



The use of fuzzy authentication integrated with convolutional neural networks in digital content protection

Junfeng Wang¹

Accepted: 16 October 2023 / Published online: 3 November 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

In order to explore the application of neural networks in the intelligent protection of digital music copyrights, this paper initially investigates the key aspects of digital music-related supply chain management (SCM). Subsequently, this paper introduces a fuzzy, comprehensive evaluation-based method for authenticating digital music infringement. Furthermore, it employs a convolutional neural network to analyze digital music's source components. Within this context, an attention mechanism and an adaptive gate model are meticulously designed. Moreover, a selective adaptive cascade is implemented to optimize the structure of the multi-resolution coder–decoder. Through fuzzy authentication, it becomes feasible to accurately and reliably track and verify copyright information pertaining to digital music, thus providing an effective protective mechanism for the digital music industry. Experimental results reveal that in comparison with the SHN-4 series models, the network model embedded with the channel space (CS) attention module achieves improvements of up to 0.04 dB in terms of performance metrics for human voice and accompaniment sources. Furthermore, it outperforms these models by up to 0.19 dB in music source separation metrics. The incorporation of the adaptive attention mechanism profoundly enhances the model's performance in audio source separation tasks, benefiting both vocal and accompaniment sources. This improvement is evident in the form of a 0.44 SDR enhancement for vocal sources and a 0.18 SDR enhancement for accompaniment sources. Additionally, the proposed two-stage music separation-gate self-attention model surpasses the two-stage music separation-gate model in metrics associated with human voice and accompaniment sources. This outcome convincingly underscores the enhanced effectiveness achieved by integrating CS, self-attention, and adaptive connection in network performance, as well as the adaptive gate structure for music copyright protection. This approach heightens the model's autonomous selectivity, minimizes outliers, stabilizes waveform characteristics, and significantly contributes to the safeguarding of digital content.

Keywords Supply chain management · Digital content protection · Fuzzy authentication · Convolutional neural network

1 Introduction

The infringement of digital music encompasses various stakeholders, including online music platforms, music player software developers, and singing software providers, ultimately resulting in the loss of interest for numerous parties [1, 5, 22]. With the continuous evolution of the Internet, conflicts among online music service providers regarding copyright infringement have grown increasingly severe. The rise of We-Media and the short video industry, combined with China's incomplete copyright laws related to music copyrights, have made the protection of digital music copyright more challenging [11, 19, 21]. Service providers on network platforms often disseminate copyrighted works in cyberspace without obtaining proper authorization, primarily driven by profit motives, making this a relatively common form of infringement [3, 17, 28]. Notably, the downloading and manipulation of digital music for the creation of short videos on platforms like Tiktok and Snack Video can potentially lead to infringements of digital music rights, thereby compromising the rights and interests of copyright owners [4, 6, 26]. Therefore, it holds practical significance to investigate copyright infringements and protection issues within the realm of digital music applications and to explore effective protection strategies.

Identifying the copyright of digital music is a challenging endeavor. This complexity arises from the fact that digital music compositions fall under the protective scope of copyright law, but their intersecting relationships with copyright, network transmission rights, and linking rights have heightened the difficulty of ascertaining digital music copyrights [8, 10, 14]. Implementing effective supply chain management (SCM) measures throughout the entire lifecycle of digital music, from creation to release, authorization, and reproduction, presents a viable approach for digital music copyright owners [23, 32, 34]. The advent of technologies like artificial intelligence (AI) and blockchain in the information age has introduced novel avenues for safeguarding digital music copyrights. Manual review and tracing of the complete usage process of digital music can significantly enhance the efficiency of music protection management and certification [20, 30]. Utilizing AI to detect song similarities to assist in determining instances of plagiarism by creators promises to expedite plagiarism identification efforts [29]. However, it is important to note that no powerful AI software can definitively ascertain whether a musician has engaged in plagiarism based solely on music scores. Moreover, within the realm of the law, plagiarism, as a form of infringement, cannot be conclusively determined by examining a musical score or a musical composition. Although the digital music industry has already become a global commercial giant, digital music copyright protection still faces significant challenges. Traditional methods of digital music copyright protection often rely too heavily on technological safeguards and lack a thorough consideration of SCM. This paper delves into the SCM related to digital music, including various aspects such as music creation, recording, distribution, and playback. This helps people gain a comprehensive understanding of the ecosystem of the digital

music industry. Secondly, this paper proposes a digital music infringement identification method based on fuzzy comprehensive assessment. Through this method, it can associate with different nodes in the supply chain to identify potential sources of infringement. This kind of association aids in tracing the origin of infringements, thereby assisting music copyright holders in better protecting their rights. Finally, this paper independently investigates digital music audio sources using a convolutional neural network (CNN). Introducing attention mechanisms and adaptive gating models optimizes the multi-resolution encoder–decoder structure. This optimization improves the accuracy and reliability of copyright information while positively impacting SCM. By tracking digital music copyright information better, the music industry can more effectively manage the music content within the supply chain.

The primary contribution of this paper lies in the integration of neural network technology with digital music copyright protection and SCM, filling a research gap in this field. This paper first delves into various aspects of SCM related to digital music, including music creation, recording, distribution, and playback. Then, it proposes a digital music infringement identification method based on a fuzzy comprehensive assessment. It applies CNN for audio source separation, optimizing model performance by introducing attention mechanisms and adaptive gating models. Ultimately, this paper tracks and verifies digital music copyright information through fuzzy authentication methods, providing a comprehensive protection mechanism that enables the digital music industry to address the challenges of copyright protection.

The paper is structured into five primary sections. Section 1 serves as the introduction, offering an overview of this paper's background, significance, and current challenges. Section 2 encompasses the literature review, which surveys existing research and achievements to delineate the present research landscape. Section 3 delves into the paper methodology, investigating the utilization of CNN in the domain of music copyright protection. Section 4 comprises experiments carried out to ascertain the effectiveness and advantages of the proposed research approach, presented both graphically and textually. Lastly, Sect. 5 encapsulates the conclusion, summarizing the paper's findings, underscoring limitations, and providing insights into prospective avenues for further investigation.

2 Overview of the work model

Hence, this work first collects and prepares a digital music dataset, including music source files and copyright information. These data are used for model training and testing. Additionally, this paper also analyzes various aspects of digital music SCM, including creation, recording, distribution, etc., to understand the industry ecosystem. Next, the paper delves into an exploration of the primary facets of SCM pertaining to digital music. Subsequently, it elucidates the viability of employing the fuzzy authentication method within the realm of safeguarding digital content. Finally, it leverages CNN for an in-depth investigation into the musical sources within the domain of digital music. This entails the design of an attention mechanism and an adaptive gate model. Additionally, a selective adaptive

cascade approach is incorporated to optimize the architecture of the multi-resolution coder–decoder. This optimization serves to enhance the feature sensitivity and accuracy of music source separation, mitigate the distortion of human voice and accompaniment, and contribute to the preservation of digital music copyright. The paper process is shown in Fig. 1.

3 Literature review

3.1 SCM

SCM, an abbreviation for SCM, is a concept embodying a chain-like structure. As per the operational model of SCM, an enterprise stands at the core, with customers located two layers downstream and two layers upstream. Alongside the customers of customers and the suppliers of the enterprise, this network collectively forms a comprehensive external supply chain catering to downstream customers. Within the confines of any internal enterprise node, the sequence of processes, ranging from order receipt to planning, procurement, production, logistics, and distribution, constitutes a more narrowly defined internal supply chain. SCM encompasses the management approach governing product manufacturing, transportation, distribution, and sales. It effectively orchestrates the cooperation of suppliers, manufacturers, distribution centers, and channel partners, all with the overarching goal of achieving the minimal total cost across the entire supply chain system while maintaining a specific level

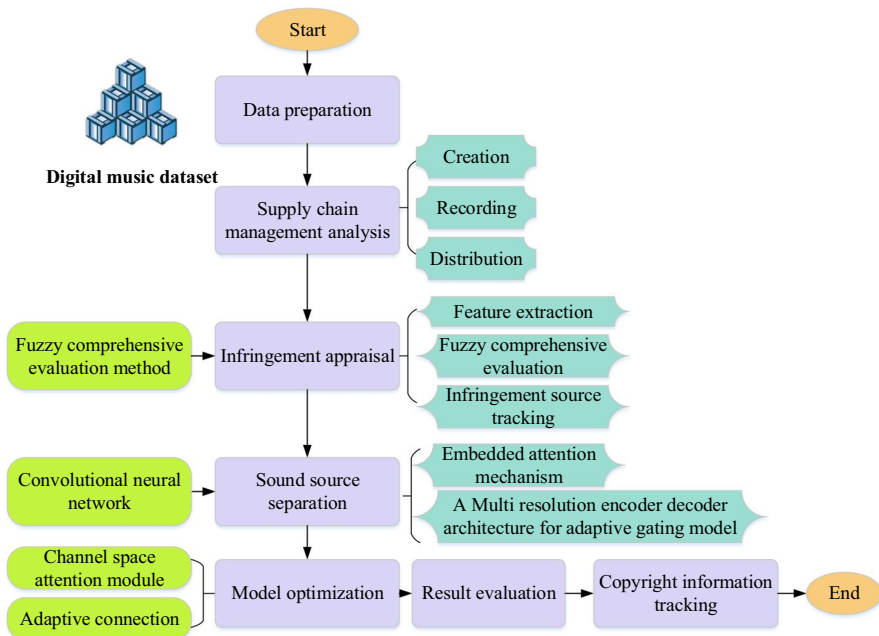


Fig. 1 Research process

of customer service. SCM comprises five fundamental elements: planning, procurement, manufacturing, distribution, and reverse logistics, which includes the return of goods.

In the paper conducted by Pournader et al. [31], they developed and validated an AI classification method. This method was subsequently utilized as a metric for bibliometrics and co-citation analysis. Jia et al. [16] introduced a framework founded on supply chain leadership, multi-level supply chain governance, multi-level supply chain structure, and supply chain learning. Their findings indicated that the combined influence of supply chain leadership and governance mechanisms had a significant impact on both supply chain structure and supply chain learning. Multinational corporations adapted their supply chain structures to foster supply chain learning. Asamoah et al. [2] delved into the impact of inter-organizational systems on an organization's SCM capabilities and supply chain performance. Utilizing a resource-based perspective, they scrutinized two critical mechanisms for enhancing supply chain performance: the effective external utilization of inter-organizational systems with network partners and the maximization of organizational management capabilities related to inter-organizational systems in SCM. Gölgeci and Kuivalainen [13] examined the roles of absorptive capacity and marketing-SCM consistency in potentially influencing social capital's impact on supply chain resilience. They empirically assessed these relationships and discovered that absorptive capacity mediated the connection between social capital and supply chain resilience. Furthermore, the link between social capital and both absorptive capacity and supply chain resilience was more pronounced when marketing and SCM exhibited a high degree of consistency.

In summary, most of the existing SCM research primarily focuses on the enterprise aspect, emphasizing the internal operations of the supply chain and corporate performance. In contrast, the paper is centered on the field of digital music content protection, which is a relatively underexplored domain. The aim is to fill the paper gap in the field of digital music content protection and provide valuable insights into this specific area. Unlike other studies, this paper aims to explore how AI and deep learning technologies, combined with SCM principles, can be applied to achieve digital music content protection. This represents a unique research direction involving interdisciplinary content and contributes to the opening up of new research areas. Additionally, this paper not only focuses on theoretical exploration but also places importance on practical application. Digital music content protection is a critical issue in the digital era that affects the sustainable development of the music industry.

3.2 Digital music copyright

Digital music copyright constitutes a niche within the realm of digital copyright, specifically referring to musical works crafted by their respective copyright owners in digital format. Its primary emphasis lies in the digital manifestation of music copyright from its inception, bearing considerable influence on both the music industry's trajectory within China and on a global scale. It plays an irreplaceable role in shaping the broader musical ecosystem's construction. The transmission medium for

digital music works relies on binary code as a digital signal, with the internet serving as the conduit. However, this reliance on the internet also introduces significant challenges in the realm of digital content protection, as these works can be readily copied and disseminated across network coverage areas. The high frequency and diverse array of infringement types further compound the complexities surrounding the safeguarding of digital music copyright.

In the digital era, network-based infringements exhibit greater audacity due to their low cost, rapid dissemination, a diverse array of actors, and a legal recourse process that proves more intricate compared to traditional domains. Consequently, copyright protection for digital music within this fragmented digital landscape emerges as a profoundly challenging subject. Chen [9] conducted a historical and socio-legal analysis of China's copyright law development within the music industry. This analysis posits that China's digital music industry has matured to a juncture marked by the convergence of three distinct business models: the cultural adaptation model, the rebellious model, and the platform ecosystem model. Kariyawasam and Palliyarachchi [18] delved into the recognition of performers' rights, who are typically contributors to musical creations under copyright law. Their work focused on the evolution of performers' rights in Australia, drawing comparisons with the United Kingdom and New Zealand. In contrast, Zhang et al. [38] designed and implemented a digital music rights management system using blockchain technology. This system leveraged blockchain for establishing evidence solidification and verifying music copyright, incorporated the Shazam algorithm to furnish authentic proof of music copyright, and harnessed smart contracts to bolster transaction security and reliability.

Shah [33] illuminated instances where the Indian film industry drew inspiration from Western proprietary works, crafting unauthorized derivatives in the process. Consequently, the need for India to enhance its copyright enforcement mechanisms, condemn infringing activities, and foster international collaborations, particularly with the United States and other developed nations, to nurture a global environment conducive to the protection of proprietary works was underscored. Cai [7] constructed a monophonic melody composition model grounded in deep generative adversarial networks and evaluated its performance using hymns as input samples. Furthermore, a multi-instrument co-authoring model founded on multi-task learning was proposed, and its composition capabilities were analyzed with actual music as input samples. Additionally, a blockchain-based digital content protection system was conceptualized. Wang et al. [36] introduced a novel deep learning neural network designed for music creation. This model incorporated a meticulously crafted reward function to adjust the probability distribution governing the generation of long short-term memory (LSTM), all while adhering to music theory principles to enable intelligent generation of specific musical styles. These works collectively underscore the persistent ambiguities within music copyright protection and advocate the utilization of neural networks as a promising avenue for resolution.

The paper of the aforementioned scholars primarily focuses on copyright laws in the music industry, music creation models, and business models in the music industry. While these studies provide valuable background information, they do not delve deeply into the field of digital music content protection, especially regarding how

AI and deep learning technologies can be applied to address issues of digital music infringement. The paper fills this gap and emphasizes technical solutions for digital music content protection. Furthermore, although previous research has emphasized legal and business models, it has not touched upon specific technological applications. This paper introduces an innovative approach to digital music content protection by combining techniques such as fuzzy comprehensive assessment and CNN. This type of technological application has not been extensively explored in digital music content protection. Finally, this paper is interdisciplinary, covering computer science, AI, and the music industry. This sets it apart from previous research, emphasizing the intersection between technology and copyright protection.

4 Research methodology

4.1 Digital music infringement authentication based on fuzzy comprehensive evaluation (FCE)

Due to the particularity of digital music itself, the identification of its infringement cannot be explained by a single indicator. Fuzzy comprehensive assessment is a multi-factor decision-making method that handles uncertainty and fuzziness in input information. In digital music infringement verification, a fuzzy comprehensive assessment determines whether a music segment is infringing. This method first requires identifying a set of evaluation indicators that reflect various characteristics of the music segment, such as spectral properties, time-domain characteristics, and rhythm. These indicators are used to quantify the attributes of the music segment. Subsequently, membership functions are constructed for each evaluation indicator. These functions define the degree of membership of each evaluation indicator in different value ranges. Typically, membership functions can be represented using shapes like triangles, trapezoids, and so on. Secondly, weights are assigned to each evaluation indicator, representing the importance of different indicators in the final decision. Weights can be determined through expert opinions, data analysis, or other methods. Finally, using the fuzzy comprehensive evaluation method, the values of each evaluation indicator and their respective weights are combined to calculate a comprehensive score indicating the likelihood of music segment infringement. This score can be a value ranging between 0 and 1, representing the probability of infringement. Depending on specific requirements, a decision threshold can be set to determine whether to classify the music segment as infringing. If the comprehensive score is higher than the threshold, infringement is considered to exist; otherwise, it is considered not to exist. The specific steps are as follows:

The initial phase entails the establishment of the factor set attributed to the evaluation subject. Hypothesis:

$$U = \{u_1, u_2, u_3, \dots, u_m\} \quad (1)$$

Let U represent the ensemble of m evaluation metrics pertaining to the evaluated entity. During the actual assessment process, primary metrics directly linked

to the evaluated entity are denoted as first-tier metrics, while secondary metrics, which may exert influence on the outcomes of the first-tier metrics, are categorized as second-tier metrics. This analogical approach can be extended to acquire tertiary-level metrics, thereby formulating a multi-tiered evaluation index system.

The subsequent phase involves defining the assessment outcomes ensemble for the evaluation entity. This ensemble encapsulates the array of conceivable assessment conclusions that the evaluator may derive following a sequence of evaluation steps concerning the evaluation subject. Hypothesis:

$$V = \{v_1, v_2, v_3, \dots, v_n\} \quad (2)$$

In Eq. (12), n represents the total count of potential assessment outcomes, with n generally adhering to the condition $n \leq 3$.

The third phase encompasses the determination of the weight vector pertaining to evaluation factors. Each evaluation metric within the evaluation framework contributes distinctively to the ultimate assessment outcome. In essence, the weight associated with each metric profoundly influences the conclusive outcome of the comprehensive evaluation. Therefore, it is imperative to quantify the weight coefficients, a process realized through the analytical hierarchy process.

$$A = \{a_1, a_2, a_3, \dots, a_m\} \quad (3)$$

$$a_i > 0 \quad (4)$$

$$\sum a_i = 1 \quad (5)$$

a_i represents the weight attributed to the i th element.

The subsequent phase involves the execution of single-factor fuzzy evaluation and the establishment of the fuzzy relation matrix denoted as R . This process hinges on the expert assessments of individual factors, which determine the degree of membership for each single factor. The matrix resulting from the assessment outcomes of each individual factor constitutes the fuzzy relation matrix R .

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{pmatrix} \quad (6)$$

In Eq. (6), r_{ij} signifies the membership degree denoting the assessed entity's affinity to the hierarchical fuzzy subset v_j within the context of the indicator u_i . r_i signifies the single-factor fuzzy evaluation matrix, delineating a fuzzy correspondence between the indicator set U and the comment set V .

The final phase entails the comprehensive evaluation of multiple metrics. Building upon the preceding stages, the fuzzy weight vector A and the fuzzy relation matrix R for each hierarchical element amalgamate to yield the fuzzy

comprehensive result vector denoted as B for the evaluation subject at this hierarchical level. The FCE model is represented by Eq. (7):

$$B = A \circ R = (a_1, a_2, \dots, a_m) \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{pmatrix} = (b_1, b_2, \dots, b_n) \quad (7)$$

During the evaluation process, the introduction of fuzzy mathematics calculations enables a comprehensive assessment of multi-level indicators and ultimately synthesizes a fuzzy evaluation result vector. This approach enhances objectivity in the results acquisition, effectively mitigating the deviations that traditional qualitative evaluations might introduce. Consequently, the adoption of the FCE method stands as an excellent solution for addressing the challenges of quantitatively analyzing digital music infringement, which can be notably complex.

4.2 Music source separation algorithm based on CNN

Within the realm of the US music industry and entertainment law, copyright infringement lawsuits often hinge on determining substantial similarity by comparing the functional spectra of two songs. This approach aids in resolving issues related to music copyright protection. Traditional identification methods frequently struggle to definitively ascertain whether two songs exhibit complete similarity or contain elements of plagiarism. In such scenarios, AI-based music source separation techniques can prove invaluable. These techniques facilitate the separation of distinct music tracks into separate audio components. Following intelligent processing, digital music can undergo further comparative analysis to determine the presence of infringement, significantly contributing to dispute resolution.

The music source separation algorithm aims to maximize the isolation of human voice sources and accompaniment sources from the audio mix. Figure 2 provides a general structural overview.

In Fig. 2, music source separation algorithms are a category of audio processing techniques designed to extract different audio sources (such as vocals and accompaniment) from mixed audio signals, allowing for the individual retrieval of each audio source. The objective of these algorithms is to automatically identify and separate different audio sources through computer-based methods, thereby enhancing applications in audio processing, music production, and music copyright protection. During the training phase, the mixed musical source is constructed by linearly superimposing the human voice source and the accompaniment source. Typically, time–frequency decomposition processes involve the utilization of time–frequency transformation techniques, which convert time-domain characteristics into spectral features. These spectral features are then incorporated into the model training alongside various distinct separation targets. Continuous training of the model using the mixed musical source facilitates the refinement of network parameters through a loss function, ultimately enhancing the separation model's performance. Given the necessity for precise measurements of each audio segment, it is essential to select

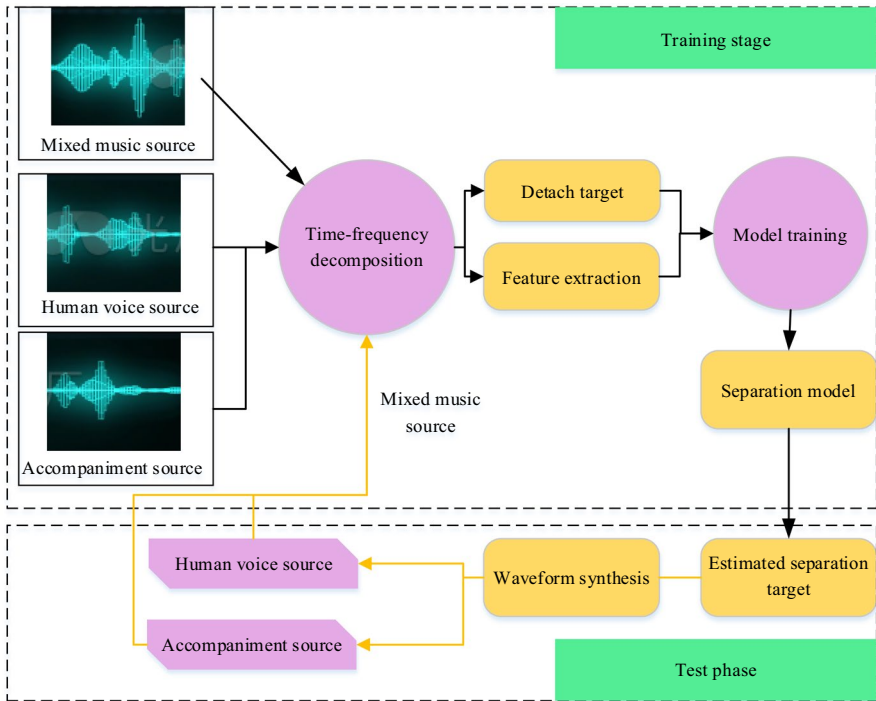


Fig. 2 Structure diagram of music source separation algorithm

suitable evaluation metrics. Different evaluation metrics offer varied perspectives on audio quality and the efficacy of the music source separation model. In this context, three metrics have been chosen to assess the disparity between each source audio after separation, specifically the source distortion ratio (SDR), source interference ratio (SIR), and source artifact ratio (SAR). The calculation methods for these metrics are presented in Eqs. (8)–(10).

$$SDR = 10 \log_{10} \frac{\|s_{target}\|^2}{\|s_{interf} + s_{noise} + s_{artif}\|^2} \tag{8}$$

$$SIR = 10 \log_{10} \frac{\|s_{target}\|^2}{\|s_{interf}\|^2} \tag{9}$$

$$SAR = 10 \log_{10} \frac{\|s_{target} + s_{noise} + s_{artif}\|^2}{\|s_{artif}\|^2} \tag{10}$$

In Eqs. (8)–(10), s_{target} represents the desired decomposition waveform of the source signal s , s_{interf} denotes the interference waveform resulting from the residual source signal in relation to the current decomposition waveform s_{target} and s_{artif} signifies artificially introduced disturbances. The variable s_{noise} accounts for errors introduced by disruptive noise, typically assuming a value of 0 in cases where disruptive noise is not taken into consideration. Higher values for these three metrics correspond to improved audio quality of the target source and enhanced model performance.

Given the utilization of models for predicting masks between sources based on masking principles, the mask code relies on the interplay between these sources. Should the frequency and energy characteristics of each digital music source bear significant similarities, the predictive mask code may diminish separation performance while surreptitiously augmenting interference between the sources. The CNN-based music source separation algorithm aims to separate different sources, such as vocals and accompaniment, from mixed audio. Firstly, a training dataset is prepared, which includes mixed audio along with corresponding individual vocal and accompaniment audio. These data are used to train the CNN model. Features are extracted from the mixed audio, typically using a short-time Fourier transform to convert the audio into a time–frequency representation. This results in input data for training. Secondly, a CNN architecture is constructed, typically consisting of convolutional layers, pooling layers, and fully connected layers, to learn the separation of different sources from the input audio. Finally, the CNN model is trained using the prepared training data through the backpropagation algorithm. The model adjusts its weights and parameters to minimize the error between the mixed audio and the known sources. Once the model is trained, it can be used to separate different sources from new mixed audio. By passing mixed audio through the model, individual vocal and accompaniment audio can be obtained. The goal of this algorithm is to improve the effectiveness of music separation, enhancing the clarity of audio sources to enhance the quality and copyright protection of digital music. It has widespread potential applications in areas such as music production, music copyright management, and music remixing. Moreover, modern coder–decoder CNN frequently integrates skip connections or similar links to facilitate the transmission of low-resolution vocal and accompaniment characteristics from input to output. However, this direct linkage may lead to input features circumventing the bottleneck layer’s filtering process, impeding the extraction of essential music source features through dimensionality reduction, ultimately diminishing music source separation performance. Consequently, an attention mechanism has been developed within this context. In order to optimize the structure of the multi-resolution coder–decoder, a selective adaptive cascade approach has been employed. This optimization aims to augment feature sensitivity and accuracy in music source separation while concurrently mitigating the distortion rate impacting human voice and accompaniment.

In order to further enhance the adaptability and simplicity of the control gate structure governing music sources, an adaptive gate model has been introduced. This model delves into the interrelationships among various music sources based on the attention mechanism. Figure 3 provides an illustration of the adaptive gate’s configuration.

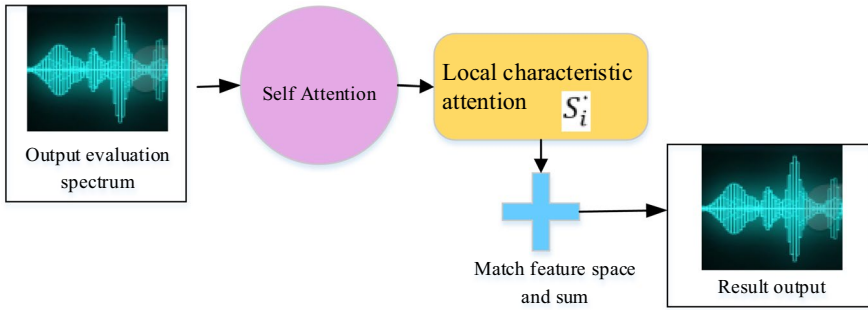


Fig. 3 Structure diagram of the adaptive gate for two-stage music source separation

In Fig. 3, the adaptive gating algorithm for two-stage music source separation refers to a music source separation method in which an adaptive gating mechanism is introduced to enhance separation performance. This algorithm typically involves a two-stage processing, where, firstly, music sources in the mixed audio are estimated and separated in some way, and then an adaptive gating mechanism is applied to improve the accuracy and quality of the separation. The diagram illustrates the division of the adaptive gate into three distinct components. The first component comprises the output evaluation spectrum diagram from the initial stage, representing the evaluation feature diagrams denoted as $S_1, S_2,$ and S_3 for each source. These diagrams conform to the requirements specified in Eq. (11):

$$S_n \in R^{H \times W} \tag{11}$$

In Eq. (11), n represents the number of evaluation sources, H signifies the frequency amplitude of the spectrum diagram, and W denotes the frame number of the spectrum diagram. Moving to the second component, denoted as S_i^{\cdot} , it reflects the attention assigned to each S_i through the utilization of the self-attention mechanism and perception of salient local frequency characteristics. i pertains to the i th evaluation source. Additionally, it is crucial to consider the low-noise relationship inherent in S_i due to the presence of other sources. The specific expressions are elucidated in Eqs. (12)–(15):

$$S_i^Q = S_i W^Q \tag{12}$$

$$S_i^K = S_i W^K \tag{13}$$

$$S_i^V = S_i W^V \tag{14}$$

$$S_{i'} = \text{Softmax} \left(\frac{s_i^Q S_i^{K^T}}{\sqrt{d}} \right) S_i^V \tag{15}$$

where

$$W^Q, W^K, W^V \in R^{W \times W} \quad (16)$$

$$S_i^Q, S_i^K, S_i^V \in R^{H \times W}. \quad (17)$$

The variable d signifies the spectrum diagram channel. Within the architecture of the coder–decoder CNN, there exist two methods of connecting the encoder and decoder. The first involves concatenating the summation of features extracted from both the encoder and the decoder, while the second entails concatenating the codec feature matrix. Each of these approaches exhibits certain limitations, such as the potential expansion of the channel count. Attempting to maintain the channel count by introducing an additional convolutional layer results in a substantial increase in model parameters and constrains the available feature space. Consequently, the decoder’s performance within the model becomes constrained. In light of these considerations, an adaptive connection structure has been devised (as depicted in Fig. 4).

Figure 4 reveals that the encoder–decoder adaptive connection CNN model is a deep learning model used for music source separation. This model typically consists of an encoder section and a decoder section. The encoder is used to transform input audio signals into latent feature representations, while the decoder is employed to reconstruct separated audio sources from these feature representations. The term “adaptive connection” refers to the presence of dynamic adjustment mechanisms within the connection between the encoder and decoder. These mechanisms adaptively adjust the connection based on the features of the input data to enhance the model’s performance and adaptability. The coder–decoder adaptive structure comprises a total of eight components. The initial component involves convolution pre-processing. Initially, a single-track mixed spectrum is input and subjected to pre-processing via five consecutive convolution kernels within the CNN. The primary objective of this stage is to facilitate the seamless transition of input spectrum features into the encoder section of the second component. The second component encompasses two modules: the channel space (CS) attention module and the down-sampling module. Within the encoder section, four down-sampling modules and five CS modules are incorporated. In order to compensate for the inherent limitation of reduced resolution in the bottleneck layer, the third component employs self-attention mechanisms to emphasize the features of the bottleneck layer, thereby extracting more critical information and enhancing the expressive features of the bottleneck layer.

5 Experimental design and performance evaluation

5.1 Dataset collection, experimental environment

The primary objective of this experiment is to validate the efficacy of the attention mechanism within the established model. In order to achieve this goal, the experiment incorporates the existing convolutional attention squeeze-and-excitation (SE)

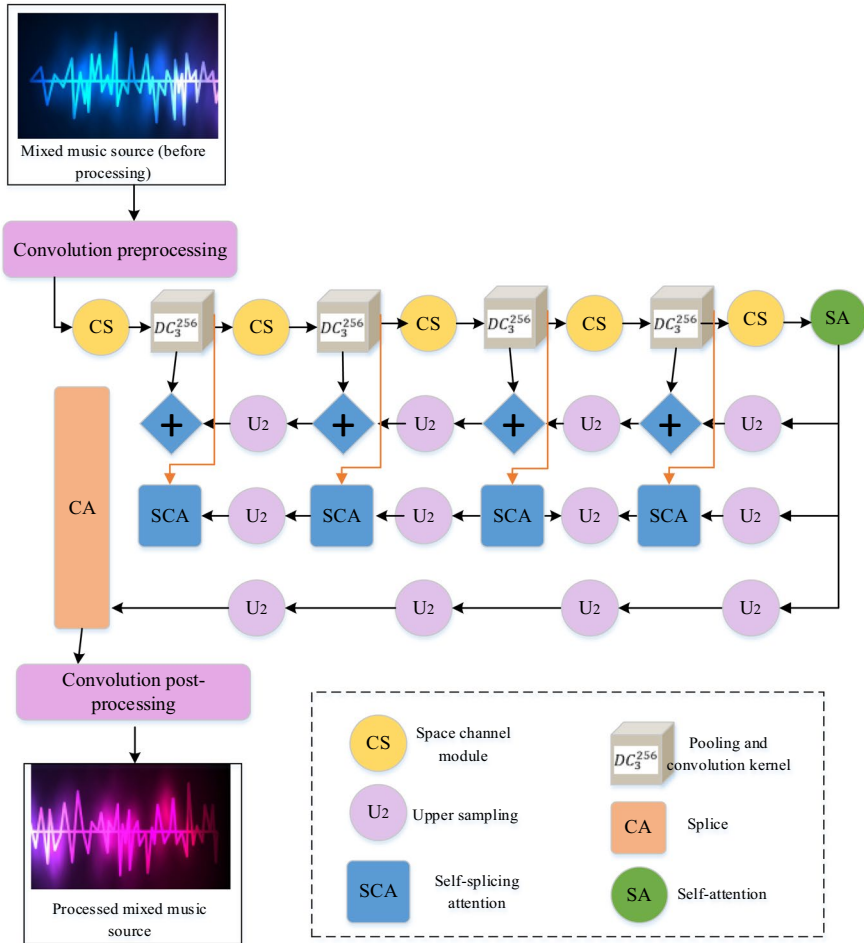


Fig. 4 CNN model diagram of the coder–decoder adaptive connection

module into two network models, namely U-Net and SHN-4. Consequently, two models, denoted as U-Net-SE and SHN-4-SE, are constructed. This experiment aims to compare the performance differences between models that utilize convolutional attention mechanisms and those that do not when dealing with different datasets, thus validating the effectiveness of convolutional attention. Furthermore, the experiment involves the utilization of the CS module to replace the SE module within the aforementioned embedded models. This results in the creation of the U-Net-CS and SHN-4-CS models. In order to achieve this goal, the CS module is applied to replace the SE module in the previously mentioned U-Net-SE and SHN-4-SE models, resulting in the establishment of U-Net-CS and SHN-4-CS models. By comparing models using the CS module and those using the SE module, the potential of the CS module in improving model performance can be evaluated. For the purpose of

verification, these models of the same type are employed as comparative models to assess whether the proposed model exhibits advantages across diverse datasets. The implementation of the model code within the experimental environment is based on Python, utilizing Pytorch, a deep learning framework developed on the Python platform. In each experimental setting, models of the same type are used as comparison models to determine whether the proposed models exhibit advantages under different datasets. These comparisons help identify which model configurations are more effective in specific scenarios, thereby providing better solutions for music source separation and copyright protection. Table 1 presents the specific operating environment parameters.

The dataset used in the experiment is based on the Multimedia Information Retrieval lab (MIR1K, 1000 song clips) [12]. The dataset comprises 1000 song segments containing music accompaniment and vocals, recorded separately in left and right channels. The dataset includes manually annotated pitch contours, indexes, and types of silent frames, lyrics, and vocal/non-vocal segments within half-tones. These recordings feature lyrics sung by the same individual who sang the songs. Each segment's duration ranges from 4 to 13 s, resulting in a total dataset length of 133 min. These clips were extracted from 110 karaoke songs, comprising mixed and instrumental tracks. The songs were freely selected from a pool of 5000 Chinese pop songs performed by researchers (8 females and 11 males) from the MIR laboratory. Most of the singers were amateurs without professional music training. The audio samples of the MIR1K dataset are stored in MP3 audio format. Dataset link is <https://sites.google.com/site/unvoicedsoundseparation/mir-1k>. The MIR1K dataset comprises recordings of vocal performances and accompanying music tracks, separately recorded in the left and right audio channels. This dataset consists of 1000 song fragments curated from 110 Chinese karaoke songs. Among these fragments, 175 clips featuring male and female vocalists, denoted as “abjones” and “amy,” serve as the training dataset, while the remainder is designated for use as the test dataset. The evaluation metric employed in this context is the segmented Median (Med.), which offers an intuitive representation of the minimum performance level achieved within 50% of the dataset. In order to characterize the dispersion of this performance distribution, the median absolute deviation (MAD) is utilized as an equivalent measure of standard deviation (SD) based on rank. It is worth noting that a larger value for Med. and a smaller value for SD correspond to more favorable model performance.

Table 1 Experimental environment

Item	Parameters	Item	Parameters
Operating system	Fedora 33	Programming language	Python
deep learning framework	PyTorch	Development platform	Pycharm
CPU performance	Inter Core i7-7700 CPU @ 4.2 Hz	GPU parameters	NVIDIA GeForce GTX 2070
Memory	128G	Hard disk drive	1T

5.2 Parameters setting

In this paper's model, the convolutional kernel size is 5×5 . The convolutional layers have a stride of 1, and a depth of 2 convolutional layers is chosen. The pooling kernel size is 3×3 . The number of neurons in the fully connected layers is 512. The learning rate is set to 0.001 and is dynamically adjusted based on training progress. In the SE module, the squeeze ratio is 2, indicating a reduction in channel count in feature maps by half. The excitation module has 64 neurons in the fully connected layer. For the U-Net model, the convolutional layer filter size is 3×3 , with a stride of 1 and 64 channels. Max-pooling is used in the pooling layers, and are 256 neurons in the fully connected layers. In the SHN-4 model, the convolutional layer filter size is 5×5 , with a stride of 1 and 128 channels. Average pooling is used in the pooling layers, and are 512 neurons in the fully connected layers. The U-Net-SE and SHN-4-SE models are extensions of the U-Net and SHN-4 models, where the SE module is added. Therefore, the squeeze ratio is 2, and are 64 neurons in the excitation module. The U-Net-CS and SHN-4-CS models are extensions of the U-Net and SHN-4 models where the SE module is replaced with the CS module. In the U-Net-CS model, the scaling factor in the CS module is 0.5, and the filter size is 3×3 with a stride of 1. In the SHN-4-CS model, the scaling factor in the CS module is 0.6, and the filter size is 5×5 with a stride of 1.

5.3 Performance evaluation

The U-Net network exhibits a holistic U-shaped architecture, encompassing a feature extraction layer on the left and an upsampling process on the right, characterized by symmetrical design from left to right. On the other hand, the Stacked Hourglass Network-4 (SHN-4) represents a network model composed of stacked Hourglass networks. Both of these models possess notable capabilities in image segmentation and classification, rendering them reference networks for the purpose of comparison. The experimental outcomes for various models in the MIR1K dataset are visually presented in Figs. 5 and 6.

In Fig. 5, in the SDR evaluation for vocal sources, the SHN-4 method outperforms the U-Net method significantly, with SDR values of 10.45 and 7.45, respectively, indicating that SHN-4 performs better in source separation. Similarly, in the source-to-interference ratio (SIR) and source-to-artifact ratio (SAR) evaluations, SHN-4 also outperforms U-Net noticeably. Compared to U-Net and SHN-4, both U-Net-SE and SHN-4-SE perform better in SDR, SIR, and SAR evaluations, indicating that the introduction of the SE module has improved performance. In the SDR, SIR, and SAR evaluations, SHN-4-CS is notably superior to U-Net-CS, suggesting that the introduction of the CS (Convolutional Attention) module has a more significant impact on performance improvement. SHN-4-CS slightly outperforms U-Net-CS in the SDR, SIR, and SAR metrics. Overall, the SHN-4 series models (SHN-4, SHN-4-SE, and SHN-4-CS) exhibit better performance on vocal sources in the MIR1K dataset, especially in terms of SDR, which is significantly superior to

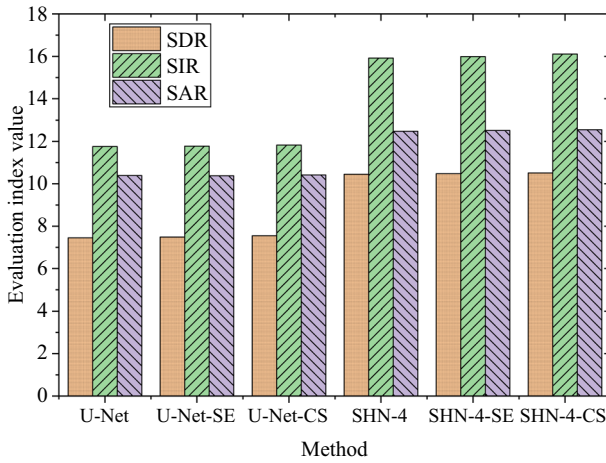


Fig. 5 Comparison results of embedded CS attention mechanism models under the MIR1K dataset (human voice source)

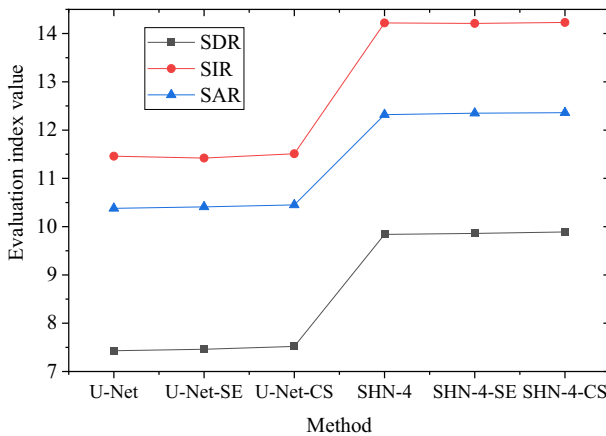


Fig. 6 Comparison results of embedded CS attention mechanism models under the MIR1K dataset (Accompaniment source)

the U-Net series models (U-Net, U-Net-SE, and U-Net-CS). The introduction of the SE and CS modules has positively contributed to performance improvement, with the CS module appearing to have a more pronounced impact. These results suggest that in the task of source separation, the SHN-4 series models, as well as the introduction of the SE and CS modules, are effective in enhancing model performance.

In Fig. 6, in the SDR evaluation for accompaniment sources, the SHN-4 method significantly outperforms the U-Net method, with SDR values of 9.84 and 7.43, respectively, indicating that SHN-4 performs better in source separation. Similarly, in the SIR and SAR evaluations, SHN-4 outperforms U-Net noticeably. Compared to

U-Net and SHN-4, both U-Net-SE and SHN-4-SE perform better in SDR, SIR, and SAR evaluations, indicating that the introduction of the SE module has improved performance. In the SDR, SIR, and SAR evaluations, SHN-4-CS is notably superior to U-Net-CS, suggesting that introducing the CS module significantly impacts performance improvement. SHN-4-CS slightly outperforms U-Net-CS in the SDR, SIR, and SAR metrics. Overall, for accompaniment sources in the MIR1K dataset, the SHN-4 series models (SHN-4, SHN-4-SE, and SHN-4-CS) still exhibit better performance, especially in terms of SDR, which is significantly superior to the U-Net series models (U-Net, U-Net-SE, and U-Net-CS). Similar to the previous analysis, the introduction of the SE and CS modules contributed positively to performance improvement, with the CS module appearing to have a more pronounced impact. These results further emphasize the effectiveness of the SHN-4 series models and the SE and CS modules in the task of source separation.

Combining the findings from Figs. 5 and 6, the SHN-4 series models, along with the SE and CS modules, demonstrate better performance in source separation tasks, particularly in SDR evaluation. The introduction of these modules helps enhance the model's ability to learn different types of audio features and their robustness, providing more reliable tools for digital music copyright protection. Table 2 displays the comparison results after adding the adaptive attention mechanism adaptive connection (AC).

Table 2 reveals that in the case of vocal sources, SHA-4-AC demonstrates a noticeable advantage over SHA-4, showing better performance in source-to-distortion ratio (SDR), SIR, and SAR evaluations. This indicates that the introduction of the adaptive attention mechanism has a significantly positive effect on enhancing model performance. In SDR evaluation, SHA-4-AC exhibits a performance improvement of 0.44 compared to SHA-4, and similar performance enhancements are observed in SIR and SAR evaluations. For accompaniment sources, it can also be observed that SHA-4-AC significantly outperforms SHA-4, demonstrating better performance in SDR, SIR, and SAR evaluations. This further confirms the effectiveness of the adaptive attention mechanism not only in vocal sources, but also in accompaniment sources, in improving performance. In SDR evaluation, SHA-4-AC shows a performance improvement of 0.18 relative to SHA-4, with similar performance improvements in SIR and SAR evaluations. In summary, the adaptive attention mechanism has a significant positive impact on improving the performance of the model in audio source separation tasks for both vocal and accompaniment sources. SHA-4-AC achieves better SDR, SIR, and SAR evaluation results for both types of sources, indicating the effectiveness

Table 2 Comparison results of adaptive attention mechanism under the MIR1K dataset

Method	Sound source	SDR	SIR	SAR
SHA-4	Human voice source	10.54	16.34	12.54
SHA-4-AC		10.98	17.23	12.69
SHA-4	Accompaniment source	9.67	14.09	12.78
SHA-4-AC		9.85	14.21	13.01

and generality of this mechanism in improving audio source separation tasks. Figures 7 and 8 show the comparison results of the two-stage music source separation under the MIR1K dataset.

The comparative models under consideration in this paper encompass the two-stage music separation-gate (TSMS-G) model and an enhanced iteration thereof, denoted as the TSMS-gate self-attention (TSMS-GSA) model, distinguished by the incorporation of a gating attention module. A comparative analysis, as depicted in Figs. 7 and 8, draws attention to the superior performance of the TSMS-GSA model vis-à-vis the TSMS-G model in terms of both human voice source and accompaniment source indicators. These findings underscore the salient impact of the combined mechanisms of CS, self-attention (SA), and AC, which collectively augment network performance and adaptive gate structures. This augmentation facilitates enhanced model self-selection, diminishes the occurrence of outliers, and imparts stability to waveform characteristics.

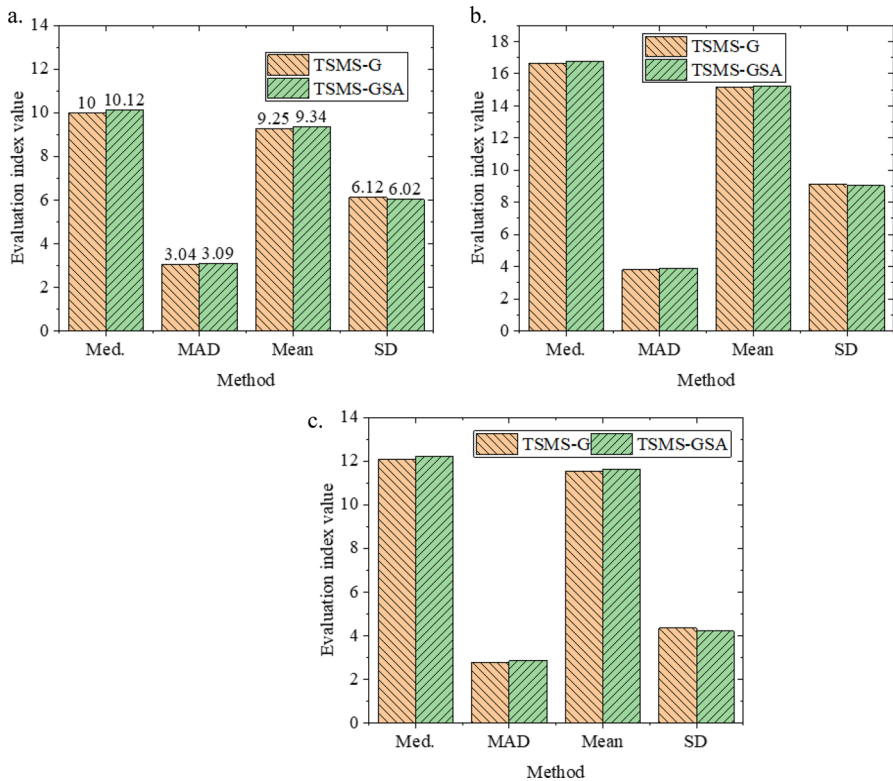


Fig. 7 Comparison results of two-stage music source separation under the MIR1K dataset (human voice source) (a SDR; b SIR; c SAR)

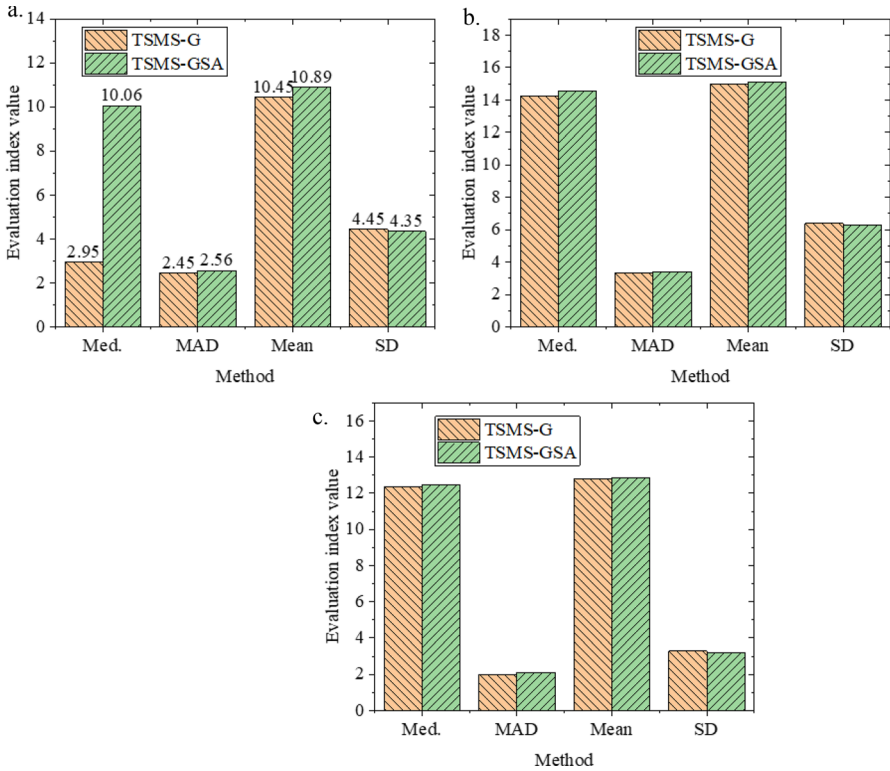


Fig. 8 Comparison results of two-stage music source separation under the MIR1K dataset (accompaniment source) (a SDR; b SIR; c SAR)

5.4 Discussion

This paper proposes a method that combines CNN and attention mechanisms for digital music copyright protection. It includes the SE module and the channel spatial (CS) module for source separation and infringement authentication. In comparison with Hu et al. [15], who constructed a ConvLSTM two-dimensional neural network to simulate long-range dependencies in the spectral domain using a two-dimensional extension architecture of LSTM, this paper’s method employs CNN and attention mechanisms to capture audio features in both the time and frequency domains. This makes the approach more flexible and capable of handling various types of infringement behaviors, not just modeling long-range dependencies. Megías et al. [27] embedded digital watermarks in audio to uniquely identify copyright information and employed a CNN classifier to test overall performance, achieving significant results in terms of accuracy, precision, F1 score, and recall. In contrast, this paper’s method introduces the SE module and the CS module for source separation and infringement authentication. These modules enhance the model’s performance and are more suitable for dealing with different

forms of digital music infringement. Wang et al. [37] proposed a shallow and deep feature fusion method to accurately describe inconsistent changes produced by original digital audio tampering operations by leveraging the complementarity of features at different levels, thereby fully utilizing electronic network frequency (ENF) information. Experimental results showed an accuracy of 88.31%, surpassing existing methods. In comparison, this paper's method introduces the SE and CS modules to improve model performance. The CS module helps capture channel correlations in audio, providing a more comprehensive feature representation, thus enhancing the accuracy of infringement detection. The method proposed in this paper is characterized by its flexibility in capturing audio features comprehensively compared to methods proposed by other researchers, leading to better performance in digital music copyright protection. The introduction of the SE module and the CS module improves both performance and robustness, enabling the approach to perform well under various types of infringement behaviors.

The aforementioned experimental findings unequivocally demonstrate the efficacy of incorporating the CS attention mechanism within the coder–decoder network, resulting in improvements across all performance metrics. Furthermore, the augmentation of the network's bottleneck layer through the utilization of self-attention mechanisms amplifies its capacity for feature representation. This augmentation selectively reinforces and complements distinctive features, ultimately enhancing the model's proficiency in music source separation—a corollary consistent with the findings of Stöter et al. [35]. Their pioneering work contributed a pre-trained model intended for end users and artists, facilitating experimentation with music source separation. Notably, Open-Unmix was conceived as a pivotal component within the open ecosystem of music separation, accompanied by open datasets, software utilities, and open evaluation protocols, all aimed at fostering replicable research initiatives. Expanding upon these advancements, Lee et al. [24] harnessed multi-level and multi-scale feature aggregation techniques to refine their model. Subsequently, they engaged in transfer learning across various music classification tasks, culminating in the visualization of CNN-learned filters within each layer. This visualization facilitated the identification of hierarchical learning features and elucidated their responsiveness to logarithmic scaling frequencies. Similarly, Li et al. [25] introduced an adaptive data fusion strategy rooted in DL, denoted as CNN with atrous convolution, for the dynamic fusion of multi-source data. These collective endeavors underscore the reliability and validity of the present research.

6 Conclusion

6.1 Research contribution

In the era of contemporary media, the subject of music dissemination has transcended the confines of traditional media, previously dominated by professional music producers and specialized outlets. With the escalating prevalence of copyright infringements, this paper initially delves into the principal methodologies of SCM related to digital music. Subsequently, it underscores the pragmatic applicability of

fuzzy authentication as a safeguard for digital music copyrights. Moreover, a CNN is employed to investigate the segregation of music sources within digital music. An attention mechanism and an adaptive gate model are meticulously crafted to pursue this goal. The adoption of a selective adaptive cascade technique is pivotal in refining the architecture of the multi-resolution coder–decoder, thereby amplifying the feature sensitivity and precision concerning music source separation. This optimization concurrently mitigates the distortion rate affecting human voice and accompaniment, thereby contributing to the safeguarding of digital music copyright. It is worth highlighting that this approach has demonstrated its feasibility and reliability as an instrument for safeguarding digital content. Nonetheless, it is imperative to acknowledge that the performance of the deep CNN-based music source separation model experiences a notable decline when applied to diverse music samples from disparate sources. This drawback is compounded by the presence of interference among multiple sources within the segregated music. It is within this context that the proposed TSMS-GSA model emerges as a potent solution to effectively tackle the aforementioned challenges encountered in the aforementioned research endeavors.

6.2 Future works and research limitations

However, it is essential to acknowledge certain limitations within the scope of these research accomplishments. For instance, the established model's validation necessitates a broader spectrum of music datasets to augment its performance. This avenue warrants exploration in forthcoming research endeavors. Furthermore, real-world music encompasses an array of genres characterized by intricacy and variability, often accompanied by layered background sounds featuring multiple instruments and vocal harmonies. Additionally, the MIR1K dataset employed contains a considerable number of silent segments. As such, experimental validation remains confined to the constraints of the MIR1K dataset. Future research endeavors are poised to concentrate on music source separation in scenarios involving intricate multivariate sources and the treatment of silent segments. Specialized processing methodologies may be contemplated to address the approximation of multiple sources and the management of silence across various fragments.

Acknowledgements The authors acknowledge the help from the university colleagues.

Author contributions The author confirms being the sole contributor of this work and has approved it for publication.

Funding This work was supported by Liaoning Educational Committee Project Research on the Legal Issue of Capital Market Supervision in the Implementation of the New Securities Law (No. LJKR0334).

Data availability The raw data supporting the conclusions of this paper will be made available by the authors, without undue reservation.

Declarations

Conflict of interest All authors declare that they have no conflict of interest.

Ethical approval This paper does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Arditi D (2019) Music Everywhere: Setting a Digital Music Trap. *Crit Sociol* 45(4–5):617–630
2. Asamoah D, Agyei-Owusu B, Andoh-Baidoo FK, Ayaburi E (2021) Inter-organizational Systems Use and Supply Chain Performance: Mediating Role of Supply Chain Management Capabilities. *Int J Inf Manage* 58:102195
3. Boshier H (2022) Sheeran Succeeds in ‘Shape of You’ Music Copyright Infringement Claim. *J Intellect Prop Law Pract* 17(7):544–546
4. Boudreau KJ, Jeppesen LB, Miric M (2022) Profiting from Digital Innovation: Patents, Copyright and Performance. *Res Policy* 51(5):104477
5. Bridy A (2019) The Price of Closing the Value Gap: How the Music Industry Hacked EU Copyright Reform. *Vand J Entertain Technol Law* 22:323
6. Burton CS (2021) “Playola” and Fraud on Digital Music Platforms: Why Legislative Action is Required to Save the Music Streaming Market. *J Bus Technol Law* 16:387
7. Cai Z (2020) Usage of Deep Learning and Blockchain in Compilation and Copyright Protection of Digital Music. *IEEE Access* 8:164144–164154
8. Chandler K (2019) The Times They are a Changin’: The Music Modernization Act and the Future of Music Copyright Law. *Tul J Technol Intell Prop* 21:53
9. Chen ZT (2021) Flying with Two Wings or Coming of Age of Copyrightisation? A Historical and Socio-legal Analysis of Copyright and Business Model Developments in the Chinese Music Industry. *Glob Media China* 6(2):191–206
10. Drott E (2021) Copyright, Compensation, and Commons in the Music AI Industry. *Creat Ind J* 14(2):190–207
11. Dunbar D (2022) Another Cog in the Machine: Digital Music Sampling and an Evaluation of Its Existing Regulatory Mechanisms in Light of the Case Act. *Tul J Technol Intell Prop* 24:181
12. Gao Y, Zhang X, Li W (2021) Vocal Melody Extraction via Hrnet-Based Singing Voice Separation and Encoder–Decoder-Based f0 Estimation. *Electronics* 10(3):298
13. Gölgeci I, Kuivalainen O (2020) Does Social Capital Matter for Supply Chain Resilience? The Role of Absorptive Capacity and Marketing-Supply Chain Management Alignment. *Ind Mark Manag* 84:63–74
14. Hergueux J, Jemielniak D (2019) Should Digital Files be Considered a Commons? Copyright Infringement in the Eyes of Lawyers. *Inf Soc* 35(4):198–215
15. Hu WS, Li HC, Pan L, Li W, Tao R, Du Q (2020) Spatial–Spectral Feature Extraction via Deep ConvLSTM Neural Networks for Hyperspectral Image Classification. *IEEE Trans Geosci Remote Sens* 58(6):4237–4250
16. Jia F, Gong Y, Brown S (2019) Multi-tier Sustainable Supply Chain Management: The Role of Supply Chain Leadership. *Int J Prod Econ* 217:44–63
17. Jing N, Liu Q, Sugumaran V (2021) A Blockchain-Based Code Copyright Management System. *Inf Process Manag* 58(3):102518
18. Kariyawasam K, Palliyaarachchi R (2021) The Song Would be Nothing Without Someone to Sing It: Copyright and Performers’ Rights in Music. *Int Rev Law Comput Technol* 35(3):222–245
19. Kaye DBV, Gray JE (2021) Copyright gossip: Exploring Copyright Opinions, Theories, and Strategies on YouTube. *Soc Media Soc* 7(3):20563051211036940
20. Kim YM, Park BC, Bang KS, Kim SY (2021) A Method of Generating Theme, Background and Signal Music Usage Monitoring Information Based on Blockchain. *J Korea Soc Comput Inf* 26(2):45–52
21. Kjus Y (2021) The Use of Copyright in Digital Times: A Study of How Artists Exercise Their Rights in Norway. *Pop Music Soc* 44(3):241–257
22. Lawrence DL (2019) Addressing the Value Gap in the Age of Digital Music Streaming. *Vand J Transnat’l L* 52:511

23. Lee JA, Li Y (2021) The Pathway Towards Digital Superpower: Copyright Reform in China. *GRUR Int* 70(9):861–870
24. Lee J, Park J, Kim KL, Nam J (2018) SampleCNN: End-to-End Deep Convolutional Neural Networks Using Very Small Filters for Music Classification. *Appl Sci* 8(1):150
25. Li S, Wang H, Song L, Wang P, Cui L, Lin T (2020) An Adaptive Data Fusion Strategy for Fault Diagnosis Based on the Convolutional Neural Network. *Measurement* 165:108122
26. Li Y, Wei J, Yuan J, Xu Q, He C (2021) A Decentralized Music Copyright Operation Management System Based on Blockchain Technology. *Procedia Comput Sci* 187:458–463
27. Megías D, Mazurczyk W, Kuribayashi M (2021) Data Hiding and Its Applications: Digital Watermarking and Steganography. *Appl Sci* 11(22):10928
28. Nurhaeni T, Nirmalasari L, Faturahman A, Avionita S (2021) Transformation Framework Design on Digital Copyright Entities Using Blockchain Technology. *Blockchain Front Technol* 1(01):35–43
29. Okonkwo IE (2021) NFT, Copyright and Intellectual Property Commercialization. *Int J Law Inf Technol* 29(4):296–304
30. Öztürk TA, Crossley N (2021) Sociological and Cultural Aspects of New Economic Models in Digital Music in Turkey. *Üsküdar Üniv Sos Bilim Derg* 13:295–332
31. Pournader M, Ghaderi H, Hassanzadegan A, Fahimnia B (2021) Artificial Intelligence Applications in Supply Chain Management. *Int J Prod Econ* 241:108250
32. Salsabila FM, Mayana RF, Rafianti L (2021) Copyright Commercialization of Songs Uploaded in TikTok Application Without the Creator's Permission. *J Sains Sosio Hum* 5(1):213–224
33. Shah U (2019) Digital Music Piracy in India: Issues and Challenges. *Int J Res Soc Sci* 9(6):709–721
34. Silfiani D (2022) Juridical Review on Collection and Distribution of Royalties for the Use of Copyrights of Song and/or Music on Over-the-Top Services in Indonesia. *Bp Int Res Crit Inst (BIRCI J) Humanit Soc Sci* 5(1):1456–1466
35. Stöter FR, Uhlich S, Liutkus A, Mitsufuji Y (2019) Open-Unmix-A Reference Implementation for Music Source Separation. *J Open Source Softw* 4(41):1667
36. Wang N, Xu H, Xu F, Cheng L (2021) The Algorithmic Composition for Music Copyright Protection Under deep Learning and Blockchain. *Appl Soft Comput* 112:107763
37. Wang Z, Yang Y, Zeng C, Kong S, Feng S, Zhao N (2022) Shallow and Deep Feature Fusion for Digital Audio Tampering Detection. *EURASIP J Adv Signal Process* 2022(1):69
38. Zhang G, Tang H, Chen J, Shen R, He Q, Huang B (2021) Digital Music Copyright Management System Based on Blockchain. *J Comput Appl* 41(4):945

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Junfeng Wang¹

✉ Junfeng Wang
wangjunfeng@lnnu.edu.cn

¹ School of Law, Liaoning Normal University, Dalian 116083, China