ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Innovation of Visual Content Dissemination Models of Intangible Cultural Heritage in Digital Media Environment

Juanning Zhang[1,2,*] *and Yaoxian Deng*[3,4]

[1] Faculty of Arts, Monash University, Melbourne, VIC 3145, Australia
[2] Taiyuan University of Technology, Taiyuan, 030000, China
[3] School of Loughborough University, E1 4EY
[4] Beijing Institute of Technology, Beijing, 100081, China

SUMMARY: *Intangible cultural heritage (ICH) is a spiritual and cultural imprint preserved by Chinese people of all nationalities in the process of historical development, and is an important part of strengthening cultural confidence and enhancing cultural soft power. Based on Harold Lasswell's "5W" theory, the study analyzes intangible cultural heritage in terms of subject, content and object; From the aspects of user's preference for different modal information and dynamic multimodal information fusion, the model of enhanced preference-aware MMGCN is proposed to achieve potentially salient feature extraction of video features through coarse- and fine-grained modeling. Construct MMGCN-based non-heritage short video communication power measurement model, and use MMGCN recommendation model to analyze the communication effect of non-heritage short video visual content for Chu opera. The data results show that the MMGCN recommendation model improves 10% to 20% on the dataset, compared to other models, and the retrieval effect is better on the dataset, which indicates the excellent performance of the recommendation model. On the dissemination effect of Chu opera video based on MMGCN recommendation, the maximum value of dissemination effect is 71.9023, and the dissemination effect is remarkable. The research provides technical support for the efficient dissemination of non-heritage visual content in accurate recommendation, which helps the inheritance and innovative dissemination of non-heritage culture.*

KEYWORDS: *dynamic multimodal fusion; intangible cultural heritage; visual communication; recommendation system; communication effect*

1 Introduction

Intangible cultural heritage (ICH) is the existence of various forms of traditional history, culture and art in intangible form that are closely related to the people and have a long history. It emphasizes the skills, spirit, emotion, experience, etc., which are centered on human beings, and is very valuable for research and dissemination [1]. And nowadays, ICH is facing the risk of being lost. On the one hand, the impact of multiculturalism in the context of globalization has led to a decrease in the audience of ICH and the gradual aging of the inheritors; on the other hand, there is a single channel of transmission of ICH, and the content of the oral transmission of the transmission method is limited, and the traditional activities of ICH are confined to the local people [2-5]. Since the new millennium, the speed of network technology development

*zhangjuanning8697@163.com
https://doi.org/10.65102/is2026233
Published: 30 April, 2026

has been accelerating, and more and more digital media have been applied in the inheritance and development of ICH.

The rapid development of digital media has greatly expanded the dissemination scope and audience of ICH, presenting ICH content in visual form [6]. Among them, visual content dissemination refers to the dissemination of information to achieve a certain purpose through image symbols, forms and behaviors, including advertisements, posters, live environments, illustrations in printed materials, movies, television, network videos, photography, image design, and visual performances in sports [7]. Through digital recording, display and dissemination, ICH can be presented to the global audience in a more vivid, three-dimensional and diversified form, arousing the interest and attention of people from different cultural backgrounds, and promoting cultural exchange and mutual appreciation [8-10]. However, in the process of practice, there are many problems in digital media-assisted ICH visual content dissemination, such as the lack of interactive experience in unidirectional dissemination, the low quality of static display and dynamic experience, and the visual content does not meet the modern aesthetics, etc. [11-13]. Therefore, with the help of digital media environment, the innovation of ICH visual content dissemination mode can help the effective inheritance, dissemination and utilization of non-heritage.

The study constructs a 5W communication model for intangible cultural heritage-based visual content mediated by short videos from five dimensions: communicator, communication audience, communication content, communication medium and communication effect. Subsequently, based on the user's preference for different modal information, coarse/fine-grained modeling is used to strengthen the influence of preference-aware modal representation learning on video feature embedding, and to achieve potentially salient feature extraction of video features. Further, the preference-aware modal representation learning method and the dynamic multimodal fusion method are proposed to further improve the performance of the existing short video recommendation system and achieve better dissemination of non-heritage content. Finally, taking Chu opera as an example, a communication effect evaluation model is constructed from the three dimensions of communication subject, communication content and communication object of intangible cultural heritage visual content videos to explore the current situation of the communication power of non-heritage short videos, so as to further promote the protection and inheritance of non-heritage culture.

2 5W-based visual content dissemination model for the intangible cultural heritage category

2.1 The “5W” communication model

The “5W” communication model in the field of communication science was proposed by Professor Harold Lasswell of Yale University in 1948, and it was the first time that the communication activities of human society were analyzed by means of the model, that is, the communicator, the message, the media, the audience and the effect of communication, with the specific structure shown in Figure 1.

As seen in Figure 1, the communicator is the starting point of the whole communication activity, responsible for collecting, processing and transmitting information in the communication activity, and is the issuer and “gatekeeper” of the whole information. Message is the content of communication, refers to a set of interrelated meaningful symbols, can express a certain complete meaning of the information. The medium is also known as the communication channel, channel, is the information in the transmission process with the help of the carrier. The audience, also known as the trustee, is the receiver and respondent of the

message. The effect is the impact and result of the communication behavior on the recipients and the society. Communicators, messages, communication media, audiences and communication effects are both independent of each other and have an impact on each other, constituting the entire chain of information dissemination activities.

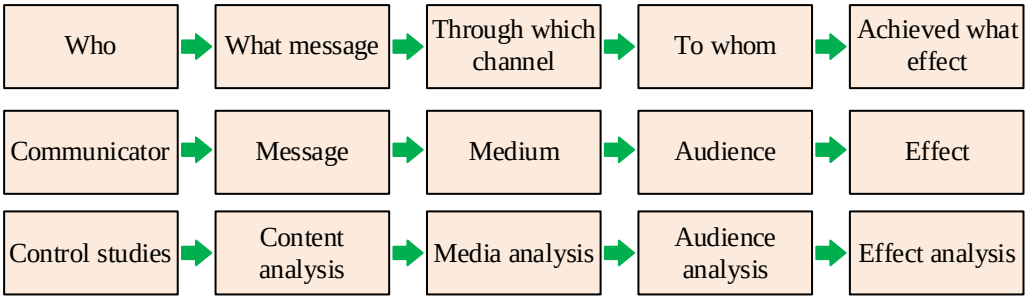


Figure 1: "5W" Model Elements Diagram

2.2 Analysis of the elements of visual content dissemination in the non-heritage category

2.2.1 Disseminators

Who is the subject of information dissemination? Commonly known as the "disseminator", also known as the information transmitter, is the initiator of communication activities, but also the "gatekeeper" of information dissemination. It is the initiator of communication activities and the "gatekeeper" of information dissemination. In the whole communication activity, it dominates the collection, filtering and screening of information. The communicator is always in a dominant position in the whole communication activity, controlling the whole communication activity. In recent years, governments at all levels have attached great importance to the promotion and protection of intangible cultural heritage, and government departments dominate and control the disseminator. The Binyang Cannonball Dragon Dance has undergone the process of conception, formation, development, maturity and improvement in its thousands of years of development, resulting in the existing long historical origins, rich cultural connotations, and harmonious coexistence with foreign cultures.

2.2.2 Dissemination of audiences

To whom? The audience, commonly known as "information receivers", is the group of people who receive information. As the object of information dissemination, the audience is both the receiver and the participant of information activities, and has an important position in the process of information dissemination activities, including local residents, tourists, scholars, news media and so on, who disseminate the wonderful moments of the Binyang Cannonball Festival and what they have seen through WeChat, microblogs, forums, and new media such as Jittery Voice, Shuttle, and so on, attracting non-live audience to like, discuss, exchange, etc., which extends the communication channels and expands the communication effect. They have extended the communication channels and expanded the communication effect.

2.2.3 Dissemination of content

What is the content of information dissemination? It is the information that is transmitted to the audience through the communication media. In the case of intangible cultural heritage, it is to publicize the rituals and processes of intangible cultural heritage, which is a cultural activity

with a national flavor, through the communication media, so as to achieve the value effect of the communication activities.

2.2.4 Media

The medium of communication refers to the tools, methods and means used for communication activities, and is an essential component for the smooth and complete realization of communication activities. Utilizing new information technology means such as computers and the Internet to digitally record, collect, process and store images, sounds, movements, videos and three-dimensional data information of intangible cultural heritage helps to promote the dissemination, protection, development and utilization of intangible cultural heritage.

2.2.5 Dissemination effects

What are the communication effects? The communication effect refers to what changes are brought to the recipients of the information after it is disseminated through diversified channels, and what effects are produced on society, economy and culture. In terms of economic effects, the sustainable development of ICH is promoted through investment attraction, development of tourism and performing arts, and diversified utilization of resources. In terms of cultural effects, strengthening the training of local inheritors of intangible cultural heritage, making full use of the fitness, fun and regional nature of traditional folk culture to cultivate the national cultural awareness of young people, so that the intangible cultural heritage can be passed on to future generations.

3 Recommendation algorithms that integrate user preference perception with non-heritage visual content

3.1 Graph Convolutional Neural Networks

Convolutional neural network is a higher level of traditional neural network, which is a feed-forward neural network with depth structure developed inspired by the visual perception mechanism of living beings. This neural network adds a convolutional layer and a pooling layer to the traditional neural network, making it easier to calculate the output values and adjust the parameters. A graph convolution network is a network that performs a convolution operation on a graph. The relevant equation (1) for the graph convolution operator is given below:

$$h_i^{l+1} = \sigma\left(\sum_{j \in N_j} \frac{1}{c_{ij}} h_j^l w_{R_j}^l \right) \tag{1}$$

In the above equation (1) i denotes the neutral point; h_i^l denotes the feature expression of node i at the l th layer; N_j denotes the neighbors of node i; c_{ij} denotes the normalization factor; R_j denotes the type of node i; $w_{R_j}^l$ denotes the transformation weight parameter of R_j type node's transform weight parameter at the l th layer.

The key to the application of the empty domain graph convolutional network algorithm is to aggregate the information of neighboring nodes. The algorithm mainly consists of two parts, which are message propagation and state update, and the v node in the null domain convolutional neural network is updated after processing the message sent by each neighbor. Then there is expression (2):

$$h_v^{k+1} = U_{k+1}\left(h_v^k, \sum_{u \in N(v)} M_{k+1}\left(h_v^k, h_u^k, X_{uv}^e\right)\right) \tag{2}$$

In the above equation (2) k denotes the k th layer of graph convolution; U_{k+1} denotes the node feature update of the k th layer. New function; M_{k+1} denotes the message passing function of the k th layer; $N(v)$ denotes the neighboring nodes of node v; and X_{uv}^e denotes the features on the edges between node v and its neighboring node u.

The frequency domain graph convolutional neural network is realized by using graph Fourier transform as a tool, through the graph Fourier transform, the frequency domain graph convolutional neural network can convert the spatial image into a frequency domain image, and use the Laplace matrix and the convolution operator in the Euclidean space to convolve the converted image, so as to realize the neural network processing of the image. The Laplace matrix A is a common graph Fourier transform, which is defined by subtracting the degree matrix L from the adjacency matrix A, i.e., L=D-A. L in a frequency domain graph convolutional neural network can be decomposed according to the following equation (3):

$$\begin{cases} L = U \wedge U^T \\ U = \left(u_1, u_2, \ldots u_n\right) \\ \wedge = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \end{cases} \tag{3}$$

The convolution operation of the features with the convolution kernel in Fig. G is the inverse transformation back to the null domain after multiplying the results of their respective Fourier transforms and Eq. Then we have Eq. (4):

$$f * g = U\left((U^T f) \odot (U^T g)\right) = U\left(U^T f \odot U^T g\right) \tag{4}$$

Treating $U^T g$ as a learning convolution kernel and setting $g_\theta = diag(U^T g)$, we have equation (5):

$$f * g = U_{g_\theta} U^T f \tag{5}$$

3.2 Short Video Recommendation Algorithm for Multimodal Graph Convolutional Networks

3.2.1 Aggregation Layer

The main purpose of the aggregation layer algorithm is to derive the higher-order information of the target vertices, which can be characterized by the attribute information of the target vertices and their neighboring vertices, in order to realize the propagation of the feature information in the topological graph from the neighboring area to the individual. The prerequisite for the realization of the aggregation layer algorithm is to construct a modal level "user-video" two-part graph as shown in Fig. 2.

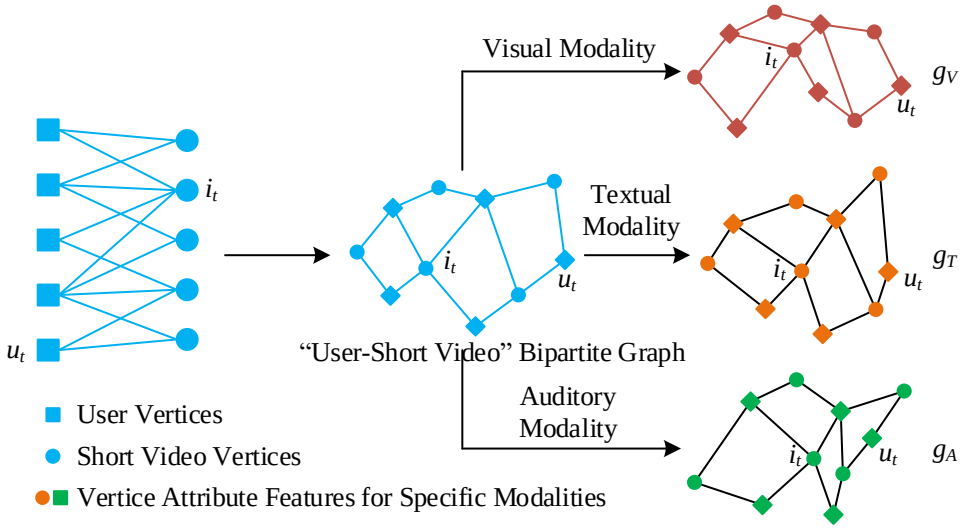


Figure 2: A Modal Hierarchy of User-Short Video Bipartite Graph

Constructed based on the interaction behaviors between users and short videos, like users commenting on short videos, users liking short videos, users playing short videos in full and incomplete. The mathematical expressions (6) and (7) are as follows:

$$G = \{(u,i) \mid u \in U, i \in I\} \tag{6}$$

$$e(u,i) = 1 \tag{7}$$

In the above equations (6) and (7), G denotes the interaction behavior between the user and the short video; u and i denote the user and the short video vertices, respectively; U and I denote the set of the user and the short video vertices, respectively; and e denotes the edge connecting the user and the short video vertices.

The attribute information of the vertices on the bipartite graph under different modes is the corresponding modal information, and the distance size between the vertices indicates the information difference under different modes. Then there is expression (8) as follows:

$$G_m, m \in M = \{T, V, A\} \tag{8}$$

In the above equation (8), G_m denotes the interaction behavior in the context of a particular modality; m denotes one of the three modalities: textual, visual and auditory; M denotes the set of the above three modalities; T is the abbreviation of Textual, denoting the textual modality; V is the abbreviation of Visual, denoting the visual modality; A is the abbreviation of Acoustic, denoting the auditory modality.

3.2.2 Integration layer

Combined with the principle of graph convolutional network algorithm, it can be seen that the main purpose of the integration layer algorithm is to integrate the attribute information of the target vertices and their higher-order neighborhood information in a specific mode. It is clearly pointed out in the above integration layer that the "homogeneous graph" constructed by the traditional graph convolutional network algorithm ignores the differences in the influence of vertex attribute information and structural information on user preferences, and inputting them

into the algorithmic model together as homogeneous information, which, to a certain extent, will lead to distortion of the computational results of the recommendation algorithm. For this reason, this paper designs the integration layer algorithm (9) based on the algorithm idea of "user-video" two-part graph construction as follows:

$$H_{m,v} = f_{merge}\left(h_{m,v}, x_{m,v}, h_{v,id}\right), m \in M, v \in V \tag{9}$$

In Eq. (9) above, $H_{m,v}$ denotes the characterization vector of vertex v in the two-part "user-short-video" graph in a particular m-modality; $f_{merge}(\cdot)$ denotes the integration function; $h_{m,v}$ denotes the higher-order characterization information of vertex feature information, which is the output of the aggregation layer of vertex v in a particular m-modality; $x_{m,v}$ denotes the zero-order information, which is the raw information of the vertex in a particular m-modality; and $h_{v,id}$ is the characterization vector of the structural information of the vertex, which is the embedding vector of vertex v.

3.2.3 Fusion and output layers

After integrating the user-short video representation information in different modalities, the fusion layer inputs it into the fusion layer for fusion and output. The mathematical expressions (10), (11) are as follows:

$$z_u = \left[H_{V,u}, H_{T,u}, H_{A,u}\right] u \in U \tag{10}$$

$$z_i = \left[H_{V,i}, H_{T,i}, H_{A,i}\right] i \in I \tag{11}$$

In Eqs. (10) and (11), z_u and z_i denote the output vectors of the integration layer of the user vertex u and the short video vertex i in the textual, visual, and auditory modalities, respectively; and U and I denote the sets of the user vertices and the short video vertices in the two-part graph of the "user-short-video".

3.3 Preference-aware multimodal recommendation algorithm for visual content

3.3.1 User and NRM visual feature representation

(1) User Characterization Representation

In order to more deeply understand and characterize the user's features and preferences, in this paper, we use the individual modal information of the short videos interacting with the user and perform multiple iterations through multilayer graph convolution to learn and enrich the representation of each user. This approach allows us to dynamically learn and update user characteristics from the perspective of the user's interaction with the short video to better meet the needs of personalized recommendation.

(2) Visual feature representation

For the visual modality of short videos, this paper first extracts the keyframe information of short videos, and then uses the pre-trained ResNet50 to extract the keyframe features, because each short video may contain a different number of keyframes, so this paper adopts the average pooling operation to obtain the global visual feature representation of short videos

$a \in \mathbb{R}^{D_e}$, where D_e is the visual modal representation dimension. In order that different modal features can be in the same mapping space, this paper utilizes a linear fully-connected layer to map the visual feature representation a into the D-dimensional space, denoted as $\breve{v}_v \in \mathbb{R}^D$.

3.3.2 Modal representation layer for preference perception

Since different users have different preferences for modal information of short videos, this paper designs two mechanisms to enhance users' preference information for each modality.

(1) Coarse-grained preference modeling

Aiming at the fact that users have different preferences for multiple modal information in short videos, this paper introduces a coarse-grained preference modeling mechanism to adaptively highlight the user's preference for specific modal information. Specifically, for the embedding of modal feature representation $\breve{v}_m \in \mathbb{R}^D \left(m \in \{v,a,t\}\right)$, this paper first calculates the importance score of each modality using the following formula:

$$g_m^{coarse} = \sigma_1\left(W_m^{coarse^1} \cdot \breve{v}_m + b_m^{coarse^1}\right) \tag{12}$$

where σ_1 is the activation function, $W_m^{coarse^1} \in \mathbb{R}^{1\times D}$, and $b_m^{coarse^1}$ denote the learned parameter vectors of the modality, respectively and bias values, respectively.

(2) Fine-grained preference modeling

This paper proposes a fine-grained preference modeling approach, which further utilizes the user's fine-grained preference information to model the modal interior.

Specifically, given the feature representation of a modality, i.e., $\breve{v}_m \in \mathbb{R}^D \left(m \in \{v,a,t\}\right)$, the fine-grained preference scores within the modality are first computed using the following equation:

$$g_m^{fine} = \sigma_2\left(W_m^{fine^1} \cdot \breve{v}_m + b_m^{fine^1}\right) \tag{13}$$

where σ_2 is the activation function, $W_m^{fine^1} \in \mathbb{R}^{D\times D}$, and $b_m^{fine^1}$ denote the learned parameter vector of the modality and the bias value, respectively .

3.3.3 Modeling User Preferences

The preference-aware short video multimodal representation is passed as input to the MMGCN backbone network to obtain the short video multimodal node representation and user node representation. For user nodes, when a user browses a short video and other users who like this video also like other short videos, this user-short-video-user-short-video third-order information as shown in Fig. 3 is a kind of higher-order interaction history information, and this kind of information is very important in video recommendation.

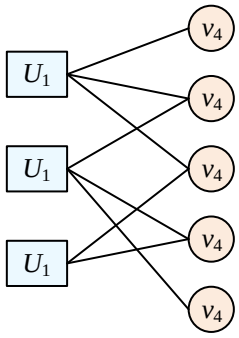


Figure 3: Example of user video interaction history

3.3.4 Dynamic multimodal fusion layer

Existing short video recommendation methods determine whether to recommend a short video to a user by calculating a similarity between a representation of the short video and a user representation. In order to obtain the final video/user representation, they usually sum or splice the multimodal representations of the video to directly obtain the final feature representation of the video. However, they ignore that different users have different preferences for different modalities of the same short video. Inspired by this, this paper proposes two mechanisms for dynamically fusing multimodal information.

(1) Coarse-grained fusion mechanism:

1) Dynamic multimodal fusion of non-legacy visual representations

For the different modal subgraphs in the model output the short video modal representation $\breve{v}_m \left(m \in \{v,a,t\}\right)$. The importance score of each modality is first calculated as follows:

$$\left[f_1^{coarse}, f_2^{coarse}, \ldots, f_M^{coarse} \right] = \sigma_3 \left(W_m^{coarse^2} \cdot \left[\tilde{v}_1 \oplus \tilde{v}_2 \oplus \cdots \oplus \tilde{v}_M \right] + b_m^{coarse^2} \right) \tag{14}$$

where M is the number of modes, σ_3 is the activation function, $\oplus$ denotes the tandem operation, $W_m^{coarse^2} \in \mathbb{R}^{M \times Md_m}$ is the learnable parameter vector, $b_m^{coarse^2}$ is the bias vector. The scores of all modes are then normalized by the $Soft\max$ function with the following formula:

$$\left[\overline{f}_1^{coarse}, \overline{f}_2^{coarse}, \ldots, \overline{f}_M^{coarse} \right] = Soft\max \left(\left[f_1^{coarse}, f_2^{coarse}, \ldots, f_M^{coarse} \right] \right) \tag{15}$$

Finally, the multimodal representation of the short video is weighted and summed to obtain the final representation of the short video:

$$v^{coarse} = \sum_{m=1}^{M} \overline{f}_m^{coarse} \tilde{v}_m \tag{16}$$

2) Dynamic multimodal fusion of user representations

For the user modal representation $\tilde{u}_m \left(m \in \{v,a,t\}\right)$ output from different modal subgraphs in the model. The user preference score for each modality is first calculated as follows:

$$\left[l_1^{coarse}, l_2^{coarse}, \ldots, l_M^{coarse} \right] = \sigma_4 \left(W_m^{coarse^3} \cdot \left[\tilde{u}_1 \oplus \tilde{u}_2 \oplus \cdots \oplus \tilde{u}_M \right] + b_m^{coarse^3} \right) \tag{17}$$

where M is the number of modes, σ_4 is the activation function, $\oplus$ denotes the tandem operation, $W_m^{coarse^3} \in \mathbb{R}^{M \times Md_m}$ is the learnable parameter vector, $b_m^{coarse^3}$ is the bias vector. The scores of all modes are then normalized by the $Soft\max$ function with the following formula:

$$\left[\overline{l}_1^{coarse}, \overline{l}_2^{coarse}, \ldots, \overline{l}_M^{coarse}\right] = Soft\max\left(\left[l_1^{coarse}, l_2^{coarse}, \ldots, l_M^{coarse}\right]\right) \tag{18}$$

Finally, the multimodal representation of the user is weighted and summed to obtain the final representation of the user:

$$u^{coarse} = \sum_{m=1}^{M} \overline{l}_m^{coarse} \tilde{u}_m \tag{19}$$

(2) Fine-grained fusion mechanism

This paper proposes a fine-grained preference modeling approach, which further utilizes the user's fine-grained preference information to model the modal interior. Similar to fine-grained preference modal representation modeling, this paper also proposes a fine-grained fusion mechanism.

1) Dynamic multimodal fusion for non-legacy visual representation

Firstly, the importance scores of $\tilde{v}_m$ on different modalities are calculated using the following formula:

$$\left[f_1^{fine}, f_2^{fine}, \ldots, f_M^{fine}\right] = \sigma_5\left(W_m^{fine^2} \cdot \left[\tilde{v}_1 \oplus \tilde{v}_2 \oplus \cdots \oplus \tilde{v}_M\right] + b_m^{fine^2}\right) \tag{20}$$

where M is the number of modes, σ_5 is the activation function, $\oplus$ denotes the tandem operation, $W_m^{fine^2} \in \mathbb{R}^{Md_m \times Md_m}$ is the learnable parameter matrix, $b_m^{fine^2}$ is the bias vector.

Finally, the scores of all modalities are normalized by $Soft\max$ on $A = \left[f_1^{fine}, f_2^{fine}, \ldots, f_M^{fine}\right] \in \mathbb{R}^{d_m \times M}$ to get the score of each modality's attention score, denoted by $\overline{f}_m^{\text{fine}}$. At this point, the final short video representation $v^{coarse} \in \mathbb{R}^{d_m}$ is available:

$$v^{fine} = \sum_{m=1}^{M} \overline{f}_m^{fine} \odot \tilde{v}_m \tag{21}$$

2) Dynamic multimodal fusion of user representations

Firstly, the importance scores of $\tilde{u}_m$ on different modalities are calculated using the following equation:

$$\left[l_1^{fine}, l_2^{fine}, \ldots, l_M^{fine}\right] = \sigma_6\left(W_m^{fine^3} \cdot \left[\tilde{u}_1 \oplus \tilde{u}_2 \oplus \cdots \oplus \tilde{u}_M\right] + b_m^{fine^3}\right) \tag{22}$$

where M is the number of modes, σ_6 is the activation function, $\oplus$ denotes the tandem operation, $W_m^{fine^3} \in \mathbb{R}^{M \times Md_{2m}}$ is the learnable parameter matrix, $b_m^{fine^3}$ is the bias vector.

Finally, the scores of all modalities are normalized by $Soft\max$ for $B = \left[l_1^{fine}, l_2^{fine}, \ldots, l_M^{fine}\right] \in \mathbb{R}^{d_m \times M}$ to obtain each modality's attention score, denoted by $\overline{l}_m^{fine}$. At this point, the final user representation $u^{coarse} \in \mathbb{R}^{d_m}$ is available:

$$u^{fine} = \sum_{m=1}^{M} \bar{l}_m^{fine} \odot \tilde{u}_m \tag{23}$$

3.3.5 Forecasting layer

Finally, in the prediction layer, the inner product is used to calculate the similarity between the user representation and the short video representation, i.e., the user's preference for the target short video. Taking the output of the fine-grained fusion mechanism as an example, by calculating the inner product of the user representation u^{fine} and the short video representation v^{fine}, a prediction score is obtained, which indicates the degree of interaction between user u and the target short video v. This can be used to predict the interaction between each user and each short video. The mathematical expression is as follows:

$$\hat{y}_{(u,v)} = u^{fine^T} v^{fine} \tag{24}$$

4 Experimentation and analysis of the effectiveness of non-legacy visual content recommendations

4.1 Experimental design

(1) Experimental Setup

In order to validate the proposed MMGCN recommendation, this section conducts a large number of experiments on three datasets, the three datasets as well as the detailed statistical information of the proposed multimodal information are shown in Table 1. For the division of the datasets, the interaction data is divided into training set, validation set, and testing set using the leave-one-out method.

Table 1: Statistics of the datasets

Dataset	#User	#Item	#Interaction	Density	#VE	#TE	V	T
Beauty	14672	8523	125748	0.00101	1067	10353	4032	290
Art	24031	9264	191358	0.00083	958	10129	4032	290
Taobao	11748	8633	82571	0.00073	1119	8358	4032	-

(1) Evaluation Metrics

In order to evaluate the model performance, two Hit Ratio (HR) and Normalized Discounted Cumulative Gain (NDCG), which have been widely adopted by the related work on recommender systems, are adopted here. Specifically, HR@k can be interpreted as the average probability that a recommendation list of size k contains a user's favorite item, and the more the hit is at the top of the recommendation list, the higher the NDCG@k score is.

4.2 Model Comparison

In order to validate the effectiveness of the proposed MMGCN recommendation, this subsection compares MHGCF with the following methods, including the classical CF methods: BPRMF, SVD++, NGCF. GCN-based methods: LightGCN, VBPR, MHGCF, MGAT, GRCN, InvRL, DMRL, MEGCF, MGCL.

The performance of all models is shown in Tables 2 and 3. The tables were analyzed by top-down approach. LightGCN using linear GCN module is significantly stronger than NGCF using nonlinear GCN in each dataset of the recommended systems HR and NDCG, which indicates

that the linear GCN module is more effective in capturing higher-order interactions.The MEGCF, MGCL, and MMGCN models perform outstandingly, and all of them significantly outperform the two models, BPRMF and SVD++, where the proposed MMGCN has excellent recommendation performance in each dataset, and the recommendation effect is improved by about 10% to 20%, which demonstrates the effectiveness of MMGCN and is beneficial to enhance multimodal user preference modeling.

Table 2: Overall performance comparison of all methods w.r.t. HR

Model	Beauty			Art			Taobao		
	HR@5	HR@10	HR@20	HR@5	HR@10	HR@20	HR@5	HR@10	HR@20
BPRMF	0.5024	0.5923	0.6981	0.7083	0.7702	0.7629	0.3965	0.4799	0.5905
SVD++	0.5334	0.6273	0.7409	0.7281	0.8175	0.7935	0.4124	0.5043	0.6216
NGCF	0.5603	0.6571	0.7561	0.7492	0.8291	0.8237	0.4325	0.5343	0.6591
LightGCN	0.5752	0.6813	0.7928	0.7564	0.8389	0.8379	0.4598	0.5643	0.6987
VBPR	0.5472	0.6422	0.7415	0.7449	0.8214	0.8252	0.4214	0.5113	0.6262
MHGCF	0.5684	0.6817	0.7866	0.7519	0.8452	0.9296	0.4399	0.5445	0.6652
MGAT	0.5763	0.6902	0.7931	0.7575	0.8449	0.9307	0.4533	0.5632	0.6948
GRCN	0.5837	0.6954	0.7991	0.7655	0.8493	0.9282	0.4615	0.5746	0.7125
InvRL	0.5883	0.6847	0.7887	0.7715	0.8498	0.929	0.4663	0.5647	0.7098
DMRL	0.5986	0.6965	0.8048	0.7722	0.8525	0.9317	0.4507	0.5967	0.7199
MEGCF	0.6189	0.7214	0.8198	0.7866	0.8652	0.9401	0.5295	0.5962	0.7266
MGCL	0.6314	0.7342	0.8272	0.7845	0.8655	0.9417	0.4856	0.5967	0.7278
MMGCN	0.6427	0.7455	0.8426	0.7924	0.8731	0.9555	0.4921	0.6091	0.7305

Table 3: Overall performance comparison of all methods w.r.t. NDCG

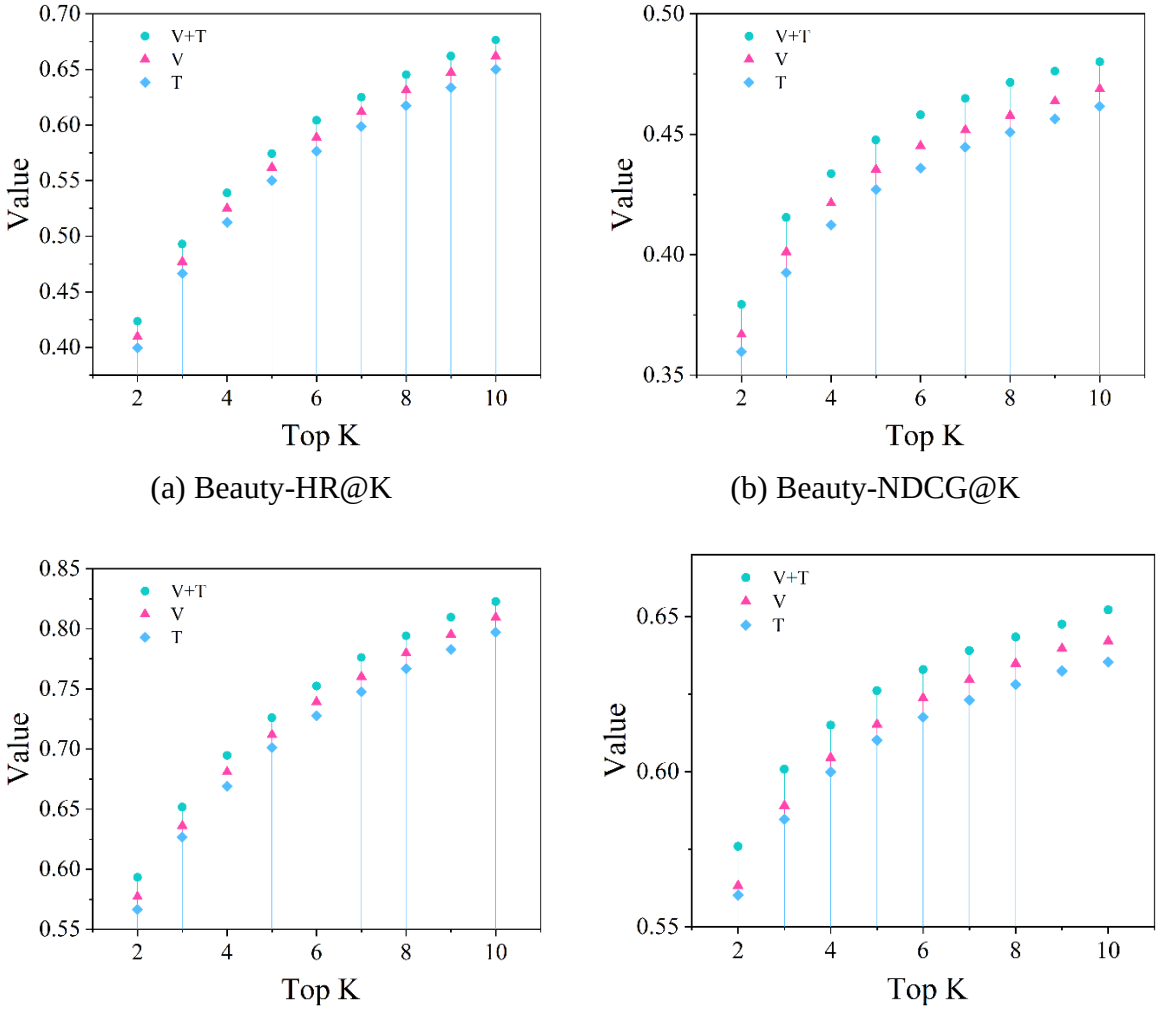
Model	Beauty			Art			Taobao		
	ND@5	ND@10	ND@20	ND@5	ND@10	ND@20	ND@5	ND@10	ND@20
BPRMF	0.4093	0.4384	0.4652	0.6347	0.6579	0.6775	0.3215	0.3483	0.3011
SVD++	0.4342	0.4645	0.4907	0.6377	0.6666	0.6884	0.3273	0.3569	0.3114
NGCF	0.4526	0.4839	0.5089	0.6632	0.6891	0.7082	0.3408	0.3736	0.3301
LightGCN	0.4557	0.4902	0.5185	0.6636	0.6903	0.7091	0.3591	0.3926	0.3515
VBPR	0.4415	0.4723	0.4974	0.658	0.6828	0.7034	0.3389	0.3678	0.3216
MHGCF	0.4464	0.4831	0.5109	0.6393	0.6695	0.6909	0.3459	0.3797	0.3216
MGAT	0.4531	0.4902	0.5181	0.6534	0.6817	0.7024	0.3574	0.3925	0.3509
GRCN	0.466	0.5022	0.5303	0.6687	0.6958	0.7157	0.3611	0.3975	0.3573
InvRL	0.4705	0.5018	0.5299	0.6736	0.6987	0.7175	0.3676	0.3994	0.3582
DMRL	0.4825	0.5146	0.5392	0.6758	0.7022	0.7225	0.3626	0.3896	0.3523
MEGCF	0.5007	0.5342	0.5588	0.6894	0.7148	0.7338	0.3717	0.4147	0.3726
MGCL	0.5072	0.5406	0.5592	0.6871	0.7109	0.7328	0.3814	0.4173	0.3795
MMGCN	0.5204	0.5538	0.5751	0.6987	0.7248	0.7498	0.3928	0.4246	0.3867

4.3 Modal analysis

(1) Effect of depth features across modalities

The results on the Beauty and Art datasets with respect to {HR,NDCG}@{1-10} are shown in Fig. 4, where the model variant V(T) is based on the removal of textual (visual) features by MMGCN in modeling content-level preferences. Note that no relevant experiments are performed here on the Taobao dataset, due to the fact that the Taobao dataset only provides deep

features for the visual modality (lacking deep features for the textual modality). The experimental results are shown in Fig. 4. According to the results in Fig. 4, firstly, V+T: MMGCN always outperforms the variant that only uses one modality, which indicates that the distribution of user preferences across modalities is significantly different, i.e., different modalities contribute differently to modeling user preferences, and thus capturing user preferences across multiple modalities at the same time can help to achieve better recommendation accuracies.



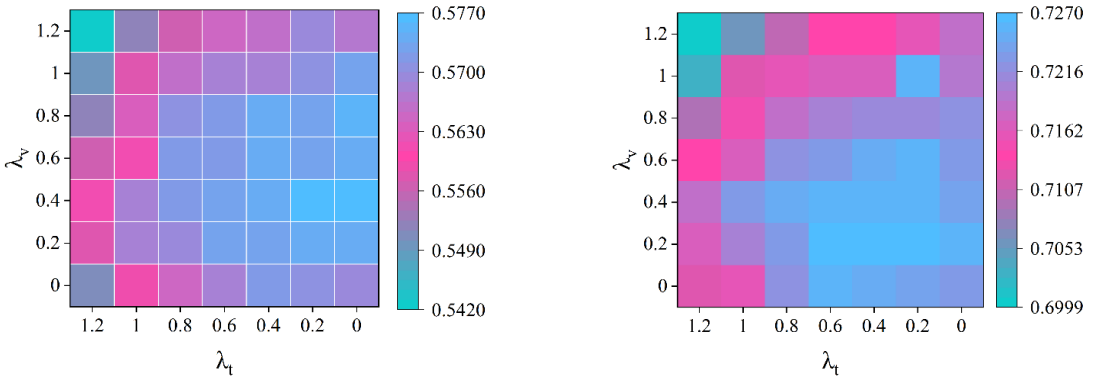
(a) Beauty-HR@K (b) Beauty-NDCG@K

(c) Art-HR@K (d) Art-NDCG@K

Figure 4: Effect of different modalities on model performance

(2) The multimodal preference weights $\boldsymbol{\lambda m}$ of the

Figure 5 illustrates the plots of model performance on Beauty and Art datasets with respect to λ_v and λ_t, and the optimal λ_v and λ_t on the Beauty dataset are 0.4 and 1, respectively. And the most λ_v and λ_t on the Art dataset are 0.2 and 1, respectively.These results again show that the textual modality is significantly more tributary than the visual modality for modeling user preferences. Therefore, when the proposed MMGCN is applied to a new dataset, it is clearly a better choice to assign higher weights to textual modal features and lower weights to visual modal features.



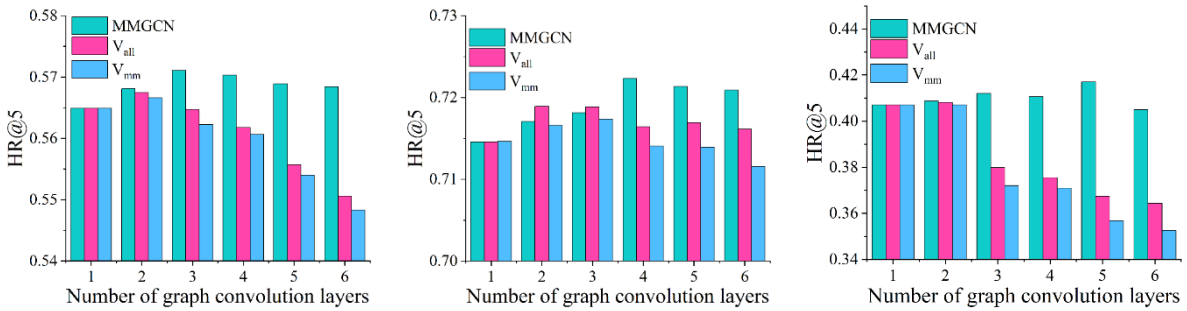
(a) Beauty-HR@5 (b) Art-HR@5

Figure 5: Effect of multimodal preference weights on model performance

4.4 Model depth analysis

For the MMGCN model, the number of layers of the propagated multimodal depth features is fixed to 1, and the effect of the depth of the GCN for the propagated ID embedding on the recommendation accuracy is investigated here. In addition, in order to further investigate whether reducing the depth of the GCN module that propagates the multimodal features can effectively reduce the noise pollution problem, two additional variants are set up here: V_{all} is the variant that keeps the depth of all the GCN modules varying consistently; V_{mm} is the variant that varies the depth of the GCN module that propagates only the multimodal features, while the depth of other GCN module depths remain unchanged. The model depth is varied in the range {1,2,3,4,5,6}. The performance of the three methods on the three datasets with respect to the model depth is shown in Fig. 6.

MMGCN shows a gradual increase in model performance as the number of graph convolution layers increases, suggesting that extending the interaction relationships to higher orders better models user preferences, and the optimal graph convolution layers on the three datasets are 3, 4, and 5, respectively, which may be due to the fact that the sparser the dataset is, the more graph convolution operations are required to ensure that sufficient synergistic relationships are captured. The fact that V_{all} always slightly outperforms V_{mm} while being significantly weaker than MMGCN suggests that the over-propagation of multimodal noise not only degrades the user preference modeling at the content level, but also dominates the recommendation results, which severely degrades the overall user preference modeling. Therefore, it is essential to set a smaller GCN depth when performing graph convolution operations on multimodal depth features.



(a) Beauty (b) Art (c) Taobao

Figure 6: Effect of different number of graph convolution layers on model performance

4.5 Visualization experiments

The Beauty dataset is selected for visualization and analysis, and Fig. 7 and Fig. 8 show the T-SNE visualization results of the images on the Beauty dataset and the text features on the Art dataset, respectively. Specifically, firstly, samples with four category labels: airplane, boat, cat and dog are randomly selected from the Beauty test set and fed into the proposed MMGCN model to obtain features. The high-dimensional features of the samples are transformed into two-dimensional features by T-SNE, and different colors are used to distinguish the categories of the samples. For each of the four categories, a set of corresponding images and texts are shown. It can be found that in Fig. 7 and Fig. 8, the dots of each color are clustered in a region, which indicates that the model has learned the discriminative information of the samples belonging to different categories.

Comparing the points of the same color in Figures 7 and 8, it is found that they appear in similar regions, which indicates that the model has achieved cross-modal semantic alignment. The distribution of yellow and green points is closer because the categories they belong to, cat and dog, are more similar. In addition, there are also some points that appear in the wrong regions, such as airplanes appearing in the region of ships. There are two main reasons for this situation, one is that airplanes and ships often appear in similar scenes, such as blue sky and sea, which leads to errors in model judgments; the other is that the selected samples may contain multiple types of objects, so a small number of mixed-color regions appear in this experiment. Overall, the proposed MMGCN model achieves the best retrieval results on both Beauty and Art datasets.

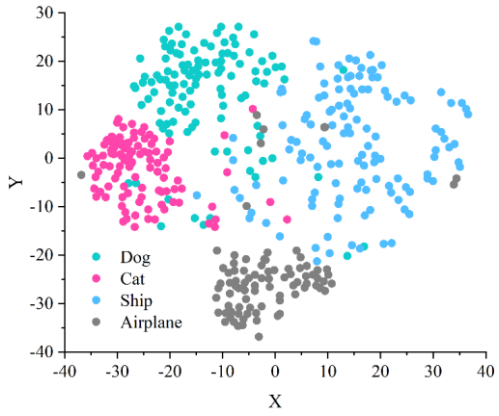


Figure 7: T-SNE visualization of image features

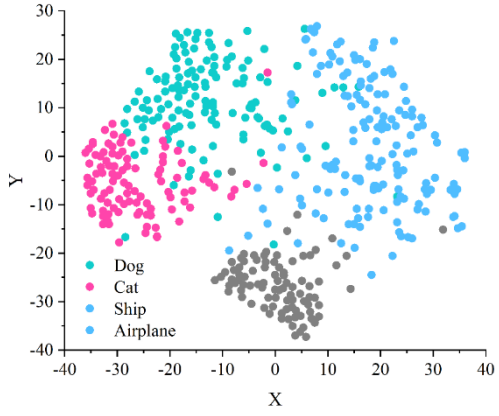


Figure 8: T-SNE visualization of text features

5 Effectiveness of dissemination of visual content on intangible cultural heritage

5.1 Selection of Indicators for Measuring the Communication Power of Non-Heritage Visual Content

5.1.1 Identification of Candidate Measurement Indicators and Quantification Methods

The measurement of the communication effect of non-heritage short videos needs to be based on a reasonable scientific and objective evaluation system. Combining the short video communication effect evaluation indexes available in the current existing research and the communication characteristics of non-heritage short videos, this paper initially selects 21 characterization indexes from three dimensions, such as examples of non-heritage, concepts of non-heritage, duration, number of words in title, and number of retweets.

The content characterization degree indicators of non-heritage short videos include non-heritage examples, non-heritage concepts, non-heritage events, non-heritage facts and non-heritage content audience. In this paper, we will analyze the text of the title introductions of NRM short videos, record the representation of the four granularities of NRM instances, NRM concepts, NRM events and NRM facts in the text, and count the number of times that the NRM instances, concepts and events appear. It should be emphasized that the types of non-heritage facts are mainly divided into 3 categories: performance, teaching clip, and dissemination and application demonstration, which can be represented by 1-performance, 2-teaching clip, and 3-dissemination and application demonstration.

Among the indicators of non-heritage short video dissemination effect characteristics, the indicators on the part of the information publisher mainly include the number of fans, the total number of likes, and the number of dynamics and the number of works, but in the Jieyin platform, the number of works is included in the number of dynamics, and the number of works refers to the user's own original works, while the number of dynamics of many users is the number of works, so these two indicators will be examined in the screening of the indicators in the following section.

5.1.2 Screening of metrics

The construction of a scientific and reasonable indicator system is an important premise and foundation for the communication power model of non-legacy short videos. In order to ensure the scientificity, rationality and representativeness of the selection of indicators in the communication power model of non-legacy short videos, this paper utilizes the minimum mean square deviation method to screen the preliminary selection of communication power indicators. The minimum mean square deviation method is an indicator analysis and screening method based on the degree of differentiation, the degree of differentiation indicates the size of the difference between the indicators, and the larger the degree of differentiation, the larger the difference between the indicators, and the more representative the indicators are. The mean square deviation represents the degree of dispersion between individuals in the group, in the minimum mean square deviation method, if the observed values under a characteristic indicator are approximately equal, even if the indicator is operable, it has no practical role and significance for the overall measurement model, so it can be sifted out by the method of the minimum mean square deviation, whose formulas are as shown in Eqs. (25) and (26):

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \tag{25}$$

$$s_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2}\ \left(j = 1, 2, \cdots, m\right) \tag{26}$$

Find the minimum mean square deviation $s_{j0} = \min_{1 \le j \le m}\{s_j\}$, if the minimum mean square deviation $s_{j0} \approx 0$, the feature indicator x_{j0} corresponding to s_{j0} can be deleted.

By calculating the mean square deviation of the standardized data of the characteristic indicators of the communication power model of non-legacy short videos, the minimum mean square deviation of the characteristic indicators is not equal to 0 and is greater than 0.05, and all indicators can be retained. Based on the above method, the indicators of the communication power measurement model were analyzed and screened, and the duplicated indicators were removed, and finally 19 feature indicators were identified from the three dimensions of content, structure and communication effect characteristics of non-legacy short videos.

5.2 Assignment of Indicators for Measuring the Communication Power of Non-Heritage Visual Content

In this paper, the entropy weight method is used to determine the weight of each indicator in the communication power measurement model of non-heritage short videos. Firstly, the original data of non-legacy short videos are collected, and the data are sorted and cleaned, with a total of 630 valid data and 19 indicators, and a $630*19$ indicator matrix is established; secondly, the data are standardized and the original matrix is standardized by the extreme value method:

$$Y_{ij} = \frac{X_{ij} - \min(X_i)}{\max(X_i) - \min(X_i)} \tag{27}$$

Information entropy was calculated for each indicator as shown in equation (28):

$$E_j = -\ln(n)^{-1}\sum_{i=1}^{n} p_{ij} \ln p_{ij} \tag{28}$$

where $p_{ij} = \dfrac{Y_{ij}}{\sum_{i=1}^{n} Y_{ij}}$ if $p_{ij} = 0$, then $\lim_{p_{ij} \to 0} p_{ij} \ln p_{ij} = 0$.

Finally, the weight of each indicator is determined, and on the basis of the information entropy obtained from the calculation, the weight value of each indicator is finally determined according to the size of the information entropy of each indicator.

$$W_i = \frac{1 - E_i}{k - \sum E_i}\left(i = 1, 2, \cdots, k\right) \tag{29}$$

By cleaning and organizing the data of the collected short videos of non-heritage, standardizing the raw data of MMGCN short videos, adopting the entropy value method to calculate the entropy value of each characteristic index, and realizing the assignment of the indexes, the calculation results are shown in Table 4.

It can be seen from the weighting results of feature indicators that the proportion of the subject latitude is the highest, at 64.64%, the proportion of the object latitude is 22.09%, and the proportion of the content feature structure is 13.27%. In the dimension of content features, specific examples of intangible cultural heritage and related concepts contribute relatively less to the dissemination power, but the inclusion of specific events related to intangible cultural

heritage has a greater impact on the dissemination power. In the structural feature dimension, the weight of duration is 3.13%, the weight of presence or absence of subtitles is 3.89%, and the weight of presence or absence of @ numbers is 9.09%. This is because the duration of short videos can reflect the amount of information they contain. The longer the content duration, the richer the information it contains, the more effective information it has, and the easier it is for users to accept and understand, and its dissemination effect is also better.

Table 4: Weight of characteristic indicators

Primary indicator	Weight	Secondary indicator	Weight
Content Feature Dimension	0.1327	Examples of intangible cultural heritage	0.0095
		The Concept of Intangible Cultural Heritage	0.0273
		Intangible Cultural Heritage Event	0.0379
		The audience for intangible cultural heritage content	0.0580
Object latitude	0.2209	Duration	0.0313
		Title word count	0.0087
		Introduction word count	0.0026
		With or without letters	0.0389
		Barrage	0.0213
		Voice match	0.0125
		Topic count	0.0147
		Have or Not @	0.0909
Principal latitude	0.6464	Forward number	0.1057
		Average number	0.0921
		Like number	0.0987
		Number of fans	0.0759
		Total number of praises	0.0894
		Dynamic number	0.0311
		Emotional Index	0.1535

5.3 Evaluation of the Communication Effect of Non-Heritage Visual Content-Taking Chu Opera as an Example

5.3.1 Distribution of video dissemination effects of Chu dramas

The distribution of Chu opera video dissemination effect score is analyzed, as shown in Table 5, the mean value of the dissemination effect of Chu opera video is 5.53, the minimum value is 0.51, the maximum value is 71.63, the standard deviation is 7.24, and there is a great difference in the extreme value of the dissemination effect of Chu opera video. Moreover, the percentile (75) of the dissemination effect of Chu opera videos is 6.38, which means that three-quarters of the dissemination effect of Chu opera videos on the dissemination medium is less than 6.38, and most of the videos have a negligible dissemination effect.

Table 5: Analysis of the Score of the Effect of Video Dissemination of Chu Opera

	Mean	SD	Least value	Crest value	Centile (25)	Centile (50)	Centile (75)
Transmissibility score	5.53	7.24	0.51	71.63	1.94	3.67	6.38

5.3.2 Analysis of the results of the video dissemination effect of Chu opera

In order to better understand the dissemination effect of Chu opera in the new media environment, this paper calculates the scores of each Chu opera video in the dimensions of dissemination content, dissemination subject, and dissemination object, as well as the overall scores, respectively, by multiplying the standardized data with the weights of each index. Among the TOP20 Chu opera video dissemination effect scores are shown in Table 6.

As can be seen from the Top20 scores of Chu opera video dissemination effect in the table, the maximum value of the current dissemination effect is 71.9023, which means that the dissemination effect is better and representative of Chu opera video in the dissemination medium, and it can continue to play a leading role in the dissemination afterward. Among the three dimensions of dissemination effect, the mean value of the score of the dissemination subject dimension of Chu opera is 19.0696, which is generally high, indicating that in the dissemination of Chinese cultural heritage videos, the number of followers, the total number of likes, and the total number of broadcasts of the dissemination subject have a greater impact on the dissemination effect of the video. In the communication object dimension score of Chu opera videos, the maximum value is 55.0068, and the minimum value is 0.9723, with a large standard deviation, indicating that even among the Chu opera videos with top20 communication effect scores, their communication objects are more different. And the mean value is 9.0313, indicating that the dissemination process of the dissemination object intangible cultural heritage video is not actively responded to and publicized. The communication content dimension of Chu opera video, the minimum value is 1.3897, the maximum value is 17.7688, and the mean value is 6.4635, which shows that even for the Top20 Chu opera videos, some of them are not complete, and it is still necessary to further improve the content description of the video, and a clear and complete description of the content of the Chu opera video facilitates the dissemination of information and the user's understanding.

Table 6: Top 20 Videos of Chu Opera with the Best Communication Effect

Number	Content characteristics	Principal latitude	Object latitude	Communication effect	Ranking
1	16.7844	0.1768	55.0068	71.9023	1
68	17.7688	23.915	2.888	42.9338	2
2	1.3897	4.113	37.5209	41.384	3
291	11.8716	23.915	3.9981	38.1467	4
55	3.9208	23.915	9.2172	35.4151	5
4	2.822	23.915	10.2972	35.3878	6
566	7.233	0.5162	25.4057	33.0464	7
446	6.7386	23.915	3.6284	32.644	8
97	3.8386	23.915	4.327	30.4426	9
205	6.1384	23.915	1.774	30.1894	10
99	5.2742	23.915	2.1551	29.7062	11
3	4.0831	23.915	3.0888	29.4489	12
52	3.0253	23.915	3.9423	29.2438	13
221	3.2286	23.915	1.9259	27.4221	14
253	2.822	23.915	1.5413	26.6403	15
160	8.3508	16.2425	3.4688	26.4242	16
34	2.6508	23.915	1.3842	26.312	17
104	2.6186	23.915	1.3938	26.2894	18
155	1.6433	23.915	0.9723	25.8599	19
21	17.0665	1.6183	6.6908	4.7376	20

Comparing the overall scores of the three dimensions, the scores of the current Chu opera video dissemination object characteristics dimension are unsatisfactory, and, the standard deviation of the dissemination object dimension is the largest, the number of times watched, the number of coins and the number of favorites, reflecting the degree of user recognition and acceptance of the video content, the dissemination object can be interacted with by liking, sending pop-ups, forwarding and posting comments, and the comment sentiment index reflects the audience's attitudes after watching the video. The great difference in the dissemination effect of videos in the category of intangible cultural heritage shows that there is a big gap between the production capacity and quality of the current Chu opera video programs, the video quality is generally low, the number of high-quality videos is not large, and the content production capacity and quality of Chu opera still need to be improved.

6 Conclusion

Based on the analysis of the content, subject and object of intangible cultural heritage from the perspective of “5W” communication model, combined with the proposed MMGCN recommendation model of preference perception, the effectiveness of MMGCN recommendation model is verified through experimental comparison and visualization analysis. The MMGCN-based short video recommendation model of intangible cultural heritage is taken as the research object, and the MMGCN-based short video communication power measurement model of intangible cultural heritage is constructed from the three dimensions of intangible cultural heritage short video content, subject and object, and the Chu opera is taken as an example to evaluate the communication effect. The research conclusions are as follows:

(1) The MMGCN recommendation model outperforms other models on the two recommended systems HR and NDCG, and the recommendation effect is improved by about 10%~20% compared with other models, which verifies the effectiveness of the model's recommendation, and achieves the best retrieval effect on both Beauty and Art datasets.

(2) As can be seen from the Top20 scores of the dissemination effect of Chu opera videos, the maximum value of the dissemination effect is 71.9023, and the dissemination effect is better. In the dimension of dissemination content, the minimum value is 1.3897, the maximum value is 17.7688, and the average value is 1.3897, the content of the video is not complete, and it is necessary to further improve the content description of the video, and a clear and complete description of the content of the Chu Opera video facilitates the dissemination of information and the understanding of users.

About the Author

Zhang Juanning received her M.A. degree in Media and Communication from Monash University, Melbourne, Australia, in 2025. She was born in 2001 in Shuozhou, Shanxi Province, China. Her research interests mainly focus on digital media culture, platform content governance, and visual communication. She is currently preparing for further academic research in these areas.

References

[1] Heredia-Carroza, J., Palma Martos, L., & Aguado, L. F. (2021). How to measure intangible cultural heritage value? The case of flamenco in Spain. Empirical Studies of the Arts, 39(2), 149-170.

[2] O'Neill, S. (2025). Extinctions: Language Death, Intangible Cultural Heritage, and Early 21st-Century Renewal Efforts. In Oxford Research Encyclopedia of Communication.

[3] Li, N., Li, X., & Xiao, W. (2025). Intangible cultural heritage threatened-level categories and criteria. Humanities and Social Sciences Communications, 12(1), 1-11.

[4] Maldonado-Erazo, C. P., Tierra-Tierra, N. P., del Río-Rama, M. D. L. C., & Álvarez-García, J. (2021). Safeguarding intangible cultural heritage: the Amazonian Kichwa people. Land, 10(12), 1395.

[5] Septiyana, I., & Margiansyah, D. (2018, December). Glocalization of intangible cultural heritage: strengthening preservation of Indonesia's endangered languages in globalized world. In International Conference on Contemporary Social and Political Affairs (IcoCSPA 2017) (pp. 85-88). Atlantis Press.

[6] Li, Y. (2022). Characteristics of Intangible Cultural Heritage Communication in the New Media Environment. Cultura: International Journal of Philosophy of Culture and Axiology, 19(1).

[7] Mezzino, D. (2023). Digital visualization for cultural dissemination. SCIRES-IT, 135-152.

[8] Shen, Y. (2024, June). Cross-media digital form design and comparative measurement of intangible cultural heritage in Zhuhai: Promoting cultural exchange and global communication. In International Conference on Human-Computer Interaction (pp. 90-104). Cham: Springer Nature Switzerland.

[9] Yuxiao, L., & Daud, K. A. M. (2025). Examining The Influence of Visual Aesthetics in Digital Storytelling on Audience Engagement for The Promotion of Intangible Cultural Heritage on Social Media. International Journal of Creative Future and Heritage (TENIAT), 13(2), 143-155.

[10] Sun, D. (2025). The application of new media technologies in intangible cultural heritage preservation: a case study of Xinjiang Uygur Muqam art. International Journal of Web Services Research (IJWSR), 22(1), 1-23.

[11] Paquienseguy, F., & Guo, Q. (2025). Douyin and the digital spread of intangible cultural heritage: Transforming cultural dissemination in the short videos age. Emerging Media, 27523543251344976.

[12] Liu, R., & Qiu, H. (2025). Ideological presentations in official promotion of intangible cultural heritage (ICH) on short-form video platforms: a multimodal content analysis. Information Research an international electronic journal, 30(iConf), 413-424.

[13] Konvit, M. (2022). New media and intangible cultural heritage: challenges and opportunities. INTANGIBLE CULTURAL HERITAGE AND DIGITAL MEDIA, 12.