions accounted for approximately 40%(Figure 5). This suggests that tourists' perception of the SY street area is average. In the XC area, the proportion of smiles and happiness was highest in the morning, reaching approximately 80%, with negative emotions accounting for about 20%. This indicates that tourists' perception of the XC street area is very positive. In the KZ area, smiles and happiness were most prevalent in the morning, accounting for approximately 50%, with negative emotions at about 25%. This suggests that tourists' perception of the KZ street area is relatively positive. In the CT area, the highest proportion of smiles and happiness occurred in the morning, reaching about 40%, while negative emotions made up around 30%. This indicates that tourists' perception of the CT street area is average. In the SQ area, smiles and happiness were most prominent in the morning, accounting for about 35%, while negative emotions reached approximately 40%. This suggests that tourists' perception of the SQ street area is relatively poor. In summary, emotional recognition revealed that the XC street area had the highest positive emotional perception (happiness and smiles), while the SQ street area had the lowest. Additionally, this method not only provides insights into the emotional perception of entire streets but can also statistically analyze the emotional distribution of specific sections within a street.



Figure 5 Spatial Distribution Of Tourist Sentiment Recognition Results

3.3 Analysis of the Correlation Between Smile Rate and Street Space

This study built upon the research methodology of Xu, Fang[11], Liu Y., et al. Aesthetic evaluation of streets and analysis of influencing factors based on deep learning and multi-source data: A case study of Shanghai. *Int. Urban Plan* 2023, 38(06), 48–58.], as well as Liu Y. et al., employing multiple linear regression analysis to explore the correlation between smile rate and street spatial elements. By utilizing the optimised YOLOv5 model, the study achieved a more precise and detailed quantification of tourists' emotional responses across various street spatial locations. Combined with (Equation 1), this methodology assigned quantitative values to distinct street spatial locations. Subsequently, the influence of street spatial indicators on the evaluation of spatial perception was analyzed using multiple linear regression through SPSS. The corresponding formula is as follows:

$$E = \beta_0 + \beta_1 * GKB + \beta_2 * KD + \beta_3 * ZSDM + \dots + \beta_{13} * FD + e \quad (1)$$

The regression analysis included coefficients through, with representing the constant term and denoting the error term. Post-regression diagnostics indicated that the tolerance values, calculated as, were all greater than 0.1, confirming the absence of multicollinearity among the independent variables. The Durbin-Watson statistic for the model was 1.429 (Table 1), suggesting that the assumption of independence in the residuals was met; The coefficient of determination ($F=5.031$, $P=0.000<0.05$) indicated that the model possessed limited explanatory power (Table 2. While a statistically valid model was constructed, its interpretability was moderate.

In summary, the analysis of the smile rate and the spatial correlation of the street reveals the following:

(1) Various elements of the spatial environment influence tourists' emotions, although the extent of this impact is relatively limited. The constructed regression model exhibits an R^2 value of 0.198, indicating that the current spatial elements account for only a small proportion of tourists' emotional perception. To enhance the explanatory power of the model, incorporating additional street spatial elements could be considered to expand the range of influencing factors. Alternatively, future studies could explore other dimensions or variables that might have a stronger impact on tourists' emotions.

(2) Based on the Beta coefficients, KJFG emerges as the primary factor influencing tourists' emotional percep-

tions, followed by KD, which also has a significant impact (Table). This finding differs from previous research. A possible explanation is that tourists tend to experience heightened expectations and positive emotions when transitioning to the next street space, leading to generally higher recognition data during the analysis. The FD factor also exhibits a strong influence, suggesting that higher commercial density within the street can enhance tourists' positive emotions. However, the Beta coefficient for ZXJCDL is -0.227, indicating a negative impact on tourists' emotions. This may be attributed to the vertically intersecting streets, which fail to evoke favorable emotions. By contrast, winding streets appear to provide a more engaging and emotionally positive experience for tourists during their visits.

In conclusion, this study's findings are based on the existing data on tourists' emotions and the spatial organization, morphology, and interface elements associated with street spaces as discussed. The analysis identified five typical street spatial elements—F, FD, QZDCLM, KD, and KJFG—that have a positive impact on tourists. These elements could be prioritized in street space construction and planning to increase their proportion, enhance functional diversity, and improve the quality of transitional spaces between streets. Such measures would contribute to enriching the street space experience and promoting the sustainable development of tourism-oriented traditional villages.

Table 1 Model summary

Model	R	R ²	Adjusted R ²	Standard Skewness Error	Durbin-Watson
1	0.445a	0.198	0.159	0.069929	1.429
a. Predicted value: (constant), F, DMTS, GKB, LMHT_WY, ZSDM, JX, KJFG, QZDCLM, FH, FD, ZXJCDL, KJFG, KD					
b. Dependent variable: E					

Table 2 Analysis of variance (ANOVA)^a

Model	Sum of Squares	df	Mean Square (MS)	F	Significance
1 Regression	0.295	12	0.025	5.031	0.000b
Residuals	1.193	244	0.005		
Statistics	1.488	256			

a. Predicted value: (constant), F, DMTS, GKB, LMHT_WY, ZSDM, JX, KJFG, QZDCLM, FH, FD, ZXJCDL, KJFG, KD

b. Dependent variable: E

Table 3 The results of linear regression analysis between tourist behavior perception and street space elements

Model	Unstandardized Coefficients	Standardized Coefficients	T	Significance	Collinearity statistics
B	Standard Error	Beta		Tolerance	VIF

	(constant)	0.183	0.052		3.486	0.001	
	F	0.002	0.005	0.022	0.369	0.712	1.110
	FD	0.087	0.089	0.119	0.976	0.330	4.518
	FH	-0.121	0.082	-0.156	-1.477	0.141	3.415
	ZSDM	-0.026	0.011	-0.156	-2.331	0.021	1.368
	ZXJC DL	-0.035	0.011	-0.227	-3.270	0.001	1.471
1	QZDCLM	0.006	0.013	0.041	0.498	0.619	2.106
	GKB	-0.010	0.006	-0.188	-1.698	0.091	3.723
	KD	0.003	0.002	0.147	1.284	0.200	3.967
	JX	-0.027	0.012	-0.168	-2.274	0.024	1.661
	DMTS	-0.005	0.012	-0.025	-0.388	0.698	1.308
	KJFG	0.032	0.013	0.184	2.531	0.012	1.600
	LMHT_WY	0.000	0.013	-0.001	-0.016	0.987	1.292
a. Dependent variable: E							

4. Conclusions

In addition to being influenced by the spatial environment, tourists' emotional perceptions are affected by various other factors, such as their personal mood, conversations with companions, and weather conditions. This study proposes a novel method for identifying emotional states through image-based deep learning, enabling a detailed depiction of tourists' emotional feedback and changes during their visits. This approach offers a new tool for evaluating and providing feedback on spatial environments. By analyzing the distribution characteristics of tourists' emotional perceptions in the street spaces of tourism villages, this study explores the correlation mechanisms between 13 spatial elements—including spatial organization, morphology, and interfaces—and tourists' emotional types and intensity preferences. Through multiple regression analysis, it was found that the spatial segmentation factor of streets had a regression coefficient of 0.184, making it a key determinant of tourists' perceptions. Additionally, five typical street spatial elements were identified as having a positive influence on tourists. Based on these findings, this study proposes optimization strategies for enhancing tourists' emotional perceptions within street spaces. These recommendations aim to improve the street experience and support the sustainable development of traditional villages.

Acknowledgement

This study was supported by the National Natural Science Foundation of China Youth Program (52208031).

References

- [1] Zeng, C., Song, Y., He, Q. and Shen, F. (2018) 'Spatially explicit assessment on urban vitality: Case studies in Chicago and Wuhan', *Sustainable Cities and Society*, 40, pp. 296–306.
- [2] Mouratidis, K. and Poortinga, W. (2020) 'Built environment, urban vitality and social cohesion: Do vibrant neighbourhoods foster strong communities?', *Landscape and Urban Planning*, 204, 103951.
- [3] Jacobs, J. (1961) *The Death and Life of Great American Cities*. New York: Vintage.
- [4] Bergmann, P., Meinhardt, T. and Leal-Taixe, L. (2019) 'Tracking without bells and whistles', in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 941–951.
- [5] Noji, S. and Kishimoto, T. (2020) 'Analysis of age, gender ratio, emotion, density of pedestrians in urban street space with image recognition: Case study mainly around Shibuya Station [画像認識による街路空間における歩行者の年齢、男女比、感情、および密度の分析渋谷駅周辺を主としたゲ - 渋谷駅周辺を主としたケーススタディ -]', *Journal of the City Planning Institute of Japan*, 55(3), pp. 1–12.

- [6] CNWest News (2024) Yuanjia Village tourism revenue exceeds 1.2 billion. Available at:<http://news.cnwest.com/sxxw/a/2024/02/18/22340660.html>(Accessed:18 February 2024).
- [7] Shungo Noji,Tatsuya Kishimoto. Analysis of Age, Gender ratio, Emotion, Density of Pedestrians in Urban Street Space with Image Recognition-Case Study Mainly Around Shibuya Station.Journal ofthe City Planning Institute of Japan, Vol.55 No.3,October; 2020
- [8] Liu, Y., Li, Z., Tian, Y., Gao, B., Wang, S., Qi, Y., Zou, Z., Li, X. and Wang, R. (2024)'A study on identifying the spatial characteristic factors of traditional streets based on visitor perception: Yuanjia Village, Shaanxi Province',Buildings, 14(6), p. 1815.
- [9] Liu, L., Silva, E.A., Wu, C. and Wang, H. (2017)'A machine learning-based method for the large-scale evaluation of the qualities of the urban environment',Computers, Environment and Urban Systems, 65, pp. 113-125.
- [10] Xu, G., Zhong, L., Wu, F., Zhang, Y. and Zhang, Z. (2022)'Impacts of micro-scale built environment features on tourists' walking behaviours in historic streets: Insights from Wudaoying Hutong, China',Buildings, 12(12), p. 2248.
- [11] Fang, Z., Liu, C., Xiao, Y. et al. (2023)'Aesthetic evaluation of streets and analysis of influencing factors based on deep learning and multi-source data: A case study of Shanghai',International Urban Planning, 38(6), pp. 48-58.