International Journal of Information Technology and Web Engineering
Volume 18 • Issue 1

Analysis of Application and Creation Skills of Story-Based MV Micro Video and Big Multimedia Data in Music Communication

Xi Zhang, Hebei Normal University for Nationalities, China
Yue Cui, Hebei Normal University for Nationalities, China*

ABSTRACT

MV is an art type that uses TV pictures to supplement information and content that cannot be covered by music. With the rapid development of micro video technology and network technology, micro video has been rapidly popularized. Combining MV and micro video with mobile devices and platforms, it breaks the limitations of traditional movies, TV, and animation and becomes a new highlight of cultural communication in the new era. This paper starts with the selection of MV songs, story script, content design, MV shooting and editing skills, and introduces the specific creation methods of story MV in combination with local culture and proposes a music genre classification algorithm model DCNN-SSA based on spectral space domain feature concern. DCNN-SSA model effectively marks genre features of different music Mel spectrograms in the spatial domain and changes the network structure, thus improving the feature extraction effect while ensuring the validity of the model, thus improving the accuracy of music genre classification.

KEYWORDS

Creative Technique, Micro Video, Music Communication, Story MV

INTRODUCTION

In English MTV stands for Music Television and MV stands for Music Video (Feng, 2022). In fact, since the birth of MV, its production concept has far surpassed the technical means of TV (Calefato et al., 2022). Today, with the increasingly developed network technology, the MV that used TV as the main communication medium in the early years has gradually "transferred" to the network medium (Karim et al., 2022). Secondly, in the award setting of some foreign professional music awards, MV is also used as the title of this art form. For example, the Grammy Awards, one of the most important awards in the US recording industry and the world music industry, take "Music Video" as one of the award settings and subdivide it into "Best Short Music Video" and "Best Long Format Music Video" (De-Deus et al., 2022). The name of MTV is unsubstantiated (Ferguson et al., 2022). The author

DOI: 10.4018/IJITWE.325213
*Corresponding Author

once asked friends living in the United States, the United Kingdom, and other places about what we call MTV in their daily life, and the answers they got were all MV or its full name Music Video. The MTV they call it there is only one thing, and that is the MTV channel (Nakano & Washizu, 2022). In the early 1990s, music TV began to be introduced to China, and MTV became its code name. After that, the United States filed a complaint against CCTV, because CCTV used its special name, MTV, without authorization from the American Music TV Channel (Liu et al., 2021). Therefore, the Chinese national TV station used the Chinese "Music TV" as the general name for MV (Comba et al., 2021). Yang Yue, deputy manager of the Performing Arts Department of the MTV Chinese Channel, pointed out in the article "The MV is habitually called MTV has a very bad influence on us" in the 12th issue of *Rolling Stone* magazine in 2006: "Everyone is accustomed to 'music video' or 'music video'. 'Music TV' is called MTV, and the detriment to the brand is huge." MV is an art type that uses TV pictures to supplement the information and content that music can't cover, and presents the interpretation of music with TV pictures at the same time. With the rapid development of micro video technology and network technology, micro video has been rapidly popularized. MV, which expresses the connotation of songs through video pictures, has been recognized and loved by vast audiences. Combining MV and micro video, with the help of mobile devices and platforms, it has broken all kinds of restrictions of traditional movies, TV, and animation, and has become a new highlight of cultural communication in the new era.

The city of Chenzhou is located in the southeast of Hunan Province, China, and in the area where the Nanling mountains and the Luoxiao mountains intersect and the Yangtze River system and the Pearl River system diverge. It has been a place from the Central Plains to the South China Coast since ancient times. It has a long history, a galaxy of talents, and rich resources. Chenzhou now has jurisdiction over one city, two districts, and eight counties. There are eleven administrative counties and 156 townships with unique historical culture, revolutionary culture, folk culture, and ecological cultural resources. The local culture is made into an MV in the form of story-like micro-videos, and the dissemination of Chenzhou local culture is promoted through elements such as pictures, music, stories, light and shadow, and text (Fahrudin & Winarni, 2021). Through investigation, it is found that the existing local songs in Chenzhou can be roughly divided into network adaptation songs and original songs (Liu, 2021). In recent years, the songs that have been widely sung include "Singing on the East River and Lake", "I am waiting for you in Mangshan," "My Chenzhou, My Home," "Chenzhou Girl, "Chenzhou,", as well as "Tongtian Beauty Waiting for You," "Jiahe Morning Wine," "Linwu Oil Tea Fragrance," "Dongjiang Lake Boating," "Flying Mountain," and other songs (Pan et al., 2020).

Through data analysis, it is found that songs made into MVs that combine songs with pictures have higher dissemination and click-through rates. And MVs that add story clues or character plots are more popular among audiences, while songs that are not made into MVs have a relatively low degree of dissemination (Saha et al., 2020). For example, "Songtai on the East River and Lake" was released on multiple online platforms, and the number of clicks on Tencent Video alone reached 248,000 (Abdelazim & Ramzy, 2022). In just a few months, the newly released "I'm Waiting for You" has reached 1,283 views (Geiko et al., 2022). The song "We Are All Chenzhou People" adapted by the Douyin platform, received 3,689 likes. "Great Chenzhou" received 39,000 likes on singer Lin Tongxue's Douyin platform (Hornung & Smolnik, 2022). The song "Chenzhou" was loved by the audience because it combined the real campus life performed by the students of Chenzhou No.2 Middle School Performance Club with micro video and music (Engin et al., 2022). Tencent alone had 1.099 million hits, and it was popular in mainstream media such as iQiYi, Tiktok, Youku, and 56.com (Garg & Kumar, 2022). Through the questionnaire survey, it was found that in the process of local culture dissemination, more than 90% of people like MV micro-video, Douyin, short video of public transportation, and other new media publicity methods And the local culture of Chenzhou is loaded into the high-flow song MV, which will make the the integration of pictures, lyrics, music, and stories, and the application of story-like MV micro-videos in the promotion of local culture conducive to the dissemination of the local culture (Yost, 2022).

New media and self-media with network information technology as the core have brought people into a new era of all-media integration; the media is no longer the exclusive communication segment of professionals, and every ordinary person can also participate, enhancing the freedom of speech and the right to express one's opinion (Wang, 2022). It is against this background that micro videos are produced. Some people think that the emergence of online micro videos has brought extra vigor to low-budget films and injected vitality into the development of the film market (Liang, 2022). As a new type of short film with low cost, short production time, and rapid dissemination, the development of micro video also benefits from the boost of the current online video dissemination platforms (Pistola et al., 2021). After decades of rapid development, the number of video playback websites has increased dramatically, and the quality and sense of technology have also improved. From the original website with only a single function as a playback platform, it now covers movies and TV series, news short videos, and advertising videos (Chang et al., 2014), and other multi-dimensional, multi-angle, multi-type information dissemination of media.

The network platform of modern technology and the huge network consumer groups in the current society have provided a channel for the rapid dissemination of many micro videos. Taking advantage of this, the development of micro videos will become more popular (Maybury, 2012). With the introduction of the Internet plus development goals issued by the state, the continuous development of the current wireless network transmission technology and the wide application of artificial intelligence terminals in daily life, people's channels for learning are no longer only through a single network platform. There are a wider range of choices, and they can use various forms such as micro videos to enrich and improve themselves (Gaeta et al., 2014).

This article is divided into three parts. Firstly, the development trends of MV and micro video and their applications in music dissemination are introduced. Secondly, the creative methods and techniques of story MV are discussed in detail, and a music genre classification algorithm model DCNN-SSA based on spectral space domain feature attention is proposed. Finally, the experimental results are analyzed and discussed, verifying the effectiveness of the model in music genre classification tasks.

MATERIALS AND METHODS

This section introduces the methods for creating MV, micro video, and electronic music recognition. When creating an MV, it is necessary to select the appropriate song, design the story outline and structure, and write a split shot script. Micro videos mainly express personal opinions, and opinions do not have equal authority, therefore they are more entertaining. The features of electronic music signals mainly include time-domain features, frequency-domain features, and Mel cepstrum coefficients, while least squares support vector machine is a commonly used music signal recognition algorithm. The specific operating methods and steps will be introduced in detail.

Creation Method

(1) Choose the right song when creating the MV.

First of all, it is necessary to select songs that can reflect the artistic conception of local culture in the lyrics (Lankshear & Knobel, 2007). When promoting the natural scenery and folk characteristics of local culture, such as landscape, architecture, humanities, and folk customs, you can choose local songs with strong representativeness such as "The Most Beautiful East River and Lake Water," "Flying Mountain," "Autumn in the East River Lake," "Jiahe Companion's Marriage Song," "My Chenzhou, My Home," "I am waiting for you in Mangshan," "Chenyang Lantern Minor," "Shanshui Hunan," and other songs that Chenzhou has. Red ruins, red stories, and red characters, you can choose red songs to create, such as "My Motherland and Me," "Sing a Folk Song for the Party," "Great China," "Heroes Praise," "Five Stars Red Flag," etc. In the case of rapid development and fashion vitality, in

addition to Chenzhou's local songs "Great Chenzhou," "Chenzhou," and "Chenzhou Girl," you can choose some adapted songs with high traffic, such as "We Are All Hunan People" Chenzhou Edition, "Make China More Beautiful" Chenzhou version, "We Are All Chenzhou People" adapted songs, etc. In addition, when promoting Chenzhou people and Chenzhou spirit, you can choose "Sincere Love," "Stubborn," "Youth," "Cheers," "Love Never Stops," and songs for hard work.

(2) Determine the theme and content of the MV story.

The MV uses the lyrics and musical mood to create a story structure and design video images in which the lyrics are more storytelling than the music. Therefore, the producer can first comprehend the theme expressed by the song, listen to the song more, understand the meaning of the lyrics, clarify the central idea of the work according to the music and the lyrics, and construct the story theme and image of the MV. Generally speaking, the content of the story can be based on real life as the material, with real people as the object of expression, and artistically processed, or it can be adapted from real stories, with fictional plots and characters, played by actors. Using the lyrics as the blueprint for creation, a storyline can be created that matches the artistic conception of the pictures provided in the lyrics, and a plot and picture can be built with unique, humanistic, and story-like characteristics, such as the newly launched "Orange Love" by Ma Danwei. Based on the adaptation of the real story, the MV describes the girl's life in the rock sugar orange orchard with her grandmother, her study in a distant place, the harvest of rock sugar orange, and her attachment to her hometown. It expresses her hope for life and love for her hometown. At the same time, it also publicizes the rock sugar orange, the main agricultural product of Yongxing County in Chenzhou City.

(3) Design the outline and structure of the story and write the storyboard script.

For music MV, the dramatic conflict and cross structure of the story always emphasize integrity and independence in content and rely more on cinematic narrative techniques in text arrangement. In view of this, in the production stage of music MV, if you want to ensure the natural and smooth switching of shots, you need to accurately analyze the melody and rhythm of songs. Only in this way can you accurately locate the editing point and then maximize the communication effect of music MV. Unique story themes and ingenious narrative techniques, through the arrangement of events and scenes, the theme, content, and ideas of the story, are vividly expressed. When designing the form of story expression, producers can design a story outline that is as closely related to the lyrics as possible according to the theme of the story. Before designing, the lyrics can be divided into paragraphs and simplified with keywords to determine the main content. The emotional points and conflict points in the lyrics and music are found, and then the rhythm, time arrangement, and conflict level of the progress of the MV story are designed.

The story structure mainly consists of a beginning, middle, and end. In these structures, specific plot points are connected in series, usually creating confrontational, conflicting, dramatic, obstructive, and detailed plot points in the middle. For example, in the MV of "Orange Love," the picture of the girl's childhood and her grandmother's life in the orangery is shown at the beginning, and the story scene is introduced. In the middle section, the picture of the girl who is away from home to study and the mother saying goodbye expresses the contradiction and confrontation of leaving and reluctance. Through the orange harvest and the picture of the girl returning home, she expresses her love for her hometown and yearning for the future.

In specific structural design, for example, when promoting historical culture, natural scenery, traditional skills, red stories, red relics, and other cultures, a time-space staggered structure can be used to break the natural order of displaying time and space, and different time and space scenes can be arranged according to the order of time and space. A certain artistic conception and logical

cross-connection combination, in order to promote the development of the plot, interweave and group together the present, the past, the future, memories, associations, dreams, and illusions. When promoting local food culture, urban life, and characters, you can choose novel, prose, comprehensive, and other structures. The novel structure is mainly characterized by describing the subtle changes in the characters' thoughts and feelings. The plot is not required to be highly concentrated, but mainly expressed through the accumulation of scenes. In prose style, there is no obvious beginning, climax, and end, and no obvious contradiction and conflict. It is mainly described by events in life. Comprehensive structure can combine drama and narrative and depict characters and events through pictures with conflicts.

For example, when promoting the red revolutionary relics, revolutionary songs can be selected as MV shots, and prose structures can be selected to depict and classify the former sites, such as the former residence of great men, the former sites of conferences, headquarters, and activities. It can be listed by category or displayed alternately, and the moving red story pictures and red characters can be interspersed in the picture process to arouse the audience's resonance.

According to the outline of the story, the content of the story, and the structure of the story, the shot splitting script can be written. Split shot script is an intermediate medium that converts words into three-dimensional audio-visual images, providing guidance for the next step of shooting. The main task of the split shot script is to process the screen content of the text script into specific images for shooting, and list the lens number of the lens in order.

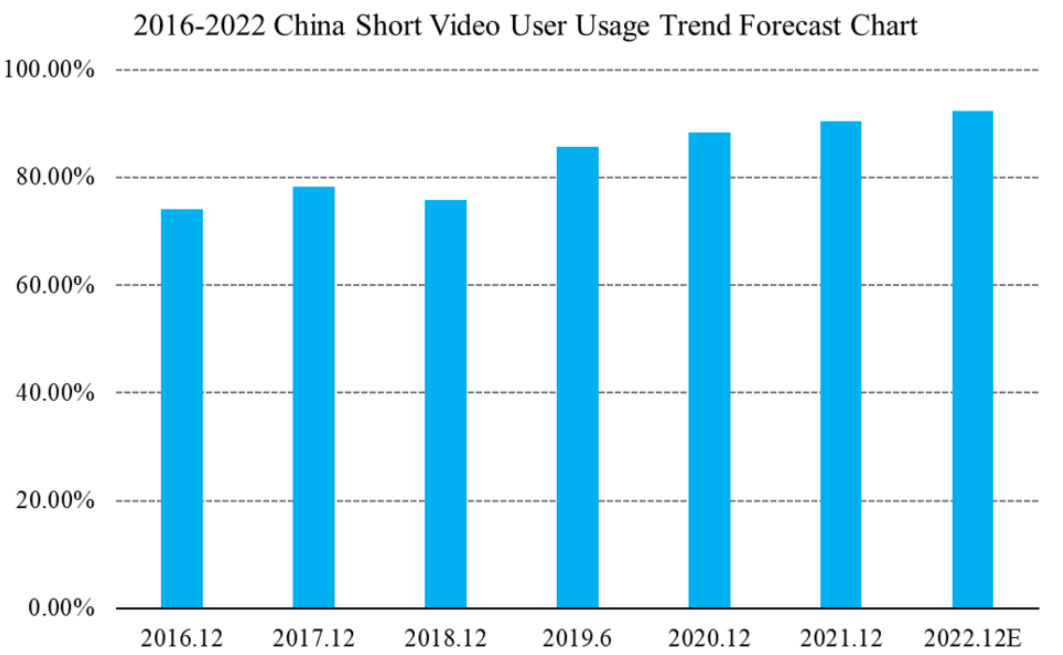
The Concept of MV

MV was born in the UK in 1975. A rock band called Queen, comprised of four boys, released their new album "One Night in the Opera House" in the same year. "Bohemian Rhapsody" was creatively made into music video (Music TV video) and played on TV media as a main song in this album. This is also the first time that M appeared in the form of commercial advertisements to promote songs and singers. As a result, the world's first real music TV video came out.

"MV" stands for "music video," which is the result of the grafting of video art onto music art. The MV studied by the author in this article is an art form dominated by music, that is, "V" is subordinate to "M". Judging from the author's viewing experience of a large number of MV works and my own practical experience in MV creation, images are created limited by music—mainly reflected in the following points. First, from the perspective of the performance, content of images, the plot, tone, light, ambient atmosphere, scene style, etc. are determined by the style of music, emotional tone, and content of lyrics. Second, the movement characteristics and the mastery of the editing rhythm are also determined by the structure and rhythm of the music. Thirdly, from the perspective of the creative process, at the shooting site of the MV, the director will keep playing the song, whether it is a singer or other performers. The performance and creation must follow the emotion and rhythm of the music. In the post-editing production, the director and the editor (sometimes the same person) also have to import the music into the work area of the production software first, and then use it as a basis to make images. Ning Hao, a young director who became famous for "Crazy Stone," has also made a lot of MVs. He thinks that compared to shooting movies, MVs will impose more constraints on the director, because the movie is his own, and others serve your ideas. The MV is for the singer to serve the music. Objectively, shooting an MV will be subject to various limitations, such as what kind of music the music is, and to create a suitable image for him, the music has become a proposition. Of course, There is no absolute single creative method, the author only listed the situation of most MV creation.

Micro video programs are filmed and produced by different cultural levels, and their environments are also different. The content of micro videos only expresses personal views and opinions, and does not represent public values and social views. As a result, they appear to be more popular and entertaining. Figure 1 shows the trend forecast of short video user usage in China from 2016 to 2022.

Figure 1. China short video user utilization trend forecast chart



Short video is the main tool for people's daily leisure, entertainment, social interaction, and information interaction, which is deeply loved by users. In recent years, the number of short video users has continued to grow. According to the 49th statistical report on China's Internet Network Development released by China Internet Network Information Center (CNNIC), as of December 2021, the number of short video users was 934 million, and the utilization rate was 90.5%. It is estimated that by December 2022, the number of short video users will reach 985 million, and the utilization rate will reach 92.4%.

Construction of Music Signal Feature Vector

Different electronic music signals have different eigenvectors. The establishment a of electronic music signal identification model is mainly used to describe the mapping relationship between eigenvectors and electronic music signal types. Electronic music signal features mainly include three categories: time domain features, frequency domain features, and Mel cepstral coefficients, so this paper extracts these three types of features for electronic music signal identification.

There are many time-domain features of electronic music signals. In this paper, short-term energy features and short-term average zero-crossing rate features are selected. For electronic music signals {z(beat)}, the calculation of short-term energy is as follows:

$$\mathrm{E}_n = \sum_{m=-\infty}^{+\infty} \left[x(m)w(n-m)\right]^2 \tag{1}$$

z (beat): the beat in electronic music signalsE_n: Short term energy of frame nx (i): value of the i-th sampling point$w$ (i): weight of the ith sampling point in the window function

The calculation of the short-term average zero-crossing rate is as follows:

$$Z_n = \frac{1}{2}\sum_{m=-\infty}^{+\infty}\left|\text{sgn}[x(m)] - \text{sgn}[x(m-1)]\right| w(n-m) \tag{2}$$

In the formula, sgn represents the symbolic function.

The frequency domain feature of the electronic music signal selects the spectral centroid, subband energy, and bandwidth. First, perform Fourier transform on the electronic music signal to obtain F (boron), and then calculate the spectral centroid, subband energy, and bandwidth, such as:

$$\text{w}_{\ 0} = \int_0^{w_0} w \mid F(w) \mid^2 dw \,/\, \int_0^{w_0} \mid F(w) \mid^2 dw \tag{3}$$

$$\text{P}_{\ 0} = \lg\left(\int_{L_j}^{H_j} \mid F(w) \mid^2\right) dw \tag{4}$$

$$B = \sqrt{\frac{\int_0^{w_0} (w - w_0)^2 \mid F(w) \mid^2 dw}{\int_0^{w_0} \mid F(w) \mid^2 dw}} \tag{5}$$

In the formula, F(w) represents the amplitude of the component with frequency w.

Extracting Mel Cepstral Coefficients of Music Signals

Mel cepstral coefficient is a key electronic music signal identification feature. Fourier transform regards the original electronic music signal as the accumulation of some electronic music signals with different sizes, frequencies, and amplitudes, and analyzes the corresponding frequency signal through the components with large amplitude. Because the noise signal has certain periodic change characteristics, it can filter out the noise in the electronic music signal.

Based on Fourier transform and cepstral analysis, each subband width of electronic music signal is:

$$\text{Mel(f)} = 2595\lg(1 + f \,/\, 700) \tag{6}$$

The specific steps of extracting the Mel cepstral coefficient feature of the music signal are as follows.

Step1: Use Fourier transform to process the electronic music signal to get the spectrum X (walk), such as:

$$\text{X(k)} = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N} \tag{7}$$

X (walk) is the spectrum of the original audio signal, and n is the number of Mel cepstrum coefficients.

Step2: Establish a filter bank of Mel cepstrum and use the Mel filter bank to process X (yes) to obtain the power spectrum Sk of the signal.

Step3: Calculate the Mel cepstral coefficients using inverse discrete cosine transform,

$$c_n = \sqrt{\frac{2}{k}} \sum_{r=1}^{K} (\lg S_k) \cos(n(k-0.5)\pi / K) \tag{8}$$

where K represents the number of filters.

Music Signal Recognition by Machine Learning Algorithm

Least squares support vector machine is a recently popular machine learning algorithm. The least squares support vector machine is an improvement of the support vector machine. Compared with the standard SVM model, the advantage of this method is to replace the inequality constraint in the standard SVM algorithm with equality constraint. The problem of solving quadratic programming is transformed into solving linear equations directly. Its learning speed is faster than that of support vector machine, and its learning effect is better than that of artificial neural network, so it is chosen to establish an electronic music signal recognition model. Let the electronic music signal identification features and signal types form a training sample set: {(xi,yi)}, i=1,2,…,n. xi and yi are the identification features and types of electronic music signals, respectively:

$$f(x) = w^T \varphi(x) + b \tag{9}$$

Transform and solve equation (9), such as equation (10).

$$\begin{aligned} &\min || w ||^2 + \frac{1}{2}\gamma \sum_{i=1}^{n} \xi_i^2 \\ &s.t. y_i - w^T \varphi(x) + b = e_i \end{aligned} \tag{10}$$

In the formula, y represents the parameters of the least squares support vector machine. *b* is the deviation term. Since the calculation process of formula (10) is very complicated, its equivalent form is established, such as formula (11).

$$L(w,b,\xi,\alpha) = \frac{1}{2} w^T w + \frac{1}{2}\gamma \sum_{i=1}^{n} \xi_i^2 + \sum_{i=1}^{n} \alpha_i (w^T \varphi(x_i) - b + \xi_i - y_i) \tag{11}$$

According to the optimization theory, such as formulas (12)-(13),

$$w = \sum_{i=1}^{n} \alpha_i \varphi(x_i), \sum_{i=1}^{n} \alpha_i = 0, \alpha_i = c\omega_i, \tag{12}$$

$$w^T \varphi(x_i) + b + \xi_i - y_i = 0 \tag{13}$$

to the electronic music signal recognition model form of least squares support vector machine, such as

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x_j) + b \tag{14}$$

$K(x_i, x_j)$ defined as:

$$k(x_i, x_j) = \exp\left(\frac{-\| x_i - x_j \|}{2\sigma^2}\right) \tag{15}$$

where σ is the radial base width.

DCNN-SSA Model

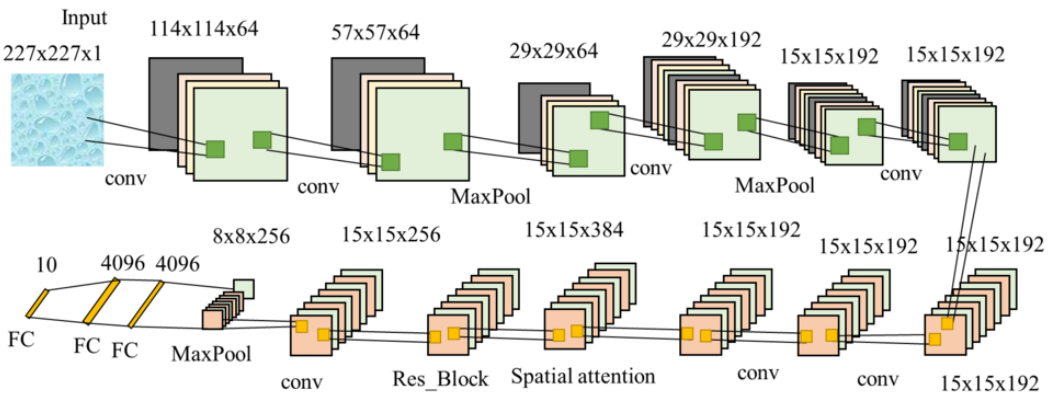
In order to enhance the adaptability of DCNN to audio spectrogram processing, this paper improves the model's extraction effect on spectrogram type features and optimizes the network structure. Through multiple experiments and comparing different model structures, the network structure of the DCNN-SSA model in this paper is finally constructed, as shown in Figure 2.

After the DCNN-SSA model extracts the Mel spectrum of the original audio signal, it restores the data dimension to refine the local features of the audio signal. The processed spectrogram is cut and input to the network. Firstly, through two small kernel convolutions and pooling operations, redundant parameters are reduced and model convergence is accelerated. Secondly, the audio signal correlation dependency of the input network is enhanced through four stacked convolutional layers, and the upcoming spatial domain annotations are preprocessed. Then, the feature-enhanced audio signal is inputted into the spatial attention module to express the genre characteristics between audio signals in the spatial domain. Finally, the labeled audio signal is inputted into the spatial attention module. The residual module fuses multi-level audio features with residual characteristics, describes the audio signals marked in the spatial domain in detail, and effectively extracts the genre characteristics of the audio signals. The model is trained once for all music files in the training set, and all music files in the validation set are validated once as a batch. Then the training model and the validation model are iterated multiple times, and the music genre classification model is outputted when the specified batch is reached.

Improvement 1. A stack consisting of four convolution layers is improved by performing multiple nonlinear mappings, increasing the Receptive field, enhancing the global representation of features in the audio spectrum, and increasing the nonlinear fitting ability of the network model.

Improvement 2. The spatial attention mechanism is introduced, and by paying attention to the spatial domain of features, the correlation dependence on the spatial dimension in the audio spectrogram is enhanced, and the extraction effect of the network model on the rhythmic features in the audio spectrogram is further enhanced.

Figure 2. DCNN-SSA network model structure



Improvement 3. After the attention of the audio spectrogram in the spatial domain, the residual idea is introduced, and the spatially labeled audio spectrogram is described in detail through the residual feature, which further enhances the expression of genre features in the audio spectrogram.

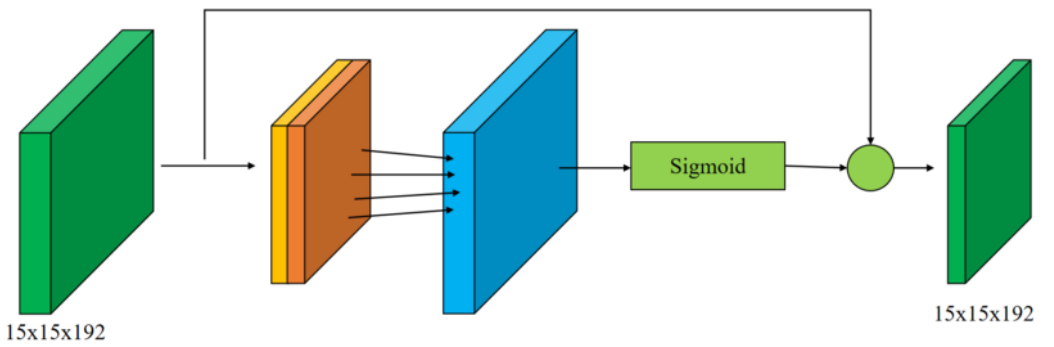
The spatial attention mechanism can perform multi-level spatial domain annotation on the global and local detail features of audio spectrograms through its spatial domain sensitive properties. Multiple convolutional layers are pre-stacked before the spatial attention module to enhance the correlation dependence of information in the spectrogram and provide preprocessing for the feature extraction of the spatial attention module. It has been verified by many experiments that the effect is the best when four convolutional layers are superimposed. Too many convolutional layers can easily cause a lot of computational redundancy and have little effect on the experimental results. The structure of the spatial attention module is shown in Figure 3.

RESULT ANALYSIS AND DISCUSSION

This section introduces the composition and partitioning of the experimental dataset, as well as the selected number of iterations and learning rate. The experimental results of 37,000 iterations are shown in Figure 4, and the experimental results of the influence of Mel spectrum on genre classification results are presented in Figure 5. Ablation experiments were conducted on the effectiveness of the main modules of the model, and the experimental results are shown in Figure 6. Finally, a comparative experiment was conducted between the DCNN-SSA model, support vector machine, and BP neural network, and the experimental results are shown in Figure 7. The experimental results indicate that the DCNN-SSA model performs better in music genre classification tasks, and its effectiveness has been verified.

The experimental dataset consists of two parts: the first part consists of 1,000 MV micro videos and multimedia big data, each of which contains different factors such as different types, styles, and music genres. These videos come from different platforms and production companies, and have undergone past reprocessing and quality evaluation to ensure the integrity and availability of the dataset. The second part is the GTZAN dataset, which contains 1,000 music tracks divided into ten different categories. These tracks come from different music genres, covering common genres such as rock, pop, classical, and jazz. Before conducting the experiment, we divided the above dataset into a training set, a testing set, and a validation set using random sampling method. Among them, the training set accounts for 80% of the dataset, with a total of 800 MV micro videos and 800 music tracks. The test set accounts for 10% of the dataset, with a total of 100 MV micro videos and 100 music tracks. The validation set accounts for 10% of the dataset, with a total of 100 MV micro videos and 100 music tracks. During the partitioning process, we strived to maintain the diversity and

Figure 3. Spatial attention module structure



representativeness of each part of the dataset in order to better evaluate the performance of the model in different scenarios and datasets. Experiments were carried out on the model validation set with a learning rate of 0.001. After 37,000 iterations, the relevant parameters of the model stabilized, as did the accuracy of genre classification. Secondly, during the ablation experiment on the impact of Mel spectrum on genre classification results, it was found that on a validation set with a learning rate of 0.001 and an iteration count of 31,000, Mel spectrum extraction of music genre features can improve the accuracy of music genre classification by 1.92%. Therefore, in order to ensure the accuracy of the experimental results, the number of iterations was chosen to be 37,000/31,000. The experimental results of 37,000 iterations are shown in Figure 4. According to the distribution of experimental results in Figure 4, it can be seen that the accuracy of genre classification increases firstly with the increase of the number of iterations, and then tends to be stable. From Figure 4, it can be considered that the relevant parameters of the model tend to be stable after 31,000 iterations.

Assuming there are n samples in the validation set, and assuming that for each sample, the predicted results of the model are the same as the real labels, then the sample is classified correctly. We record the number of correctly classified samples as k, and the accuracy can be expressed as:

$$Accuracy=k/n$$

Among them, n is the number of samples, and k is the number of correctly classified samples.

To illustrate the influence of Mel spectrum on genre classification results, a feature preprocessing ablation experiment was performed on the validation set with a learning rate of 0.001 and an iteration number of 31,000. The experimental results of feature preprocessing ablation are shown in Figure 5.

By analyzing the experimental results in Figure 5, it can be seen that the extraction of music genre features by Mel spectrum can improve the accuracy of music genre classification by 1.92 percentage points on the validation set compared with the traditional Fourier transform. In order to

Figure 4. Experimental results of 37,000 iterations

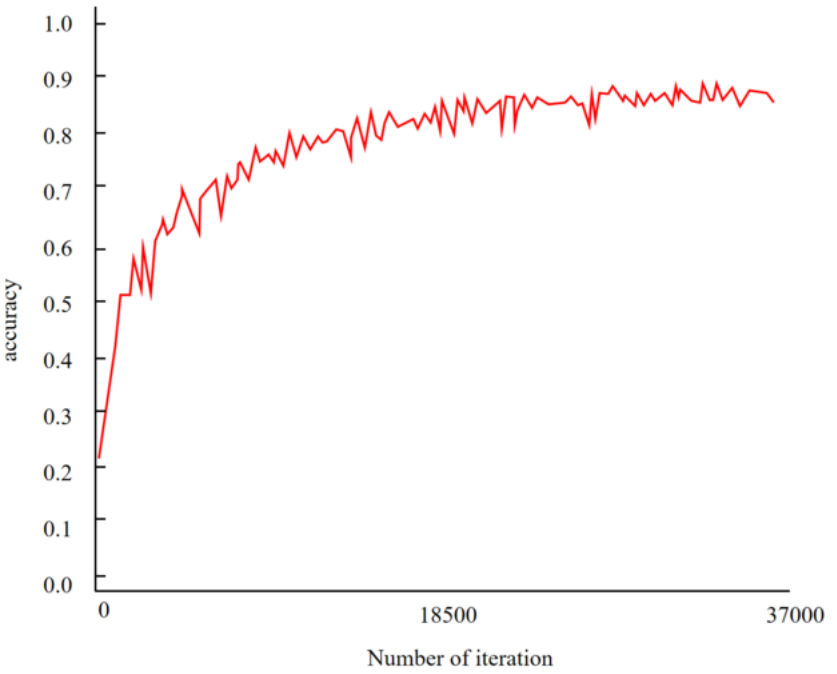
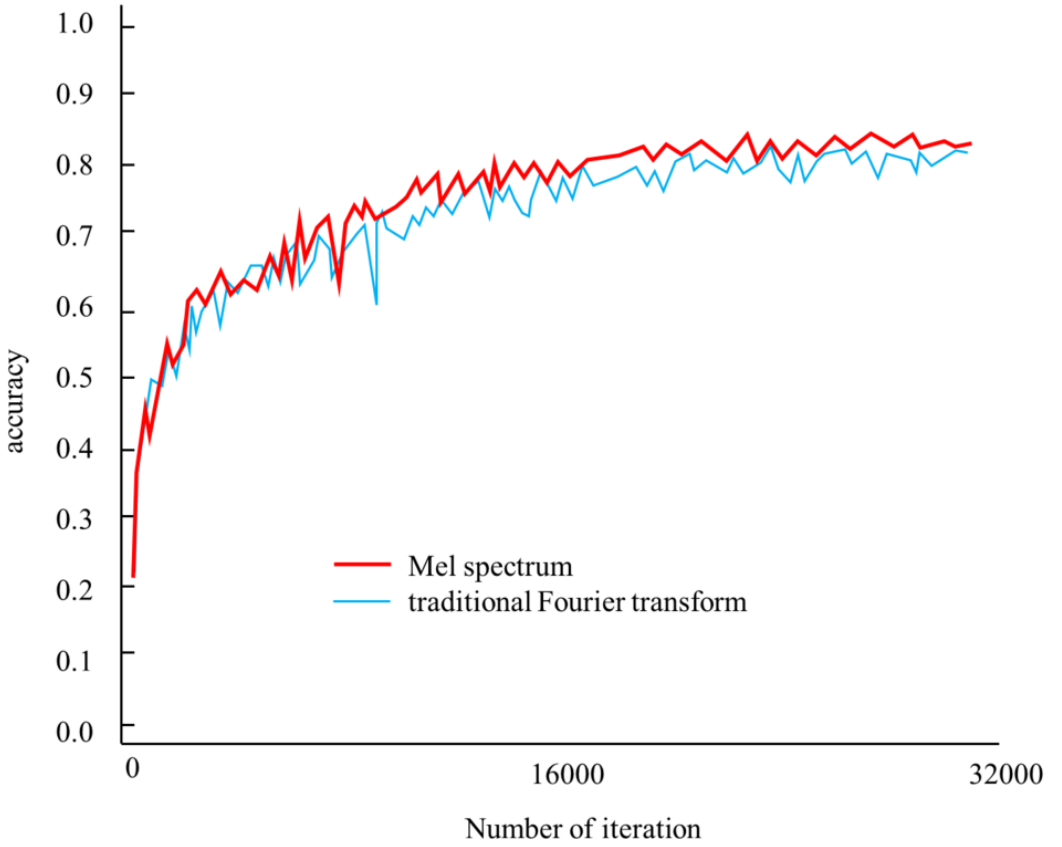


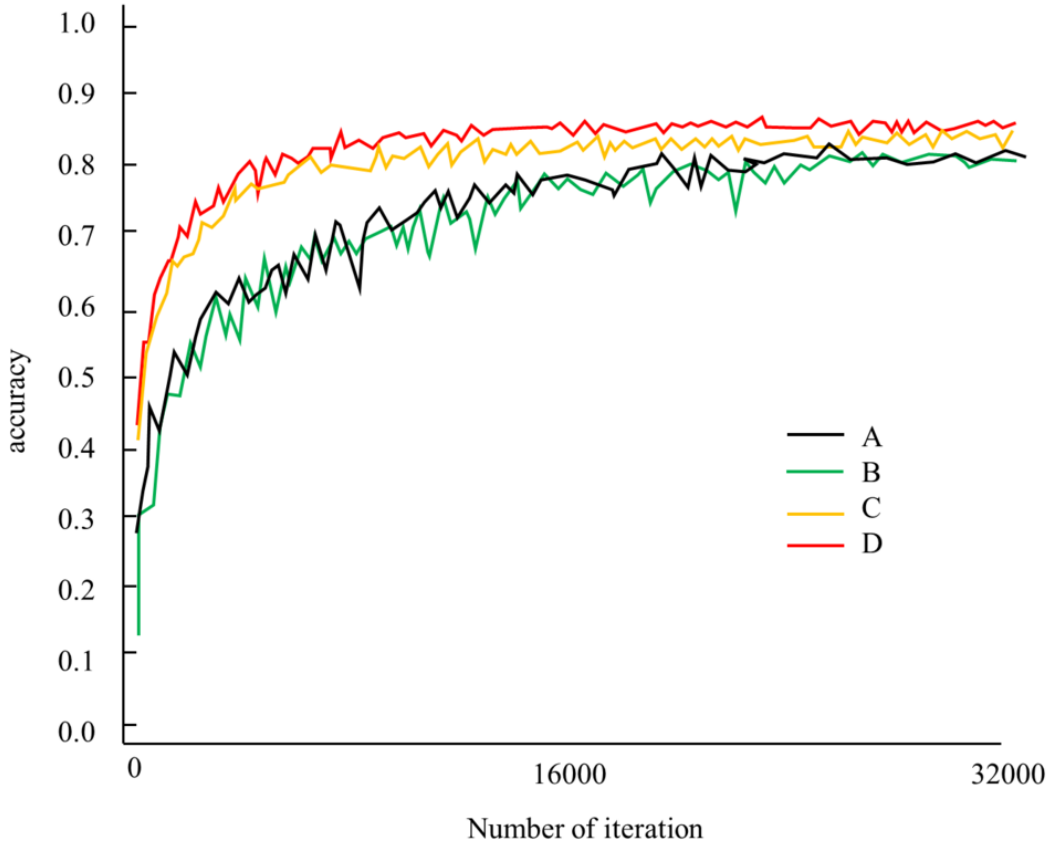
Figure 5. Experimental results of feature preprocessing ablation



verify the effectiveness of the main modules of the DCNN-SSA model, ablation experiments were performed on the validation set with a learning rate of 0.001 and an iteration number of 31,000. The ablation experiment results of the main modules of the validation set model are shown in Figure 6. Experiment a is the original model, experiment b is the enhanced spatial domain attention model, experiments c and d are the auxiliary enhanced models before and after the spatial attention module, and experiment d is the enhanced model. By analyzing the experimental results in Figure 6, comparing experiments a and b, it can be seen that the spatial domain attention of the feature has a gain of 1.27% in the classification effect of music genres. Comparing experiments b and c with experiments b and d respectively, it can be seen that preprocessing the features through quadruple convolution before the spatial attention module can improve the classification effect of music genre by 0.63 percentage points, after the spatial attention module through the residual error. The features after the module refinement and annotation can improve the music genre classification effect by 1.72 percentage points. Comparing experiments e and c with experiments e and d respectively, it can be seen that adding auxiliary enhancement modules before and after the spatial attention module improves the classification effect of music genres by 1.52% to 2.61%. It can be seen that the spatial attention module has improved the accuracy of music genre classification, and both the residual module and the quadruple convolution have an auxiliary enhancement effect on the spatial attention module. And at the same time, the spatial attention module can be enhanced to achieve better results.

In order to verify the effectiveness of the DCNN-SSA model, this paper conducted comparative experiments on two commonly used machine learning algorithms, Support Vector Machine (SVM)

Figure 6. Ablation experiment results of the main modules of the model

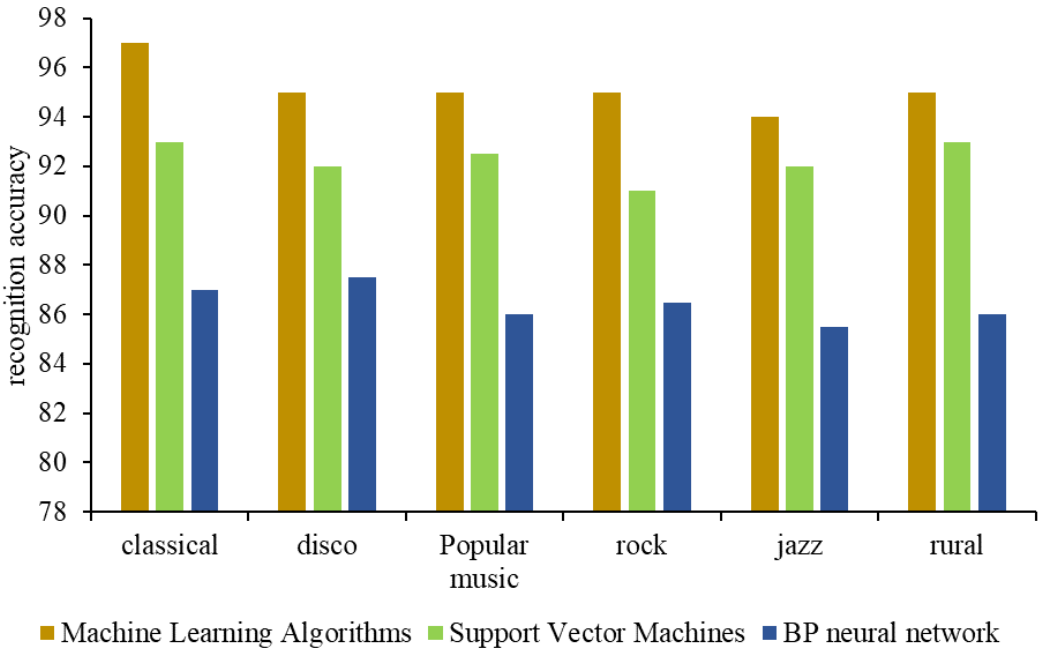


and BP Neural Network (BPNN), and counted their recognition accuracy on six types of electronic music signal validation sample sets and compared them., as shown in Figure 7.

Through the comparative analysis of experimental results in Figure 7, it can be concluded that the DCNN-SSA model proposed in this paper performs better in music genre classification tasks. In the case of Mel spectral feature extraction, DCNN-SSA can achieve higher average classification accuracy compared to support vector machines and BP neural networks, reaching 89.4%. At the same time, the classification performance for six categories has also been significantly improved, with an accuracy rate of over 92%. This proves the effectiveness of the music genre classification algorithm model based on spectral space domain feature attention, and helps to solve the problem of current music signal recognition results not being ideal.

This article aims to explore the application and creative skills analysis of story based MV micro videos and multimedia big data in music dissemination, and proposes a music genre classification algorithm model DCNN-SSA based on spectral space domain feature attention. This study found that the issues and challenges that need to be addressed in the field of digital music communication include multidimensional personalized recommendations and mining user needs from massive data. These challenges are brought about by the impact of digital music dissemination in the era of big data. The dissemination model of digital music proposed by Yang and Hu (2013) provides a deeper analysis and understanding of the application of story-based MV micro videos and multimedia big data in music dissemination. The DCNN-SSA model proposed in this article can effectively solve the problem of music genre classification, which is also one of the important issues that need to be

Figure 7. Music signal is the correct rate



addressed in the field of digital music dissemination. Therefore, this article can be considered as a supplement and support to Yang and Hu's (2013) research, providing a more comprehensive and in-depth exploration, which helps to further understand the dissemination of digital music in the era of big data. In future research, more innovative ways and technological methods can be explored to further tap into the potential of combining MV and micro video, and improve the effectiveness and quality of music dissemination. At the same time, the DCNN-SSA model can be further improved to improve its accuracy, stability, and generalization ability to meet the needs of more practical application scenarios.

CONCLUSION

MV is more easily accepted and recognized by the public through the use of music, lyrics, and video images for narrative and emotional communication. It is a process of Music Visualization and visual musicalization to carry out narrative lyricism through the rhythm of music and the rhythm of image pictures. Chenzhou has a long history and rich local cultural resources. It is necessary to deeply explore the story, character, history, and object resources in the local culture, depict and restore them from such factors as images, sketches, characters, light and shadow, colours, sounds, and environment, and truly integrate the local culture into the audio-visual elements and stories of MV, so that songs, images, history, humanities, and other arts can blend and penetrate and create story MV micro videos close to life and people's emotional needs, spread Chenzhou's local culture, and arouse people's attention to life and thinking about life. This article starts with the selection of songs, story scripts, content design, MV shooting, editing skills, and other aspects of the MV, and introduces the specific creation methods of the story MV in combination with the local culture of Chenzhou.

This paper analyzes and designs a DCNN-SSA model based on a deep convolutional neural network. In the aspect of music genre feature expression, the Mel spectrum restored by dimension is divided and inputted into the network as a training set, so that the training sample can improve

its own feature expression while improving the calculation speed. In terms of music genre feature extraction and quadruple convolution, the DCNN is augmented with a residual structure-assisted spatial attention mechanism. The quadruple convolution can improve the correlation dependence of the internal features of the audio spectrogram in advance, so that stronger rhythm characteristics can be obtained when labeling in the spatial domain. The residual structure can improve the detailed expression of the audio spectrogram after spatial annotation after the spatial attention module. The spatial attention mechanism can effectively enhance the correlation of music spectral features, enhance the extraction of rhythm features in the audio spectrogram, and enhance the directivity of the network for the extraction of audio spectrogram features, thereby improving the application effect.

DATA AVAILABILITY

The figures used to support the findings of this study are included in the article.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING STATEMENT

This work was supported by the Hebei Normal University for Nationalities (No. HQ2021001).

ACKNOWLEDGMENT

The authors would like to show sincere thanks to those techniques who have contributed to this research.

REFERENCES

Abdelazim, A. H., & Ramzy, S. (2022). Application of different quantitative analytical techniques for estimation of aspirin and omeprazole in pharmaceutical preparation. *BMC Chemistry*, *16*(1), 1–8. doi:10.1186/s13065-022-00854-6 PMID:35971129

Calefato, F., Iaffaldano, G., Trisolini, L., & Lanubile, F. (2022). An in-depth analysis of occasional and recurring collaborations in online music co-creation. *ACM Transactions on Social Computing*, *4*(4), 1–40. doi:10.1145/3493800

Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, *63*, 67–80. doi:10.1016/j.dss.2013.08.008

Comba, A., Baldi, A., Tempesta, R. M., Vergano, E. A., Alovisi, M., Pasqualini, D., & Scotti, N. (2021). Influence of curing mode and layering technique on the 3D interfacial gap of bulk-fill resin composites in deep class-I restorations: A micro-CT volumetric study. *The Journal of Adhesive Dentistry*, *23*(5), 421–428. PMID:34549925

De-Deus, G., Souza, E. M., Silva, E. J. N. L., Belladonna, F. G., Simões-Carvalho, M., Cavalcante, D. M., & Versiani, M. A. (2022). A critical analysis of research methods and experimental models to study root canal fillings. *International Endodontic Journal*, *55*(S2), 384–445. doi:10.1111/iej.13713 PMID:35226760

Engin, B., Willis, S. A., Malaikah, S., Sargeant, J. A., Yates, T., Gray, L. J., & King, J. A. (2022). The effect of exercise training on adipose tissue insulin sensitivity: A systematic review and meta-analysis. *Obesity Reviews*, *23*(7), e13445. doi:10.1111/obr.13445 PMID:35319136

Fahrudin, S. H., & Winarni, R. (2021). Analysis of learning speaking skills using the WhatsApp application in elementary schools. In *Journal of Physics: Conference Series*. IOP Publishing. doi:10.1088/1742-6596/1808/1/012033

Feng, H. (2022). Analysis and research on the techniques of digital painting creation. In *Frontier Computing: Proceedings of FC 2021*. Springer Nature Singapore.

Ferguson, S. A., Cheng, K., Adolphe, L., Van de Zande, G., Wallace, D., & Olechowski, A. (2022). Communication patterns in engineering enterprise social networks: An exploratory analysis using short text topic modelling. *Design Science*, *8*, E18. doi:10.1017/dsj.2022.12

Gaeta, M., Loia, V., Mangione, G. R., Orciuoli, F., Ritrovato, P., & Salerno, S. (2014). A methodology and an authoring tool for creating Complex Learning Objects to support interactive storytelling. *Computers in Human Behavior*, *31*, 620–637. doi:10.1016/j.chb.2013.07.011

Garg, G., & Kumar, R. (2022). Analysis of image types, compression techniques and performance assessment metrics: A review. *Journal of Information and Optimization Sciences*, *43*(3), 429–436. doi:10.1080/02522667.2022.2037282

Geiko, P. P., Korolkov, V. A., & Tatur, V. V. (2022). Development and implementation of UV absorption gas analysis techniques for ecological monitoring of the atmosphere. *Atmospheric and Oceanic Optics*, *35*(4), 443–449. doi:10.1134/S1024856022040030

Hornung, O., & Smolnik, S. (2022). At the mercy of our emotions? A multi-dimensional analysis of emotions in knowledge management research. *Knowledge and Process Management*, *29*(2), 109–120. doi:10.1002/kpm.1701

Karim, S., Hussain, E., Khan, J. A., Hameed, A., Ougahi, J. H., & Iqbal, F. (2022). Spatiotemporal investigation of soil salinity using geospatial techniques: A case study of Tehsil Toba Tek Singh. *Communications in Soil Science and Plant Analysis*, *53*(15), 1960–1978. doi:10.1080/00103624.2022.2070189

Lankshear, C., & Knobel, M. (2007). Researching new literacies: Web 2.0 practices and insider perspectives. *E-Learning and Digital Media*, *4*(3), 224–240. doi:10.2304/elea.2007.4.3.224

Liang, H. (2022). The application of multi-source big data mining techniques in the analysis of basketball economic management. *Mathematical Problems in Engineering*, *5362900*, 1–10. Advance online publication. doi:10.1155/2022/5362900

Liu, L. (2021). Realistic independent documentary: Four springs. *Education Reform and Development*, *3*(2), 70–73. doi:10.26689/erd.v3i2.3596

Liu, Q., Chen, X., Dai, X., Liu, X., Xu, F., & Peng, P. (2021). Comparative analysis of five inspection techniques for the application in the diagnosis and treatment of osteoarticular tuberculosis. *International Journal of Infectious Diseases*, *112*, 258–263. doi:10.1016/j.ijid.2021.09.019 PMID:34536611

Maybury, M. T. (2012). *Multimedia information extraction: Advances in video, audio, and imagery analysis for search, data mining, surveillance and authoring*. John Wiley & Sons. doi:10.1002/9781118219546

Nakano, S., & Washizu, A. (2022). *Creation and application of the 2015 input-output table for analysis of next-generation energy systems: Analysis of the effects of introducing carbon tax* (No. 2103) [Doctoral dissertation, Research Institute for Environmental Economics and Management]. Waseda University.

Pan, H., Li, H., Yan, Y., Mao, W., Li, G., & Yin, C. (2020). Analysis on new techniques of secondary and auxiliary control system in intelligent substation. *Hunan Electric Power*, *40*(4), 68.

Pistola, T., Diplaris, S., Stentoumis, C., Stathopoulos, E. A., Loupas, G., Mandilaras, T., & Kompatsiaris, I. (2021). Creating immersive experiences based on intangible cultural heritage. In *2021 IEEE International Conference on Intelligent Reality (ICIR)*. IEEE. doi:10.1109/ICIR51845.2021.00012

Saha, V., Mani, V., & Goyal, P. (2020). Emerging trends in the literature of value co-creation: A bibliometric analysis. *Benchmarking*, *27*(3), 981–1002. doi:10.1108/BIJ-07-2019-0342

Wang, D. (2022). Research on the art value and application of art creation based on the emotion analysis of art. *Wireless Communications and Mobile Computing*, *2022*, 1–10. doi:10.1155/2022/2435361

Yang, L., & Hu, D. (2013). The dissemination model of digital music in big data era. In *2013 International Conference on Computer Sciences and Applications*. Wuhan, China: IEEE. doi:10.1109/CSA.2013.195

Yost, R. A. (2022). Why tandem mass spectrometry for trace analysis: Concepts of tandem analytical techniques. *Rapid Communications in Mass Spectrometry*, *36*(13), e9310. doi:10.1002/rcm.9310 PMID:35384102

Xi Zhang works in the department of Music and Dance, Hebei Normal University for Nationalities.

Yue Cui is a professor of the Department of Music and Dance, Hebei Normal University for Nationalities.