ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Research on Branding Red Tourism Collaboration Between Universities and Rural Areas Based on Intelligent Visual Computing and Cultural Logic

Wei Xu[1,*], *YiminShen*[1] *and Jianqin Shi*[1]

[1] Jiaxing Nanyang Polytechnic Institute; Jiaxing City, Zhejiang Province, 314031

SUMMARY: *In order to support the collaborative brand building of universities and rural red tourism resources, this paper proposes an intelligent visual computing framework guided by cultural logic. A multi-modal data set for cross-scene brand collaborative analysis is constructed by focusing on the images of university exhibition halls, practical activity images, course texts, rural ruins images, village space images, route node texts and spatial attribute information. Then, the narrative semantics, visual symbols and rural context clues in heterogeneous scenes are aligned through the cultural logic constraint module. Finally, the brand consistency and cross-scene correlation strength are determined by combining image, text and spatial attributes. Experimental results show that the proposed method achieves 92.3% accuracy, 90.8% Macro-F1 value and 0.918 cross-scene consistency on the final brand collaborative discrimination task, which is superior to the comparison models in terms of representation stability and collaborative brand building effectiveness. This framework provides a computable path for data-driven red tourism brand construction in the university-rural integrated cultural communication scenario.*

KEYWORDS: *Intelligent visual computing; Cultural logic constraint; Cross-scene semantic alignment; Red tourism brand collaboration*

1 Introduction

With the expansion of the digital communication scene of red tourism, the linkage mode between university cultural resources, rural spatial resources and red narrative is shifting from experience organization to data-driven expression. Brand building is no longer limited to logo design, event promotion and text introduction. Visual organization, cross-scene semantic cohesion and audience perception calculation have become components of collaborative communication modeling. Universities have content production capabilities and youth communication groups, and villages carry red sites, local memories and cultural elements. The collaborative relationship between them in red tourism brand communication has multi-modal characteristics, involving the joint calculation of images, texts, scene structures, geographical clues and cultural symbols. Establishing a computable, discriminative and verifiable brand collaboration model for this scenario can not only enhance the consistency of red tourism brand expression, but also help to improve the semantic connectivity between communication nodes.

Existing research forms a technical foundation in visual perception of tourist destinations, image classification, and multimodal content understanding. Xiao X et al. studied the

*vianbibabo@163.com
https://doi.org/10.65102/is2026710 Published: 30 April, 2026

relationship between the visual content of tourism photos and the quantitative analysis of destination image, and constructed a differentiated marketing analysis framework based on the visual content of photos, which provided data support for the computational expression of tourism image [1]. Cho N et al. proposed a tourist photo classification method based on deep learning model, and explored the path of tourist destination image recognition based on it, indicating that visual features can directly participate in the cognitive modeling of tourism scenes [2]. Wang X et al. studied the extraction mechanism of visual cues in inbound tourists' photos, and proposed an analysis path to perceive the destination image based on the user's photo content, which expanded the application boundary of tourism visual computing [3]. Nixon L J B compared the effect difference between the fine-tuned deep learning model and the existing measurement methods in visual destination image recognition, and verified the measurement ability of the deep model in tourism image perception [4]. Qian L et al. studied the application of user-generated photos in the image analysis of dark tourism destinations, and used deep learning methods to complete visual representation and image recognition [5]. Ma S et al. proposed a tourism demand prediction method combined with OTA platform user images, making image information transfer from static display content to dynamic demand calculation process [6]. Hu T and Geng J studied the tourism destination image perception method driven by multimodal user-generated content, and showed that more stable perception results could be formed after text, image and scene information fusion [7]. Wen T and Xu X proposed the BERT-BiLSTM-CNN-Attention model for tourism destination image perception analysis, which strengthened the coupling expression of semantic information and visual information [8]. Yuan X applied the BERT-enhanced deep learning model and BP algorithm to the evaluation of rural tourism development level, indicating that rural tourism scenes can enter the deep representation and computational analysis framework [9].

Red tourism brand collaborative computing has obvious cross-source heterogeneity. The university side data mainly came from course materials, theme exhibitions, practical activity images and explanatory texts, while the rural side data included site space images, village landscapes, oral narratives, local symbols and route information. The two types of data are not consistent in acquisition perspective, symbol density, semantic granularity and scene background. Directly using general image classification or regular text matching, it is easy to misjudge visual approximation as brand consistency, and it is easy to separate samples with semantically related but large appearance differences, which is difficult to form a unified representation for collaborative shaping. In order to make the brand modeling process computable, this paper constructs a method framework from three levels: visual feature organization, cultural semantic mapping and cross-scene collaborative discrimination. Firstly, the commemorative facilities, logo colors, character activities and spatial composition in the red tourism image are hierarchically encoded, and then the revolutionary narrative, local memory and rural culture clues are mapped into a unified semantic space. Finally, the image, text and geographical attributes are combined to determine the synergy strength and output the brand consistency. Such a processing path integrates university communication content and rural local resources into the same computing link, and also provides a clear feature basis and discrimination basis for subsequent experimental verification.

2 Related work

Under the background of the continuous advancement of tourism image recognition, recommendation modeling and cultural content digitization research, the existing results provide a clear technical reference for the collaborative computing between universities and

rural red tourism brands. Wang X studied the optimization analysis of rural tourism images under the conditions of Internet of things and deep learning, and proposed a rural tourism visual expression method combining network perception and deep feature analysis, so that the image information in rural scenes can enter the structured calculation process [10]. Yoon J H and Choi C studied real-time context-aware recommendation in tourism scenarios, and proposed a real-time recommendation system for dynamic environments, so that location, behavior and time information can be synchronized to participate in tourism decision calculation [11]. Alenezi T and Hirtle studied the attraction personality representation in the tourism recommendation system, and proposed the normalized travel personality representation method to enhance the matching ability of the recommendation results from the perspective of user preference modeling [12]. Nan X and Wang X studied the process design of personalized tourism recommendation, and proposed a recommendation system implementation method based on data mining and collaborative filtering, which strengthened the organization of behavior data in tourism service modeling [13].

Li Y studied digital tourism recommendation and route planning calculation, and proposed a recommendation and path design model based on RippleNet and improved genetic algorithm, so that tourism knowledge association and path search were included in the unified reasoning link [14]. Liu X studied the visual recommendation task of tourism destinations, and proposed a recommendation method combining bag of visual words model and support vector machine classification, indicating that visual representation and classifier design can jointly support tourism scene recognition and recommendation output [15]. This kind of research shows that tourism computing has extended from single information matching to visual understanding, behavior modeling and path reasoning, and the correlation between image features, context information and recommendation mechanisms is constantly strengthening.

At the same time, cultural content computing research also provides a methodological basis for red tourism brand building. Li H and Liu D studied the innovative development of arts and crafts intangible cultural heritage in artificial intelligence decision support system, and proposed a cultural content organization method based on intelligent decision support, so that the dissemination of cultural resources can realize knowledge reorganization through computational structure [16]. Wang Q studied the digital transmission of intangible cultural heritage supported by neural network vision, and proposed a digital expression method for inheritance and dissemination, so that the recognition, coding and presentation of cultural images form a continuous technical link [17]. Liu Y, Cheng P and Li J studied the design of Chongqing intangible cultural heritage application interface supported by deep learning, and proposed the realization path of the integration of cultural content interface and visual recognition, which enhanced the adaptability between cultural symbols and digital interaction [18].

Angheluș ă l M, Popovici A I and Ratoiu L C studied the visualization of multimodal imaging data in cultural heritage asset documents, and proposed a 3D display system based on Web platform, which provided a reusable platform for the structured presentation of complex cultural objects [19]. Sha S, Li Y, and Wei W et al. studied the classification and restoration of ancient textile images supported by convolutional neural networks, and proposed a visual processing method for joint classification and restoration, which extended cultural heritage image computing from static recognition to reconstructed expression. Related studies have shown that the digitization of cultural resources is no longer limited to image storage at the display level, but has gradually entered the computational stage of visual recognition, semantic organization and interactive expression in parallel [20].

To facilitate the comparison of the differences in application objects, technical paths, and

computational focus of the above studies, this paper organizes the related work as Table 1.

Table 1: Summary of related work

Reference	Author(s)	Research Content	Method or System Characteristics
[10]	Wang X	Optimization analysis of rural tourism images	Combines IoT sensing with deep learning to enhance the visual expression of rural tourism
[11]	Yoon J H, Choi C	Real-time tourism recommendation	Constructs a context-aware recommendation system incorporating location and time-period information
[12]	Alenezi T, Hirtle S	Tourism personality representation	Enhances recommendation matching by using normalized travel personality representations of attractions
[13]	Nan X, Wang X	Personalized tourism recommendation system	Integrates data mining and collaborative filtering for system implementation
[14]	Li Y	Digital tourism recommendation and route planning	Uses RippleNet and an improved genetic algorithm to unify recommendation and path search
[15]	Liu X	Visual recommendation for tourism destinations	Combines the bag-of-visual-words model with SVM classification for recommendation output
[16]	Li H, Liu D	Intelligent decision support for intangible cultural heritage content	Organizes cultural content through an artificial intelligence decision support system
[17]	Wang Q	Digital dissemination of intangible cultural heritage	Uses neural-network-based vision methods to achieve inheritance-oriented digital expression

It can be seen from Table 1 that the existing research has formed a relatively complete foundation in tourism image recognition, recommendation computing, digitization of cultural resources and multimodal display. Research on tourism focuses more on visual representation, context awareness and personalized recommendation, while research on culture focuses more on image coding, digital communication and visual presentation. The two kinds of results provide a referable algorithm basis and system idea for the construction of red tourism brand, and also show that visual computing and cultural content analysis have a strong combination space.

However, most of the existing work focuses on general tourism scenes, single-type cultural objects or single-platform recommendations, and the joint modeling between university course images, practical activity records, rural site space, local narrative text and route nodes has not yet formed a unified computational expression. Based on this research basis, this paper introduces intelligent visual computing and cultural logic into the collaborative modeling of universities and rural red tourism brands, which is used to realize cross-scene brand representation, consistency discrimination and collaborative communication calculation.

3 The collaborative method of university and rural red tourism brands based on intelligent visual computing and cultural logic

3.1 Visual feature extraction of red tourism brand based on intelligent visual computing

The visual expression of college and rural red tourism brands includes information such as memorial buildings, exhibition layout, activity scene, logo color and character behavior. It is difficult for single-scale image description to stably cover the structural clues required for brand building. To this end, this paper constructs a hierarchical visual feature extraction method, which inputs the university side pavilion images, practical teaching images, rural side site space images, and village environment images into the unified coding link. Firstly, local texture sampling is completed, and then global semantic aggregation is performed to generate the brand visual representation that can be used for subsequent collaborative modeling.

As shown in Fig. 1, red tourism brand visual feature extraction does not simply recognize the original image, but establishes a hierarchical visual coding link around the brand shaping task. The input terminal receives university side display images, practical activity images, rural side site space images, and village environment images at the same time, so that visual information from different sources, different perspectives, and different semantic densities enter a unified processing framework. After size normalization, brightness correction and region cropping, the background disturbance in the image is compressed, and the memorial facility contour, logo color distribution, human activity pose and spatial composition relationship are more clearly preserved. The two-branch convolutional encoding stage extracts the local detail response and the overall scene semantics in parallel, so that the model can not only identify the fine-grained differences of red symbols, but also maintain the overall narrative structure of the exhibition layout and rural spatial form.

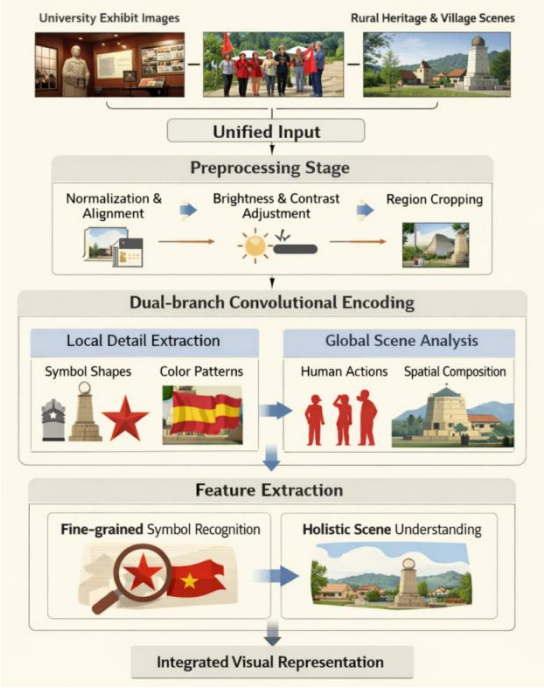

Figure 1: Red tourism brand visual feature extraction process

In the basic convolution mapping stage, the model performs convolution operation and nonlinear activation on the input image patch to obtain the feature response of the corresponding layer:

$$F_i = \sigma(W_i * X_i + b_i) \tag{1}$$

where X_i represents the input image block or feature block of the i layer; W_i represents the convolution kernel parameters of this layer; b_i represents the bias term; $*$ denotes the convolution operation; Let $\sigma(\cdot)$ denote the nonlinear activation function; F_i represents the feature map output from layer i. Formula (1) is used to extract the edge, texture and structure information in the red tourism image, which provides the basic visual representation for subsequent multi-scale feature aggregation.

After obtaining the multi-layer feature map, the model calculates the spatial average of each channel response to measure the contribution strength of different semantic channels in the visual expression of the brand:

$$\alpha_c = \frac{1}{HW} \sum_{m=1}^{H} \sum_{n=1}^{W} F_c(m, n) \tag{2}$$

Here, $F_c(m, n)$ represents the response value of the c channel at the spatial position (m, n). H and W represent the height and width of the feature map; Let α_c denote the average response weight of the c channel. Equation (2) is used to complete the channel-level feature recallification, so that key visual cues such as memorial facilities, logo colors, event scenes, and spatial composition are more stably retained.

As shown in Fig. 2, multi-scale visual feature aggregation is not a simple concatenation of features at different levels, but a cross-scale reorganization and semantic reorganization under unified constraints. The low-level edge features retain local structural information such as monument contour, architectural interface, flag texture and character action, while the high-level scene semantic features carry the overall content such as exhibition order, activity organization, space atmosphere and narrative direction. After the two types of features are scaled in the alignment layer, the channel recalibination module redistributes the contribution strength of different semantic channels, so that the visual elements with brand recognition receive higher weights and the redundant background responses are suppressed synchronously. After entering the semantic aggregation layer, the scattered visual responses are further integrated into stable brand prototype vectors.

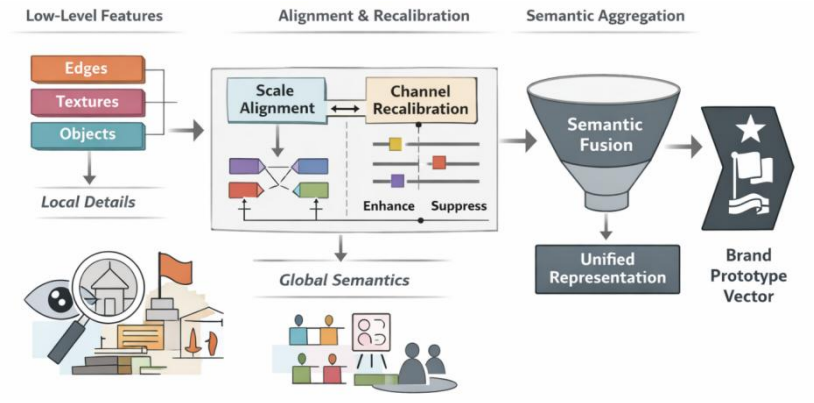

Figure 2: Multi-scale visual feature aggregation and brand semantic mapping mechanism

This paper further constructs the brand prototype representation vector:

$$z = \sum_{c=1}^{C} \alpha_c G_c \tag{3}$$

Here, G_c represents the aggregated feature vector corresponding to the c channel. C represents the total number of channels; z represents the final generated brand visual prototype vector. Formula (3) is used to compress the multi-scale visual response into a unified brand representation result, so that the university museum images and rural heritage images can enter the same discriminant space.

Through the above extraction process, the course display, activity organization and exhibition structure in the university end image, and the site outline, street shape and environmental symbols in the rural end image can be converted into vector expressions suitable for cross-scene comparison. This method provides a feature basis for the semantic alignment under the subsequent cultural logic constraints, and also provides a unified input for the collaborative modeling of universities and rural red tourism brands.

3.2 Cross-scene semantic alignment method based on cultural logic constraints

There are obvious differences between universities and rural red tourism brands in visual source, narrative context and scene organization. The images on the university side mainly represented course display, exhibition hall display, practical activities and youth participation, while the images on the village side carried more site space, village texture, memorial facilities and local narrative. If the two types of images are directly fed into the same discriminator, it is easy to cause semantic response offset, which makes the representation of brand clues unstable in cross-scene transmission. Therefore, based on the results of visual feature extraction, this paper introduces cultural logic constraints to construct a cross-scene semantic alignment method, which maps the educational expression in the university communication context and the historical expression in the rural spatial context into the same semantic space, so as to form a unified representation suitable for brand collaborative computing.

As shown in Fig. 3, the alignment process does not rely on a single distance compression. Instead, the semantic responses corresponding to revolutionary sites, memorial facilities, theme colors, character activities and narrative backgrounds are extracted by the cultural symbol coding module, and then the features of the university side and the rural side are mapped to the shared space through the projection layer. The mapped features continue to enter the distribution alignment layer and the logical constraint layer, where the former deals with the scene source difference, and the latter deals with the cultural expression order, symbol co-occurrence relationship and narrative consistency, so that the cross-scene features have stable comparability while maintaining difference information.

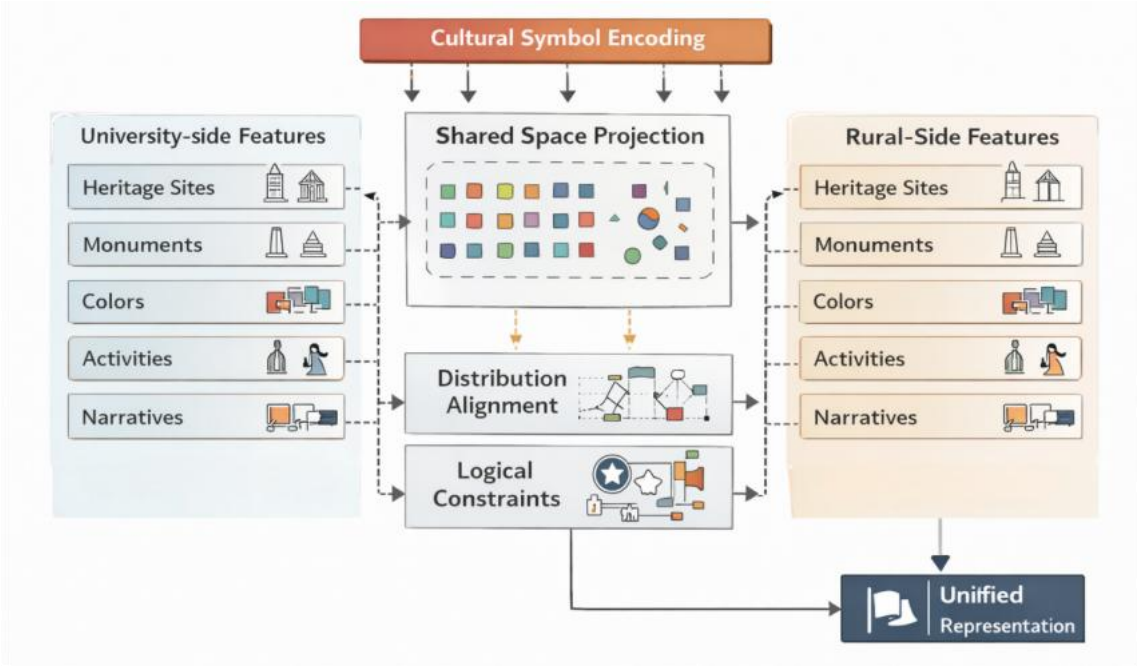

Figure 3: Cross-scenario semantic alignment process under cultural logic constraints

Let the visual feature matrix of the university scene be $U \in \mathbb{R}^{m \times d}$, the visual feature matrix of the rural scene be $V \in \mathbb{R}^{n \times d}$, and the semantic projection function be denoted as $\phi(\cdot)$. The base projection representation is defined as follows.

$$H_u = \phi(U)W_u, \quad H_v = \phi(V)W_v \tag{4}$$

Here, W_u and W_v represent the projection parameter matrix of the university side and the rural side, respectively, and H_u and H_v show the latent semantic representation after mapping. Equation (4) is used to transform the original visual features into low-offset representations that can participate in semantic alignment, so that the two types of scene features enter the unified computational domain.

After the initial projection, we apply the weighted mean difference constraint to compress the overall distribution distance of the two types of scenes in the shared space:

$$\mathcal{L}_m = \left\| \frac{1}{m} \sum_{i=1}^{m} H_u^{(i)} - \frac{1}{n} \sum_{j=1}^{n} H_v^{(j)} \right\|_2^2 \tag{5}$$

Here, $H_u^{(i)}$ represents the projection vector of the i university sample, $H_v^{(j)}$ represents the projection vector of the j rural sample, and m and n represent the number of samples of the two categories, respectively. Equation (5) is used to measure the overall distribution center difference so that the cross-scene semantic representation becomes stable at the global level.

The compressed mean distance alone is still not enough to express the cultural logical relationship in the red tourism brand. Therefore, this paper further constructs the symbolic correlation matrix C to constrain the semantic adjacent structure:

$$\mathcal{L}_c = \sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij} \left\| H_u^{(i)} - H_v^{(j)} \right\|_2^2 + \lambda \|C\|_F^2 \tag{6}$$

Here, C_{ij} represents the cultural symbol association strength between university sample i and rural sample j, λ represents the regularization coefficient, and $\|C\|_F^2$ represents the Frobenius norm of the association matrix. Equation (6) is used to strengthen the symbol co-occurrence relationship, so that samples with common revolutionary narratives, common spatial symbols, or common activity themes maintain higher semantic proximity after alignment.

As shown in Fig. 4, the global distribution constraint acts together with the symbol association constraint within the shared semantic space. The former controls the overall projection direction, while the latter modifies the local semantic relations. After logical consistency correction, monumental, educational, and local expressions in brand images are reorganized into aligned vectors that can be used for collaborative discrimination.

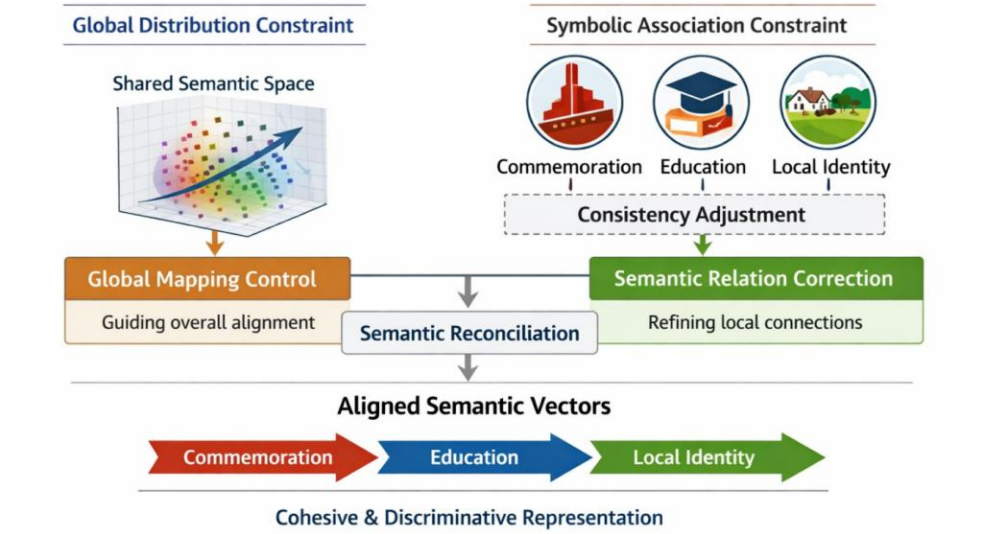

Figure 4: Cross-scenario semantic alignment and cultural logic constraint collaboration mechanism

In order to further reflect the role of narrative order in cultural logic, this paper adds adjacent preservation term to the sample sequence:

$$\mathcal{L}_s = \sum_{t=1}^{T-1} \|P_{t+1} - AP_t\|_2^2 \tag{7}$$

Here, P_t represents the semantic state of the t narrative segment, A represents the state transition matrix, and T represents the total number of narrative segments. Equation (7) is used to maintain the semantic continuity in the exhibition narrative, activity organization, and site visit sequence, so that the cross-scene alignment results not only preserve the symbolic similarity, but also the consistent structure of the expressive link.

In terms of local structure preservation, this paper further introduces the graph constraint term:

$$\mathcal{L}_g = \mathrm{Tr}(Z^\top L Z) \tag{8}$$

Here, Z represents the joint semantic representation matrix, L represents the graph Laplacian matrix constructed from the sample adjacency relation, and $\mathrm{Tr}(\cdot)$ represents the matrix trace operation. Formula (8) is used to maintain the local geometric relationship of adjacent samples in the projection space, so that the active image in the university scene and the site image in the rural scene still maintain stable adjacency when they are logically similar.

Considering the above constraints, the goal of cross-scene semantic alignment is written as follows.

$$\mathcal{L} = \alpha\mathcal{L}_m + \beta\mathcal{L}_c + \gamma\mathcal{L}_s + \eta\mathcal{L}_g \tag{9}$$

Here, α, β, γ and η denote each loss weight. Formula (9) is used to jointly adjust the global distribution, cultural symbol, narrative sequence and local structure, so that the university and rural red tourism images form a consistent semantic representation without losing discrimination in the shared space.

Through the above alignment process, the university end course display images and practical activity images can be stably mapped with the village end site images and village space images under the constraints of cultural logic. This method not only compreses the cross-scene visual differences, but also enhances the semantic continuity in the brand narrative, provides a unified feature basis that can be directly invoked for subsequent multi-modal collaborative discrimination, and provides a more stable semantic support for brand collaborative scoring.

3.3 Collaborative modeling of university and rural red tourism brand based on multi-modal collaborative discrimination

After the completion of visual feature extraction and cross-scene semantic alignment, the model also needs to further determine the brand synergy strength between the university end communication content and the rural end local resources. Instead of categorizing a single image, the process jointly models image, text, and spatial attributes around the branding task to output brand coherence results that can be used for collaborative decision making. This paper constructs a multi-modal collaborative discriminant model, which organizes the university side display images, course texts, activity records, and rural side site images, narrative texts, and route nodes into a unified input, and completes brand collaborative modeling through modal fusion, relationship constraints and collaborative scoring.

As shown in Fig. 5, the multimodal collaborative discrimination process consists of an input layer, a fusion layer, a relational constraint layer, and an output layer. The input layer receives the visual prototype vector, the semantically aligned text representation, and the spatial attributes such as geographical location, route connectivity, and scene category. The fusion layer completes the mapping between modalities. The relationship constraint layer depicts the collaborative relationship between universities and rural samples. The output layer gives the brand consistency score and collaboration level.

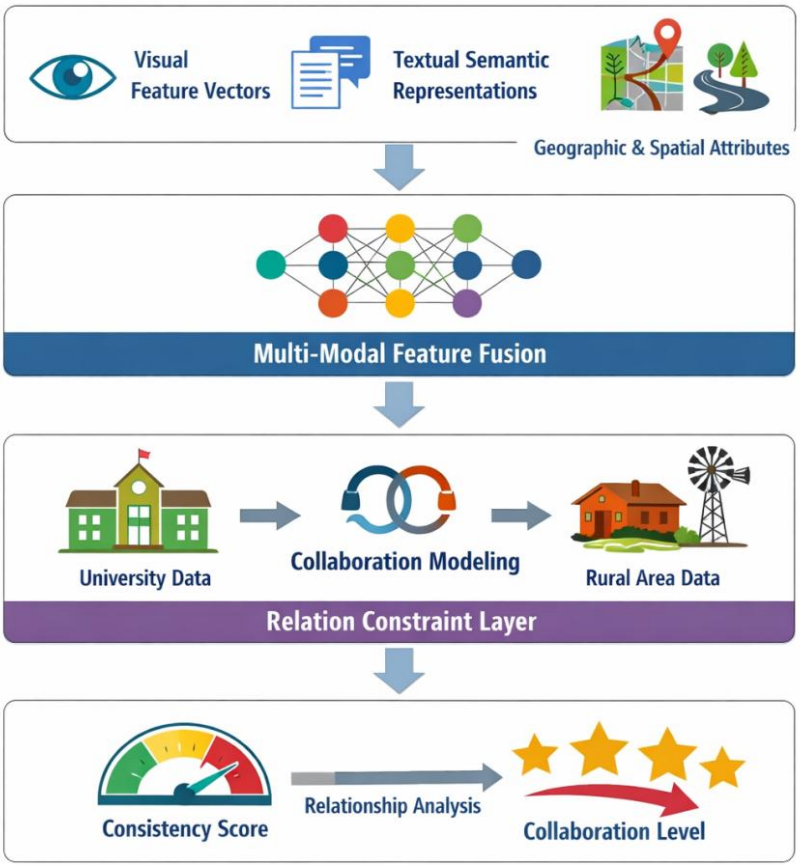

Figure 5: Multimodal collaborative discrimination process between universities and rural red tourism brands

In order to ensure that the expression strength of different modalities in the unified space is adjustable, this paper first performs a linear mapping of each modal representation:

$$M_v = X_v W_v, \quad M_t = X_t W_t, \quad M_s = X_s W_s \tag{10}$$

where X_v, X_t and X_s represent the visual, textual and spatial attribute input matrices, respectively. W_v, W_t and W_s denote the projection parameter matrices of the corresponding modes, respectively. M_v, M_t, and M_s denote the mapped modal representation. Equation (10) is used to eliminate the differences in dimensions and dimensions of different modalities so that the three types of inputs enter a unified discriminant space.

After completing the single-modal mapping, a gated fusion function is constructed to adjust the contribution ratio of different modalities:

$$g = \mathrm{softmax}\left([M_v \| M_t \| M_s] W_g + b_g\right) \tag{11}$$

Here, $[\cdot \| \cdot \| \cdot]$ represents the vector concatenation operation. W_g stands for gating weight matrix; b_g represents the bias term; g represents the normalized weight vector corresponding to the three classes of modes. Equation (11) is used to dynamically assign the effect strength of visual, textual and spatial attributes according to the sample content, so that brand collaborative modeling can retain the modal dominant features in different scenarios.

The fused representation is defined as follows.

$$Z = g_v \odot M_v + g_t \odot M_t + g_s \odot M_s \tag{12}$$

Here, g_v, g_t and g_s denote the components of the weight vector g corresponding to visual, textual and spatial attributes, respectively. $\odot$ for element-wise multiplication; Z represents the joint representation after multi-modality fusion. Formula (12) is used to form the core input of brand collaborative discrimination, so that the exhibition content of the university end and the site resources of the rural end can be jointly expressed in the same vector space.

After the single-modal mapping and gated fusion, the visual modality retained the explicit features such as memorial buildings, exhibition formats, activity scenes and logo colors, the textual modality retained the semantic clues such as revolutionary narratives, course explanations and activity themes, and the spatial modality depicted the structural information such as route organization, point connection and scene adjacent. The gated weights are generated by the content of the sample itself, so that images and texts from different sources can maintain their respective expression focus when entering a unified space.

Only relying on fusion features is still not enough to stably depict the collaborative structure, so this paper constructs a relationship matrix between the university sample set and the rural sample set. The matrix is not a simple similarity table, but a comprehensive relationship representation that simultaneously absorbs visual proximity, semantic echo and spatial correlation. After normalization, the relationship strength between any university sample and rural sample can reflect their corresponding degree in the brand communication chain. The introduction of relationship matrix makes the model no longer stay at single sample judgment, but can describe brand collaboration from two levels of pairwise connection and overall structure. The relationship strength of any sample pair (i, j) is defined as follows.

$$R_{ij} = \frac{\exp\left(Z_i^{\top} Q Z_j\right)}{\sum_{k \in B} \exp(Z_i^{\top} Q Z_k)} \tag{13}$$

Here, Z_i represents the joint representation of the i sample on the university side. Let Z_j denote the joint representation of the j sample on the rural side; Q represents the learnable relational parameter matrix; B is the rural sample set. R_{ij} denotes the normalized relationship strength of the sample pair (i, j). Equation (13) is used to describe the matching degree between the university communication content and rural scene resources, so that the model can perceive the direction of collaborative connection as a whole.

In order to enhance the structural consistency in brand building, this paper further introduces the collaborative constraint loss:

$$\mathcal{L}_r = \sum_{i \in A} \sum_{j \in B} R_{ij} \left\| Z_i - Z_j \right\|_2^2 + \mu \sum_{i \in A} \sum_{j \in B} \left(R_{ij} - Y_{ij} \right)^2 \tag{14}$$

where A represents the sample set of colleges and universities; Y_{ij} denotes the collaborative prior matrix generated by manual annotations or rules; μ represents the constraint balance coefficient; $\| \cdot \|_2^2$ denotes the squared Euclidean distance. Equation (14) The former term is used to compress the representation differences between high synergy samples, and the latter term is used to make the learned relationship matrix consistent with the brand prior, thus enhancing the model's ability to identify the red tourism brand synergy structure.

In order to facilitate the explanation of the input composition and discriminant function of the model, the main inputs are arranged as Table 2 in this paper.

Table 2: Input composition of multimodal collaborative discriminant model

Modality	Input Content	Function
Visual	Exhibition images, activity images, heritage site images	Represents brand symbols and scene structure
Text	Explanatory texts, course texts, narrative labels	Represents red narratives and cultural semantics
Spatial	Coordinates, route nodes, scene categories	Represents local associations and organizational relationships

It can be seen from Table 2 that the visual modality is responsible for preserving the explicit forms in the university pavilion and rural ruins, the text modality is responsible for strengthening the revolutionary narrative and cultural expression, and the spatial modality is responsible for presenting the point organization and route connection. When the three types of inputs enter the joint representation space together, brand collaboration no longer stays at the single modal level, but can form a consistent judgment from three levels of content, semantics and structure. The collaborative relationship established in this way is not only suitable for binary classification, but also suitable for subsequent ranking screening and combination recommendation.

In the final output stage, the model uses a two-layer discriminant function to generate brand synergy scores:

$$s_i = \sigma\left(W_o^{(2)}\rho\left(W_o^{(1)}Z_i + b_o^{(1)}\right) + b_o^{(2)}\right) \tag{15}$$

where,$W_o^{(1)}$ and $W_o^{(2)}$ represent the two-layer discriminant parameter matrices, respectively. $b_o^{(1)}$ and $b_o^{(2)}$ denote the bias terms, respectively; Let $\rho(\cdot)$ denote the intermediate activation function; $\sigma(\cdot)$ is the Sigmoid function. s_i represents the brand synergy score of the i sample pair. Equation (15) is used to map the multimodal joint representation to the 0 to 1 interval to form a directly comparable synergy result.

After integrating the classification loss, relational constraint loss, and parameter regularization term, the overall objective function of the model is written as follows.

$$\mathcal{L} = \lambda_1\mathcal{L}_{\text{cls}} + \lambda_2\mathcal{L}_{\text{r}} + \lambda_3\|\Theta\|_2^2 \tag{16}$$

Here $\mathcal{L}_{\text{cls}}$ represents the collaborative rank classification loss; $\mathcal{L}_r$ denotes the relational constraint loss; Θ represents all the trainable parameters of the model; $\|\Theta\|_2^2$ denotes the parameter regularization term;λ_1, λ_2, and λ_3 denote the weight of each part, respectively. Formula (16) is used to jointly optimize brand level discrimination, cross-scene relationship consistency and model complexity, so that the collaborative output of university and rural red tourism brands is more stable.

Through the above modeling process, the course display, exhibition images and activity records at the university end and the site space, local narrative and route nodes at the rural end can be jointly distinguished in the same framework. The proposed model not only preserves the visual identifiability of red tourism brands, but also preserves cultural semantics and spatial relationships, which provides direct input for the analysis of collaborative results in subsequent experiments.

4 Experimental analysis of collaboration between universities and rural red tourism brands based on intelligent visual computing and cultural logic

4.1 Selection of evaluation indicators

The evaluation metrics selected for this experiment include precision, recall, Macro-F1, cross-scenario agreement, and root mean square error. The accuracy represents the proportion of samples that are correctly predicted in the brand collaborative discrimination results output by the model, which is used to measure the stability of the multimodal collaborative discrimination model at the overall classification level. The recall rate represents the proportion of samples correctly identified by the model among all real collaborative samples, which can reflect the model's ability to capture the effective connection relationship between universities and rural red tourism brands. Macro-F1 comprehensively considers the precision and recall rate of each category, and is suitable for evaluating the collaborative recognition performance when the sample distribution is different. The higher the value is, the more stable the comprehensive performance of the model in brand consistency discrimination.

Cross-scene consistency is used to measure the matching degree between university end dissemination content and rural end local resources in a unified semantic space. This index does not only reflect surface visual proximity, but also reflects the correspondence strength between revolutionary narrative, activity theme, spatial organization and cultural symbols, so it is more suitable to describe the brand synergy state in this study. The higher the value, the better the model can maintain a stable mapping between samples from different sources, different views, and different expression densities. Root mean square error is used to measure the degree of deviation between the predicted synergy score and the manual annotation score, which can reflect the fine expression ability of the model at the continuous score level.

The above indicators evaluate the performance of the model from five aspects of overall discrimination, effective recognition, comprehensive balance, semantic consistency and score deviation, which can completely describe the experimental effects of collaborative modeling of universities and rural red tourism brands under the constraints of intelligent visual computing and cultural logic. For this study, brand synergy includes both clear category differentiation and continuous changes in synergy strength, so the classification index and error index need to be retained at the same time. Such a combination of indicators can not only correspond to the accuracy, Macro-F1 and cross-scene consistency results in the summary, but also provide a unified basis for subsequent experimental analysis. At the same time, the recall rate can reflect the model's ability to detect highly collaborative samples and reduce the interference of the overall sample distribution difference on the evaluation results, so as to reflect the model's recognition level of key collaborative relationships more truly.

4.2 Analysis of brand visual feature extraction effect based on intelligent visual computing

In this section, the experimental analysis is carried out around the brand visual feature extraction module, and the goal is to verify the representation ability of the intelligent visual computing method in the red tourism scene of colleges and villages. The experimental sample is composed of university pavilion images, practical activity images, rural ruins images, village space images and supporting information. After deduplication, cropping and unified annotation, the training set and test set are formed. In the training phase, five-fold cross validation is used, and stratified sampling is carried out according to the source of the scene

and the category label, in which 80% samples are used for training and 20% samples are used for testing. The visual coding part is implemented based on PyTorch, the backbone network uses a two-branch convolution structure with channel recali-ation, the optimizer uses AdamW, the initial learning rate is set to 0.0003, the batch size is 32, and the number of training rounds is 120. Under the same experimental conditions, ResNet50 and MobileNetV3 are selected to compare with the proposed method to observe the differences of different feature extractors in brand scene recognition.

In order to compare the performance of different visual feature extraction methods in college and rural red tourism brand scenes more intuitively, this paper counts the recognition accuracy of seven types of brand scenes under the three methods, and the results are shown in Table 3. The results in the table can reflect the recognition stability of different models for memorial facilities, event organization, spatial composition and comprehensive brand scenes.

Table 3: Comparison of feature extraction accuracy for seven categories of brand scenes

Scene Category	ResNet50/%	MobileNetV3/%	Proposed Method/%
Memorial Hall Scene	89.4	85.7	94.2
Heritage Site Exterior Scene	90.1	84.9	95.1
Village Spatial Scene	87.6	83.4	92.8
Activity Organization Scene	88.3	82.7	93.4
Participant Scene	86.9	81.5	91.7
Symbolic Sign Scene	85.8	80.2	92.3
Comprehensive Branding Scene	89.0	83.1	95.8

As can be seen from Table 3, the recognition results of the proposed method in seven types of scenes are higher than those of the comparison models, and the advantages are especially obvious in the scene of site appearance, activity organization and comprehensive brand. ResNet50 maintains a relatively stable recognition level on building contours and character activity samples, but is susceptible to background texture interference in logo symbols and comprehensive brand scenes. MobileNetV3 has a lighter model structure and faster inference speed, but it is weak in maintaining the details of memorial facilities, the narrative structure of the scene, and the combination of brand symbols. In contrast, the proposed method jointly models through dual-branch convolutional coding and channel recalitioning, while preserving local edges, color organization, and overall composition information, thus showing stronger cross-scene representation ability between university images and rural images.

In Fig. 6(a), the diagonal recognition rates of ResNet50 on seven types of scenes are 89%, 91%, 86%, 84%, 87%, 85% and 88%, respectively. In Fig. 6(b), MobileNetV3 corresponds to 85%, 88%, 82%, 81%, 84%, 81% and 86%; In Fig. 6(c), the proposed method is improved to 93%, 95%, 90%, 88%, 94%, 91% and 96%. Among them, the comprehensive brand scene reaches 96%, the memorial scene and the site appearance scene are 95% and 94%, respectively, indicating that the recognition of the proposed method in the composite brand scene is more concentrated and the category distinction is more stable.

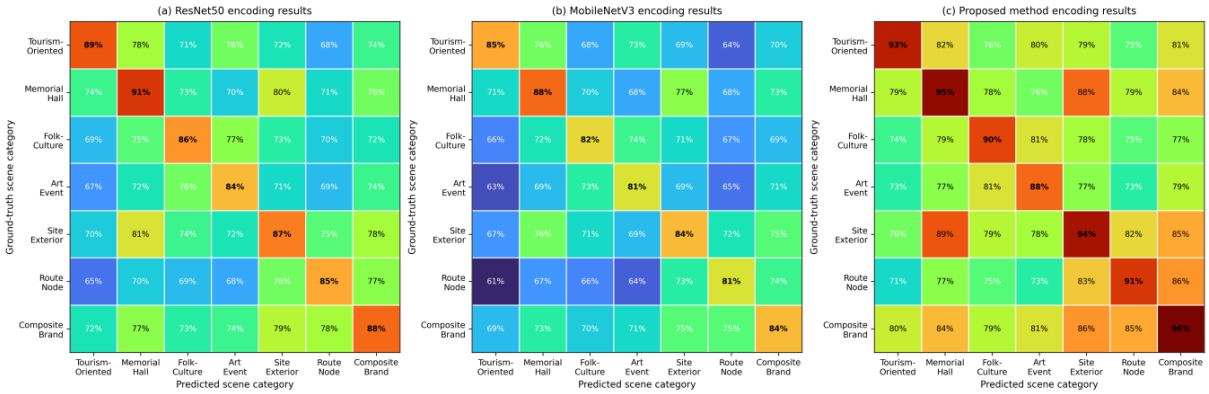

Figure 6: Comparison of recognition accuracy for seven categories of brand scenes

Based on the category results reflected in Fig. 6, the experiment continues to count the overall performance of the three methods in terms of average recall, Macro-F1 and precision, and the results are shown in Fig. 7. Fig. 7(a) shows that the average recall rate of the proposed method is 91.6%, which is higher than 86.3% of ResNet50 and 82.5% of MobileNetV3. Fig. 7(b) shows that the average Macro-F1 of the proposed method is 92.4%, which is higher than 88.1% of ResNet50 and 83.7% of MobileNetV3. Fig. 7(c) shows that the average accuracy of the proposed method is 93.2%, which is also higher than ResNet50 and MobileNetV3. The results show that the proposed method not only performs better in single-class scene recognition, but also has a more stable feature extraction ability under the overall sample distribution.

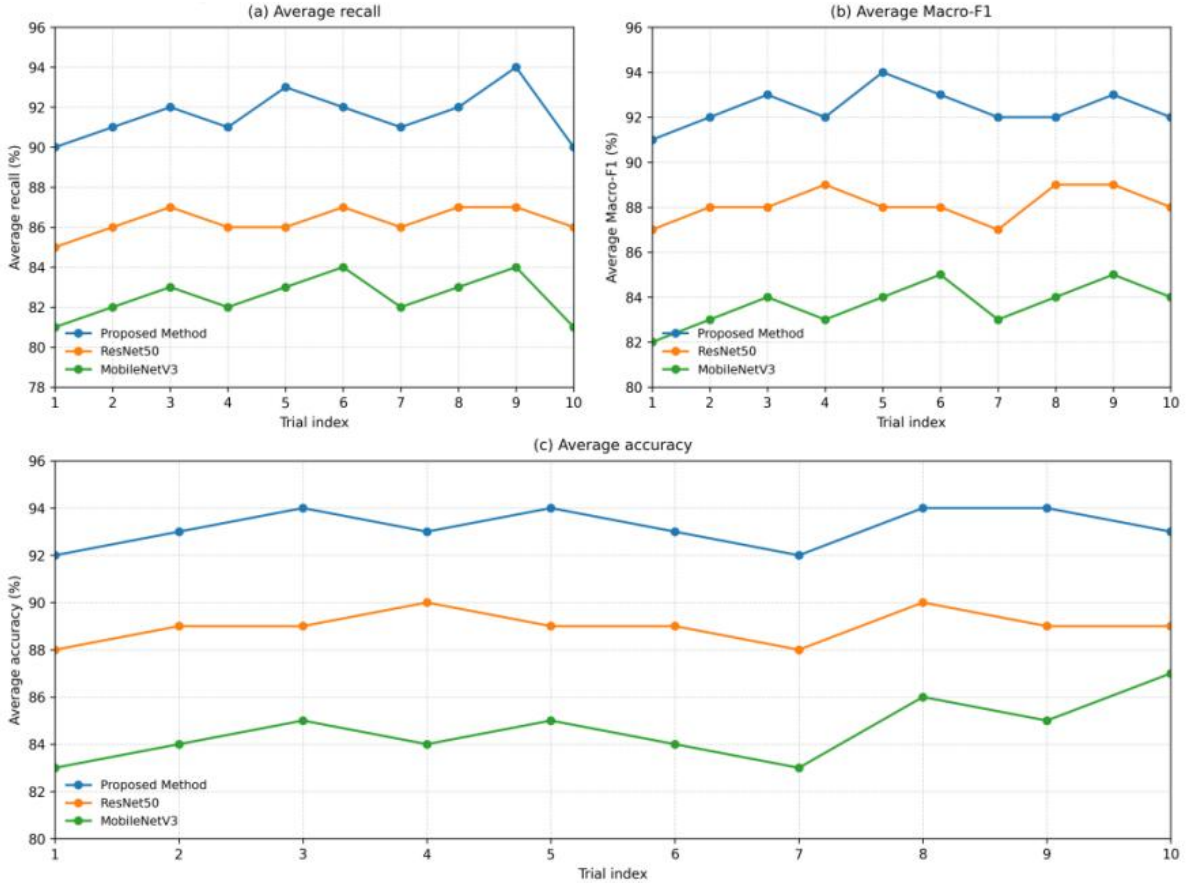

Figure 7: Comparison of the average performance of three visual feature extraction methods

Combined with Table 3, Fig. 6 and Fig. 7, it can be judged that the brand visual feature extraction method proposed in this paper forms a more stable unified representation between university and rural red tourism images. The course display, activity organization and exhibition structure of the university end, the outline of the site, the shape of the street and the sign node of the village end can maintain high separability and strong comparability in the same visual space. This feature representation result provides a reliable input for subsequent cross-scene semantic alignment under cultural logic constraints, and also lays a foundation for the stable output of the collaborative discriminant model.

4.3 Effect analysis of cross-scene semantic alignment based on cultural logic constraints

This section mainly tests whether the cross-scene semantic alignment module under the cultural logic constraint can effectively compress the representation difference between the college end and the rural end samples. The experimental objects include university exhibition hall images, practical activity images, course texts, as well as rural site images, village space images, and route node texts. Before alignment, the model only uses the original features output by the visual extractor. After alignment, cultural symbol constraints, narrative order constraints and local structure preservation terms are further added to the shared semantic space. In order to ensure the interpretability of the comparison results, the experiment uniformly uses the same number of training rounds, the same batch size and the same test set. After each round, the average distance, inter-class divergence and cross-scene consistency scores of university samples and rural samples in the latent space are recorded.

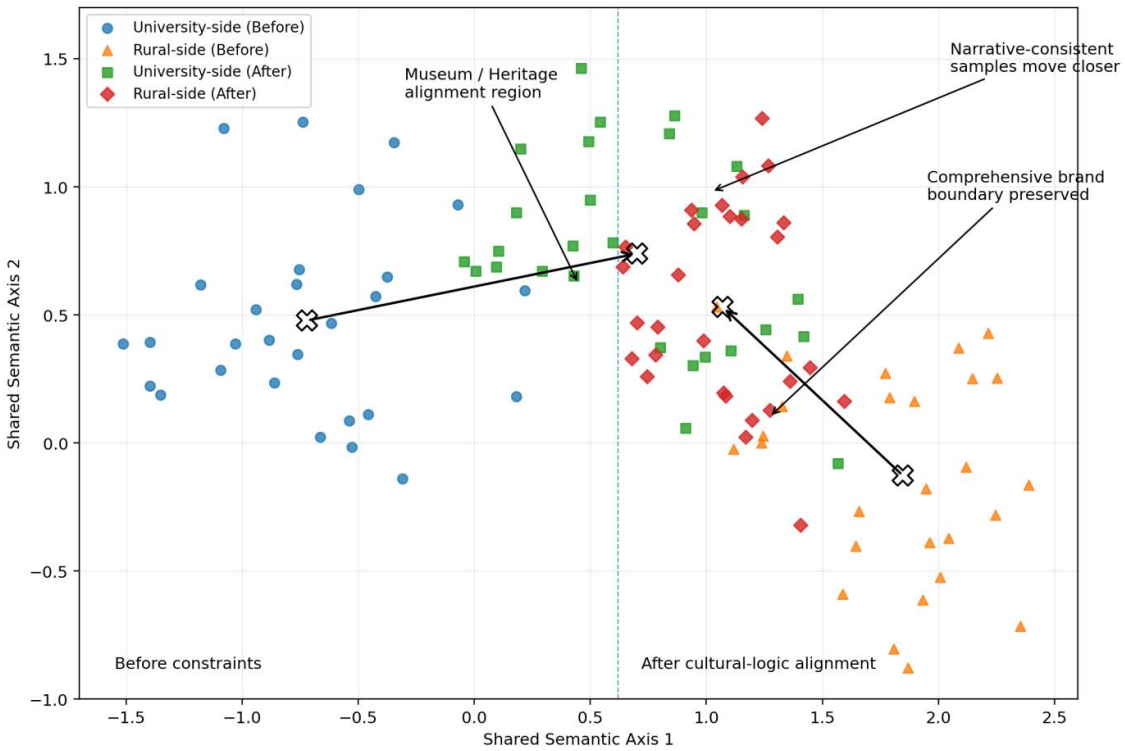

Figure 8: Comparison of feature spatial distributions before and after cultural logic constraints

As shown in Fig. 8, when the cultural logic constraint is not added, there is a clear separation between the university side samples and the rural side samples in the shared space.

The university display images are more concentrated in the area dominated by color and composition, and the rural ruins images are more distributed in the area dominated by texture and spatial boundary. Although there is a certain comparability between the two types of samples, the semantic center of gravity is still deviated, which leads to the unstable response of the association between memorial exhibition, course activities and rural sites. After adding the cultural logic constraint, the distribution centers of the samples at the university end and the rural end were significantly closer to each other, and the samples with consistent revolutionary narrative, close spatial symbol and connected activity theme gradually gathered into adjacent areas, while the samples with similar background textures but different cultural orientations were further separated. It shows that the module does not simply compress all sample spacing, but enhances the cross-scene proximity relationship with real brand collaboration significance while maintaining category discrimination.

From the quantitative results, the average distribution distance between college samples and rural samples decreased from 0.417 to 0.268 after semantic alignment, a decrease of 35.7%. The between-class scatter increased from 1.124 to 1.386, with an increase of 23.3%. The cross-scene consistency is improved from 0.842 to 0.918, an increase of 9.0%. The decrease of the average distance indicates that the semantic deviation of the two types of scenes in the shared space is controlled, the increase of the inter-class divergence indicates that different brand scenes still maintain good separability, and the increase of the consistency index indicates that the semantic mapping between university communication content and rural land resources is more stable. From the category level, the clustering boundaries of the memorial scene, the site appearance scene and the comprehensive brand scene are clearer after alignment, and the pairing stability between the university end activity samples and the rural end route node samples is also enhanced synchronously.

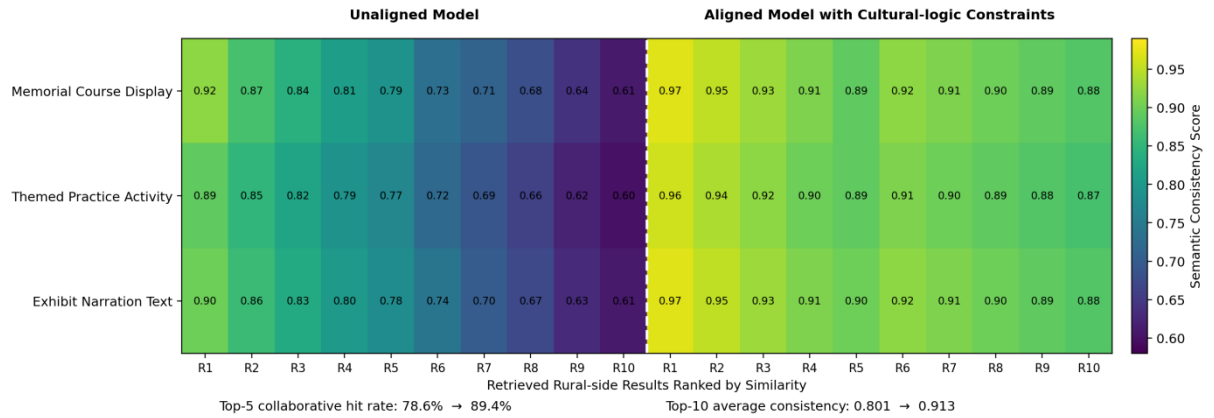

Figure 9: Comparison of collaborative retrieval results before and after semantic alignment across scenes

To further verify the role of this module in practical tasks, the experiment continues to compare the performance of the unaligned model and the aligned model in collaborative sample retrieval. In the test stage, the memorial course display, theme practice activities and exhibition interpretation texts are randomly selected from the university end as query input, and then the site images, village node images and route narrative fragments are retrieved from the rural end. As shown in Fig. 9, although the unaligned model can return some visually similar rural samples, the top-ranked results are still mixed with some nodes that are only close in color and composition, but deviate in narrative theme. After adding the cultural logic constraint, more rural resources that are consistent with the course theme, commemorative significance and spatial expression at the Top of the ranking are reserved. The top-5

collaborative hit rate is increased from 78.6% to 89.4%, and the Top-10 average consistency is increased from 0.801 to 0.913. Among the three groups of query examples, the consistency of the first result of the memorial course display is improved from 0.92 to 0.97, the theme practice activity is improved from 0.89 to 0.96, and the interpretation text of the exhibition is improved from 0.90 to 0.97. It shows that cultural logic constraints not only improve the distribution structure in the shared space, but also improve the consistency of the first result of the memorial course display. It also enhances the front aggregation ability of highly relevant samples in collaborative retrieval, and provides a more stable semantic foundation for subsequent brand collaborative discrimination.

4.4 Analysis of collaborative results between universities and rural red tourism brands based on collaborative discriminant model

This section analyzes the output effect of the collaborative discriminant model in the task of university and rural red tourism brand building. The test set is composed of university display images, activity records, course texts, village site images, node descriptions and route attributes. The evaluation is still based on accuracy, recall, Macro-F1, cross-scene consistency and root mean square error. To verify the effectiveness of the proposed model, ResNet50+MLP, BiLSTM-Fusion and CrossScene-GAT are selected as comparison methods in the experiment, and the results are shown in Table 4.

Table 4: Comparison of brand synergy discrimination results for different models

Model	Accuracy/%	Recall/%	Macro-F1/%	Cross-scene Consistency	RMSE
ResNet50+MLP	86.5	84.7	84.8	0.792	0.462
BiLSTM-Fusion	88.1	86.9	87.3	0.834	0.418
CrossScene-GAT	90.4	89.2	89.1	0.881	0.356
Proposed Model	92.3	91.6	90.8	0.918	0.318

Table 4 shows that the proposed model achieves the best results on the five core indicators. Compared with ResNet50+MLP, the accuracy is improved from 86.5% to 92.3%, Macro-F1 is improved from 84.8% to 90.8%, cross-scene consistency is improved from 0.792 to 0.918, and root mean square error is reduced from 0.462 to 0.318. Compared with CrossScene-GAT, the proposed model maintains obvious advantages in terms of consistency and error, indicating that multimodal collaborative discrimination not only improves the category discrimination ability, but also enhances the stable mapping between university content and rural resources. This result indicates that the joint modeling of visual features, cultural logic and spatial attributes in a unified framework can more fully express brand collaborative relationships.

To further examine the contribution of each component module to the results, this paper continues to carry out ablation experiments, and the results are shown in Table 5. The complete model maintains the highest performance, and the accuracy drops to 89.1% after removing the visual branch, indicating that brand symbols, scene composition and site contours are still important bases for collaborative discrimination. After removing the cultural logic constraint, the cross-scene consistency decreases from 0.918 to 0.861, indicating that semantic order, narrative direction and symbol association play a core role in connecting the university and rural samples. After removing the spatial attributes, the root mean square error rises to 0.367, indicating that the route node, geographical location and scene organization information have an obvious supporting effect on the collaborative strength score.

Table 5: Ablation experimental results of the collaborative discriminant model

Model Variant	Accuracy/%	Macro-F1/%	Cross-scene Consistency	RMSE
Full Model	92.3	90.8	0.918	0.318
Without Visual Branch	89.1	87.9	0.876	0.351
Without Cultural Logic Constraints	88.7	87.2	0.861	0.359
Without Spatial Attributes	89.4	88.1	0.873	0.367

Based on Table 4 and Table 5, it can be judged that the collaborative discriminant model proposed in this paper shows strong stability in the three levels of overall classification, cross-scene mapping and continuous scoring. The complete model can not only identify the effective connection between the content transmitted by the university end and the rural end resources, but also maintain a high consistency output under the condition of multi-source heterogeneous input, which provides a direct basis for the analysis of the applicability and deployment feasibility of the model in the subsequent discussion section.

5 Discussion

The experimental results show that after the joint modeling of intelligent visual computing and cultural logic, the brand collaborative relationship between university end communication content and rural end local resources can be more stably identified. The results in 4.2 show that the average accuracy of the visual feature extraction method proposed in this paper in seven types of brand scenes reaches 93.2%, the average recall rate is 91.6%, and Macro-F1 is 92.4%, which are higher than ResNet50 and MobileNetV3. The recognition effect is particularly obvious for the memorial hall scene, the site appearance scene and the comprehensive brand scene. The results show that the dual-branch convolutional coding and channel recalitioning can simultaneously retain local texture, logo symbols and overall composition information, so that the university exhibition images and rural ruins images maintain high separability in a unified space. The results in 4.3 further show that after adding cultural logic constraints, the average distribution distance between university samples and rural samples decreases from 0.417 to 0.268, the between-class scatter increases from 1.124 to 1.386, and the cross-scene consistency increases from 0.842 to 0.918. It shows that revolutionary narrative, spatial symbol and activity theme have stable constraint function in semantic mapping. The collaborative discrimination results in 4.4 show that the Accuracy of the full model reaches 92.3%, the Recall reaches 91.6%, the Macro-F1 reaches 90.8%, and the RMSE decreases to 0.318, which are better than those of the contrast model and the ablation model. It can be seen that the joint modeling of visual modality, textual modality and spatial modality in a unified framework not only enhances the expressive power of brand connection relationship, but also improves the stability of collaborative scoring output. This indicates that the computational support in red tourism brand building cannot stay at the single image recognition level, but should consider visual identification, cultural consistency and spatial organization relationship at the same time.

6 Conclusions

Focusing on the task of university and rural red tourism brand building, this paper constructs a computational framework composed of visual feature extraction, cultural logic constraint semantic alignment and multi-modal collaborative discrimination. Experimental results show

that the proposed method maintains a high level of accuracy, Macro-F1 and cross-scene consistency, indicating that university exhibition content, course activities, rural ruins and route nodes can form a stable mapping in a unified semantic space. This result shows that the joint modeling of intelligent visual computing and cultural logic can enhance the reliability of brand symbol recognition, narrative association analysis and collaboration strength assessment. The limitations of this paper are mainly reflected in two aspects. First, the sample sources are still mainly labeled images and structured texts, and complex oral data and dynamic videos have not been fully included. Second, the model training scene is based on fixed regional data, and there is still room for continued testing of cross-regional transfer ability. In addition, cultural logic constraints are mainly established based on explicit narrative labels and symbol co-occurrence relationships, and the description of implicit emotional expression and fine-grained historical context can still be further deepened. Subsequent research can further introduce video sequences, voice explanations and spatio-temporal trajectory data to build a more complete multimodal red tourism data system. At the same time, combined with lightweight deployment and incremental update mechanism, the real-time response ability and cross-regional adaptation ability of the model in the actual platform can be enhanced, and more stable computing support can be provided for the collaborative brand communication between universities and villages. The model output has both discrimination accuracy and deployment flexibility.

Funding

Project Source: Major Humanities and Social Sciences Research Projects in Zhejiang higher education institutions, Grant/Award Number: 2024QN022

Project level: Provincial and ministerial level project

Project Name: Research on the Mechanism and Practice of College Assisted Rural Red Tourism Brand Management Design

Grant/Award Number: 2024QN022

References

[1] Xiao X, Fang C, Lin H, et al. A framework for quantitative analysis and differentiated marketing of tourism destination image based on visual content of photos[J]. Tourism Management, 2022, 93: 104585.

[2] Cho N, Kang Y, Yoon J, et al. Classifying tourists' photos and exploring tourism destination image using a deep learning model[J]. Journal of Quality Assurance in Hospitality & Tourism, 2022, 23(6): 1480-1508.

[3] Wang X, Mou N, Zhu S, et al. How to perceive tourism destination image? A visual content analysis based on inbound tourists' photos[J]. Journal of Destination Marketing & Management, 2024, 33: 100923.

[4] Nixon L J B. Do deep learning models accurately measure visual destination image? A comparison of a fine-tuned model to past work[J]. Information Technology & Tourism, 2024, 26(3): 377-406.

[5] Qian L, Guo J, Qiu H, et al. Exploring destination image of dark tourism via analyzing user generated photos: A deep learning approach[J]. Tourism Management Perspectives,

2023, 48: 101147.

[6] Ma S, Li H, Hu M, et al. Tourism demand forecasting based on user-generated images on OTA platforms[J]. Current Issues in Tourism, 2024, 27(11): 1814-1833.

[7] Hu T, Geng J. Research on the perception of the terrain image of the tourism destination based on multimodal user-generated content data[J]. PeerJ Computer Science, 2024, 10: e1801.

[8] Wen T, Xu X. Research on image perception of tourist destinations based on the BERT-BiLSTM-CNN-attention model[J]. Sustainability, 2024, 16(8): 3464.

[9] Yuan X. Evaluation of rural tourism development level using BERT-enhanced deep learning model and BP algorithm[J]. Scientific Reports, 2024, 14(1): 25748.

[10] Wang X. The analysis of rural tourism image optimization under the internet of things and deep learning[J]. Scientific Reports, 2024, 14(1): 29898.

[11] Yoon J H, Choi C. Real-time context-aware recommendation system for tourism[J]. Sensors, 2023, 23(7): 3679.

[12] Alenezi T, Hirtle S. Normalized attraction travel personality representation for improving travel recommender systems[J]. IEEE Access, 2022, 10: 56493-56503.

[13] Nan X, Wang X. Design and implementation of a personalized tourism recommendation system based on the data mining and collaborative filtering algorithm[J]. Computational Intelligence and Neuroscience, 2022, 2022(1): 1424097.

[14] Li Y. Digital tourism recommendation and route planning model design based on RippleNet and improved GA[J]. Informatica, 2024, 48(10).

[15] Liu X. Tourism destination recommendation based on bag of visual word combined with SVM classification[J]. Informatica, 2024, 48(17).

[16] Li H, Liu D. Innovative development of intangible culture of arts and crafts in artificial intelligence decision support system[J]. Mobile Information Systems, 2022, 2022(1): 1123356.

[17] Wang Q. The digitisation of intangible cultural heritage oriented to inheritance and dissemination under the threshold of neural network vision[J]. Mobile Information Systems, 2022, 2022(1): 6323811.

[18] Liu Y, Cheng P, Li J. Application interface design of Chongqing intangible cultural heritage based on deep learning[J]. Heliyon, 2023, 9(11).

[19] Angheluță L M, Popovici A I, Ratoiu L C. A Web-based platform for 3d visualization of multimodal imaging data in cultural heritage asset documentation[J]. Heritage, 2023, 6(12): 7381-7399.

[20] Sha S, Li Y, Wei W, et al. Image classification and restoration of ancient textiles based on convolutional neural network[J]. International Journal of Computational Intelligence

Systems, 2024, 17(1): 11.