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

AI Technology Supports Mechanisms for Enhancing Digital Music Composition and Arrangement Efficiency

Chuanli Liu[1] *and Yifan Chen*[1,*]

[1] Music and Dance Academy, Nanchang Vocational University, Nanchang, Jiangxi, 330500, China

SUMMARY: *Recent advancements in deep learning technology have made significant strides in the field of musical creation based on GAN architectures. GAN algorithms consist of two parts: generator and discriminator, where through their mutual competitive training process, the generator develops techniques in creating musical pieces closer to the properties of the real musical pieces. In this project, a classical multi-track music generation architecture named MuseGAN will be used to increase the context similarity of generated musical phrases through modifications in generator's temporal architecture. Additionally, the use of a feature extractor, along with other modifications during the training process, will improve the smoothness of the transition between notes. The evaluation of the MuseGAN generated music samples uses several parameters: distribution density of notes, chord matching score, harmony, and BLEU score as well as Self-BLEU. The results prove that the generated music samples using the MuseGAN algorithm show high harmony score and good musicality. The mutual interaction between generator and discriminator improves music created by artificial intelligence, making it more diverse and realistic.*

KEYWORDS: *Generative Adversarial Networks; MuseGAN; BLEU score; music generation technology*

1 Introduction

The empowerment of artificial intelligence in arts involves coding the fundamentals behind human music production as well as the use of massive datasets in training intelligent agents. This can be considered as a symbolic representation of human cognitive symbolism and experience in AI [1, 2]. Intelligent music composition marks a highly evolved stage of smart music technology and acts as a significant yardstick for the progression of AI from fundamental information perception to intelligent creation [3]. Currently, within the applications realm, AI is not only proficient in chord progressions, musical form, and arrangement principles, but also creates complex arrangements of musically rich compositions and exhibits intelligent voice synthesis and processin [4-6]. AI music composition will develop from materiality through machinery into intelligent virtuality. The current state-of-the-art innovations include virtual instruments, virtual audio-visual synthesis, and AI music in the metaverse [7, 8].

The conventional methods for arranging a piece of music require rich knowledge of music theory and experience in composition. On the contrary, the intelligent arrangement system identifies suitable harmonies, instruments, and other musical elements with the help of learning in large collections of musical pieces to support musicians' creativity [9, 10]. The intelligent

*chenyifan092813@163.com
https://doi.org/10.65102/is2026487
Published: 30 April, 2026

system created by Huang et al., which uses the note-trimming approach, is able to perform the automatic arrangement of certain compositions. Musical quality assessment of such an arrangement was conducted using the Turing test to show the efficiency of the system in composing the piano music [11]. Later, Cao tried to overcome the shortcomings of the conventional way to compose piano music using the distributed sensor technique to facilitate creative processes. Tests conducted in terms of style, melody, and timbre revealed that the K-Nearest Neighbors classifier integrated into such a technique resulted in the more stable piano arrangement [12]. Additionally, Liu examined the application of GANs in the intelligent development of music pieces. The suggested improved multi-track music generation model can generate not only the bass but also drums in the music piece, where the fragments of such music seem smooth and pleasant [13]. In his study, Li applied several intelligent algorithms for music chord recognition and generation. The Hidden Markov Model reached 81.8% of accuracy in recognizing chords played on the piano, while multi-style music chord generation scored higher than three points in classical, folk, and pop music styles [14]. Finally, Sun et al. offered an efficient intelligent music composition system, based on deep learning, and which incorporates information about the musical work for providing inspirations and ideas for creators [15]. Furthermore, Zhu et al. introduced a cross-generative model for rhythm and melody based on chords and suggested a multi-instrument collaborative arrangement algorithm based on multi-task learning as a multi-track music arrangement model [16]. It has been proven that their approach allows achieving successful intelligent pop music generation on the real data set. Consequently, the combination of learning-based intelligent algorithms and artificial intelligence technologies opens a lot of promising perspectives for the intelligent music creation and arrangement [17].

According to the theory of multiple intelligences by Gardner, musical intelligence is considered one of the forms of intelligence [18]. Intelligent music composition is very complicated because it involves many components like instruments, melodies, percussions, and chords that need to be coordinated properly to create harmony and emotion [19]. Artificial intelligence and machine learning have paved new ways in creating intelligent music with remarkable advantages to the music industry [20].

This paper addresses the issues of music technology based on artificial intelligence technology, music creation concepts based on digital technology, and music generation concept based on artificial intelligence. This work highlights the shift in AI technology in moving from an instrument that depends only on humans for music creation to a collaborative partner in music composition by humans. The impact of artificial intelligence technology innovations on the music industry will be addressed through two different approaches, which are related to the creative and production processes. Two models, namely, the generative adversarial network (GAN) and multitrack music MuseGAN, will be suggested. Through improving the MuseGAN model, the recurrent feature generative adversarial network (RFGAN) is suggested. Three different metrics for evaluating generated music samples will be used; they include note density distribution, chord matching rate, and harmony.

2 Development of AI Music Technology

2.1 AI Music Technology

2.1.1 Vowel-Consonant Separation

Audio source separation uses AI technology and machine learning algorithms to separate various sound sources from the mixed audio track, such as vocal and instrumental sounds. Such

technology helps solve the problem of missing multi-track recordings while mixing, enhancing possibilities for sound production in the fields of music and film.

2.1.2 Music Generation

Music generating software allows people who have not studied music to compose music automatically by conforming to predetermined musical styles and parameters. Similarly, with the fast development of music generation technology, there is an expansion of possible uses of this software in films, television music scores, video games, and advertisement technology.

Despite being able to mimic musical structures and patterns, artificial intelligence fails in creating compositions that carry true depth of emotion and are innovative and unique. It fails in providing compositions that are imbued with emotion and artistic qualities and are therefore capable of conveying true emotion and personality into them through performance techniques, dynamics, and rhythm, the capabilities of human composers. This is because music is beyond a collection of sounds or notes; its essential function is in expressing emotion.

2.1.3 Music Mixing

There has been remarkable advancement in music mixing technology, where analysis of spectrum, dynamics, and timbre have been utilized to automate levels and mixes for improved results. With regards to making films and games, there are possibilities of using artificial intelligence to create music specifically suited for particular situations, thus improving the artistic nature of the pieces.

Despite the good performance of the AI algorithms, there may arise some difficulties in the full replacement of human judgment by the algorithm in dealing with difficult or specialized music mixes. For instance, the difficulty of teaching AI how to interpret the semantic and emotive aspects of music and to attain better natural mixing results may pose a problem. However, with continued advancement in technology and research, such problems will definitely be solved incrementally.

2.2 Digital Music Composition and AI-Generated Technology

In terms of music creation, AI music is currently widely accepted and recognized by the public in the broad sense as AI-generated music.

2.2.1 Digital Music Creation Paradigms

Prior to this, digital audio technology had already reached a highly mature stage in music composition and production. Taking songwriting as an example, the creative process can be divided into three phases: pre-production, production, and post-production.

During pre-production, the composer conceives the melody, style, and overall presentation of the music, completing the melody composition. Some composers may further annotate harmonic function requirements based on the melody.

The mid-stage involves the arrangement and production phase. This is primarily accomplished using digital audio workstations (DAWs), such as Logic Pro commonly used on Apple Mac OS systems or Cubase commonly used on Microsoft Windows systems, among others. These DAWs can load an extremely rich array of musical resources and materials: diverse electronic sound libraries, libraries simulating authentic acoustic instrument tones, and digital samplers that can synthesize an infinitely varied digital orchestra. With this model, complete musical presentation can be achieved according to the vision of the creator. In cases where there are rigorous arrangements, the production phase might include recording of live sections such as solos, strings, and drums. It is with this integration that the production becomes

artistically superior.

After arranging the tracks, the vocalists proceed to record their lines. After this, the process moves into post-production which includes mastering and mixing. With this, the spatial positioning of each sound will be optimized. A song is only considered complete when it passes through all stages from its conception to production.

The creation process with digital technology provides for an extremely sophisticated system. In all phases, the human factor takes precedence since it requires a certain degree of expertise on the part of the creator.

2.2.2 AI Music Generation Paradigms

AI music generation platforms have radically changed the way a musical composition is converted into production through creation. Entering keywords of musical style and instruments on the interface of the platform will generate a song in less than half a minute. In addition to being able to come up with songs these platforms are capable of generating orchestral arrangements as well.

Udio is an AI music generation service Take the case of the AI music generation platform Udio. Upon inputting the music style keywords, which are folk music and epic music, and the instrument keywords pipa, dizi, piano, and violin, and hitting the create button, it will take less than a minute to produce two 30-second audio clips of an orchestrated music in under a minute. With the new 1.5v advanced model, one can at the same time generate four audio clips that will sum up to 130 seconds in total. In any generation mode chosen, the music duration can be further extended via the One-Click Continue option.

2.3 AI Technology Collaboration and Music Creation

However, despite the fact that AI music generation platforms can provide the most convenient experience ever in generating musical material, each and every significant step of the music-making process, including entering key words as prompts and the overall AI-based music production process, still depends on human artistic intuition.

In the course of this procedure, AI music platforms are merely technical instruments that human creators use without any autonomous consciousness or will. The professional field's assessment of whether AI can serve as a human collaborator in musical art creation hinges on its “perception” and ‘agency’ of music. For years, scientists and artists in music technology have explored computer intelligence systems capable of independently completing the cycle of music content perception, musical data computation, and musical content presentation. This enables such systems to function as “machine collaborators” in the creative and performance processes of human artists. After all, for many musicians, the goal of achieving high-level automation through machines is not to remove themselves entirely from the creative process. Instead, establishing an “intelligent collaborator” operating in parallel to explore the inaccessible realms of human thought and performance presents a challenge full of surprises.

In this type of creation, the relationship between humans and AI systems resembles two relatively independent yet interconnected musical entities, constantly listening to each other. They assess each other's musical output and decide their subsequent performance accordingly. Through parallelism and interaction, they form a musical complementarity—the computer's role is more akin to that of an independent ‘machine performer.’ Its response to the human performer is based on complex cognition of the performance content, not merely dependent on the extraction of low-dimensional audio features. “Each real-time output from the system is a new, unique response—it may mimic, oppose, complement, or ignore the human improviser.” The entire period of that process has seen the advanced and highly trained AI system demonstrating a significant amount of independent agency. However, its actions are essentially

based on human intentions. In fact, the better trained it is, the more it resembles human musical values, thus turning into an actual human machine cooperator.

2.4 Impact of AI Technological Innovation on the Music Industry

(1) Democratization and Diversification of the Creative Process

The emergence of the age of digital information has pushed the music industry to develop in more diverse ways. Music production within the conventional music business was typically restricted to dedicated hardware and settings, which were usually controlled by somewhat stable creative groups. The model was clearly a significant barrier to entry in the creation of music as it was hard to find opportunities to show off your musical skills when you are talented but have few resources at your disposal. Nevertheless, due to the fast development of digital technology, music creation is experiencing unprecedented democratization and diversification. Currently, even individuals and small groups may use advanced technology and music software to create quality music. The trend of democratization has greatly reduced the barriers to music creation, creating more opportunities to take part in the process. Professional musicians and amateur enthusiasts can all freely express their musical ideas and feelings in these tools. Certainly, the development of digital technology has brought music creation to a new era of increased democratization and diversity. In this new age, each individual can be a music creator and bring the world more beautiful melodies and enthralling music to the world using their musical talent and creativity.

(2) Digitalization and Automation of Production Workflows

The massive implementation of digital technology in music production has significantly changed the conventional production methods used in the music industry. All stages, including recording, mixing and mastering are being involved in the digital revolution which has brought about much improvement in the effectiveness and quality of music creation. With their powerful capabilities and easy-to-use interfaces, music production workflows have been simplified significantly by digital tools, including Digital Audio Workstations (DAWs). Musical artists may now carry out all of the tasks involved in the workflow, including the composition, production, and so forth, in one integrated setup without having to move across various devices and software. This consolidation will not only increase productivity but also create uniformity in sound quality and coherence of style among musical pieces. At the same time, the use of automation technology is becoming more and more popular in the music production. The development of sophisticated technologies, such as AI algorithms, allows musicians to easily find appropriate choices out of large sound and rhythm sample libraries, making the music-making process much quicker. This smart help does not just save the time of search and filter, but also suggests more suitable materials depending on the creative purpose of the artist and stylistic preferences. Actually, the combination of digital and automation technologies are driving music production into a new age characterized by increased efficiency, accuracy, and intelligence. In this age, musicians can concentrate more closely on the creative process itself, turning inspiration and ideas into compelling musical compositions.

3 AI Music Generation Using Generative Adversarial Networks

3.1 Generative Adversarial Network Model

Generative adversarial networks (GANs) consist of a generator (G) and a discriminator (D), which achieve a dynamic equilibrium between data generation and discrimination capabilities through an adversarial game [21].

The architecture of the generative adversarial network is shown in Figure 1. The generator takes random noise z (typically following a Gaussian distribution $z \sim N(0,1)$) as input. It generates images $G(z)$ through a multi-layer deconvolution network, aiming to approximate the distribution of real data $pdata(x)$ with the generated samples. The discriminator is a binary classification network that receives real images x and generated images $G(z)$, outputting the probability $D(x)$ or $D(G(z))$ that a sample belongs to the real data. Its task is to accurately distinguish between the two. This competitive process can be described by a minimax objective function.

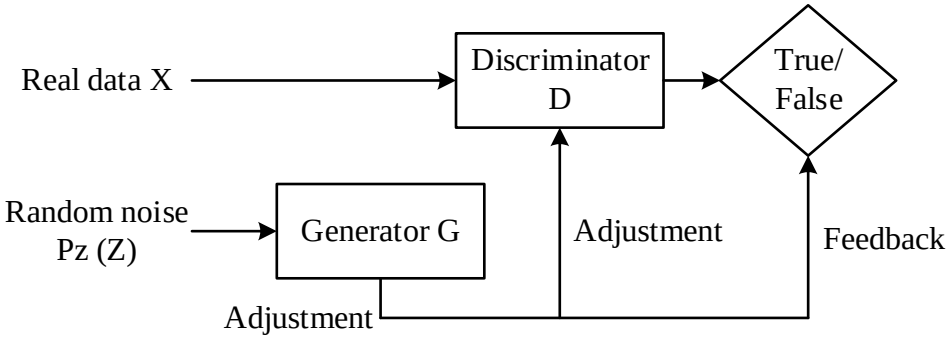

Figure 1: Generate an architecture against the network

The objective function is as follows:

$$min_G E_{z\sim p_z}[\log(1-D(G(Z)))] \tag{1}$$

The objective function is as follows:

$$max_D E_{x\sim p_{data}}[logD(x)] + E_{z\sim p_z}[\log(1-D(G(Z)))] \tag{2}$$

The overall objective function is as follows:

$$\overset{minmax}{G\ D}V(D,G) = E_{x\sim Pdata}\left[\log D(x)\right] + E_z \sim P_z\left[\log(1-D(G(z)))\right] \tag{3}$$

During training, the generator and discriminator are updated alternately. First, the generator is fixed while the discriminator parameters are updated using real sample $\{x_i\}$ and generated sample $\{G(z_i)\}$ to enhance its classification capability. Subsequently, the discriminator is fixed while samples are generated using noise $\{z_i\}$ to update the generator parameters, enabling it to produce images that more effectively fool the discriminator. This process iterates

cyclically until both components reach a dynamic equilibrium.

3.2 Music Generation Models

3.2.1 Music Generation Models Based on Generative Adversarial Networks

Here we introduce a classic multi-track music model—MuseGAN. The model adopts DCGAN as its foundational framework and incorporates the optimization algorithm from the WGAN-GP model to enhance training efficiency. Compared to the original generative adversarial network (GAN) model, it features a more stable training mechanism. By synthesizing various classic optimization approaches for GANs, it improves training effectiveness and can generate multi-track musical phrase samples.

Batch Normalization (BN) is also incorporated into the model. Developed by Google, BN accelerates training and improves model performance by effectively addressing issues like vanishing gradients and exploding gradients.

3.2.2 Generative Adversarial Networks for Generating Cyclic Features

This paper proposes a series of improvements based on the MuseGAN model. The enhanced model structure is termed the Recurrent Feature Generative Adversarial Network (RFGAN).

(1) Recurrent Pattern in the Generator

To address the lack of contextual relevance in generated musical phrases and to incorporate the essential repetition characteristics found in musical structures, this paper transforms the original unidirectional GAN generator into a recurrent structure. This resembles the architecture of a Recurrent Neural Network (RNN). The generator in the adversarial network is treated as a processing unit within an RNN. Mimicking the RNN's recurrent pattern, the generator's previous output is concatenated with random noise and fed back into the generator's input.

(2) Model's Temporal Structure

To enhance contextual coherence between musical phrases, this paper refines the temporal structure of the generative model. Building upon this foundation, a novel temporal generation mechanism is introduced.

The temporal pattern expression is shown below. Equation (4) represents the temporal pattern expression, while Equations (5) and (6) denote different data represented by y for varying numbers of training rounds. Equation (5) corresponds to the y function for the initial training round, and Equation (6) corresponds to the y function for subsequent training rounds. The expressions are as follows:

$$G(z_i, y) = \{G^o(E(y), z)\}^T \tag{4}$$

$$y = G_{bar}(G_{temp}(z)^T), s = 1 \tag{5}$$

$$y = G_{s-1}(z, y), s > 1 \tag{6}$$

z_i is a time-independent random vector, z is a time-dependent random vector, G_{temp} is a temporal structure generator, G_{bar} denotes a bar generator, $G_{s-1}\left(z, E\left(y\right)\right)$ represents samples generated from the previous training round, E denotes a decoder based on a convolutional neural network, y is the decoder's input, and subscript s indicates the training round number.

During the first training round, y represents the output generated by the T1 temporal

model. In subsequent training, y denotes the result samples from the previous round. After incorporating random noise z, these serve as input to the T2 temporal model. Decoder E adopts a mirrored network architecture of the generator model, capable of learning either monophonic music tracks generated by the Generation from scratch temporal model or the result samples from the preceding training round. The decoder output $E(y)$ is then combined with noise z and fed into the accompaniment pattern generator T2 sequence model. The result $G_s(z, E(y))$ serves as the output sample for the entire model's current training iteration.

(3) Feature Extractor

To further enhance the overall consistency of generated music samples, this paper considers that the data format of the trainable music sample dataset resembles image representations. Therefore, it draws inspiration from convolutional neural networks' techniques for extracting image features and applies them to this specialized form of musical expression.

(4) Average Pooling Layer

This paper incorporates average pooling layers into both the temporal and pitch output layers of the generative model. Music samples generated by the model are in Pianoroll format. The average pooling layer reduces the amplitude of fluctuations between values within regions of the resulting sample matrix. This further organizes and filters the generated data, smoothing the generated music data. After conversion to MIDI format, transitions and connections between musical notes become more fluid.

(5) Overall Network Architecture of RFGAN

In summary, the aforementioned improvements have led to the development of a novel multi-track generation model—the Recurrent Feature Generative Adversarial Network (RFGAN). The first input serves as the main melody generation component, functioning as the prior condition for the generative model. During the initial training phase, the T1 mode generates monophonic music samples, which are fed into the decoder E as the main melody. Subsequent training utilizes the resultant sample $F\left(G_{s-1}\left(z, E\left(y\right)\right)\right)$ from the previous round. After feature extractor TE extracts feature information $F\left(G_{s-1}\left(z, E\left(y\right)\right)\right)$ from the resultant sample, decoder E decodes this feature information. This decoded information is then input into different convolutional layers of the generative models (with identical input array matrix shapes), thereby influencing the musical output generated by the models.

The generated sample $G_s\left(z, E\left(y\right)\right)$ from this training round will not only be input into the discriminative model D alongside the actual training set x to complete training of model D, but will also be used in the next round of generative model training.

The random noise input Z in the T2 temporal mode consists of four components: intra-track and inter-track random noise. One component constrains music generation between tracks. Another component comprises intra-track random noise with and without time series z_t and z, respectively, which influences music generation within tracks. That is:

$$y = G_{bar}(G_{temp}(z_i)^T), s = 1 \tag{7}$$

$$y = G_{s-1}(z, y), s > 1 \tag{8}$$

$$G(z, y) = \{G^o(E(f(G_{s-1}(z, y)), \overline{z})\}^T \tag{9}$$

Equation (7) represents the expression for generating the overall model. Equations (8) and (9) respectively denote the y expression functions for different training iterations.

3.3 Experiments and Results Analysis

3.3.1 Experimental Environment

The most basic and important step to success in any experiment is setting up the experimental conditions. Consequently, this section will give an elaborate overview on how to set up the experimental setup. The experiment falls into the larger sphere of deep learning, so it uses Python 3.8 as the language of development and toolkits like TensorFlow-GPU and NumPy. It also places some requirements on computer hardware configuration.

The experimental environment specifications are outlined in Table 1.

Table 1: Experimental environment configuration

CPU	Intel(R) Xeon(R) CPU E7-2640v4
GPU	GeForce GTX 1080 T1
CPU Frequency	2.40 GHZ
GPU memory	24GB
Memory	1TB
Operating system	Ubuntu 17.01.1 LTS
System kernel	GNU/Linux 4.4.0-88-generic x86_64
Opening language	Python 3.8
Development framework	Tensorflow-gpu=1.10.1
Python	Numpy, Music22, Pretty……

3.3.2 Objective Evaluation

The present section will analyze 10 randomly chosen pieces out of 200 musical works created by RFGAN model to affirm that it has the ability to produce high quality music. To compare with, 10 samples are also chosen on compositions created by Tr-MTMG, HRNN, and MuseGAN models. These samples are judged on the basis of the following criteria.

Tr-MTMG model is based on a learning network that includes mainly a cross-track attention mechanism which is one of the improvements of the Transformer architecture. The mechanism is applied to learn the information across various instrumental tracks. It uses text continuation ability of the GPT model in the generation stage to prolong the music.

The HRNN model is a multi-track music generation model that is based on a hierarchical recurrent neural network. It produces melodies using a lower-level RNN and subsequently a higher-level RNN to produce an accompaniment based on that melody, leading to multi-track music.

MuseGAN is a multi-track music generation system that is based on generative adversarial networks (GANs). After being trained on a sample larger than 100,000 songs, it can create multi-track compositions with bass, guitar, piano, drums and strings. It is done in two ways: randomly or using a particular starting note.

(1) Note Distribution Density: Note distribution density evaluates whether notes appear reasonably within a given range. This metric can be obtained through the probability distribution $T(i)$ of notes, as shown in Formula (10):

$$T(i) = P(I \geq i) = 1 - F(i) = ci^{-D} \tag{10}$$

i represents the pitch of the note, c represents the coefficient, and D represents the slope corresponding to the frequency of the note's occurrence. This means that when the slope

is a straight line, it suggests that the frequency of occurrence of every note is fairly even and causes a more desirable outcome.

Figure 2 shows the experimental findings on note density distribution of the four models, which are RFGAN, Tr-MTMG, HRNN, and MuseGAN. The values on the horizontal axis indicate the number of notes and the values on the vertical axis indicate the value of the logarithmic function of note occurrence frequency.

In order to make sure that the experiment is valid and fair, similar variables were used in this section. As illustrated, the slope of models Tr-MTMG and MuseGAN sharply declines within the 0 to 5 range. Compared to RFGAN, HRNN exhibits a non-linear slope between 15 and 25, indicating uneven note distribution with significant fluctuations in Tr-MTMG, HRNN, and MuseGAN. This suggests the generated melodies lack harmonic coherence.

In contrast, the music generated by the RFGAN model exhibits a trend closer to a straight line overall, indicating that the frequency of note occurrence in this model is relatively uniform. This figure further reveals that the RFGAN model generates fewer distinct note types, suggesting a higher note density. Thus, the model in this paper achieves the best generation results.

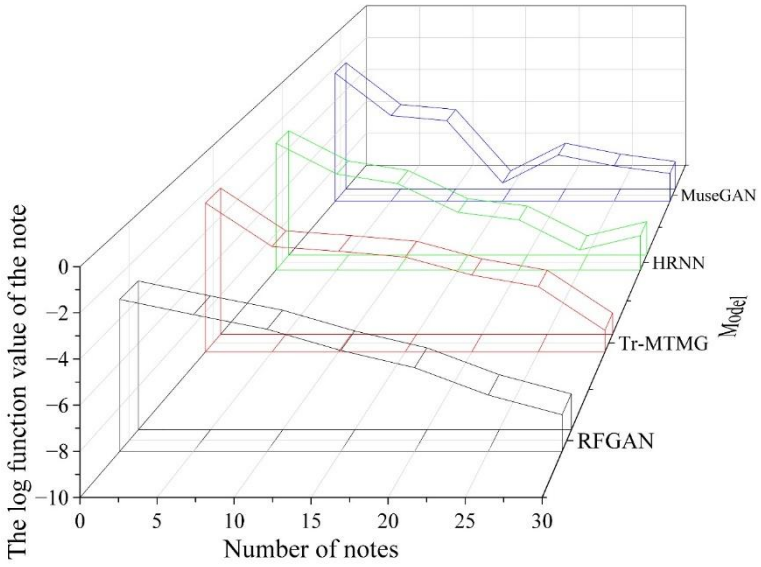

Figure 2: The slope corresponding to the frequency of the note

(2) Chord Matching Accuracy: To evaluate whether incorporating the learning network improves chord prediction accuracy, this paper analyzes the accuracy of chords generated by the RFGAN model.

Chord matching accuracy is defined as the similarity between chords produced by model samples and those input from real samples. Specifically:

$$\text{Chord match degree} = \frac{\sum_{m=1}^{p} E(y_m, \bar{y}_m)}{P} \tag{11}$$

$$E(y_m, \tilde{y}_m) = \begin{cases} 1, & \text{If } \ y_m = \tilde{y}_m \\ 0, & \text{If } \ y_m \neq \tilde{y}_m \end{cases} \tag{12}$$

Here, P denotes the number of musical phrases, y_m indicates the m chord in the

generated music, and y_m corresponds to the m chord in the actual music.

The chord matching scores for each model are shown in Figure 3.

After comparing the chord matching performance across the four models, it was found that the compositions generated by the RFGAN model achieved the highest chord matching scores. The general chord matching curve of the RFGAN model showed slight variations whereby the score remained high throughout and did not go below 0.6. It is indicated here that the RFGAN model has excellent learning abilities when it works with real sample data, which plays an essential part in the learning of musical composition.

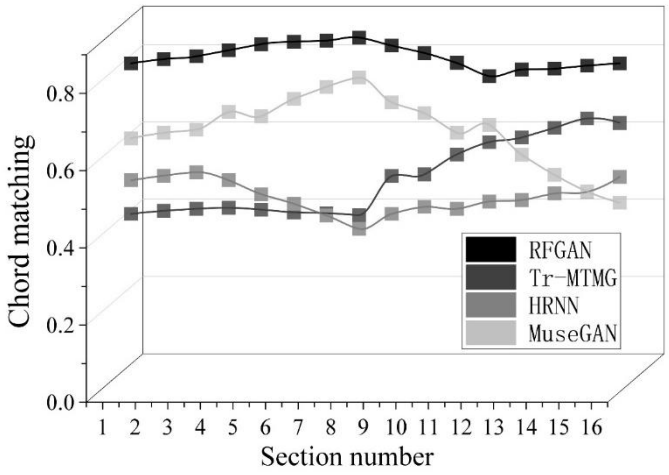

Figure 3: Chord matching of each model

(3) Harmony Analysis: To investigate whether RFGAN model parameters influence experimental outcomes, this study analyzes results from the perspective of iteration counts. Specifically, we evaluate the harmony of RFGAN-generated music by setting different training iteration thresholds.

Harmony is defined as the presence of similar chord progressions between tracks. That is, if chord progressions are similar, the music is harmonious. This section formalizes chord similarity to evaluate harmony among multi-track compositions. Specifically:

$$\text{Degree of harmony} = \frac{\sum_{p=1}^{p} \delta(cap_{k=1}^{K} C_p^k)}{KP} \tag{13}$$

$$\delta(a) = \begin{cases} 1, & \text{If} \quad a \neq \varnothing \\ 0, & \text{If} \quad a = \varnothing \end{cases} \tag{14}$$

In the formula, P represents the number of musical phrases, K denotes the number of instruments, and C_P^K indicates the chord appearing in the P th phrase of the K -track arrangement.

As shown in Figure 4, the RFGAN model achieves the highest musical harmony scores across 5K, 10K, 15K, and 20K iterations. Furthermore, as the number of iterations increases, the model's training effectiveness improves, resulting in multi-track musical compositions with higher levels of harmony.

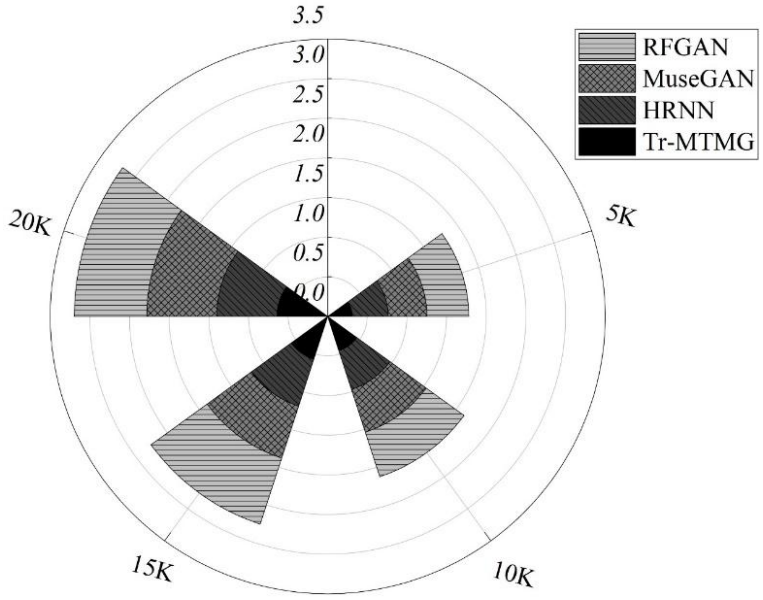

Figure 4: Comparison of music harmony in multi-track music

3.3.3 Arrangement Dimension

Due to the varying approaches to modeling musical data in melody generation research, existing models often focus on specific tasks such as standalone melody generation or accompaniment generation, resulting in significant differences in their data modeling components. Therefore, this paper selects LSTM-GAN as the baseline model and also incorporates classical generative models and probability distribution-based random models for comparison.

Random Baseline Model: Given the known distribution of real data across music attributes, attributes are randomly assigned to the baseline model based on their probability distribution, yielding data that is random yet conforms to the true music attribute distribution. Serves as one of the comparison models.

LSTM: An LSTM network trained using maximum likelihood estimation serves as one of the comparison models.

GRU: The GRU network, a variant of recurrent neural networks, serves as one of the comparison models. Compared to LSTM, it features fewer parameters and is easier to train.

Seq2Seq[22]: The Seq2Seq model comprises an encoder and a decoder, serving as a common conditional generative model and one of the comparison models.

RMC: Employs a standalone Relational Memory Core (RMC) for generative tasks without a discriminator or adversarial training, serving as one of the comparison models.

LSTM-GAN: This model consists of an LSTM-based generator and discriminator, achieving the task of generating melodies from lyrics text. It serves as one of the comparison models.

The BLEU and Self-BLEU metrics used in text generation tasks are also the most widely recognized metrics in current melody generation tasks. This paper employs these two metrics to evaluate the quality of generated melodies.

BLEU results are shown in Figure 5. First, a higher BLEU score indicates better generation quality. Our model RFGAN achieves the best performance on BLEU-2 (0.982) and BLEU-9 (0.962). LSTM performs best on BLEU-2 (0.825) and BLEU-3 (0.799). This indicates that these two models excel respectively in shorter and longer phrase dimensions, meaning they are most likely to generate melodies of relatively high quality.

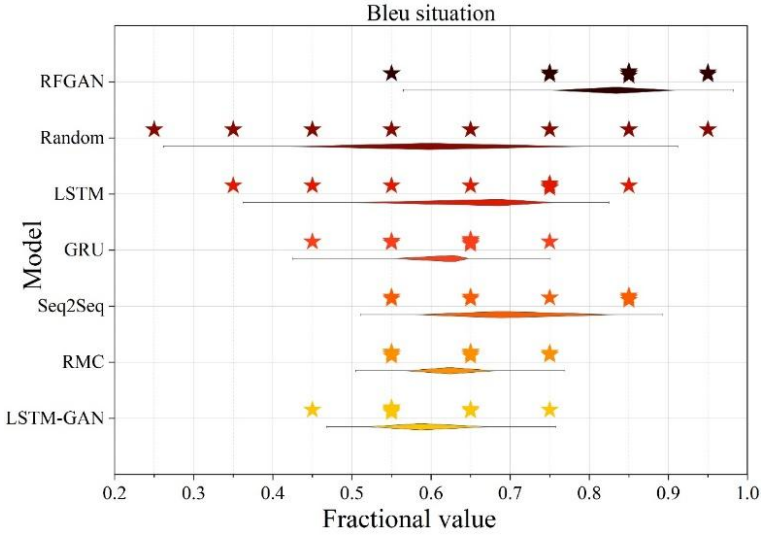

Figure 5: Bleu situation

When BLEU quality is sufficiently high, we hope the model can also balance Self-BLEU, i.e., diversity. A lower Self-BLEU value indicates richer diversity.

Self-BLEU results are shown in Figure 6. Our proposed algorithm RCGAN maintains a low Self-BLEU—indicating a good level of diversity—while generating pitch distributions that are already quite accurate and close to real data. This represents a favorable trade-off between quality and diversity.

Although the Random model exhibits high diversity, its performance on BLEU is mediocre. This is because it relies on probabilistic random selection without considering semantic context, resulting in generated melodies that are diverse yet random and often inaccurate.

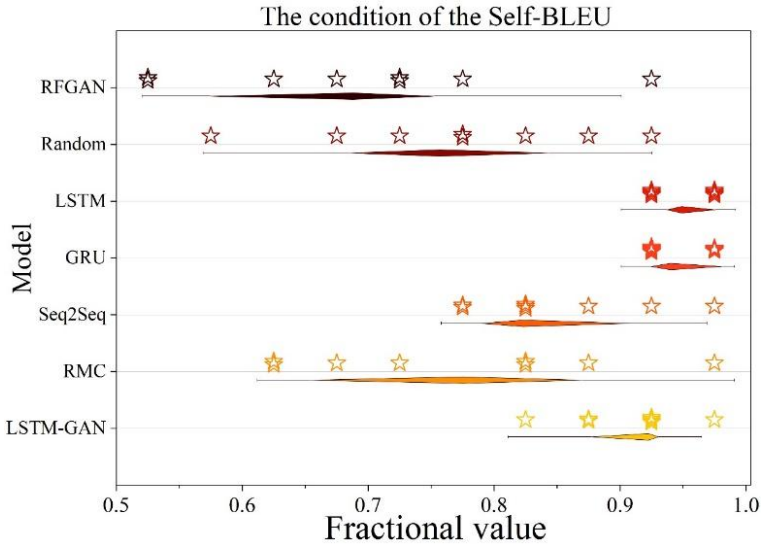

Figure 6: The condition of the Self-BLEU

4 Conclusion

This paper proposes a Recurrent Feature Generative Adversarial Network (RFGAN) by enhancing the multi-track music model (MuseGAN), enabling AI-generated music to form logical generation patterns. The created music is assessed based on the distribution of note

density, chord matching, and harmonic consistency tests, and then the RFGAN model is analyzed in terms of its arrangement.

Observe the density distribution curves of music samples produced by reference models Tr-MTMG, HRNN, and MuseGAN. The overall trend of notes used in the music produced by the RFGAN model is closer to a straight line, which means that the frequency distribution of notes is relatively more even. Less variety of note types and greater note density are associated with better model performance. The chord matching curve of the RFGAN model has low variation; it always stays at a match rate of over 0.6. Concerning the harmony of multi-track music, the effectiveness of the RFGAN model in training is increased as the number of iterations increases, and the harmony level of the multi-track music created by the model is also raised.

The RFGAN model has both a high BLEU and Self-BLEU score. In general, the RFGAN model is capable of producing music pieces having great harmony and high degree of musicality.

References

[1] Jiang, Y., & Sun, Z. (2025). Intelligent Music Content Generation Model Based on Multimodal Situational Sentiment Perception. Informatica, 49(5).

[2] Mycka, J., & Mańdziuk, J. (2025). Artificial intelligence in music: recent trends and challenges. Neural Computing and Applications, 37(2), 801-839.

[3] Kaliakatsos-Papakostas, M. A., Floros, A., & Vrahatis, M. N. (2013). Intelligent music composition. In Swarm Intelligence and Bio-Inspired Computation (pp. 239-256). Elsevier.

[4] Dey, M. T., Patra, S., & Mitra, S. (2025). Enhancing music education with innovative tools and techniques: the role of artificial intelligence in musical works. In Enhancing Music Education With Innovative Tools and Techniques (pp. 19-50). IGI Global Scientific Publishing.

[5] Fernández, J. D., & Vico, F. (2013). AI methods in algorithmic composition: A comprehensive survey. Journal of Artificial Intelligence Research, 48, 513-582.

[6] Zhang, Y. (2025). Increasing Emotional Perception in Academic Singing During Vocal Performance: The Use of AI Solutions. International Journal of Human–Computer Interaction, 1-9.

[7] Kyriakou, T., de la Campa Crespo, M. Á., Panayiotou, A., Chrysanthou, Y., Charalambous, P., & Aristidou, A. (2024, May). Virtual instrument performances (vip): A comprehensive review. In Computer Graphics Forum (Vol. 43, No. 2, p. e15065).

[8] Zhu, H. (2025). Music Transmission and Performance Optimization Based on the Integration of Artificial Intelligence and 6G Network Slice. International Journal of Network Management, 35(1), e70000.

[9] Doornbusch, P. (2010). Algorithmic composition: Paradigms of automated music generation. Computer Music Journal, 34(3), 70-74.

[10] Wang, A. A. (2024, July). Intelligent Music Generation: Reducing Anxiety in High School Students. In 2024 20th International Conference on Natural Computation, Fuzzy

Systems and Knowledge Discovery (ICNC-FSKD) (pp. 1-8). IEEE.

[11] Huang, J. L., Chiu, S. C., & Shan, M. K. (2012). Towards an automatic music arrangement framework using score reduction. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 8(1), 1-23.

[12] Cao, C. (2022). Creation characteristics of music piano arrangement based on distributed sensors. Mobile information systems, 2022(1), 3086542.

[13] Liu, W. (2023). Literature survey of multi-track music generation model based on generative confrontation network in intelligent composition. The Journal of Supercomputing, 79(6), 6560-6582.

[14] Li, F. (2024). Chord-based music generation using long short-term memory neural networks in the context of artificial intelligence. The Journal of Supercomputing, 80(5), 6068-6092.

[15] Sun, G., & Wang, H. (2025). Deep Learning-based Intelligent Music Composition System: Assisting Composition and Arrangement. WSEAS Transactions on Computer Research, 13, 342-348.

[16] Zhu, H., Liu, Q., Yuan, N. J., Zhang, K., Zhou, G., & Chen, E. (2020). Pop music generation: From melody to multi-style arrangement. ACM Transactions on Knowledge Discovery from Data (TKDD), 14(5), 1-31.

[17] Hu, X. (2023). Reinforcement learning-based algorithms for music improvisation and arrangement in sensor networks for the Internet of Things. Scalable Computing: Practice and Experience, 24(3), 499-510.

[18] Wen, Y. W., & Ting, C. K. (2022). Recent advances of computational intelligence techniques for composing music. IEEE Transactions on Emerging Topics in Computational Intelligence, 7(2), 578-597.

[19] Hernandez-Olivan, C., Hernandez-Olivan, J., & Beltran, J. R. (2022). A survey on artificial intelligence for music generation: Agents, domains and perspectives. arXiv preprint arXiv:2210.13944.

[20] Tabak, C. (2023). Intelligent music applications: innovative solutions for musicians and listeners. Uluslararası Anadolu Sosyal Bilimler Dergisi, 7(3), 752-773.

[21] Yongjun He & Shijie Zhang. (2025). Enhancing art creation through AI-based generative adversarial networks in educational auxiliary system. Scientific Reports,15(1),29202-29202.

[22] Haoyue Zhang,Chunmei Zhao,Zhengbin He & Tianming Ma. (2025). A seq2seq model based on autocorrelation attention for long-term orbit prediction using two-line element. Advances in Space Research,76(3),1729-1739.