Data Engineering and Applications

Review Paper

Data-Driven Generative Adversarial Networks for Music Composition: Models, Data Representation, and Application

Dorota Urban[1, *], Joanna Tomaszewski[1] and Emilia Wisniewski[2]

[1] Faculty of Electronics and Information Technology, AGH University of Science and Technology, 30-059 Krakow, Poland
[2] Faculty of Automatic Control, Electronics and Computer Science, Lublin University of Technology, 20-618 Lublin, Poland
*Corresponding author: d.urban@agh.edu.pl

Abstract. With the rapid advancement of artificial intelligence technology, generative adversarial networks (GANs) have demonstrated immense potential in the field of music composition. This paper reviews the theoretical foundations, model research progress, application analysis, and challenges faced by GANs in music composition, while also outlining future research directions. First, the basic structure of GANs, music data representation methods, and the principles of GAN-generated music are introduced, laying the theoretical foundation for subsequent research. Next, the current state of GAN-based music composition models—including sequence generation, image-to-sequence conversion, and models integrating other technologies—is elaborated in detail, with comparative analyses of these approaches. Subsequently, the application of GANs in music style imitation and transformation, musical improvisation, and music creation assistance is explored in depth, demonstrating their broad prospects in the musical domain. However, GANs in music composition also face challenges such as unstable model training and generated music lacking emotional expression. Future research can focus on improving model architecture, integrating multimodal information, and enhancing music semantic understanding to advance GANs in the field of music composition.

Keywords: *Generative Adversarial Networks; Music Composition; Model Research; Application Analysis; Challenges and Prospects*

Received on 22 October 2024, Accepted on 2 Jan 2025, Published on 12 Jan2025

Introduction

Amidst the dazzling constellation of human cultural arts, music transcends geographical and cultural boundaries with its unique charm and infectious power, touching the depths of human souls [1]. From ancient folk ballads to grand symphonic works, from rousing rock melodies to tender pop songs, musical creation has always been the crystallization of human wisdom and emotion, embodying the dedication and inspiration of countless creators [2]. The traditional music composition process typically requires composers to undergo lengthy and arduous professional training, accumulate profound theoretical knowledge and practical experience, and invest substantial time and effort into conceptualization, composition, revision, and refinement [3]. However, with the rapid advancement of technology, new techniques and approaches continuously emerge in the field of artificial intelligence, infusing fresh vitality and possibilities into this ancient and creative endeavor. Generative Adversarial Networks (GANs) stand out as a particularly brilliant "new star" among these innovations [4].

In recent years, GANs have achieved remarkable accomplishments in numerous fields such as image generation and natural language processing, demonstrating formidable generative capabilities and the ability to learn complex data distributions. This has sparked significant interest in their immense potential for application in music composition [5-7]. Currently, research on GANs in music composition has made considerable progress, with numerous research teams worldwide actively exploring this frontier field [8]. Researchers have designed and implemented diverse GAN-based music composition models from various perspectives, conducting extensive experiments and trials across different music creation tasks and styles [9]. Reference [10] focuses on imitating a single musical style, aiming to generate music fragments highly similar to those of classical composers like Mozart and Bach; Reference [11] blends elements from multiple musical styles to create works with distinctive characteristics; Reference [12] targets specific application scenarios, such as generating background music for videos or games, or assisting composition instruction in music education. These achievements fully demonstrate GANs' broad applicability and powerful potential in music composition, laying a solid foundation for subsequent research and development [13-15].

However, despite these achievements, the application of GANs in music composition remains in an exploratory and evolving stage, facing numerous challenges that require urgent resolution [16]. Regarding musical quality, while GANs can generate pieces with discernible style and structure, they still lag significantly behind human-composed music in emotional expression, artistic merit, and innovation [17]. In music semantic understanding, how to enable GANs to better grasp musical grammar rules—such as the correctness of harmonic progressions [18], the coherence of melodic development [19], and the rationality of musical structure [20]—remains an unsolved problem. This limitation constrains the artistic integrity and professionalism of generated musical works.

Given the current state of GAN research in music composition, its challenges, and its vast potential applications, this paper aims to provide a systematic and in-depth review of the field. First, it will detail the fundamental theoretical foundations of GANs in music composition. Next, it will thoroughly analyze research progress in GAN-driven music composition models, offering detailed descriptions and comparative analyses of various models based on different technical approaches and application scenarios, showcasing current research achievements and innovations. Simultaneously, it will explore the practical applications of GANs in music composition, analyzing their advantages, limitations, and potential future development trends. Furthermore, it will comprehensively examine the challenges and issues GANs face in music composition, conduct in-depth root cause analyses of these problems, and discuss possible solutions and development directions, providing valuable references and insights for subsequent research.

Theoretical Foundations of GANs in Music Composition

Basic Structure of GANs

Generative Adversarial Networks (GANs) have garnered significant attention within the artificial intelligence and machine learning communities due to their unique approach to data generation. Unlike traditional generative models that rely on explicit probability distributions or reconstruction-based objectives, GANs introduce a competitive framework that drives two neural networks to improve through mutual confrontation. This adversarial learning paradigm not only enables the discovery of intricate data patterns but also fosters the synthesis of highly realistic samples across various domains. The versatility of GANs lies in their ability to adapt to a wide range of data types, from images and text to audio and symbolic sequences, making them particularly well-suited for creative applications such as music composition. Through iterative training, these networks evolve to capture the subtle statistical properties and structural nuances of real-world datasets. As a result, GANs have emerged as a powerful tool for generating novel content that closely mirrors authentic data, opening new avenues for innovation in computational creativity.

GANs consist of two neural networks: a Generator and a Discriminator [21]. The Generator aims to map random noise vectors into the data space, producing samples that are as realistic as possible. The Discriminator is responsible for distinguishing between generated samples and real data samples. During training, the Generator and Discriminator engage in an adversarial process, continuously adjusting their parameters until they reach a Nash equilibrium. At this point, the samples generated by the Generator align with the distribution of real data, making it impossible for the Discriminator to distinguish between real and synthetic samples.

The fundamental training objective of GANs can be formally described as a minimax game, which mathematically captures the adversarial relationship between the generator and the discriminator. The generator strives to

DOI: 10.64972/dea.2025.v3i1.33

produce data that is indistinguishable from real data, while the discriminator attempts to correctly distinguish between real and generated samples. This process is expressed by the following objective function:

$$\min_{G}\max_{D} V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_{x}(z)}[\log(1 - D(G(z)))] \quad (1)$$

In this equation, G denotes the generator, D the discriminator, x the real data sampled from the true data distribution p_{data}, and z the random noise vector sampled from a prior distribution p_z. The generator G maps noise z to the data space, while the discriminator D outputs the probability that a given sample is real. The minimax game leads both networks to improve iteratively: D maximizes its ability to distinguish real from fake, and G minimizes its loss by generating more convincing samples. This adversarial setup continues until a Nash equilibrium is reached, ideally when generated samples are indistinguishable from real data.

The network architectures of the Generator and Discriminator can be designed according to specific tasks and data types. For music composition tasks, common Generator architectures include fully connected layers, recurrent neural networks (RNNs), and convolutional neural networks (CNNs) [22]. Fully connected layers capture global features in musical data, recurrent neural networks process temporal information in music sequences, and convolutional neural networks extract local features from musical data [22-24]. The discriminator's architecture typically mirrors the generator's: if the generator uses LSTMs, the discriminator may also employ LSTMs or convolutional neural networks to evaluate and distinguish generated music sequences.

To further clarify the adversarial learning mechanism of GANs as applied to music composition, it is helpful to visualize the fundamental architecture and training process. The following diagram illustrates the interaction between the generator and discriminator, highlighting how each network is updated through their alternating competition. By depicting the data flow and feedback loops, this schematic provides a clear overview of the core components and their roles in generating music that closely resembles authentic compositions. Understanding this structure is essential for grasping how GANs achieve realism and diversity in music generation tasks.

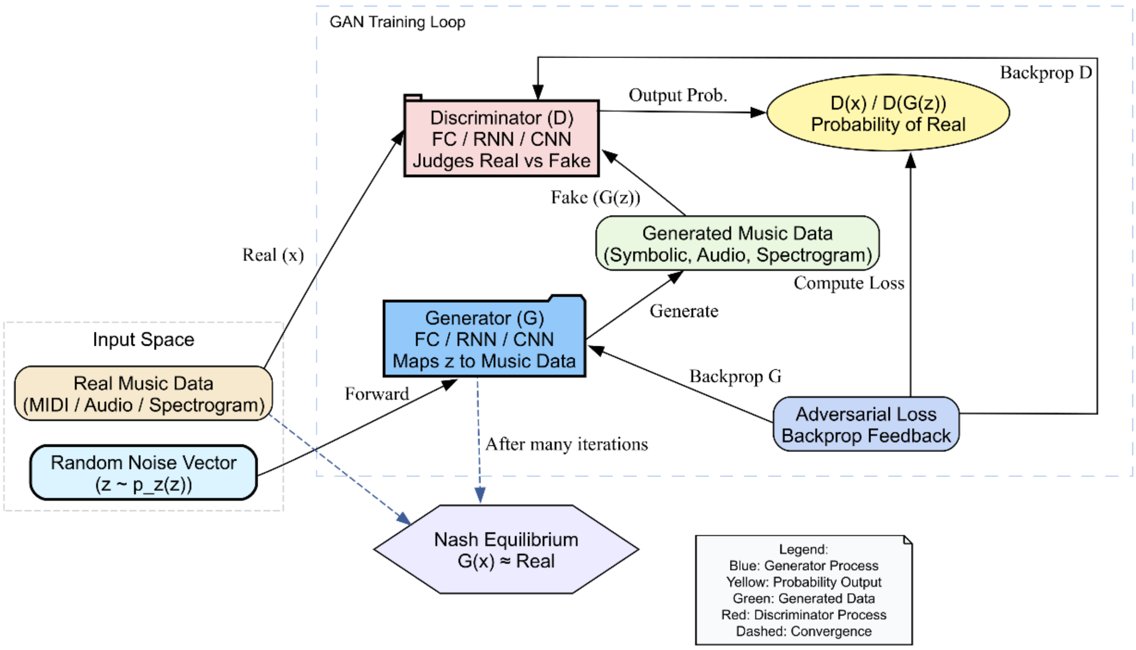

Figure 1. Basic structure and adversarial process of a Generative Adversarial Network (GAN) in music composition.

The city’s architectural landscape reflects a harmonious blend of historical heritage and modern development. As dusk settles in, the interplay of natural light and artificial illumination highlights the distinct contours of each building. The vibrant atmosphere suggests a dynamic urban life, where tradition and innovation coexist. This setting provides not only a picturesque backdrop but also a glimpse into the city's evolving identity, shaped by its people, culture, and aspirations.

Music Data Representation Methods

Before discussing how Generative Adversarial Networks (GANs) can be applied to music composition, it is essential to recognize that the successful training of any neural network depends heavily on the quality and form of its input data. Music, as a highly structured and multidimensional art form, presents unique challenges for machine learning models. Unlike images or text, music embodies rich temporal dynamics, intricate harmonic relationships, and subtle expressive nuances that must be accurately captured and represented. The complexity of musical information requires careful consideration when designing data formats suitable for neural network learning. The choice of representation not only affects the learning efficiency of the model but also determines the degree to which musical structure, semantics, and expressiveness can be preserved throughout the generative process. In this context, selecting an effective method for encoding musical data becomes a foundational step toward enabling GANs to model and generate musically meaningful content.

Before employing GANs for music composition, music data must be converted into suitable representations for neural network learning and processing [25]. Common music data representation methods include notation-based representation [26], audio waveform representation [27], and spectral representation [28], with detailed analysis presented in Table 1.

Table 1. Advantages and Disadvantages of Different Music Data Representation Methods

Representation Method	Advantages	Disadvantages
Symbolic Representation	Can precisely represent the structure and semantic information of music Data size is relatively small	Requires symbolic encoding and decoding May lose some audio details
Audio Waveform Representation	Can fully preserve audio information High quality of generated music	Large data size High processing difficulty Requires large models and computing resources
Spectral Representation	Effectively represents frequency and energy information Suitable for music generation and classification	Requires Fourier transform and other operations May cause spectral distortion issues

The choice of music data representation critically influences the effectiveness of GAN-based music composition models. Symbolic representations, such as MIDI, strike a balance between computational efficiency and musical structure preservation, making them a preferred choice for many symbolic music generation tasks. However, the abstraction involved in symbolic encoding may lose expressive nuances present in original audio. On the other hand, audio waveform representations maintain the richest level of musical detail, which is essential for capturing timbral and expressive characteristics, albeit at the cost of increased computational demands. Spectral representations offer a middle ground, enabling models to leverage frequency-based features for style transfer and classification. Selecting the appropriate representation method is therefore fundamental to aligning model capabilities with specific compositional objectives and available computational resources.

Symbolic representation encodes musical elements such as notes, rhythms, and chords using symbols or numerical values, specifically in MIDI format [29]. MIDI files contain information about musical notes, durations, volumes, instruments, and more, which can be converted into tensor form suitable for neural network input using specialized libraries and tools [30]. Audio waveform representation directly samples the audio signal of music to obtain waveform data. Spectral representation converts the audio signal into a spectrogram, effectively capturing the frequency components and energy distribution of music, making it suitable for tasks like music classification and generation.

In addition to these widely used methods, hybrid approaches have emerged that leverage the strengths of multiple representations to enhance the expressiveness and accuracy of music data modeling. For instance, some systems combine symbolic representations with spectral features, allowing neural networks to access both high-level musical structure and detailed timbral information during training. This multi-faceted perspective enables more nuanced music generation and analysis, as models can simultaneously learn about the relationships between notes, rhythmic patterns, and the underlying acoustic characteristics of the sound. Furthermore, recent advances in data preprocessing have made it possible to encode additional musical

DOI: 10.64972/dea.2025.v3i1.33

attributes—such as articulation, dynamics, and phrasing—into the input tensors, further enriching the information available to generative models. The choice of representation is often influenced by the specific requirements of the composition task, the style of music being targeted, and the capabilities of the neural network architecture in use. By carefully selecting and potentially integrating different representation strategies, researchers can more effectively capture the multifaceted nature of music, paving the way for more sophisticated and musically convincing generative systems.

Principles of GAN Music Generation

The successful application of Generative Adversarial Networks (GANs) to music composition relies heavily on understanding both the underlying data structure and the unique characteristics of musical information. Unlike images or text, music encompasses intricate temporal patterns, harmonic relationships, and expressive dynamics that present additional challenges for generative modeling. Therefore, designing effective representations—such as symbolic sequences, audio waveforms, or spectrograms—is a crucial first step in adapting GAN frameworks for music generation tasks. Once the data is suitably formatted, the essential adversarial process can be employed to synthesize new musical material that captures the salient features of the input domain.

The process of generating music with GANs is analogous to generating other types of data [31]. During training, the generator receives random noise vectors as input and generates musical data through forward propagation of the network. The discriminator simultaneously receives real music data and music data generated by the generator, learning to distinguish between the two through training. Based on feedback from the discriminator, the generator continuously adjusts its parameters to gradually align the generated music data with the distribution of real data. As training progresses, both the generator and discriminator improve their capabilities, eventually reaching an equilibrium state where the generator produces music indistinguishable from real compositions [32].

Throughout this adversarial process, careful tuning of hyperparameters, network architecture, and data preprocessing is essential to achieve musically meaningful outputs. Researchers often employ techniques such as data augmentation, regularization, and early stopping to prevent overfitting and mode collapse, thereby ensuring that the generated music retains both diversity and coherence. In addition, evaluation metrics—both objective and subjective—are used to assess the quality of the generated music, considering factors like melodic structure, harmonic accuracy, and listener satisfaction. By iteratively refining the training process and incorporating domain expertise, GAN-based systems are increasingly capable of producing compositions that reflect the complexity and artistry of human-created music.

GAN Training Challenges and Solutions

Despite the promising advancements made in the application of Generative Adversarial Networks (GANs) to music composition, the process of training such models remains an intricate and highly sensitive endeavor. The adversarial nature of GANs, which relies on the delicate balance between the generator and the discriminator, often introduces significant complexity into the optimization dynamics. Minor adjustments in model architecture, hyperparameters, or data preprocessing strategies can yield disproportionately large impacts on the resulting quality and stability of generated music. Furthermore, the inherently high dimensionality and temporal structure of musical data amplify the risk of instability, as models must learn to capture both local nuances and global coherence within musical sequences. These factors collectively contribute to a landscape where repeated experimentation and fine-tuning are necessary, yet even well-designed systems may encounter unforeseen difficulties during training. As a result, the development of practical GAN-based music generation systems frequently requires not only technical expertise but also iterative exploration and a deep understanding of both machine learning principles and musical structure.

Despite GANs' immense potential in music composition, training these models presents significant challenges. A primary issue is pattern collapse, where the generator produces only a limited set of musical samples, failing to capture the diversity of real data [33]. Additionally, the GAN training process is often unstable, prone to stagnation or divergence [34].

Another significant obstacle in GAN training for music composition is the sensitivity to hyperparameter settings and network architecture choices. Small modifications in learning rates, batch sizes, or layer configurations can lead to drastically different results, often requiring extensive empirical tuning and experimentation. Furthermore, the adversarial nature of GANs makes them susceptible to issues such as vanishing or exploding gradients, which can disrupt the delicate balance between the generator and discriminator. This instability is exacerbated by the high dimensionality and complex temporal dependencies inherent in musical data, making it challenging to maintain consistent learning dynamics throughout the training process. Researchers have attempted to mitigate these difficulties by introducing regularization techniques, careful initialization strategies, and alternative optimization schemes. Nevertheless, achieving reliable and reproducible results remains a persistent challenge. Moreover, the evaluation of generated music quality itself is inherently subjective, further complicating efforts to systematically improve model performance and stability. As a result, the development of robust training protocols and objective assessment metrics continues to be an active area of research within GAN-driven music composition.

To address these issues, researchers have proposed various approaches. Methods to enhance GAN performance in music composition include Conditional GANs [35], WGAN [36], and WGAN with gradient penalties [37], as detailed in Table 2.

Table 2. Improved Methods for Enhancing GAN Performance in Music Composition

Improvement Method	Description	Advantages
Conditional GAN	Introduce conditional information into generator and discriminator to constrain generated music with specific conditions	Can control style and content of generated music Enhances diversity and specificity of generated music
WGAN	Use Wasserstein distance to replace traditional Jensen-Shannon divergence as loss function Improves training stability and generation quality	More stable training Generates higher quality music
WGAN with Gradient Penalty	Add gradient penalty term on top of WGAN further improving training stability and generation quality	Training process more stable better generation quality and diversity

Each improvement method offers unique advantages that address specific challenges in GAN training for music composition. Conditional GANs provide a powerful mechanism for guiding the generation process, enabling the production of music that conforms to user-specified styles, genres, or moods. Wasserstein-based approaches, by redefining the loss function, significantly enhance training stability and sample diversity, which are crucial for generating music with varied structures and emotional expressions. The application of gradient penalties further refines model robustness, helping to mitigate issues such as mode collapse. By leveraging these improvements, researchers can develop GAN models that are better suited to the intricacies of musical data, ultimately enhancing the quality and usability of generated compositions.

Analysis of GAN Loss Functions

The loss function plays a crucial role in the training process of GANs, determining the learning direction and optimization objectives of the generator and discriminator [38]. Common GAN loss functions include the original GAN loss function [39], the least squares loss function [40], and the Wasserstein loss function [41], with a comparative analysis shown in Table 3.

Table 3. Comparative Analysis of Different GAN Loss Functions

Loss Function	Features	Advantages and Disadvantages
Original GAN Loss	Based on Jensen-Shannon divergence Minimizes distribution difference between generated and real data	Simple implementation Theoretically good generation performance But prone to gradient vanishing in late training causing training stagnation
Least Squares Loss	Minimizes mean squared error between generated and real data Improves training stability	Mitigates gradient vanishing to some extent But may introduce mode collapse risks Limits diversity of generated music
Wasserstein Loss	Based on Earth Mover's Distance Measures difference between generated and real data more accurately Improves training stability and generation quality	More stable training Better generation quality But requires Lipschitz condition Implementation is relatively complex
Wasserstein Loss with Gradient Penalty	Adds gradient penalty term on top of Wasserstein loss Further improves training stability and generation quality	Good training stability High generation quality But computational cost is higher Implementation complexity is increased

DOI: 10.64972/dea.2025.v3i1.33

The selection of an appropriate loss function is pivotal in shaping the learning dynamics of GANs for music generation. Each loss function brings distinct strengths and weaknesses, impacting both the convergence behavior during training and the diversity of generated outputs. For instance, while the original GAN loss function is straightforward to implement, it can lead to training stagnation in practice. Least squares and Wasserstein losses help alleviate some of these issues, each offering different trade-offs between stability, diversity, and implementation complexity. A thorough understanding of these loss functions enables practitioners to tailor GAN models to the specific requirements of music composition tasks, balancing the goals of realism, diversity, and expressive potential.

The original GAN loss function is based on Jensen-Shannon divergence, training the generator and discriminator by minimizing the distributional difference between generated and real data [39]. However, this loss function is prone to vanishing gradients during late training stages, causing model training to stall.

Generative adversarial networks (GANs) have revolutionized the field of generative modeling by introducing a competitive framework between two neural networks: the generator and the discriminator. The core objective is to train the generator to produce samples indistinguishable from real data, while the discriminator aims to distinguish between authentic and generated samples. The choice of loss function plays a pivotal role in guiding this adversarial process and significantly influences both the stability and quality of training outcomes.

The least-squares loss function enhances training stability by minimizing the mean squared error between generated and real data [40]. It mitigates gradient vanishing to some extent but may introduce additional risks of mode collapse.

Nevertheless, while the least-squares loss addresses certain deficiencies associated with traditional adversarial losses—such as unstable gradients—it is not without its trade-offs. Researchers have continued to explore alternative loss equations and regularization strategies to further improve convergence and model diversity. The ongoing investigation into loss functions underscores the importance of balancing training stability with the need to generate diverse and high-fidelity samples.

The Wasserstein loss function, based on the Earth Mover's Distance, more accurately measures the difference between generated and real data, enhancing both training stability and generation quality [41]. Furthermore, the Wasserstein loss with gradient penalty further improves the model's training performance.

The specific loss functions used for the discriminator and generator in the standard GAN framework further clarify the learning objectives for each component. The discriminator's goal is to maximize the probability of correctly identifying real and generated data, while the generator aims to maximize the probability that the discriminator classifies generated samples as real. The loss functions are defined as follows:

$$\begin{gathered} L_D = -\left(\mathbb{E}_{x \sim p_{\text{lata}}(x)}[\log\, D(x)] + \mathbb{E}_{z \sim p_x(z)}[\log\,(1 - D(G(z)))]\right) \\ L_G = -\mathbb{E}_{z \sim p_z(z)}[\log\, D(G(z))] \end{gathered} \tag{2}$$

In equation (2), L_D represents the discriminator's loss, which penalizes it when it misclassifies real or generated samples. L_G is the generator's loss, which penalizes it when generated samples are correctly identified as fake by the discriminator. During training, the two networks are updated alternately: the discriminator's parameters are adjusted to minimize L_D, and the generator's parameters are updated to minimize L_G. This iterative optimization process is central to the adversarial learning mechanism of GANs, and it underpins the gradual improvement of both the generator's output quality and the discriminator's classification accuracy during training.

DOI: 10.64972/dea.2025.v3i1.33

Research Progress on GAN-Driven Music Composition Models

Sequence-Generation-Based GAN Models

Sequence-generation-based GAN models primarily focus on music's temporal and structural characteristics, modeling and generating music as sequential data [42]. These models typically employ recurrent neural networks (RNNs), such as LSTMs or GRUs [43], as the network architecture for both generators and discriminators to capture long-term dependencies within musical sequences.

Among sequence-generation-based GAN models, much attention has been devoted to those that effectively capture the temporal evolution and underlying structure of music. These models typically leverage recurrent neural networks (RNNs) as their primary architecture, given RNNs' intrinsic ability to model long-range dependencies in sequential data. By processing musical information as a series of events—such as notes, chords, or rhythmic patterns—these architectures are well-suited for generating musically meaningful sequences that maintain coherence across time. Researchers have explored various encoding schemes for representing musical elements, ranging from simple one-hot encodings to more nuanced continuous representations that preserve important details of musical expression.

C-RNN-GAN is a classic sequence-based GAN model that represents notes using quadruplets of continuous pitch, duration, intensity, and onset time, learning music sequence generation through adversarial training [44]. C-RNN-GAN has achieved notable results in classical music generation tasks, producing music with coherent and plausible melodies and harmonies.

Building on the foundational work of C-RNN-GAN, subsequent research has sought to address its limitations and further enhance the quality and diversity of generated music. Efforts have included refining the data representation to capture more subtle aspects of musical performance, experimenting with alternative network structures for both the generator and discriminator, and integrating additional regularization strategies to stabilize training. Furthermore, evaluation of generated sequences often involves not only objective measures, such as statistical similarity to real music, but also subjective assessments by musicians and listeners to ensure artistic validity. The ongoing development of sequence-based GAN models continues to push the boundaries of automated music composition, demonstrating the potential for artificial intelligence to contribute meaningfully to creative domains.

SeqGAN combines reinforcement learning and GAN concepts for learning discrete sequence data [45]. In music generation, SeqGAN treats musical sequences as a series of decision processes. The generator produces a note or note combination at each step, while the discriminator evaluates the entire sequence and provides feedback signals to guide the generator toward producing higher-quality musical sequences. Based on the above analysis, a comparative summary of sequence-based GAN models is presented in Table 4.

Table 4. Comparison of Sequence-Based GAN Models

Model Name	Generator Network Structure	Discriminator Network Structure	Features
C-RNN-GAN	LSTM Network	CNN Network	Uses continuous quadruplets to represent notes Can capture continuity and dynamics of music sequences
SeqGAN	RNN or LSTM Network	CNN Network	Combines reinforcement learning Treats music generation as sequence decision process Can generate music sequences with long-term structure

Sequence-based GAN models have proven particularly effective in capturing the temporal dependencies and structural regularities inherent in music. By modeling music as a sequence, these models are capable of generating compositions that exhibit coherent melodic progression and harmonic continuity. The use of recurrent architectures, such as LSTMs, allows the generator to remember musical context over long time spans, which is critical for producing musically meaningful phrases. Comparative studies of these models highlight the importance of balancing sequence modeling capabilities with computational tractability, ensuring that generated music maintains both technical correctness and artistic appeal.

When generating musical sequences, the probability of generating a complete sequence X composed of elements $(x_1, x_2, \dots, x_T)$ is typically factorized according to the chain rule of probability. This approach models the sequential nature of music, where each note or event depends on the preceding elements. The following equation illustrates this decomposition:

$$P(X) = \prod_{t=1}^{T} P(x_t \mid x_1, x_2, \ldots, x_{t-1}) \qquad (3)$$

Here, $P(X)$ indicates the probability of the entire musical sequence, and $P(x_t \mid x_1, \ldots, x_{t-1})$ denotes the conditional probability of generating the t-th element given all previous elements. This equation allows the model to capture temporal dependencies and complex structures inherent in musical compositions, enabling the generation of coherent and musically plausible sequences.

Image-to-Sequence GAN Model

This model analogizes musical sequences with image data, translating the structure and features of music into image form. It then leverages GAN's strengths in image generation to convert the image back into a musical sequence, thereby achieving music generation [46].

This innovative approach effectively bridges the domains of computer vision and music information retrieval, allowing models to capitalize on the spatial processing capabilities of convolutional neural networks (CNNs). By transforming musical representations—such as piano roll matrices or spectrograms—into image-like formats, the model enables the exploitation of well-established image generation techniques within the GAN framework. This not only enhances the model's ability to capture and reproduce intricate musical textures and patterns but also facilitates the exploration of cross-modal relationships between visual and auditory data. The process typically involves the generator crafting realistic piano roll images from random noise or latent codes, while the discriminator evaluates the authenticity of these images relative to real musical data. Once generated, the synthesized images are systematically decoded back into symbolic or audio-based musical sequences, ensuring the preservation of essential musical attributes such as rhythm, harmony, and melodic contour. This image-to-sequence paradigm opens up new possibilities for music generation, supporting the creation of stylistically diverse and structurally coherent compositions. Moreover, it provides a flexible foundation for further integration with other modalities, such as lyrics or performance dynamics, thereby enriching the creative potential of GAN-driven music composition systems.

Image-to-Music GAN models treat piano roll representations of music as images. Using convolutional neural networks as both generator and discriminator, they learn the mapping relationship from image to sequence [47]. The generator produces piano roll images based on input random noise, while the discriminator judges whether the image is a genuine piano roll. The resulting piano roll can ultimately be converted into a music sequence. Beyond Image-to-Music GAN, image-to-sequence GAN models include Pix2Pix Music GAN [48], MusicVAEGAN [49], Audio Spectrogram GAN [50], and MusicStyleGAN [51]. A comparative analysis is presented in Table 5.

Table 5. Comparison of Image-to-Sequence GAN Models

Model Name	Generator Network Structure	Discriminator Network Structure	Features
Image-to-Music GAN	CNN Network	CNN Network	Treats piano roll as image Utilizes CNN image processing ability to generate music sequences from images
Pix2Pix Music GAN	U-Net Network	PatchGAN Network	Uses conditional GAN to generate music sequences from input images
MusicVAEGAN	VAE-GAN Hybrid Network	CNN Network	Combines VAE probabilistic modeling with GAN generation Improves diversity and quality of generated music
Audio Spectrogram GAN	U-Net Network	PatchGAN Network	Generates music audio from audio spectrograms Retains frequency and energy information of music
MusicStyleGAN	StyleGAN Network	StyleGAN Network	Uses StyleGAN style control to achieve diverse music style generation

Image-to-sequence GAN models introduce an innovative paradigm by treating musical representations as images, thereby leveraging advances in image generation for music composition. This approach enables the exploitation of convolutional architectures' strengths in capturing spatial patterns, which translates into improved modeling of harmonic and rhythmic structures in music. By converting between image-like representations and sequential musical data, these models facilitate new forms of cross-modal creativity and enable novel applications in music arrangement and style transformation. The integration of image-based techniques thus expands the expressive range of GANs in music generation and opens pathways for interdisciplinary research.

In order to concretely demonstrate how image-to-sequence GAN models are utilized for music generation, it is valuable to present the step-by-step process involved in transforming musical data. The following figure visualizes the workflow in which a piano-roll image—serving as an intermediate, image-like representation of music—is generated by the GAN and subsequently converted back into a symbolic music sequence. This approach effectively leverages powerful image processing techniques for music creation, bridging the gap between visual and auditory data domains. The diagram offers an intuitive understanding of the data transformations and neural network operations that underlie this innovative method, thus enhancing comprehension of the model's practical implementation in the context of AI-driven music composition.

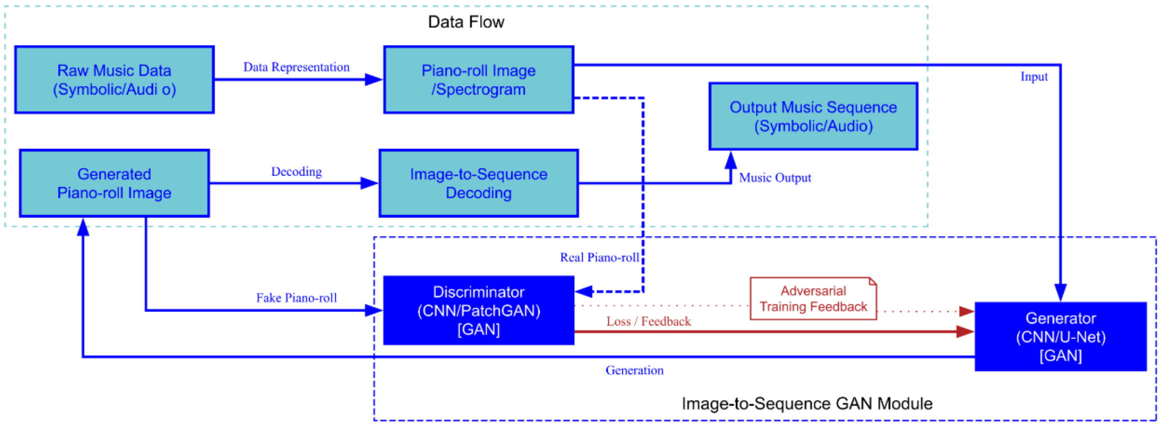

Figure 2. Process of converting piano-roll images to music sequences using image-to-sequence GAN models.

The carefully arranged instruments in the laboratory demonstrate the meticulous approach adopted by researchers. Each piece of equipment plays a vital role in ensuring the accuracy and reliability of experimental results. The environment is designed to minimize external influences, thereby maintaining consistency throughout the research process. This setup exemplifies the dedication to scientific rigor and attention to detail that are fundamental to experimental studies in this field.

GAN Models Combined with Other Technologies

In recent developments within music generation research, the inherent limitations of standalone GAN architectures have prompted researchers to explore more synergistic frameworks. Single-model approaches, while capable of producing stylistically convincing outputs, often struggle with issues such as limited diversity, mode collapse, and insufficient structural coherence in extended compositions. As a result, there is a growing trend toward combining the strengths of different deep learning models to address these challenges and push the boundaries of algorithmic creativity.

To further enhance GAN performance and effectiveness in music composition, GANs can be integrated with other technologies such as Variational Autoencoders (VAE) [52], Transformers [53], and Attention Mechanisms [54], leveraging their respective strengths.

The integration of these advanced architectures introduces several advantages. For example, coupling GANs with VAEs enables the model to exploit a well-structured latent space, improving both the diversity and controllability of generated music. Similarly, the adoption of Transformer-based designs and attention mechanisms empowers the system to capture long-range dependencies and intricate relationships within musical sequences, leading to more coherent and expressive outputs. Such hybrid models not only inherit the generative power of GANs but also benefit from the representational richness and learning stability offered by VAEs and attention-driven architectures. This multidimensional approach represents a promising direction for overcoming persistent obstacles in automated music composition, paving the way for more sophisticated, flexible, and musically meaningful generative systems.

In recent years, the intersection of generative adversarial networks (GANs) with other deep learning paradigms has emerged as a promising direction for advancing the field of algorithmic music composition. Among these, the fusion of GANs with variational autoencoders (VAEs) stands out as an innovative approach for overcoming some of the inherent limitations faced by traditional GAN architectures. By leveraging the strengths of both frameworks, hybrid models aim to address challenges such as mode collapse and limited diversity in generated musical sequences.

The GAN-VAE hybrid model combines GAN's generative capabilities with VAE's probabilistic modeling power [55]. VAE learns the latent distribution of musical data, encoding music sequences as vector representations in a latent space; GANs then perform generation and discrimination within this latent space, generating more diverse and realistic music sequences through adversarial training.

This integration enables the model to benefit from the structured latent space provided by the VAE, which facilitates smooth interpolation between different musical ideas and supports the generation of novel compositions that retain meaningful musical attributes. Meanwhile, the adversarial component encourages the output to closely mimic the stylistic and structural characteristics found in authentic musical works. As a result, the GAN-VAE hybrid is capable of capturing both the high-level abstractions and the fine-grained details essential for compelling music generation.

Moreover, such hybrid models have demonstrated effectiveness in generating coherent musical passages that maintain logical progression and stylistic consistency. The capacity to explore the latent space allows for intuitive manipulation of musical features, such as mood, tempo, or instrument timbre, thereby broadening the creative possibilities available to composers and researchers. These advancements not only enhance the technical quality of generated music but also open new avenues for interactive and user-guided composition systems.

Transformer-GAN leverages the powerful sequence modeling capabilities of Transformer models to capture long-range dependencies and global structural information within music sequences [56]. The self-attention mechanism in Transformers enables the generator to fully consider the relationships between preceding and following notes during music generation, producing more coherent and logically consistent musical compositions. The discriminator also adopts a Transformer architecture to evaluate and distinguish generated music sequences. Based on the analysis of the aforementioned model techniques, a comparison of GAN models incorporating other technologies is presented in Table 6.

Table 6. Comparison of GAN Models Incorporating Other Technologies

Model Name	Generator Network Structure	Discriminator Network Structure	Features
GAN-VAE Hybrid Model	VAE	Utilizes VAE probabilistic modeling to generate diverse music sequences Learns latent distribution of music data	Relatively complex model structure Training difficulty is high
Transformer-GAN	Transformer	Strong ability to capture long-range dependencies and global structure Generates coherent and logical music compositions	Requires large training data High computation resource consumption
Attention-GAN	Attention Mechanism	Automatically focuses on key parts of music sequences Improves quality and coherence of generated music	High training complexity Requires large labeled datasets
Adversarial Variational Autoencoder	Variational Autoencoder	Combines VAE generation ability with GAN adversarial training Improves quality and diversity of generated music	Training process is complex Needs to balance VAE and GAN loss functions

The integration of GANs with other advanced machine learning technologies has enabled the development of hybrid models that combine the strengths of different architectures. For instance, combining VAEs with GANs allows for efficient latent space exploration while maintaining adversarial training's generative power. Similarly, Transformer-based GANs leverage attention mechanisms to model long-range dependencies, enhancing the coherence and expressiveness of generated music. Although these hybrid approaches present additional

DOI: 10.64972/dea.2025.v3i1.33

challenges in terms of model complexity and training, they offer promising avenues for tackling persistent issues such as diversity, structure, and semantic richness in music composition.

Analysis of GAN Applications in Music Composition

Music Style Imitation and Conversion

The application of GANs in music composition has opened new frontiers for creative expression and stylistic exploration. By leveraging the powerful generative capabilities of adversarial learning, researchers have developed models capable of capturing subtle stylistic nuances and replicating complex musical patterns. Unlike traditional rule-based approaches, GANs learn directly from data, enabling them to internalize a wide array of musical features such as melody, harmony, rhythm, and instrumentation. This data-driven methodology allows for flexible adaptation to various genres, supporting the synthesis of music that reflects both existing conventions and innovative stylistic blends. Moreover, the ability of GANs to generalize from diverse training examples makes them especially suited for tasks that require imitation or transformation of musical styles, positioning them as valuable tools for composers, producers, and music technologists seeking to push the boundaries of algorithmic creativity.

GANs possess unique advantages in music style imitation and conversion [57]. By training the GAN model with diverse musical data, the generator learns the characteristics and patterns of various musical styles, enabling it to produce compositions in specific genres. Reference [58] trained a GAN model to mimic Baroque-era musical styles, generating melodies featuring intricate counterpoint and ornate embellishments; Reference [59] achieved style transfer from classical to pop music, offering novel approaches and inspiration for musical creation.

In addition to these achievements, GAN-driven models have demonstrated the ability to capture and synthesize a wide range of stylistic nuances across different genres and composers. Through adversarial training, generators can internalize the melodic contours, harmonic progressions, and rhythmic patterns that define a particular style, allowing for the creation of music that is not only structurally consistent but also artistically expressive. Researchers have further explored the use of conditional GANs, where additional information such as genre labels or composer identity is provided as input, thus enabling more precise control over the stylistic attributes of the generated output. This approach facilitates the seamless blending of multiple musical influences within a single composition, encouraging creative experimentation and cross-genre fusion. Moreover, subjective evaluations by listeners and expert musicians have consistently highlighted the capacity of GANs to produce music that faithfully reflects the intended stylistic characteristics, underscoring their growing role as valuable tools for both musicological research and practical composition. As GAN models continue to evolve, their applications in style imitation and conversion are expected to broaden the horizons of automated music creation and inspire new forms of artistic innovation. The comparative results of musical style imitation and conversion are shown in Table 7.

Table 7. Comparative Results of Musical Style Imitation and Conversion

Application Scenario	Method	Effect
Music Style Imitation	Style-specific GAN Model	Generated music highly consistent with target style in melody harmony rhythm Accurately imitates specific composers or periods
Music Style Conversion	Conditional GAN CycleGAN	Converted music retains main structure and content of original while integrating target style features Creates novel musical works

The application of GANs in music style imitation and conversion demonstrates their versatility in handling complex musical transformations. Style-specific models can faithfully reproduce distinctive features of targeted genres or composers, while conditional models facilitate seamless transitions between musical styles. These capabilities not only enrich the creative toolkit available to composers and producers but also provide valuable resources for music education and analysis. The ability to generate stylistically consistent or blended compositions underscores the potential of GANs to bridge gaps between tradition and innovation in the musical arts.

To better illustrate the capabilities of GANs in the domain of music style imitation and style transfer, the following figure presents representative results produced by style-specific and conditional GAN models. This visual example showcases how GAN-generated music fragments can faithfully replicate the melodic, harmonic, and rhythmic features of target genres or composers, or seamlessly blend characteristics from different musical styles. By providing concrete output cases, the figure helps to elucidate the practical effectiveness of GAN-based approaches in capturing stylistic nuances and achieving creative transformations between genres. Such visualizations are instrumental in demonstrating the real-world impact of GANs on expanding the horizons of automated music creation and stylistic innovation.

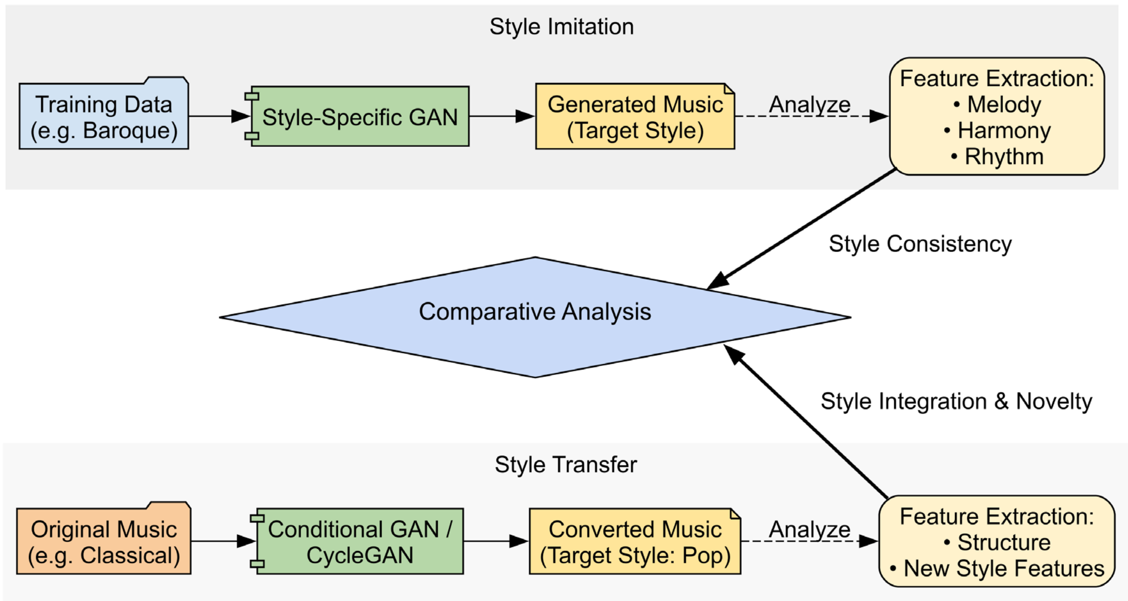

Figure 3. Example results of music style imitation or style transfer using GAN-based models.

The workflow diagram provides a clear overview of the sequential steps involved in the proposed approach. By visualizing the process, it becomes easier to understand the logical connections between each stage. This representation not only clarifies the methodology but also highlights the systematic nature of the research design. Through such visual aids, complex procedures are rendered more accessible, facilitating comprehension and further discussion among readers.

Musical Improvisation

Among the various creative domains influenced by artificial intelligence, musical improvisation occupies a unique position, characterized by spontaneity, interaction, and expressive freedom. Traditional approaches to improvisation often rely heavily on the individual skills, intuition, and experience of musicians, with the creative process unfolding in real time and shaped by both personal style and immediate context. The integration of generative adversarial networks (GANs) into this sphere introduces new dimensions of collaborative creativity, offering algorithmically generated prompts and motifs that can serve as starting points or catalysts during live performance. Unlike static composition tools, GAN-powered systems are capable of dynamically adapting to evolving musical cues, making them particularly well-suited for interactive environments where responsiveness and adaptability are essential. By supplementing human improvisation with algorithmic diversity and surprise, these systems foster an expanded creative dialogue between performer and machine, enriching both the artistic process and the resulting musical output.

In the realm of musical improvisation, GANs can serve as real-time music generation tools, providing inspiration and assistance for musicians during improvisation [60]. The generator can rapidly produce a series of coherent and creative musical fragments based on input initial musical motifs or random noise. Musicians can then

DOI: 10.64972/dea.2025.v3i1.33

improvise and create based on the generated music during performance, broadening the scope and possibilities of improvisation [61]. Furthermore, GANs can synthesize musical fragments blending multiple styles by integrating characteristics from diverse genres, offering musicians novel creative elements [62]. Through real-time interaction, GANs not only respond to musicians' performances but also stimulate their creativity, enriching improvisational expressions and delivering unprecedented experiences and inspiration for musical composition [63]. The analysis of advantages in musical improvisation is summarized in Table 8.

Table 8. Analysis of Advantages in Musical Improvisation

No.	Advantage Aspect	Description
1	Real-time	Can quickly generate music fragments Responds to musician's performance
2	Creativity	Provides novel musical elements and style fusion Stimulates musician's creativity
3	Interactivity	Interacts with musicians in real time Broadens improvisation ideas and possibilities

The strengths of GANs in musical improvisation are evidenced by their ability to generate spontaneous and contextually relevant musical material. This capability supports musicians in exploring new creative directions and overcoming creative blocks. By enabling real-time interaction and adaptive responses, GAN-powered tools foster dynamic collaborations between human performers and AI systems. This synergy enhances the improvisational process, offering fresh perspectives and expanding the expressive possibilities of live performance and composition.

Assisting Music Composition

In the evolving landscape of algorithmic creativity, the intersection of artificial intelligence and music composition has gradually transformed the traditional creative process. Generative Adversarial Networks (GANs), in particular, have emerged as powerful tools capable of augmenting and reshaping how composers approach their work. By leveraging the ability of GANs to learn complex musical structures and generate original content, composers are no longer limited to manual experimentation or conventional rule-based systems. Instead, they can interactively collaborate with intelligent models, exploring a broader spectrum of musical ideas and possibilities. This synergy between human intuition and machine-driven suggestion opens new pathways for creative exploration, enabling artists to break free from habitual patterns and discover unexpected musical directions. Whether working independently or as part of a creative team, music creators can harness GANs not only to accelerate the ideation process but also to infuse their compositions with fresh, diverse, and innovative elements.

For music creators, GANs can also serve as auxiliary tools to aid in music composition [64], with specific application scenarios shown in Table 9. During the compositional process, music creators can utilize GAN models to generate musical elements such as melodies, chord progressions, and rhythmic patterns. They can then select, modify, and combine these elements according to their creative vision and requirements, thereby enhancing both the efficiency and quality of their work [65]. Additionally, GANs can be employed for music arrangement and orchestration, generating suitable accompaniments and tracks for different musical passages and instruments [66-68]. Through these methods, GANs not only stimulate creators' inspiration but also expand possibilities and innovation in music creation. This drives music composition toward greater diversity and personalization, delivering novel musical experiences to audiences [69].

Table 9. Application Scenarios for Assisting Music Composition

No.	Application Scenario	Description
1	Melody Generation	Generate coherent and creative melodies
2	Chord Progression	Provide reasonable chord progression schemes
3	Rhythm Pattern Creation	Generate diverse rhythm patterns
4	Music Arrangement	Generate suitable accompaniment and tracks for different musical sections and instruments

GAN-based tools for assisting music composition have gradually become valuable assets in the creative process. By automating the generation of melodic, harmonic, and rhythmic elements, these tools can significantly reduce the time and effort required for ideation and arrangement. Moreover, the ability to customize generated material according to specific artistic intentions allows composers to maintain creative control while benefiting from AI-driven inspiration. The integration of GANs into composition workflows thus supports both efficiency and innovation, empowering musicians to explore diverse musical landscapes.

DOI: 10.64972/dea.2025.v3i1.33

Challenges and Issues Faced

Over the course of recent advancements in artificial intelligence, the application of Generative Adversarial Networks (GANs) in music composition has garnered widespread scholarly attention. Researchers from diverse backgrounds have explored the integration of GANs into the creative process, seeking to harness their generative capabilities for both traditional and contemporary musical forms. The breadth of work in this field has spanned a variety of compositional tasks—ranging from melodic generation and harmonic progression to style transfer and improvisation support. Through a combination of theoretical exploration, experimental validation, and practical deployment, the community has witnessed a surge in innovative models that demonstrate the artistic potential of GAN-driven music creation. These efforts have not only broadened the understanding of algorithmic composition but have also inspired new interdisciplinary collaborations between computer scientists, musicians, and cognitive researchers. Such progress has laid a solid foundation for the continued evolution of automated music generation, showcasing the transformative power of GANs within the broader landscape of computational creativity.

Despite significant achievements in novel music composition research, GANs still face several challenges and issues, specifically:

(1) Model training stability issues. During training, GANs are prone to pattern collapse and training instability, resulting in poor-quality generated music or failure to converge [70]. This primarily stems from the complex adversarial relationship between the generator and discriminator, as well as the inherent complexity and high-dimensional nature of musical data.

(2) Insufficient emotional expression in generated music. Currently, GAN-generated music often lacks emotional depth, struggling to convey the rich emotional range and nuanced shifts characteristic of human-composed music [71]. This limitation arises because GANs primarily learn patterns and regularities from data, making it difficult for them to comprehend and capture the emotional essence of music.

(3) Difficulties in understanding and modeling musical semantics. Music is an art form rich in semantic and structural elements. However, enabling GANs to deeply comprehend musical semantic information—such as the rules of harmonic progression, the developmental logic of melodies, and the layout of musical forms—and effectively integrate this understanding into the music generation process remains a challenging problem.

(4) Incomplete evaluation criteria. Assessing musical quality involves subjectivity and diversity, and there is currently no comprehensive, objective evaluation framework to accurately measure the quality and artistic value of GAN-generated music. This complicates model training and optimization, limiting GANs' further development in music composition.

In the absence of standardized evaluation protocols, existing studies often rely on a combination of objective metrics and subjective assessments to judge the quality of generated music. Objective indicators may include statistical measures such as pitch class distributions, note density, or rhythmic variance, which provide some insight into structural aspects of the music. However, these metrics frequently fall short in capturing the nuanced dimensions of musicality, such as emotional expressiveness, creativity, and aesthetic coherence. Subjective evaluations, on the other hand, usually involve expert or listener-based judgments, but these are inherently influenced by personal taste, cultural background, and prior musical exposure, leading to inconsistencies and limited reproducibility. The lack of unified benchmarks not only hampers the direct comparison of different models but also makes it challenging to track progress in the field. As a result, the development of robust, multidimensional evaluation frameworks—capable of balancing quantitative analysis with qualitative appreciation—remains a crucial task for advancing the reliability and artistic relevance of GAN-generated music.

Future Research Directions and Outlook

In recent years, the application of Generative Adversarial Networks (GANs) in music composition has attracted growing attention from both academia and industry. While significant strides have been made in generating stylistically coherent and structurally sound musical pieces, several limitations still impede the widespread adoption and effectiveness of GAN-based systems in creative domains. Among these, issues related to model robustness, expressive depth, and evaluation standards remain particularly prominent. As the field continues to evolve, there is a pressing need to address these challenges holistically, ensuring that future advances are not only technically innovative but also musically meaningful.

Addressing the aforementioned issues and challenges, future research directions for novel music composition methods based on GAN networks primarily focus on improving GAN model architectures, integrating multimodal information, enhancing musical semantic understanding, and developing new evaluation metrics [73], as detailed below: 1) Improving GAN Model Architecture. Designing more stable GAN architectures, such as introducing novel regularization techniques and optimizing loss functions, to enhance training stability and the quality of generated music [74]. 2) Integrating multimodal information. Fusing music with information from other modalities enables GANs to generate music with richer emotional expression and semantic richness. Joint modeling of lyrics and music ensures generated compositions better align with the emotional tone and content of the lyrics. 3) Strengthen music semantic understanding. Deepen research into musical syntax rules and semantic structures, developing feature representations and model architectures capable of capturing musical semantic information. This enables GANs to generate works with greater artistic merit and listenability by adhering to music's inherent logic and rules [70]. 4) Develop new evaluation metrics. Establish a comprehensive evaluation framework that considers multiple aspects of music—including melody, harmony, rhythm, and emotion—to more accurately assess the quality and artistic value of GAN-generated music, thereby guiding model training and optimization [74].

In addition to these focal areas, it is crucial to encourage interdisciplinary collaboration among researchers, musicians, and technologists. Such collaborations can bridge gaps between technical innovation and practical musical application, ensuring that advancements in GAN-based music generation are informed by both computational rigor and artistic sensibility. Furthermore, expanding open-source datasets and benchmarking platforms will facilitate more transparent and reproducible research, helping to standardize progress across the field. By systematically addressing these directions, the future of GAN-driven music composition holds the promise of producing music that not only meets technical criteria but also resonates deeply with listeners, reflecting the nuanced interplay of creativity and technology.

Conclusion

As a cutting-edge technology in the field of artificial intelligence, Generative Adversarial Networks (GANs) demonstrate immense potential and broad prospects for application in music composition. Through systematic review research, this paper delves into the theoretical foundations, model development progress, application analysis, and challenges faced by GANs in music composition, while also outlining future research directions. First, the fundamental structure of GANs, methods for representing musical data, and the principles of generating music form the theoretical foundation for their application in music composition. However, GANs still face numerous challenges in this domain. Future research should focus on improving GAN model architectures, integrating multimodal information, enhancing musical semantic understanding, and developing new evaluation metrics to address these issues.

While GAN research in music composition has achieved significant results, it remains in a state of continuous development and refinement. Through future research and innovation, it is anticipated that current challenges will be overcome, propelling further advancements in GAN applications for music composition. This will unlock greater possibilities and creative space for musical creation, offering audiences entirely new musical experiences.

References

[1] Kaliakatsos-Papakostas, M., Floros, A., & Vrahatis, M. N. (2020). Artificial intelligence methods for music generation: a review and future perspectives. Nature-Inspired Computation and Swarm Intelligence: Algorithms, Theory and Applications, 217–245. DOI:10.1016/B978-0-12-819714-1.00024-5

[2] Zhang, B., & Lin, J. (2018). Elements of music based on artificial intelligence. Advances in Intelligent Systems and Computing, 613, 137–144. DOI:10.1007/978-3-319-60744-3_15

[3] Shi, J., & Ning, Q. (2022). Research on the Innovation of Music Teaching in Universities Based on Artificial Intelligence Technology. Security and Communication Networks, 2022(2), 022051 (6pp). DOI:10.1155/2022/3140370

[4] Luo, L. (2022). Practical Exploration on the Construction of Theoretical Courses of Composition Technology in the Age of Artificial Intelligence. Mobile Information Systems, 2022(Pt.25), 3099312.1-3099312.14. DOI:10.1155/2022/3099312

[5] Chen, F., & Meng, H. (2022). The Use of Wireless Network Combined with Artificial Intelligence Technology in the Reform of Music Online Teaching System. Wireless Communications and Mobile Computing, 2022. DOI:10.1155/2022/5957708

[6] Sun, P. (2023). Application of Artificial Intelligence System in Chamber Music Teaching. Lecture Notes on Data Engineering and Communications Technologies, 155, 323–330. DOI:10.1007/978-981-19-9373-2_35

[7] Ryu, E. (2023). A Contrastive Study on Artificial Intelligence's Music Creation and its Usability. The Korean Society of Human and Nature, 4(1), 167–179. DOI:10.54913/hn.2023.4.1.167

[8] Tanberk, S., & Tukel, D. B. (2021). Style-Specific Turkish Pop Music Composition with CNN and LSTM Network. SAMI 2021 - IEEE 19th World Symposium on Applied Machine Intelligence and Informatics, Proceedings, 181–185. DOI:10.1109/SAMI50585.2021.9378654

[9] Yang, Y., Dolly, R. J., Alassafi, M. O., Slowik, A., & Alsaadi, F. E. (2023). MULTI-SOURCE and HETEROGENEOUS ONLINE MUSIC EDUCATION MECHANISM: AN ARTIFICIAL INTELLIGENCE-DRIVEN APPROACH. Fractals, 31(6). DOI:10.1142/S0218348X23401540

[10] Chang, Y. H., & Yao, S. N. (2021). Artificial intelligence on the identification of Beiguan music. Archives of Acoustics, 46(3), 471–478. DOI:10.24425/aoa.2021.138139

[11] Siphocly, N. N. J., El-Horbaty, E. S. M., & Salem, A. B. M. (2021). Top 10 artificial intelligence algorithms in computer music composition. International Journal of Computing and Digital Systems, 10(1), 373–394. DOI:10.12785/ijcds/100138

[12] Liang, J. (2023). Harmonizing minds and machines: survey on transformative power of machine learning in music. Frontiers in Neurorobotics, 17(000), 15. DOI:10.3389/fnbot.2023.1267561

[13] Carnovalini, F., & Rodà, A. (2020). Computational Creativity and Music Generation Systems: An Introduction to the State of the Art. Frontiers in Artificial Intelligence, 3. DOI:10.3389/frai.2020.00014

[14] Liu, Y. (2023). Design of Graphic Design Assistant System Based on Artificial Intelligence. International Journal of Information Technologies and Systems Approach, 16(3). DOI:10.4018/IJITSA.324761

[15] Cádiz, R. F. (2020). Creating Music With Fuzzy Logic. Frontiers in Artificial Intelligence, 3. DOI:10.3389/frai.2020.00059

[16] Tigre Moura, F., & Maw, C. (2021). Artificial intelligence became Beethoven: how do listeners and music professionals perceive artificially composed music? Journal of Consumer Marketing, 38(2), 137–146. DOI:10.1108/JCM-02-2020-3671

[17] Civit, M., Civit-Masot, J., Cuadrado, F., & Escalona, M. J. (2022). A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. Expert Systems with Applications, 209, 118190. DOI:10.1016/j.eswa.2022.118190

[18] Yang, T., & Nazir, S. (2022). A comprehensive overview of AI-enabled music classification and its influence in games. Soft Computing, 26(16), 7679–7693. DOI:10.1007/s00500-022-06734-4

[19] Gioti, A. M. (2021). Artificial Intelligence for Music Composition. Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity, 53–73. DOI:10.1007/978-3-030-72116-9_3

[20] Chen, W. (2022). Design of Music Teaching System Based on Artificial Intelligence. Mathematical Problems in Engineering, 2022. DOI:10.1155/2022/2627395

[21] Mo, Y. (2022). Designing an Automatic Piano Accompaniment System using Artificial Intelligence and Sound Pattern Database. Mobile Information Systems, 2022. DOI:10.1155/2022/7875818

[22] Bergomi, M. G., & Baratè, A. (2020). Homological persistence in time series: an application to music classification. Journal of Mathematics and Music, 204–221. DOI:10.1080/17459737.2020.1786745

[23] Qin, W., & Lin, Q. (2022). Research on the Fusion Model of Professional Vocal Music Performance Voice Care and Artificial Intelligence Technology in Intelligent Medical Treatment. Wireless Communications and Mobile Computing, 2022. DOI:10.1155/2022/2947554

[24] Zhou, D., Liu, L., Tang, T., Huang, Y., Lee, Y., & Hong, J. (2020). Design on intelligence music system in the cultural center based on IoT. Personal and Ubiquitous Computing, 24(3), 319–332. DOI:10.1007/s00779-019-01245-w

[25] Zhang, Z., Li, Z., Pan, J., Chen, W., & Bai, Q. (2022). Artificial Intelligence Development and Music Education System Reform in the Context of 5G Network. Wireless Communications and Mobile Computing, 2022. DOI:10.1155/2022/2384794

[26] Zhang, L. (2022). Music Morphology Interaction under Artificial Intelligence in Wireless Network Environment. Computational Intelligence and Neuroscience, 2022, 9002093. DOI:10.1155/2022/9002093

[27] Wen, Y. W., & Ting, C. K. (2023). Recent Advances of Computational Intelligence Techniques for Composing Music. IEEE Transactions on Emerging Topics in Computational Intelligence, 7(2), 578–597. DOI:10.1109/TETCI.2022.3221126

[28] Yu, X., Ma, N., Zheng, L., Wang, L., & Wang, K. (2023). Developments and Applications of Artificial Intelligence in Music Education. Technologies, 11(2). DOI:10.3390/technologies11020042

[29] Anfilova, S., & Mykhailova, O. (2020). Naples As Seen by Musicians: Intus Et Extra Cultura. Journal of History Culture and Art Research, 9(1), 236. DOI:10.7596/taksad.v9i1.2415

[30] Murray, S., & Ventura, D. (2012). Algorithmically flexible style composition through multi-objective fitness functions. AAAI Workshop - Technical Report, WS-12-16, 55–62. DOI:10.1609/aiide.v8i4.12557

[31] Loui, P. (2021). Neuroscience of Musical Improvisation. Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity, 97–115. DOI:10.1007/978-3-030-72116-9_5

[32] Yu, J., Chen, Z., & Hai, H. (2023). Multi-style Chord Music Generation Based on Artificial Neural Network. Journal of Donghua University (English Edition), 40(4), 428–437. DOI:10.19884/j.1672-5220.202203009

[33] Xu, N., & Zhao, Y. (2021). Online Education and Wireless Network Coordination of Electronic Music Creation and Performance under Artificial Intelligence. Wireless Communications and Mobile Computing, 2021. DOI:10.1155/2021/5999152

[34] Harrison, P. M. C., & Pearce, M. T. (2020). A computational cognitive model for the analysis and generation of voice leadings. Music Perception, 37(3), 208–224. DOI:10.1525/MP.2020.37.3.208

[35] Wiafe, A., & Fränti, P. (2023). Affective algorithmic composition of music: A systematic review. Applied Computing and Intelligence, 3(1), 27–43. DOI:10.3934/aci.2023003

[36] Wang, Y. (2021). Music Composition and Emotion Recognition Using Big Data Technology and Neural Network Algorithm. Computational Intelligence and Neuroscience, 2021, 5398922. DOI:10.1155/2021/5398922

[37] Hou, Y. (2020). Discussion on Practice and Promotion Strategy of College Music Education Informatization Based on Artificial Intelligence. Proceedings - 2020 International Conference on Computers, Information Processing and Advanced Education, CIPAE 2020, 114–116. DOI:10.1109/CIPAE51077.2020.00038

[38] Zheng, H., & Dai, D. (2022). Construction and Optimization of Artificial Intelligence-Assisted Interactive College Music Performance Teaching System. Scientific Programming, 2022. DOI:10.1155/2022/3199860

[39] Su, W., & Kim, H. (2022). Case Analysis and Characteristics of Popular Music Creative Activities Using Artificial Intelligence. The Journal of Humanities and Social Sciences 21, 13(2), 1937–1948. DOI:10.22143/hss21.13.2.136

[40] Sandred, Ö. (2021). Constraint-Solving Systems in Music Creation. Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity, 327–344. DOI:10.1007/978-3-030-72116-9_12

[41] Williams, D., Hodge, V. J., & Wu, C. Y. (2020). On the use of AI for Generation of Functional Music to Improve Mental Health. Frontiers in Artificial Intelligence, 3. DOI:10.3389/frai.2020.497864

[42] Gioti, A. M. (2020). From Artificial to Extended Intelligence in Music Composition. Organised Sound, 25(1), 25–32. DOI:10.1017/S1355771819000438

[43] Singh, M., Sachdev, S., Singh, A., & Raj, D. (2022). Relationship between Different Types of Intelligence and Effect of Music on Intelligence. CHRISMED Journal of Health and Research, 9(3), 172–176. DOI:10.4103/cjhr.cjhr_106_21

[44] Della Ventura, M. (2023). Human-Centred Artificial Intelligence in Sound Perception and Music Composition. Lecture Notes in Networks and Systems, 646 LNNS, 217–229. DOI:10.1007/978-3-031-27440-4_21

[45] Ben-Tal, O., Harris, M. T., & Sturm, B. L. T. (2021). How music AI is useful engagements with composers, performers and audiences. Leonardo, 54(5), 510–516. DOI:10.1162/leon_a_01959

[46] Zhang, Y. (2023). Utilizing Computational Music Analysis and AI for Enhanced Music Composition: Exploring Pre- and Post-Analysis. Journal of Advanced Zoology, 44(S6), 1377–1390. DOI:10.17762/jaz.v44is6.2470
[47] Min, J., Wang, L., Pang, J., Han, H., Li, D., Zhang, M., & Huang, Y. (2022). Application and Research of Monte Carlo Sampling Algorithm in Music Generation. KSII Transactions on Internet and Information Systems, 16(11), 3355–3372. DOI:10.3837/tiis.2022.10.008
[48] Shi, N., & Wang, Y. (2020). Symmetry in computer-aided music composition system with social network analysis and artificial neural network methods. Journal of Ambient Intelligence and Humanized Computing, 1–16. DOI:10.1007/s12652-020-02436-7
[49] Guo, R., Ding, J., & Zang, W. (2021). Music Online Education Reform and Wireless Network Optimization Using Artificial Intelligence Piano Teaching. Wireless Communications and Mobile Computing, 2021. DOI:10.1155/2021/6456734
[50] Palmov, S. V., & Skakun, O. O. (2023). Use of Artificial Intelligence By Music Services. The Applied Economic Researches Journal, 1, 198–203. DOI:10.47576/2949-1908_2023_1_198
[51] Chang, C. Y., & Chen, Y. P. (2020). AntsOMG: A framework aiming to automate creativity and intelligent behavior with a showcase on cantus firmus composition and style development. Electronics (Switzerland), 9(8), 1–19. DOI:10.3390/electronics9081212
[52] Ferreira, P., Limongi, R., & Fávero, L. P. (2023). Generating Music with Data: Application of Deep Learning Models for Symbolic Music Composition. Applied Sciences (Switzerland), 13(7). DOI:10.3390/app13074543
[53] Zhang, S. (2022). Online Music Performance Evaluation System Based on Apriori Algorithm. Proceedings - 2022 International Symposium on Advances in Informatics, Electronics and Education, ISAIEE 2022, 673–676. DOI:10.1109/ISAIEE57420.2022.00142
[54] Helmuth, M., & Schedel, M. (2022). Links between sonification and generative music. Interdisciplinary Science Reviews, 47(2), 215–242. DOI:10.1080/03080188.2022.2035106
[55] Dai, C. (2022). An Intelligent Evaluation Algorithm for the Matching Degree of Music Lyrics Based on LabVIEW Digital Image. ACM International Conference Proceeding Series, 34–38. DOI:10.1145/3561518.3561524
[56] Zhou, W., & Fu, M. (2022). AMRM: Machine Learning Based Intelligent Recommendation of Preference of Music. 2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications, AEECA 2022, 531–535. DOI:10.1109/AEECA55500.2022.9918925
[57] Zheng, Y., Tian, T., & Zhang, A. (2022). Training Strategy of Music Expression in Piano Teaching and Performance by Intelligent Multimedia Technology. International Transactions on Electrical Energy Systems, 2022(000), 14. DOI:10.1155/2022/7266492
[58] Melchiorre, A. B., Penz, D., Ganhör, C., Lesota, O., Fragoso, V., Fritzl, F., Parada-Cabaleiro, E., Schubert, F., & Schedl, M. (2023). Emotion-aware music tower blocks (EmoMTB): an intelligent audiovisual interface for music discovery and recommendation. International Journal of Multimedia Information Retrieval, 12(1). DOI:10.1007/s13735-023-00275-8
[59] Liu, H. (2023). A novel Internet of things-enabled intelligent piano musical identification system using support vector machine. Soft Computing, 27(24), 19125–19142. DOI:10.1007/s00500-023-09322-2
[60] Li, Q., Liu, H., & Zhao, X. (2023). IoT Networks-Aided Perception Vocal Music Singing Learning System and Piano Teaching with Edge Computing. Mobile Information Systems, 2023(000), 9. DOI:10.1155/2023/2074890
[61] Nand, R. (2023). Mood Music Recommendation Using Emotion Detection. International Journal for Research in Applied Science and Engineering Technology, 11(5), 3423–3427. DOI:10.22214/ijraset.2023.52385
[62] Liu, W., & Shapii, A. B. (2022). Study on Aesthetic Teaching Methods in Ethnic Music Teaching in Universities in the Context of Intelligent Internet of Things. Scientific Programming, 2022(000), 1. DOI:10.1155/2022/7901855
[63] He, R. (2022). The Intervention of Music Therapy on Behavioral Training of High-Functioning Autistic Children under Intelligent Health Monitoring. Applied Bionics and Biomechanics, 2022. DOI:10.1155/2022/5766617
[64] Ventura, M. Della. (2023). A Self-learning Musical Tool to Support the Educational Activity. Lecture Notes in Networks and Systems, 543 LNNS, 49–67. DOI:10.1007/978-3-031-16078-3_3
[65] Tang, Z. (2023). Application Model Construction of Emotional Expression and Propagation Path of Deep Learning in National Vocal Music. International Journal of Advanced Computer Science and Applications, 14(11), 1055–1062. DOI:10.14569/IJACSA.2023.01411107

[66] Zhang, J. (2022). Intelligent Layout of Music and Cultural Facilities Based on Heterogeneous Cellular Network. Mobile Information Systems, 2022. DOI:10.1155/2022/2049905

[67] de Souza Jeronimo, B., de Albuquerque Wheler, A. P., de Oliveira, J. P. G., Melo, R., Bastos-Filho, C. J. A., & Kelner, J. (2022). Comparing Social Robot Embodiment for Child Musical Education. Journal of Intelligent and Robotic Systems: Theory and Applications, 105(2), 1–16. DOI:10.1007/s10846-022-01604-5

[68] Guo, Y., & Tang, Y. (2023). The assessment model of online vocal music teaching quality under the optimized DL model. Intelligent Systems with Applications, 20. DOI:10.1016/j.iswa.2023.200276

[69] Su, J. H., Hong, T. P., Chen, Y. T., & Chin, C. Y. (2022). High-performance content-based music retrieval via automated navigation and semantic features. Engineering Applications of Artificial Intelligence, 115. DOI:10.1016/j.engappai.2022.105267

[70] Aiswarya, M. A., Sinith, M. S., & Rajan, R. (2023). Automatic Tonic Pitch Estimation in South Indian Classical Music using Frequency- ratio Method. Proceedings of the 2023 International Conference on Intelligent Systems for Communication, IoT and Security, ICISCoIS 2023, 527–532. DOI:10.1109/ICISCoIS56541.2023.10100503

[71] Luo, C. (2022). Waveform Feature Extraction of Intelligent Singing Skills under the Background of Internet of Things. Mobile Information Systems, 2022. DOI:10.1155/2022/4638801

[72] Seng, W. (2022). Agile Teaching Framework for Music Education Software Development Curriculum Based on Deep Learning. Mobile Information Systems, 2022(000), 8. DOI:10.1155/2022/5243849

[73] Xu, J., Gan, M., & Zhang, X. (2023). MMusic: a hierarchical multi-information fusion method for deep music recommendation. Journal of Intelligent Information Systems, 61(3), 795–818. DOI:10.1007/s10844-023-00786-0

[74] Gan, L., Wang, D., Wang, C., Xiao, D., Zhang, M., Wang, Z., & Li, F. (2023). RETRACTED: Design and implementation of multimedia teaching platform for situational teaching of music appreciation course based on virtual reality (International Journal of Electrical Engineering & Education, (2021)). International Journal of Electrical Engineering and Education, 60(1_suppl), 2597–2612. DOI:10.1177/0020720920986090