International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-16

https://doi.org/10.70917/ijcisim-2026-0381
Article

Modeling Multicultural Interaction in Silk Road Ethnomusicology Education Based on Deep Learning Framework

Xue Zhao [1,*] and Dan Shen [1]

[1] School of Music and Dance, Harbin University, Harbin, Heilongjiang, 150086, China
* Correspondence author: S123456ddd123@163.com

Abstract: The countries along the Silk Road are rich in music resources, and the unique multicultural interaction of ethnic music education is an important trend to carry out the living heritage and innovative development of ethnic music culture. In this paper, based on analyzing the dynamic interaction between technology and ethnic music culture, a Trans-GAN model is constructed by combining Transformer Network and GAN model to generate diversified ethnic music. Using the generated diversified ethnic music as teaching resources, we established an ethnic music teaching model by combining the concept of realm pulse. A teaching experiment was designed with the teaching model to verify the feasibility of the model to enhance the multicultural interaction level of ethnic music with the students majoring in ethnic music in G Nationality University as the research object. The results show that the deep learning model can generate more diversified ethnic music of the Silk Road, and the students' ethnic music performance and cultural interaction level can be significantly improved. The integration of deep learning technology into folk music education to realize the diversification of teaching resources provides technical support to help students enhance the level of multicultural interaction, and also provides a new path for the inheritance and innovation of Silk Road folk music.

Keywords: Transformer network; GAN model; deep learning; Silk Road; ethnic music

1. Introduction

Nowadays, the terms of multicultural music and cultural pluralism have a high frequency of use, which also means that the focus of people's research is gradually expanding, and people are paying more and more attention to the value change, content improvement and style change of folk music, looking at the development of music from the perspective of the development of the times, perceiving the richness and colorfulness of the culture, and promoting the development of the cause of folk music [1-4]. Based on this, multiculturalism and cultural pluralism in the study of folk music is not only on the surface of the theoretical discussion, more and more scholars have launched the activities of the practice, with the help of multicultural music to enrich people's spiritual world, so that the music has a substantial significance of the inquiry [5-8].

The “Silk Road” began in the Western Han Dynasty, running through the East and West, connecting China, Central Asia, Jiangnan Asia, the Middle East, the Mediterranean coast and other places. In the long-term exchanges, the musical cultures of different countries and regions mingled and influenced each other [9-11]. For example, Chinese guqin, ruan, flute and other musical instruments, as well as the Chinese meter, in the East to Central Asia, South Asia, West Asia and other regions, affecting the development of local national music culture. And instruments such as the guitar in Central Asia, the drop zither in South Asia, and the harp in the Mediterranean coastal region were also introduced to

China through the Silk Road, enriching the musical instrumentation and musical styles of Chinese music [12-15]. In addition, the commerce and trade activities of the Silk Road also promoted the exchange of ethnic music and culture. Merchants, travelers, scholars and missionaries came and went from all over the world, bringing the music of their respective countries to foreign lands, colliding and fusing with local cultures, and creating new forms of music. In this process, Eastern and Western musical cultures borrowed from and influenced each other, forming a unique musical and cultural phenomenon [16-19].

Ethnic music melts the hearts of the people on the Silk Road, promotes mutual cultural appreciation among countries, enhances mutual cultural identity and mutual trust and respect, and joins hands to compose the colorful music of the new Silk Road that manifests the spirit of the Silk Road. In this paper, we first sort out the new paradigm of national music education for the Silk Road, and explore the dynamic interaction mechanism between technology and national music culture. Then, a Trans-GAN model is constructed based on the Transformer Network and combined with the Generative Adversarial Network to generate the ethnic music of the Silk Road. Finally, the model-generated ethnomusic is used as a teaching resource, and an ethnomusic teaching model is constructed by combining the concept of realm pulse, and students of G University of Nationalities are selected to conduct teaching experiments, which illustrates the feasibility of the deep learning-supported teaching model to improve the level of students' multicultural interaction in ethnomusic.

2. Ethnic Music Education for the Silk Road

Under the historical condition of building the “Silk Road Economic Belt”, the interaction and exchange of national music culture is facing the double responsibility of inheritance and innovation, and under the situation of globalization, the interaction and exchange of national music culture is more unstoppable, which is an important means of cooperation and exchange between groups, nationalities and countries. Ethnic music education should seize the opportunity of the times, continuously improve the cooperation ability of Chinese ethnic music culture, infiltrate the highly charming and characteristic Chinese ethnic music culture into the economic construction of the Silk Road, and promote the prosperity of the world music culture.

2.1. Silk Road Music Culture

The Silk Road is an east-west transportation route connecting China with River China and India, mediated by the economic transactions of the Silk Road. In the vast area radiated by this civilization road through Eurasia, it contains an extremely wide range of political, economic and cultural exchanges of significance, which formed the growth conditions of the western and eastern cultures of China and the West, making the Silk Road culture one of the important sources of Chinese culture [20].

Through the history and reality of multi-directional exchanges, people deeply perceive the Silk Road across Europe and Asia so that the musical culture of various ethnic groups interdependent on each other, forming a whole with a variety of intrinsic links, realizing the academic significance of the chiseling, the formation of the world's cultural pluralism and integration of the pattern. The growth of regional music culture along the Silk Road cannot be separated from the projection, fusion and nourishment of neighboring national music culture. Music as the most important “catalyst” in the road, the history of the development of music culture of various ethnic groups, but also directly led to the development of Chinese and Western culture. The civilization of musical instruments imported along the Silk Road also added color to the prosperity of Chinese civilization.

2.2. Ethnic music cultural interaction

As a borderless art form, the spread and exchange of music along the Silk Road was particularly remarkable. Unlike material cultural exchanges such as silk, porcelain and tea, the spread of folk music relied more on human interaction. The travelers, merchants, emissaries and cultural messengers of the time not only carried their respective musical instruments and musical skills on their long journeys, but also carried their love and respect for music. Different cultures and nationalities have their own unique musical languages, but the emotions and charms of music are universal and can cross the barriers of language and geography to touch the hearts of people. Therefore, in the context of globalization, music has become a bridge for cultural exchanges among countries, enabling the spread and integration of different musical styles, techniques and ideas [21].

In the future, with the further development of the Internet and digital technology, the dissemination and communication of music will be more convenient and instant. Both Eastern classical music and Western popular music can be quickly spread around the world, attracting listeners from different cultural backgrounds. This not only provides a broader stage for musicians and composers, but also

brings a richer and more varied experience of national music to ordinary listeners. In this kind of exchange, different folk music cultures will influence and inspire each other, giving birth to brand new styles and forms of folk music.

2.3. New paradigm in folk music education

According to the needs of the Silk Road construction and the requirements of the "One Belt, One Road" education initiative issued by the Ministry of Education, the Silk Road Ethnic Music Education Charter will be established to give full play to the resource advantages of the ethnic music academics to promote the construction of the Silk Road, and transform it into a long-term mechanism for the establishment of the Silk Road Ethnic Music Talent Cultivation, Academic Research, Music Creation and Performance as well as International Music Education and Exchange Activities. It will be transformed into a long-term mechanism for the cultivation of national music talents, academic research, music creation and performance, as well as international music education and exchange activities.

The countries along the Silk Road are rich in music resources, they all have their own unique national music culture, excellent composers, and a large number of national music and works are the resources and objects of our research. Through in-depth research, we can promote our mutual understanding and knowledge, mutual appreciation of various cultures, and cultural identity and mutual respect among each other. To use tradition for reality, and to lead the cause of folk music and folk music education into the future in the connection between tradition and reality.

From the perspective of folk music education, it is necessary to establish a world view of music, that is, to shift from the national, and the past with the Western music as the center, to the present global and world view of music. Through the platform of the Music Union, we will establish a world view of music from a global perspective, hold the Silk Road International Music Festival and organize regional music weeks, masters' workshops, world music training courses, etc., to bring together the excellent national music cultures of the countries along the Silk Road, and conduct in-depth research on the mode of national music education along the Silk Road. To build a world ethnomusicology discipline, to highlight the world's colorful ethnomusicology culture, and to enable the people of the world to appreciate the exquisite ethnomusicology culture of each country.

3. Ethnic Music Generation Model under Deep Learning Framework

Folk music on the Silk Road is characterized by diversity and requires attention and active inheritance. Ethnic music on the Silk Road represents an important stage in the development of ethnic music in various countries, and determines the direction and depth of contemporary ethnic music development. Music is an essential spiritual substance and spiritual support in our life, and the study of ethnic music cultural interaction on the Silk Road has an important value and role in improving the quality of life and promoting social development. The diversity and richness of ethnic music reflect the long social and cultural accumulation of various historical periods, ethnic groups and regions, which show obvious characteristics in terms of relativity, stability and integration. Based on this, this paper introduces a deep learning framework to establish an ethnomusic generative model, which aims to realize the two-way interaction between technology and culture, and to provide new opportunities for the expansion and dissemination of ethnomusic.

3.1. Interaction between technology and ethnomusicological culture

3.1.1. Technology influences mechanisms of cultural expression

(1) Mechanisms of Technological Influence on Ethnic Music Cultural Expression

Technological progress has provided new ways and means for ethnic music cultural expression. In the field of ethnic music, the emergence of deep learning generation technology has made ethnic music creation no longer limited to the traditional human creation mode. Technology has expanded the boundaries of ethnomusicological art and culture by providing new musical languages and forms of expression. In addition, technology has changed the way ethnic music is disseminated, enabling the melodies of ethnic music culture to be rapidly spread to a wider audience through digital media.

(2) The Path of Ethnic Music Culture Shaping the Development of Technology

Ethnic music cultural needs and values have a guiding role in the development of technology. In the field of ethnomusicology education, artificial intelligence is reconstructing the education model. However, there is no doubt that machines can never replace the influence of teachers in the spiritual world of students, nor can data, algorithms and application modules provide all the rational tools for the

realization of precise education. In the process of generating folk music melodies, the requirement of cultural inheritance prompts deep learning technology to not only pursue technological innovation, but also focus on the depth and accuracy of folk music cultural content. The values and historical significance of folk music culture need to be properly reflected and respected in the creation.

3.1.2. Dynamic interaction between technology and culture

Empowering AI for artistic creation, digitally programming the laws of human music creation, and using massive data as learning resources for intelligences, is an abstract description of AI's symbolic and empirical description of human intelligent cognition. The mathematical logic of music creation is presented in the quantization of physical vibrational sound waves, and music has a natural translatable relationship with computational science. Of all art forms, the most vulnerable to the impact of science and technology may be music. This is because both the input and output of music lend themselves to precise mathematical descriptions; the input is the mathematical pattern of sound waves, and the output is the electrochemical reaction pattern of a neural storm. Algorithms might learn how to predict that a certain input produces a certain output just by going through millions of musical experiences. At the same time, it is also possible to develop a machine's auditory memory and emotional computation through deep machine learning, resulting in somewhat relatively independent programming templates for AI music melody writing and arranging efforts. Figure 1 shows the dynamic interactive process of technology and music culture, which makes full use of technology to mine the cultural connotations of national music as a way to realize technological empathy.

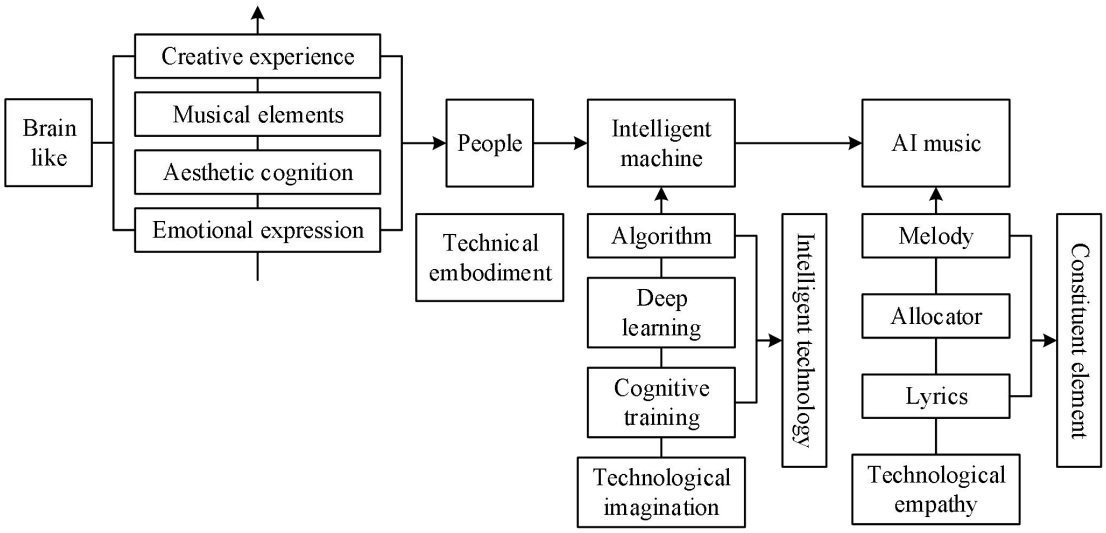


Figure 1. The dynamic interaction between technology and music culture

However, no matter the algorithmic technology or deep learning, it is all based on the music created by human beings as the basic template, to carry out a large amount of analysis, extraction and learning, and ultimately return to the interpretation of human behavior by artificial intelligence music, and return to the original point of technical embodiment with human subjects. From the perspective of technical phenomenology and technical philosophy, the development path of AI music research can be roughly focused on the three aspects of technical embodiment, human-computer interaction imagination and intelligent empathy, technical embodiment is the foundation of AI music generation, human-computer interaction imagination is the trend of AI music development, and intelligent empathy is the driving force for the continuous amplification of the value of technically created music.

3.2. Generative Modeling of Multicultural Ethnomusicology

3.2.1. Transformer network

Transformer is a neural network architecture based on the attention mechanism, Transformer has the advantage of parallel computing to process sequence data more efficiently.The encoder-decoder structure of the Transformer model is very modular, and the model consists of two parts, a multilayer encoder and a decoder, and the module of each layer contains a multi-head self-attention mechanism and a feed-forward neural network, each layer has the same structure but different parameters, and the

decoder has an encoding-decoding attention layer in between the two [22].

Transformer uses a self-attention mechanism that allows the network to focus on different locations in the input sequence while processing it. This mechanism allows the network to capture the dependencies between different elements in the input sequence without having to process the sequences sequentially as in traditional recurrent neural networks, which are not restricted to a fixed local context window.

For the input sequence $X=[x_1,\ldots,x_N]\in R^{D_x\times N}$, respectively, the dot product operation with the weight matrix $W_q\in R^{D_k\times D_x}$, $W_k\in R^{D_k\times D_x}$,$W_v\in R^{D_v\times D_x}$ yields the query vector matrix $Q=[q_1,\ldots,q_N]$, the key vector matrix $K=[k_1,\ldots,k_N]$, and the value vector matrix $V=[v_1,\ldots,v_N]$, and the mapping computation can be expressed as:

$$Q=W_qX\in R^{D_k\times N} \tag{1}$$

$$K=W_kX\in R^{D_k\times N} \tag{2}$$

$$V=W_vX\in R^{D_v\times N} \tag{3}$$

The computational expression for the final attention score is:

$$Attention(Q,K,V)=softmax(\frac{K^TQ}{\sqrt{D_k}})V \tag{4}$$

For the computation of multi-head attention, it is the output of each attention head that is concatenated to get the final output. To wit:

$$head_i=Attention(QW_i^Q,KW_i^K,VW_i^V) \tag{5}$$

$$MultiHead(Q,K,V)=Concat(head_i,\ldots,head_h)W^o \tag{6}$$

In the Transformer model, since the self-attention mechanism is not concerned with the forward and backward ordering of sequence elements, positional encoding needs to be introduced to provide the model with information about the relative positions of the elements in the sequence. Positional encoding is a way of adding learnable positional information to each position in an input sequence so that the model can distinguish between elements at different positions.A common form of positional encoding used in Transformer can be represented as:

$$Positional\ \ Encoding_{(pos,2i)}=\sin(\frac{pos}{10000^{2i/d}\,\mathrm{model}}) \tag{7}$$

$$Positional\ \ Encoding_{(pos,2i+1)}=\cos(\frac{pos}{10000^{2i/d}\,\mathrm{model}}) \tag{8}$$

where pos denotes the position in the input sequence, i denotes the dimension of the position encoding, and d_{model} denotes the embedding dimension of the model, i.e., the dimension of the input vector. The formula uses sine and cosine functions to generate different position encodings for elements at different positions. The encoding of each position is a vector of length d_{model} that is summed with the input vector, thus allowing the model to take into account the relative positions of the elements when processing the sequence.

3.2.2. Generating Adversarial Networks

Generative Adversarial Network (GAN) structure consists of two parts, generator and discriminator, the generator's task is to generate false samples while the discriminator's task is to judge the generated samples to be true or false [23]. The GAN structure usually has the following three processes:

(1) The generator generates brand new sequences based on random numbers.

(2) The true sequence and the sequence generated by the generator are fed into the discriminator.

(3) The discriminator outputs the probability of determining whether the sequence is true or false, the probability value is between 0 and 1, the closer the value is to 1 means the sequence is more true, the closer the value is to 0 means the sequence is more false.

The design idea of GAN is to make two neural networks compete with each other in a zero-sum

game, when the quality of one network increases, the quality of the other network decreases.The purpose of GAN is that the generator creates non-existent samples, and improves the quality of the generated samples through the continuous confrontation with the discriminator, thus deceiving the discriminator and making it unable to distinguish the real samples from the generated samples.The standard objective of GAN can be expressed as follows function can be expressed as:

$$\begin{aligned} &\min_G \max_D V(G,D) \\ &= \min_G \max_D \mathbb{E}_{x\sim P_{data}}\left[\log D(x)\right] + \mathbb{E}_{z\sim P_z}\left[\log\left(1-D(G(z))\right)\right] \end{aligned} \tag{9}$$

where is G the generator, $G(z)$ denotes the sample generated by the generator, D is the discriminator, and $D(x)$ denotes the probability output of the discriminator. $P_{data}(x)$ denotes the probability distribution of the real data x defined in the data space χ and $P_z(z)$ denotes the probability distribution of the z defined on the latent space z. The generator G maps z from z into elements of χ, and the discriminator D takes input x distinguishing whether x is a real sample or a pseudo-sample generated by G. For D, it is desired to obtain a maximized output for samples from real data and a minimized output for samples from G. Also since G goal is to deceive D, it tries to maximize D the output when providing pseudo-samples to D, so D wants to maximize $V(G,D)$ and G wants to minimize $V(G,D)$.

3.2.3. Trans-GAN Music Generation Modeling

In order to realize the communication and interaction of ethnic music on the Silk Road, this paper combines the Transformer model with GAN to construct the Trans-GAN music generation model, and the specific structure of the model is shown in Figure 2. The music data used in this paper are all in MIDI format, and all files contain different types of ethnic music tracks. Firstly, the three tracks are encoded into time sequences respectively, and the internal information of a single track sequence is learned and the state of the next moment is generated by three generators (G_p, G_g, G_b). Secondly, the track sequences are learned two by two by using six CT-Transformer modules, and the ethnomusic sequences after the learning of y_t ethnomusic track sequences are spliced to obtain new ethnomusic sequences, and the remaining music sequences are learned in the same way as the ethnic music sequences. Finally, the real sample sequence and the generated sample sequence are discriminated by the discriminator D_φ.

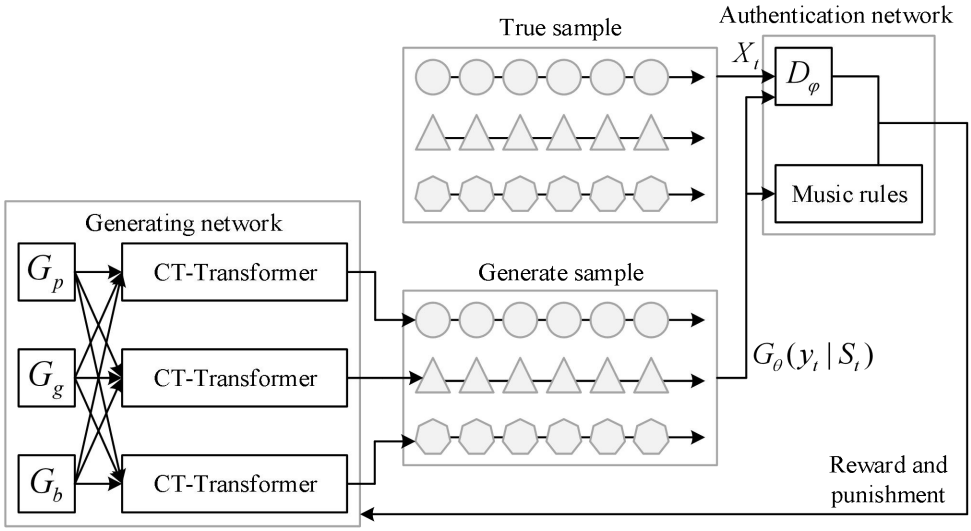


Figure 2. Framework of multi-track music generative adversarial network

(1) Using the Transformer decoding part in the generative network as part of the sequence generation network, the input character sequence is represented by mapping it into Embedding via an embedding matrix. This embedding sequence is then coupled with a positional embedding, which is passed through N_g self-attention modules that mask out the information after moment k, ensuring

that the character can only learn information up to moment k. The output of the last self-attention module is first mapped to the lexical space and then activated by the SoftMax layer to obtain the output character distribution. In the pre-training phase, the track generator is trained to minimize the cross-entropy loss between the predicted characters and the input which is a real character. In the generation phase, the generator generates characters one by one in an autoregressive manner, and the generation can start from the beginning or be given a starting sequence to begin with [24].

(2) In the discriminative network this paper uses the SoftMax layer as the output layer, i.e., the output is the probability distribution of the notes, and therefore the loss function is constructed using cross entropy. When generating multi-track music, by training the model cross entropy can greatly optimize the parameters and improve the quality of musical works. Namely:

$$D_{\varphi} = -\frac{1}{I}\sum_{t=1}^{I}\left[y_t \log \hat{y}_t + (1-y_t)\log(1-\hat{y}_t)\right] \tag{10}$$

In order to obtain music that conforms to the music theory knowledge, this paper mathematically models the music theory rules and feeds back different reward and punishment values to the generative network according to the importance of the music theory knowledge in the music, so as to guide the music generation with the music theory rules. In the piano track, guitar track and bass track of folk music, the notes need to be among A2-C5, E-C3 and E-1-E1 respectively. Namely:

$$R^{m_1}(S_{1,t}, y_t) = \begin{cases} 0.1, & y_t \in \left[y_{\min}, y_{\max}\right] \\ -0.6, & y_t \notin \left[y_{\min}, y_{\max}\right] \end{cases} \tag{11}$$

where $y_{\min}$ and $y_{\max}$ are the lowest and highest notes set in advance according to the key of the music, y_t is the pitch at moment t, and $R^{m_1}(S_{1;t}, y_t)$ indicates the reward value of the state at moment t out of the first t moments.

3.3. Folk music generation model validation analysis

3.3.1. Ethnic Music Dataset Construction

In order to better realize the communication and interaction of ethnic music culture along the Silk Road, this paper collects the characteristic ethnic music of the countries along the Silk Road through the data crawler technology, which provides data support for the verification of the ethnic music generation effect of the Trans-GAN model. In this paper, a total of 720 pieces of ethnic music along the Silk Road are collected, which are divided into 10 categories, and converted into WAV format with a sampling rate of 16KHz, sampling precision of 16bit and mono channel by Goldwave software. Then all the ethnic music audios were sliced at 15s and 30s lengths respectively, following which 19238 15s segments and 9431 30s segments were obtained. For the convenience of data analysis later, it was named as Folks dataset.

After the frame-splitting operation on the Folks clips, each clip was divided into many tiny audio frames, and each Folks clip was divided into many audio frames. Multiple features, such as short-time energy, MFCC, etc., are first extracted at the frame level, with a total of 36 dimensions, so that for each audio segment, these features form a 36-dimensional feature matrix instead of a vector. In order to extract the information of the features more efficiently, the mean value is taken for each column of data in this matrix, and its standard deviation is calculated, which is sequentially combined into a 72-dimensional vector. That is, it corresponds to the feature vector on the frame level of each folk music clip, which is used as the sample data for classification. 70% of the Folks dataset produced in this paper is randomly selected as a training set to train the Trans-GAN model, and the remaining 30% is used as a test set to validate the model's effectiveness in generating ethnic music.

3.3.2. Model Training and Results Analysis

The hardware environment for this experiment is set up as a computer with a memory size of 256 GB, a video memory size of 32 GB, and a compute card NVIDIA P106. The Manjiro Linux operating system is used, the software compilation environment is PyCharm, the parallel computing architecture is CUDA, the deep learning acceleration library is cuDNN, the programming language is Python, and the deep learning framework is TensorFlow-GPU.

In order to better achieve the fast optimization of the model, this paper seeks to select the optimal optimizer before model training, for which four representative model optimization methods, namely

Adam, Adagrad, RMSProp and SGD, are first selected, and their initial parameters and learning rates are set respectively. The 30% Folks dataset is used for training and testing, and Figure 3 shows the convergence process of different optimizers. After the experimental comparison, it can be seen that the Adam optimizer has a faster descent rate during the iteration process, which means the convergence of the optimizer is faster, followed by Adagrad, and the RMSProp optimizer and SGD optimization converge relatively slowly. Therefore, when carrying out the validation of the effectiveness of the Trans-GAN model, this paper chooses the Adam optimizer for model optimization as a way to improve the model's ethnic music generation effect.

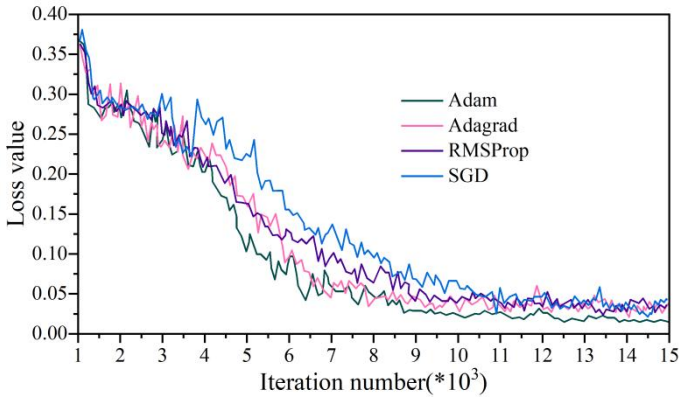


Figure 3. The convergence of different optimizers

Based on the Folks dataset established in the previous paper, 70% are randomly selected for training the Trans-GAN model, with the Dropout ratio set to 0.4, the maximum paradigm weight constraint set to 5, optimized using the better Adam optimizer, the initial learning rate set to 0.001, and the training time is 24 hours. In order to verify the effectiveness of the model in this paper, RNN, GRU, LSTM, BiLSTM, 3D-RNN, 3D-GRU, 3D-GCN are selected as the comparison models, and the accuracy, cross-entropy loss, precision and recall are used as the evaluation indexes, and the comparison results of different models are obtained as shown in Table 1. As can be seen from the table, the accuracy rate and cross-entropy loss of the Trans-GAN ethnic music generation model constructed in this paper are 99.37% and 0.624, respectively, which is 1.66% higher than the accuracy rate of the sub-optimal 3D-GCN model in terms of performance, and the value of the cross-entropy loss is 19.59% lower than it. This fully demonstrates that the model in this paper, when generating ethnic music, fully exploits the characteristics of Silk Road ethnic music by using the Transformer network, and then uses the generative adversarial network to create a new ethnic music, which provides a reliable technical support for the cultural exchange and interaction and dissemination of Silk Road ethnic music. And the precision of the Trans-GAN model reaches up to 0.498, although the recall rate is low, but it also shows that the model can significantly learn the spatial probability distribution between ethnic notes in ethnic music generation, i.e., the wholeness of the local bars of the music between musical segments and the specialization of the music grammar between musical segments as a whole.

Table 1. Comparison results of different models

Model	Accuracy	Cross entropy loss	Precision	Recall
RNN	62.31	0.865	0.179	0.516
GAN	75.46	0.831	0.181	**0.552**
LSTM	74.12	0.829	0.185	0.538
BiLSTM	78.34	0.817	0.024	0.531
3D-RNN	83.29	0.814	0.028	0.389
3D-GRU	95.53	0.805	0.076	0.215
3D-GCN	97.75	0.776	0.423	0.237
Trans-GAN	**99.37**	**0.624**	**0.498**	0.232

3.3.3. Comparison of objective evaluation indicators

After passing the model training and testing, in order to further validate the effectiveness of the Trans-GAN ethnomusicogenesis model, this paper proposes objective evaluation methods, i.e., EB, UPC, and QN.EB refers to the ratio of the number of empty bars without notes to the total number of bars of the generated music samples in the music bars of a track.UPC refers to the pitch class contained

in each bar within the track of a music sample The UPC refers to the number of pitch grades contained in each measure within the track of the music sample, ranging from 0 to 12. QN refers to the ratio of the number of notes in the measure accounted for by qualifying notes in the measures of the generated music sample, which is determined as when the duration of a note is less than three standard time steps (32 cents notes), then it is judged as a non-qualifying note. Table 2 shows the comparative results of the objective metrics, with bolded data being excellent. In the quantitative results of each music track, the closer the value is to the real data set indicates the better the samples generated by ethnic music.

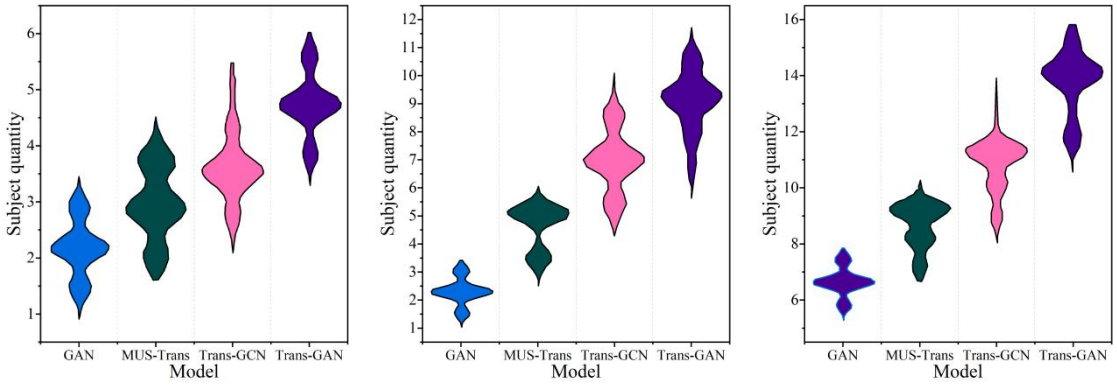
As can be seen from the table, the performance of the ethnic music sample data generated by the Trans-GAN model designed in this paper is closer to that of the real music training dataset in each objective index, which objectively illustrates the effectiveness of this paper's model in generating ethnic music. In the performance in Empty Bar Rate (EB), the ethnic music samples generated by the Trans-GAN model perform better in Guitar and Strings tracks, and the gap between the GAN and the real training set is 28.11% in Guitar notes, and the gap between the Trans-GAN model and the real training set is only 0.18%, and the Trans-GAN model performs better than the GAN model with 22.09% improvement in ethnic music generation performance. In the other tracks the same or close samples were produced with the GAN model. In the UPC evaluation data results, it is observed that only in the Piano track, the objective metrics of the Trans-GAN model performs slightly worse than the GAN model by a reduction of 0.116, and all other tracks perform better than the GAN model. It improves 32.72%, 41.31%, 21.62%, and 50.54% in Bass, Drums, Guitar, and String tracks, respectively. This indicates that the Trans-GAN model is closer to the distribution of pitch types of real music in the number of pitch classes than the GAN model. In the QN metric, the Trans-GAN model improves 5.60%, 2.06%, and 2.27% than the GAN model in Guitar, Piano, and Strings tracks, respectively. This indicates that the scores of the ethnic music samples generated by the Trans-GAN model are all closer to the Folks real music training set by a small margin. This shows that the note duration lengths in the ethnic music samples generated by the Trans-GAN model are closer to the notes in real music, and the note distribution is more reasonable than the GAN, and the noise in the music samples generated by the model after adding the average pooling layer is reduced.

Table 2. Comparison results of objective indicators

Index	-	Bass	Drums	Guitar	Piano	Strings
EB (%)	Real	4.815	12.024	20.175	2.847	13.927
	GAN	26.328	**18.741**	25.847	18.361	17.538
	Ours	21.637	20.895	**20.138**	18.059	**14.072**
UPC	Real	2.351	2.379	2.031	1.823	3.039
	GAN	1.672	3.958	2.715	**1.857**	2.204
	Ours	**2.219**	**2.323**	**2.128**	1.741	**3.318**
QN (%)	Real	1.024	-	91.342	93.245	96.584
	GAN	1.024	-	84.659	90.991	94.273
	Ours	1.024	-	**89.735**	**92.868**	**96.409**

In addition, in order to explore the ability of the Trans-GAN model to generate musical theme fragments when composing folk music of different lengths, this paper chooses GAN, MUS-Trans, and Trans-GCN as comparison models to generate 20 pieces of 24-, 48-, and 96-bar folk music, respectively, and records the number of theme fragments for each piece of folk music. Fig. 4 shows the distribution of the number of themes of ethnic music generated by different models, where Fig. 4(a)~(c) shows the distribution of the number of themes of ethnic music with 24 bars, 48 bars and 96 bars, respectively.

Comparing the distribution of the number of ethnic music themes of different models, it can be found that with the doubling of the number of music bars, the number of themes of the music generated by all the models does not show a corresponding multiplication trend, which indicates that the ability of all the models to generate long-term structured music is inferior to the ability to generate short-term structured music. In addition, a side-by-side comparison between the Trans-GAN model and other models reveals that regardless of the number of music bars, the distribution of the number of themes of ethnic music generated by the Trans-GAN model is relatively higher, which indicates that the Trans-GAN model has a stronger ability of generating themes of ethnic music, which can help to realize the communication and interaction of ethnic music and culture of the Silk Road.



(a) 24 sections (b) 48 sections (c) 96 sections

Figure 4. Different models generate the number of national music topics

In addition, in order to evaluate the quality of ethnomusic generated by Trans-GAN models, in this paper, we let each model generate 20 ethnomusic excerpts of 48 bars, and then record the Scale Consistency (SC) and the Pitch Class Entropy (PCE) of each ethnomusic and compute the average of their results. Pitch class entropy can be used to characterize the uniformity and complexity of pitch distribution in music. Lower pitch class entropy values indicate a more regular and orderly musical structure, reflecting that the distribution of pitches in music is relatively concentrated and biased toward specific pitches, while higher pitch class entropy values indicate a more complex and diverse musical structure, with a more uniform or irregular distribution of pitches. Scale consistency refers to the largest scale proportion among all scales, which can intuitively reflect whether the notes and chords in the model-generated music are consistent with the scales and modes of the musical works. When the scale consistency of the model-generated music is closer to the scale consistency of the music in the dataset, it proves that the model-generated music has more harmony and unity, and is more in line with the real music composition law. Table 3 shows the comparison results and the results of the ablation experiment.

Compared with other music generation methods, the scale consistency and pitch class entropy of the ethnic music generated by the Trans-GAN model are the closest to the real music, and the difference between them and the real dataset constructed in this paper is only 0.92% and 0.52%, respectively. This indicates that the folk music generated by the Trans-GAN model has better rhythmic and structural regularity than other models, and the created folk music is more in line with the compositional laws of the real folk music of the Silk Road. Moreover, the results obtained from both the GAN and Transformer models alone are not satisfactory, while the Trans-GAN model obtained from the combination of the two is more effective, indicating the effectiveness of the two structures in the generation of folk music.

Table 3. Comparison results and ablation experiment results

Comparison result			**Ablation experiment**		
Model	SC	PCE	Model	SC	PCE
GAN	2.842	0.989	GAN	2.781	0.987
MUS-Trans	2.916	0.982	Transformer	2.842	0.985
Trans-GCN	2.954	0.974	Trans-GAN	**3.128**	**0.961**
Trans-GAN	**3.128**	**0.961**	Folks' dataset	3.157	0.956
Folks' dataset	3.157	0.956	-		

4. Ethnic Music Teaching Practices Combined with Deep Learning

The ethnomusicological culture of the Silk Road is a long river of history in which countless creators used their wisdom to constitute the harmony of the mind. The ethnomusicological relics of the Silk Road have precipitated the artistic spirit of the Chinese nation with their richness, continuity and exemplary nature. The purpose of using the Silk Road folk music acquired by the Deep Learning Framework as a foundation and integrating it into folk music education is to enhance students' pride in folk music culture and better promote multicultural interaction in folk music.

4.1. Teaching Objects and Teaching Process Design

4.1.1. Folk Music Teaching Model Construction

As a kind of artistic and emotional education, the essence of ethnic music education lies in promoting students' emotional expression and communication, creativity and aesthetic ability through musical experience. Introducing deep learning technology into the field of Silk Road ethnomusicology education helps to understand the background and process of Silk Road ethnomusicology learning in a more comprehensive way, so as to construct a teaching mode that is more in line with students' cognitive laws and emotional needs. Based on the deep learning framework, the teaching concept of realm pulse is introduced to construct the ethnomusic teaching model, and its specific framework is shown in Figure 5. The specific framework is shown in Figure 5. It is driven by the problems of ethnic music teaching, combining with the existing experience to stimulate students' interest in ethnic music learning and emotional experience, and realizing the communication, sharing and cooperative learning of ethnic music culture in teacher-student interaction and student-student interaction.

In the practice of ethnic music classroom, the application of realm pulse is mainly reflected in the following aspects:

1) “Enlightenment of the realm”, combining the knowledge of ethnic music with students' life experience, through the creation of real or simulated and articulated ethnic music situations, to produce beautiful perception and reaction to the ontology of ethnic music, and to stimulate the students' interest in learning and emotional experience.

2) “Driving Situation”, centering on clear teaching objectives, creates interlocking learning activity situations driven by clear questions, and guides students to enter the deep learning state in artistic performance tasks such as singing, moving, playing and acting through different teaching strategies.

3) “Creating Context”, with the support of learning tools, emphasizes the learning mode of strong autonomy, high degree of participation, and jumping out of the ontological knowledge structure, and achieves the advanced learning goal through the cooperative learning of teacher-student interaction or the communication and sharing learning of student-student interaction.

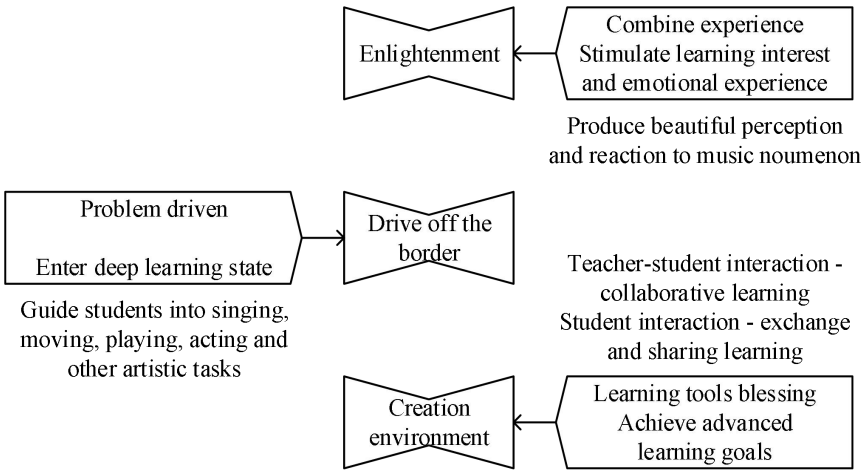


Figure 5. National music teaching mode

4.1.2. Selection of targets for teaching folk music

In order to verify the effectiveness of the Silk Road Ethnic Music Teaching Mode based on the deep learning framework, a 72-credit hour teaching experiment was conducted with 100 students majoring in ethnic music at G University of Nationalities to explore the teaching effect of the teaching mode on students' theoretical performance in ethnic music, learning level, learning attitudes, interest in learning, and communication and interaction in the education of ethnic music, so as to provide references for the reform of the informationization of the education of ethnic music. Opinions.

The experimental subjects have the same ethnic music foundation before the teaching experiment, and there is no obvious difference in the ethnic music learning attitude, learning interest and music culture communication and interaction level. For this reason, the students will be grouped according to their academic numbers, the number of people will be kept the same, and the experimental group (EG) and the control group (CG) will adopt the realm pulse teaching mode and the traditional teaching mode for teaching ethnic music respectively. According to the teaching work plan of the 2023 grade of ethnic music major in the School of Music of G National University, it will be uniformly arranged in the

second semester of the 2023~2024 semester, with a total of 12 weeks of teaching practice in the course, 3 classes per week (one class is 45 minutes, 2 credit hours) for a total of 72 credit hours, and practical and theoretical examinations will be arranged during the 13th and 14th weeks, respectively.

Through EXCEL software, the data of test content data before and after the experimental intervention of 50 students of Class 2023 Ethnic Music Major 1 in the School of Music of G University for Nationalities were organized, and SPSS statistical software was used to statistically and analytically analyze the data of ethnic music literacy and questionnaires (learning attitudes, learning interests) of the students of the experimental group and the control group before and after the experiments. As well as the two groups' theoretical knowledge test scores after the experiment were statistically and scientifically analyzed, and the data were presented in the form of mean ± standard deviation. The experimental and control groups were analyzed using independent samples t-test, and before and after the experiment were analyzed using paired samples t-test.

4.2. Effectiveness and Analysis of Cultural Interaction Teaching

4.2.1. Comparison of Students' Achievement in Ethnic Music

The effectiveness of folk music teaching model based on the deep learning framework and the concept of realm chakra was collected from the students' performance in folk music in the experimental group and the control group after conducting a semester-long teaching experiment. It was mainly tested in six dimensions: singing and dictation, association and structure, activity and experience, transfer and application, value and evaluation, and culture and communication. Five different levels of choices were included, with 1 to 5 representing not at all, not at all, not at all, generally, relatively, and completely. Statistics were conducted for the students' ethnic music performance after the beginning of teaching, and Table 4 shows the results of the independent samples t-test for the students' ethnic music performance before teaching.

As can be seen from the table, the total ethnic music scores of the experimental group and the control group before carrying out the teaching experiment were 18.197±0.741 and 18.093±0.752, respectively, and the results of the independent samples t-test showed that there was no significant difference in ethnic music scores between the two classes (Sig>0.05), and there was no significant difference in the dimensions (Sig>0.05). Therefore, the ethnic music education model established in this paper can be utilized to carry out teaching experiments as an illustration of the effectiveness of the teaching model in the Silk Road ethnic music education and cultural exchange and interaction.

Table 4. The former student national music achievement-before

Index	Group	M±SD	T value	Sig.
Singing and listening	EG	2.815±0.824	-0.147	0.426
	CG	2.749±0.759		
Association structure	EG	2.927±0.908	-0.338	0.507
	CG	2.893±0.872		
Activity and experience	EG	3.105±0.594	-0.582	0.329
	CG	3.114±0.601		
Migration and application	EG	3.138±0.745	-0.619	0.175
	CG	3.127±0.752		
Value and evaluation	EG	3.059±0.674	-0.423	0.102
	CG	3.062±0.681		
Culture and communication	EG	3.153±1.175	-0.537	0.348
	CG	3.148±1.169		
Total	EG	18.197±0.741	-0.369	0.513
	CG	18.093±0.752		

At the end of the teaching experiment, the data related to the ethnic music achievement of the students in the two groups were collected and entered into EXCEL for the independent samples t-test, whose specific results are shown in Table 5, where *** indicates that there is a significant difference at the 1% level. The total scores of students' ethnic music achievement in the experimental group and the control group at the end of the teaching experiment were 25.161±0.406 and 18.592±0.684, respectively, and the results of their independent sample t-tests indicate that there is a significant difference between the two groups of students' ethnic music achievement at the 1% level (t=6.541, Sig=0.005<0.01). This suggests that the Silk Road ethnic music teaching model under the deep learning framework helps to enhance students' ethnic music achievement and also has a contributing effect on students' ability to

learn ethnic music in depth. In terms of specific dimensions, the enhancement of singing and listening, and culture and communication is relatively large, with the students in the experimental group enhanced by 54.99% and 42.78%, respectively, compared with those before the experiment. This indicates that the ethnic music generated by deep learning has diversity, which can significantly enhance students' singing and listening skills when integrated into the teaching process of ethnic music, and can also realize the communication and interaction of different ethnic music cultures. This shows that the teaching of folk music under the framework of deep learning has a positive impact on students' ability to interact with different ethnic music cultures, which promotes the progressive development of students' core literacy in folk music and provides an opportunity for the cross-cultural communication development of folk music on the Silk Road.

Table 5. The former student national music achievement-after

Index	Group	M±SD	T value	Sig.
Singing and listening	EG	4.363±0.127	5.469	0.000***
	CG	2.853±0.459		
Association structure	EG	3.948±0.506	6.903	0.003***
	CG	2.951±0.615		
Activity and experience	EG	4.124±0.315	5.254	0.007***
	CG	3.208±0.427		
Migration and application	EG	4.065±0.643	4.713	0.006***
	CG	3.238±0.749		
Value and evaluation	EG	4.159±0.571	5.259	0.002***
	CG	3.127±0.634		
Culture and communication	EG	4.502±0.247	8.365	0.000***
	CG	3.215±0.753		
Total	EG	25.161±0.406	6.541	0.005***
	CG	18.592±0.684		

4.2.2. Evaluation of the effectiveness of interactive cultural learning

To address the learning effect of multicultural interaction in Silk Road folk music education, this paper designed a questionnaire from three dimensions: aesthetic perception, interactive performance and cultural exchange, and its specific content is shown in Table 6. The evaluation levels of the questionnaire are able, basic able and unable, i.e. Level A~C. The questionnaires were distributed to students in two classes, and a total of 50 questionnaires were recovered.

Table 6. Cultural interaction learning effect questionnaire

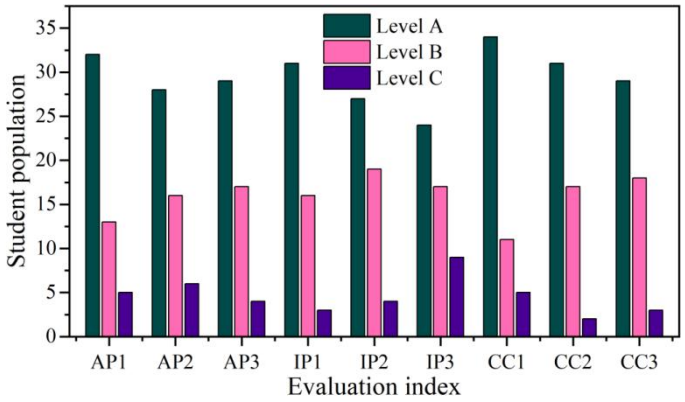
Dimension	Evaluation content	Code
Aesthetic perception	Can you tell different types of national music	AP1
	Can you understand the connotation of the humanities through the national music	AP2
	Understand the style and performance of the national music	AP3
Interactive performance	Able to carry out the interaction of national music	IP1
	Ability to actively participate in creative activities	IP2
	Can understand the excellent music of different nationalities	IP3
Cultural communication	Can understand the excellent music of different nationalities	CC1
	Can you respect the cultural diversity of music	CC2
	Can we carry forward the excellent national music	CC3

The data of the questionnaire were counted to get the interactive learning effect of students' ethnic music culture as shown in Fig. 6, in which Fig. 6(a)~(b) shows the interactive learning effect of students' ethnic music culture in the experimental group and the control group respectively.

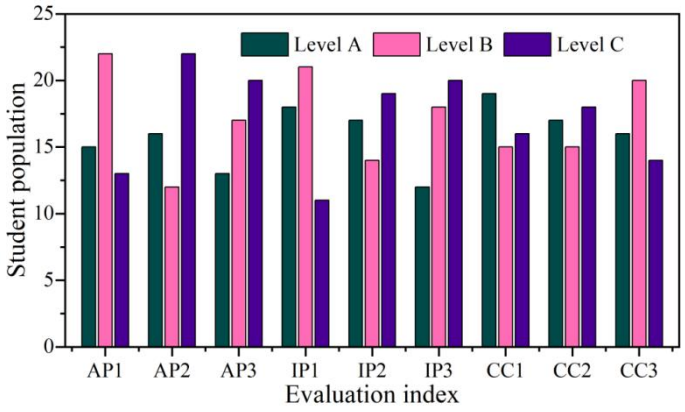
(1) From the feedback data of aesthetic perception, it can be seen that the students in the experimental group are stronger than the students in the control group in their aesthetic perception ability. Under the guidance of the teacher, students in the experimental group were able to better analyze different types of Silk Road folk music from the perspective of musical elements, and were more willing to explore the unique humanistic connotations behind the Silk Road folk music, and to better understand the styles and expressive methods of Silk Road folk music. Students in the experimental group have strong thinking logic of ethnic music, students are more likely to conduct in-depth learning and thinking, and their music aesthetic perception ability is a bit stronger.

(2) The teaching method of combining the realm pulse under the deep learning framework pays more attention to teaching in the form of problem leading, and in the process of learning, more attention is paid to the students' focusing on problems and solving problems. The learning process emphasizes more on students' activities around problems and solutions. It can be seen that students in the experimental group are more willing to participate in practice in this teaching method, focusing on the interactive emotional experience of folk music. Students will actively participate in the creative activities of ethnic music on the Silk Road, and actively use their own language to express the characteristics of ethnic music works of different countries in the process of practical creation.

(3) In order to better promote students' in-depth learning, the teaching model designed in this paper hopes that students can take the initiative to transfer knowledge, connect new and old knowledge, think about problems and solve them in the process of learning folk music. Students in the experimental group are more able to understand the excellent musical works of China and other countries and nations in the teaching mode of ethnic music, to look critically at the aesthetic value of music from different countries and regions, and to form correct aesthetic judgment of ethnic music. At the same time, they are more willing to synthesize what they have learned to pass on and carry forward more excellent ethnic music of the Silk Road, and to provide support for the promotion of multicultural interaction and exchange of ethnic music of the Silk Road.



(a) Experimental group



(b) Control group

Figure 6. Interactive learning effect of national music culture

5. Conclusion

The article proposes a Silk Road folk music teaching model combined with a deep learning model, which is based on the folk music generated by the Trans-GAN model and combined with the concept of realm pulse to carry out folk music teaching. Students majoring in folk music at G Nationalities University were selected as the research subjects, and teaching experiments were conducted to analyze the effect of teaching folk music on the Silk Road.The cross-entropy loss of the Trans-GAN model is 0.624, which is 19.59% lower than the sub-optimal performance of the 3D-GCN model, and better results were obtained in terms of the entropy of the pitch class and the consistency of the pitch scale.

After the implementation of the ethnomusicology teaching model, the total ethnomusicology achievement score of the students in the experimental group reaches 25.161±0.406 points, which is 6.964 points higher than that before the experiment, and the students' ethnomusicology multicultural interaction ability has been significantly enhanced. Combining deep learning technology with Silk Road ethnic music education can effectively enhance students' ethnic music cultural interaction level, and lay a solid foundation for promoting cross-cultural exchanges in Silk Road ethnic music.

References

1. Volk, T. M. (2017). The history and development of multicultural music education as evidenced in the Music Educators Journal, 1967–1992. In Critical essays in music education (pp. 481-500). Routledge.
2. Du, J., & Leung, B. W. (2022). The sustainability of multicultural music education in Guizhou Province, China. International Journal of Music Education, 40(1), 131-148.
3. Ronström, O. (2019). On the meaning of practicing folk music in the 21st Century. PULS: Journal for Ethnomusicology and Ethnochoreology, 4, 10-25.
4. Tullberg, M., & Sæther, E. (2022, December). Playing with tradition in communities of Swedish folk music: Negotiations of meaning in instrumental music tuition. In Frontiers in Education (Vol. 7, p. 974589). Frontiers Media SA.
5. Drandić, D., Gortan-Carlin, I. P., & Jadan, E. (2021). Traditional (Folk) Music in Intercultural Education of Students. FACULTY OF EDUCATION JOSIP JURAJ STROSSMAYER UNIVERSITY OF OSIJEK, 141.
6. Zhang, X. (2023). Digital Communication of Folk Music in Social Music Culture. Frontiers in Art Research, 5(14).
7. Roberts, V. S. (2017). Folk Music. The Bloomsbury Handbook of Religion and Popular Music, 260.
8. Campbell, P. S. (2017). Music, education, and diversity: Bridging cultures and communities. Teachers College Press.
9. Whitfield, S. (2015). Life along the silk road. Univ of California Press.
10. Frankopan, P. (2016). The silk roads: A new history of the world. Vintage.
11. Tang, K., Li, Z., Li, W., & Chen, L. (2017). China's silk road and global health. The Lancet, 390(10112), 2595-2601.
12. Kurin, R. (2024). The silk road: the making of a global cultural economy. Explorations in Cultural Anthropology: A Reader, 145.
13. Levin, T. (2016). The Silk Road as Jam Session, Then and Now. Revue des Traditions Musicales, 10, 109-27.
14. Li, Q. (2019). Silk Road: The Study of Drama Culture (Vol. 3). World Scientific.
15. Li, M. (2020). A musical journey along the Silk Road—Encounter, discovery and change. China and the New Silk Road: Challenges and Impacts on the Regional and Local Level, 147-154.
16. Boltaev, O. (2024). BUKHARA'S CARAVAN TRADE AND ITS ROLE ON THE SILK ROAD. Analytical Journal of Education and Development, 4(10), 293-297.
17. Arakawa, M. (2016). The Silk Road trade and traders. Memoirs of the Research Department of the Toyo Bunko, 74, 29-59.
18. Blaydes, L., & Paik, C. (2021). Trade and political fragmentation on the silk roads: The economic effects of historical exchange between china and the muslim east. American Journal of Political Science, 65(1), 115-132.
19. Li, W., & Zhang, Y. (2025). Symbol and Narrative in Religious Music: A Cross-Cultural Comparative Analysis. Cultura: International Journal of Philosophy of Culture and Axiology, 22(3).

20. Li Yang & Sun Ruoran. (2023). Foreign Communication of Chinese Music Through Transmitting the Spatiotemporal Context of the Old and New Silk Roads: A Modern Approach. Journal of psycholinguistic research(4),1249-1261.
21. Jia Luo. (2024). Culture and Tourism Integration Strategy of Zhanjiang Folk Music: Double Wheel Drive of Inheritance and Innovation. Journal of Asian Research(3).
22. Shengming Dong, Yue Meng, Shubin Yin & Xianli Liu. (2025). Tool wear state recognition study based on an MTF and a vision transformer with a Kolmogorov-Arnold network. Mechanical Systems and Signal Processing112473-112473.
23. Cheng Qian, Wenzhong Tang & Yanyang Wang. (2025). RGAnomaly: Data reconstruction-based generative adversarial networks for multivariate time series anomaly detection in the Internet of Things. Future Generation Computer Systems107751-107751.
24. R. Zhang,W. Ni,N. Fu,L. Hou,D. Zhang & Y. Zhang. (2025). DP-LTGAN: Differentially private trajectory publishing via Locally-aware Transformer-based GAN. Future Generation Computer Systems107686-107686.