IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 7, May 2024

Harmonic Alchemy: Exploring Musical Creation through GANs

Asst. Prof Tabassum Khan, Aditi Sharma, Ayush Parate, Bhuvanesh Takod, Deepasha Khapre, Harsh Shrikundawar, Om Bangar
Artificial Intelligence Engineering
G. H Raisoni Institute of Engineering and Technology, Nagpur, India

Abstract*: The integration of cutting-edge technologies has opened up new avenues for innovation and exploration. Generative Adversarial Networks (GAN), a type of artificial intelligence, has changed the process of generating diverse and novel content. The article looked at how GANs are used in the context of music composition. The purpose of the study is to understand the potential of GANs to facilitate the creation of new and unique musical compositions. The article seeks to shed light on the power of technology in the creative field by exploring the capabilities of GANs in generating music. Despite the excitement surrounding GANs in music, it is important to acknowledge the challenges and limitations that come with their application. As we delve deeper into the realm of music creation through GANs, it is necessary to critically examine the implications, limitations, and ethical considerations that accompany this innovative approach. Our goal is to showcase the possibilities that GANs offer in music creation, but also to reflect on the nuances and complexities involved in using artificial intelligence for artistic endeavors. The field of music generation has been revolutionized by the application of Generative Adversarial Networks (GANs), which have demonstrated the ability to create new and unique compositions based on existing datasets. This review provides a comprehensive overview of the key concepts, approaches and challenges involved in music generation using GANs. First, the basics of GANs are introduced, the contentious process and the roles of generator and discriminator networks are explained. The application of GANs in music generation is then explored, highlighting various methods and architectures such as MuseGAN, Wavenet, and Pix2Pitch that have been developed to address the unique challenges of music generation. The review also discusses the importance of adapting GANs for music generation, allowing music to be generated from various information sources such as images or human sentiment. In addition, it deals with the evaluation of generated music, emphasizing the need for user studies and statistical analysis to validate the results. Finally, the review concludes with a discussion of ethical considerations and the potential impact of automated music generation on the music community. The review aims to provide a valuable resource for researchers and practitioners in the field of music generation using GANs, highlighting the potential of the technology while acknowledging the challenges and ethical implications.*

Keywords: Harmonic Alchemy, cutting-edge technology, musical creation, GANs, Technology, realm, AI, music generation, wave net, music composition, pix2pitch, generating, utilized

I. INTRODUCTION

The application of Generative Adversarial Networks (GANs) in music generation, highlighting their fundamentals, role in the adversarial process, and the development of methods and architectures like MuseGAN, Wavenet, and Pix2Pitch. It provides a comprehensive overview of the key concepts, approaches, and challenges associated with GANs, highlighting the unique challenges faced in music generation. Generative Adversarial Networks (GANs) are a promising field for music generation, enabling the creation of unique compositions based on existing datasets. GANs train two neural networks, a generator and a discriminator, in an adversarial process to generate new musical sounds from various sources, such as raw waveforms, spectrograms, or paintings.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 7, May 2024

1.1 Overview:

Generative Adversarial Networks (GANs) are a promising field for music generation, enabling the creation of unique compositions based on existing datasets. GANs involve training two neural networks, a generator and a discriminator, in an adversarial process to generate new musical sounds. GANs can generate music from various sources of information, such as raw waveforms, spectrograms, or paintings. For example, MuseGAN generates music segments from random noise, which are then mapped to the final score over multiple iterations

1.2 Problem Statement:

GANs present a promising avenue for innovative music composition, yet they face several challenges. These include the need to accommodate diverse musical data formats such as MIDI and waveforms, the acquisition of large-scale datasets while navigating issues of data privacy, copyright, and bias, and the evaluation of generated music in terms of quality, coherence, and creativity. Striking a balance between stylistic fidelity and creative novelty remains a significant challenge, requiring techniques for style transfer, interpolation, and fusion to maintain coherence in generated compositions. Moreover, ethical concerns surrounding authorship, ownership, and cultural representation are becoming increasingly salient, particularly in the context of the democratization of music composition through AI. Addressing these challenges necessitates critical discourse and ethical reflection to ensure responsible and inclusive practices in AI-generated music systems

1.3 Research Objectives:

The main objectives of the research are: -

- Generate Novel and Creative Musical Content
- Utilize GANs to generate original and innovative musical compositions.
- Explore style transfer and adaptation to mimic specific genres or create fusion music
- Support music composition and production by providing inspiration and aiding in musical development
- Enhance Human-Computer Collaboration:
- Facilitate collaboration between humans and AI algorithms in music composition.
- Develop interactive systems where human input influences the generation process..
- Enable real-time exploration and co-creation of music between humans and AI.
- Research Hypotheses and Ethical Considerations:.
- Investigate patch-based discriminators and texture loss for faster model optimization
- Utilize labelled emotion data and incorporate rhythmic data for improved output quality
- Explore ethical implications of automation in creative activities and its impact on the artist's role.
- Develop tools to democratize music production while preserving the artist's creative input

1.4 Scope of the Research

A novel kind of neural network that is superior to existing networks in producing music is the DCG_GAN Model. The model creates more harmonious and pleasing songs by combining the concepts of music theory with a GAN model that restricts the use of chords. The DCG_GAN model has remarkable performance in both subjective and objective evaluations, garnering high ratings from listeners and demonstrating notable improvements in evaluation metrics like the qualified note ratio (QNR) .

GANs, or generative adversarial networks, are a tool used in GANSynth, a method for producing realistic sound. The ability of GANSynth to simultaneously generate entire audio sequences, in contrast to autoregressive models, expedites the audio synthesis process. It makes it possible to independently adjust timbre and pitch, showcasing advancements in audio production.

II. RELATED WORK

The search results reveal a study on harmonic alchemy and musical creation using Generative Adversarial Networks (GANs). GANSynth+ is a method for generating high-fidelity audio using GANs, allowing faster synthesis and

DOI: 10.48175/IJARSCT-18638

independent control over pitch and timbre. It was developed on the NSynth dataset of musical instrument notes. A research paper explores using GANs for music generation, addressing challenges in resource-constrained environments. The results suggest structuring the input space with conditional constraints and using a patch-based discriminator to improve the system's ability to generate music that conforms to musical standards. The paper also proposes a similarity loss to reduce mode collapse and stabilize the training process. GAN applications are discussed, including their use in generating synthetic data, high-quality results, and versatility across various domains like image, text, and audio synthesis. Challenges of GAN training, such as instability, mode collapse, and computational cost, are also discussed. The search results underscore the potential of GANs for music generation and the ongoing research efforts to address these unique challenges.

Music Generation with GANs

Project by teomotun: This project evaluates the performance of different GAN architectures such as FCN-RNN and C-RNN GANs for music generation. It emphasizes the complexity of modeling musical sequences and the nuances of notes, pitches, and tones. The project aims to generate melodies using a continuous representation of music rather than a discrete symbolic representation

Music Generation with Generative Adversarial Networks

Varun Rawal's Work: This study explores advanced deep learning techniques like GANs and VAE-GANs for generating music pieces. The experiments were conducted using Beethoven's music in MIDI format. The study highlights the challenges of generating music due to the lack of global structure and the need for extensive data to prevent overfitting. It also discusses the use of VAE-GAN architectures to improve the quality of generated music

GANSynth: Making Music with GANs

Magenta's GANSynth: Introduced by Google AI's Magenta team, GANSynth generates high-fidelity audio using GANs. Unlike autoregressive models, GANSynth generates entire audio sequences in parallel, significantly speeding up the process. It uses a Progressive GAN architecture to synthesize audio from a single latent vector, allowing for independent control of pitch and timbre

III. METHODOLOGY

The "Harmonic Alchemy: Exploring Musical Creation through GANs" project uses Generative Adversarial Networks (GANs) to generate music. GANs consist of a generator and a discriminator, with the generator creating new data instances and the discriminator evaluating them. Both networks are continuously improved through adversarial training. A GAN consists of two neural networks locked in an adversarial competition

- Generator: This network acts like a composer. It takes a random noise vector as input and transforms it into a piece of music, typically in MIDI format. The goal of the generator is to create music that fools the discriminator into thinking it's real.
- Discriminator: This network acts like a critic. It receives two inputs: real music pieces from the training dataset and music generated by the generator. The discriminator's job is to analyze each piece and determine whether it's authentic or a fake created by the generator.
- Recurrent Neural Networks (RNNs): These networks are adept at handling sequential data like music. They can be integrated with GANs to improve the coherence and structure of the generated pieces.

Deep Convolutional GAN (DCGAN):

Uses convolutional layers in both the generator and discriminator to improve the quality and stability of the generated outputs.

Convolutional layers help the model learn spatial hierarchies, which are crucial for structured data like images or music.

- Cycle GAN Architecture: The project utilized a CycleGAN architecture, which is a type of GAN that can perform bidirectional style transfer between two domains. In this case, the two domains were different music genres.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 7, May 2024

- Conditional GANs (CGANs): These GANs allow for additional control over the generation process. You can provide extra information like genre or mood as input, guiding the generator towards a specific style.
- Transfer in Two Directions: Each GAN focuses on transferring music from one genre to another. GAN_A aims to turn Genre B music into something resembling Genre A, while GAN_B does the opposite.

IV. ARCHITECTURE

In a more detailed introduction, a Generative Adversarial Network (GAN) architecture for music creation involves a sophisticated interplay between two neural networks, the generator, and the discriminator, to produce novel and authentic musical compositions. This process leverages the power of deep learning to learn the intricate patterns and structures present in music data, enabling the generation of new musical pieces that exhibit coherence and creativity.

Detailed Overview:

- Data Representation: Music data, typically in MIDI format, is preprocessed and represented as sequences of notes, chords, and timing information. This data is then transformed into a format suitable for input into the GAN model, such as numerical vectors or matrices.
- Generator Network: The generator network takes random noise vectors as input and learns to map them to meaningful musical sequences. Through multiple layers of neural networks, the generator generates music that progressively improves in quality and complexity.
- Discriminator Network: The discriminator network acts as a critic, distinguishing between real music data from the training set and generated music sequences from the generator. It provides feedback to the generator on how to improve its output, guiding the learning process towards creating more realistic music.
- Training Process: The GAN architecture undergoes a training process where the generator and discriminator networks are trained iteratively. The generator aims to produce music that can fool the discriminator, while the discriminator learns to differentiate between real and generated music effectively.
- Loss Functions: Various loss functions, such as binary cross-entropy or Wasserstein loss, are used to measure the performance of the generator and discriminator during training. These functions guide the networks towards generating high-quality music compositions.
- Hyper parameter Tuning: Fine-tuning hyper parameters like learning rates, batch sizes, and network architectures is crucial for optimizing the performance of the GAN model and achieving desirable results in music generation.
- Evaluation Metrics: Generated music pieces are evaluated based on criteria like melody coherence, harmonic progression, rhythm consistency, and overall musicality. Human evaluation and feedback may also be incorporated to assess the subjective quality of the generated music.

GAN architecture

- Define a generator network: Design a generator network to take random noise vectors as input and generate musical sequences. The generator network should consist of multiple layers of neural networks, such as fully connected layers, convolutional layers, or recurrent layers, depending on the chosen architecture. The output of the generator should conform to the format of pre-processed MIDI data such as piano roll matrices.
- Design Discriminator Network: Create a discriminator network to distinguish between real music from a dataset and generated music from a generator. A discriminative network should also consist of multiple layers of neural networks, similar to a generator network. The discriminator takes either real or generated music as input and outputs a probability indicating whether the input is real or fake.
- Incorporate Conditioning: To control the music generation process and control specific musical attributes, incorporate conditioning information into the GAN architecture. Conditioning can be achieved by providing an additional input to both the generator and the discriminator network, such as genre, key, or time signature. This allows the GAN to generate music that adheres to the desired characteristics.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 7, May 2024

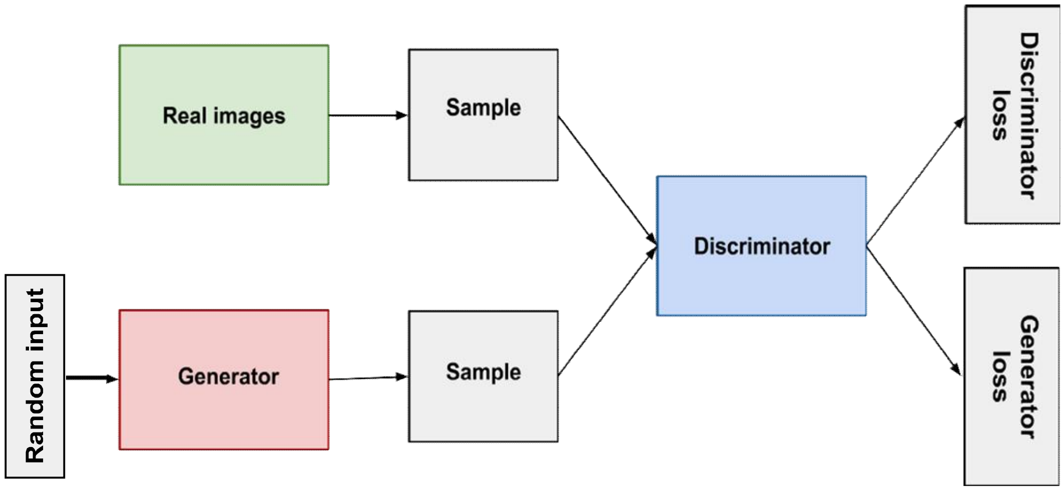

Figure 1: Architecture of GAN Model

V. IMPLEMENTATION

Data preparation

- Collect a MIDI dataset: Collect a comprehensive dataset of MIDI files representing the desired musical style or genre, such as classical, jazz or pop music. The dataset should contain a sufficient number of high-quality MIDI files to ensure that the GAN model effectively learns the underlying patterns and structures
- MIDI data preprocessing: Convert the collected MIDI files into a suitable format for training the GAN model. One common approach is to represent MIDI data as a piano roll matrix, where each row corresponds to a note and each column represents a time step. Piano roll matrices can be binary, indicating the presence or absence of a note, or they can contain additional information such as note velocity or duration.
- Split dataset: Split the pre-processed MIDI dataset into training and validation sets. The training set will be used to train the GAN model, while the validation set will be used to monitor the performance of the model during training and avoid overfitting.

GAN architecture

- Define a generator network: Design a generator network to take random noise vectors as input and generate musical sequences. The generator network should consist of multiple layers of neural networks, such as fully connected layers, convolutional layers, or recurrent layers, depending on the chosen architecture. The output of the generator should conform to the format of pre-processed MIDI data such as piano roll matrices.
- Design Discriminator Network: Create a discriminator network to distinguish between real music from a dataset and generated music from a generator. A discriminative network should also consist of multiple layers of neural networks, similar to a generator network. The discriminator takes either real or generated music as input and outputs a probability indicating whether the input is real or fake.
- Incorporate Conditioning: To control the music generation process and control specific musical attributes, incorporate conditioning information into the GAN architecture. Conditioning can be achieved by providing an additional input to both the generator and the discriminator network, such as genre, key, or time signature. This allows the GAN to generate music that adheres to the desired characteristics.

GAN training

- Training Channel Setup: Create a training channel by defining loss functions and optimization algorithms for both the generator and discriminator networks. Common loss functions used in GAN training include binary cross entropy loss or Wasserstein loss. Choose suitable optimizers such as Adam or RMSProp and set the learning rate for stable and efficient training.

- GAN model training: Train the GAN model using the prepared data set, monitor the training progress and stability. During training, the generator and discriminator networks are iteratively updated, with the generator aiming to produce music that can fool the discriminator, and the discriminator learning to effectively distinguish real music from the generated music.
- Use stabilization techniques: Incorporate techniques to stabilize the training process and mitigate problems such as mode collapse where the generator produces limited diversity in its output. One approach is to use a patch-based discriminator that evaluates the music in smaller fields instead of the entire sequence. Another technique is to incorporate similarity loss, which encourages the generator to produce music that is similar to real music in terms of musical elements such as pitch, rhythm, or harmony.

Generating music

- Music Sequence Generation: Use a trained network of generators to generate new music sequences by providing random noise vectors as input. The generator converts the noise vectors into musical sequences that resemble the real music in the training dataset

  Condition Generation: Control the generated music by adjusting the generator to specific musical attributes such as genre, key, or time signature. This allows you to generate music that adheres to the desired characteristics and fits into a specific musical context.
- Explore the generation space: Experiment with different noise vectors and conditioning to explore the space of possible musical compositions. By varying the input noise and editing, you can create a diverse set of musical tracks that exhibit different styles, emotions, and textures.

Evaluation and refinement

- Assess the generated music: Evaluate the generated music for quality, coherence and adherence to the target musical style. This can be done by analyzing the generated music using musical features such as pitch, rhythm, harmony and structure and comparing it to the real music in the dataset.
- Collect human feedback: Collect feedback from musicians, composers and listeners to further refine the generated music. Human evaluation can provide valuable insights into the subjective quality and creativity of generated music that may not be captured by automated metrics.
- Iterate and Improve: Based on evaluation results, iterate on the GAN architecture, training process, and dataset to improve the quality and variety of music generated. This may involve modifying network architectures, hyperparameters, or training data to better capture the desired musical characteristics.

Development

- Package the model: Package the trained GAN model, including the generator network and all necessary preprocessing and postprocessing steps, for deployment.
- Develop a user interface: Create a user-friendly interface that allows users to interact with the music generation system. This can include functions such as selecting musical attributes, providing input noise vectors, and generating and listening to the resulting music.
- Cloud or on-premises server deployment: Deploy the packaged model and user interface to a cloud platform or on-premises server and ensure the system is available and scalable based on the target user base and usage patterns.
- Monitoring and Updates: Continuously monitor the performance, stability and user feedback of the deployed system. Regularly update the model, user interface, and deployment infrastructure to incorporate improvements, fix issues, and adapt to changing user requirements.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 7, May 2024

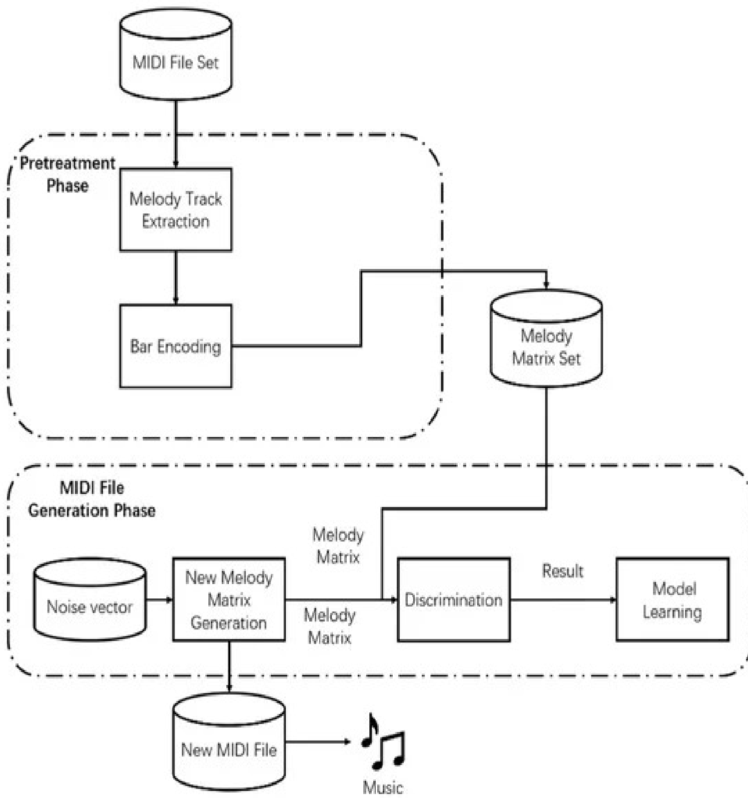

Figure 2: Flow chart of the project

VI. RESULTS

Generating music using Generative Adversarial Networks (GANs) is an emerging field in the intersection of artificial intelligence and music composition. GANs, a type of deep learning model, consist of two neural networks - a generator and a discriminator - which are trained simultaneously in a competitive manner. The generator creates new instances of data, in this case, music, while the discriminator evaluates these instances for authenticity. Through this adversarial training process, GANs can learn to produce music with characteristics similar to those in the training data.

Research in music generation using GANs has shown promising results in various aspects:

- Composition Style Imitation: GANs can learn to mimic the style of a particular composer or genre by training on a dataset of existing compositions. By capturing the statistical patterns and structures of the music, GANs are capable of generating new pieces that resemble the input style.
- Creative Exploration: GANs offer a platform for creative exploration in music composition. By adjusting input parameters or introducing randomness into the generation process, users can influence the output and generate novel musical ideas.
- Interactive Music Generation: Some research explores interactive interfaces where users can provide real-time feedback to guide the music generation process. This interactive framework allows for collaboration between the user and the GAN model, enabling personalized and responsive music generation.
- Multi-modal Generation: GANs can be extended to generate not only audio but also other modalities such as music scores or accompanying visuals. This multi-modal approach enriches the creative possibilities and opens up new avenues for artistic expression.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 7, May 2024

VII. OUTPUTS

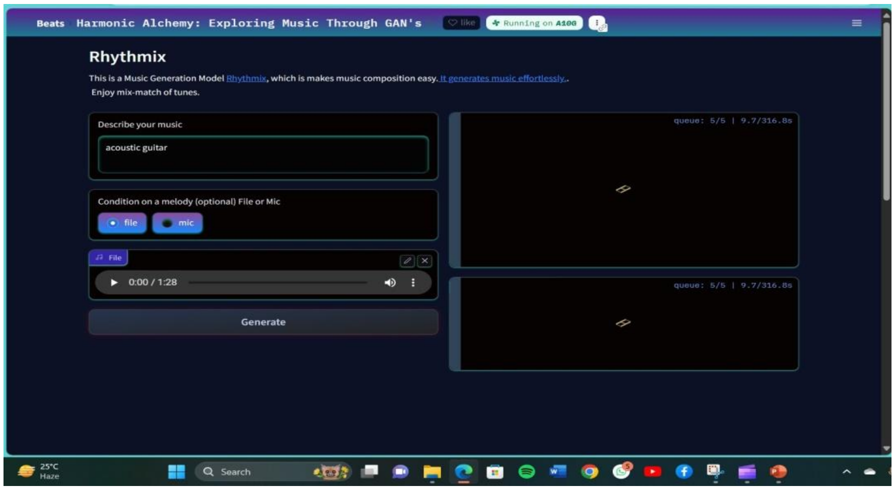

Figure 3: Output of the Project Home Screen

Generator: The generator network uses random noise as input to generate musical compositions. It transforms this noise into meaningful musical sequences through neural network layers. During training, the generator learns to mimic statistical patterns in the training data, such as melodic motifs and chord progressions. It adjusts its parameters through backpropagation and gradient descent optimization to minimize differences between the generated output and real music samples. Once trained, the generator generates new musical data in MIDI or audio formats for evaluation.

Discriminator: The discriminator network in the GAN framework is a critic that evaluates the authenticity of musical compositions. It distinguishes between real and fake samples, providing feedback to the generator to adjust its output. Both networks are trained simultaneously in an adversarial manner, enhancing their performance. The discriminator network consists of layers of neural network units that extract relevant features from the input musical data. Its optimization objective is to optimize a loss function, such as binary cross-entropy loss, which quantifies its ability to correctly classify real and fake sample formatting and data types.

Adversarial Training: In a GAN setup, both the generator and discriminator networks are trained simultaneously, with the generator learning to produce convincing compositions and the discriminator learning to distinguish between real and fake music samples. This adversarial dynamic creates a feedback loop, with the generator aiming to produce output that fools the discriminator into classifying it as real music, and the discriminator becoming more adept at detecting inconsistencies. The goal is to reach a point of convergence, where the generated music is high quality and realism

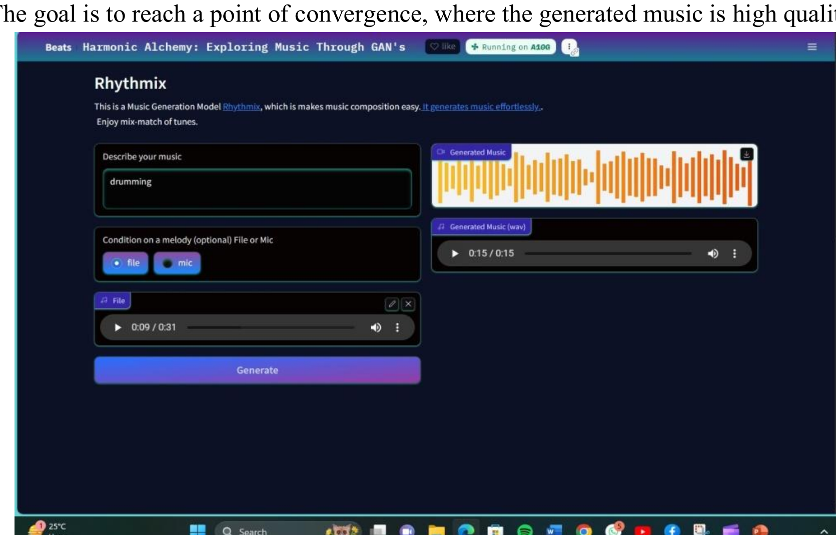

Figure 4: Output of the Project Generated Screen

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 7, May 2024

A Generalized Adversarial Network (GAN) is trained to find optimal parameters for both the generator and discriminator networks by minimizing a predefined loss function, often called binary cross-entropy loss. This loss function measures the difference between the discriminator's predictions and the ground truth labels assigned to the samples. The discriminator trains by classifying real samples as real or fake, while the generator trains by generating fake samples from random noise inputs. The training process alternates between updating the parameters of the generator and discriminator networks, creating a feedback loop where both networks continually improve their performance. The goal is to reach a state of equilibrium where the generator generates realistic music samples, and the discriminator cannot reliably differentiate between real and fake samples.

VIII. LIMITATIONS

While Generative Adversarial Networks (GANs) have shown promise in generating music, especially in capturing patterns and styles, they also have several limitations:

- Lack of Long-term Structure: GANs often struggle to generate music with coherent long-term structure. They can produce short musical phrases or loops effectively, but creating pieces with extended musical development and coherence remains challenging.
- Limited Understanding of Music Theory: GANs operate based on patterns in data but lack understanding of music theory concepts such as harmony, melody, rhythm, and form. As a result, the generated music may lack musicality or adherence to established compositional principles.
- Mode Collapse: GANs are prone to mode collapse, where they generate repetitive or similar outputs, failing to explore the full diversity of the training data. This can lead to the generation of generic or uninteresting musical sequences
- Training Data Dependency: The quality of generated music heavily depends on the quality and diversity of the training data. If the training dataset is limited or biased, the GAN may struggle to produce varied and original compositions.
- Limited Control and Interpretability: GAN-generated music often lacks user control over specific musical attributes such as mood, tempo, or instrumentation. Additionally, understanding and modifying the inner workings of the GAN to achieve desired musical outputs can be challenging due to its complex architecture.
- Computational Resources: Training GANs for music generation requires significant computational resources, including high-performance GPUs and large amounts of memory. This can limit accessibility for individuals or organizations with limited computational resources.
- Evaluation Metrics: Assessing the quality of generated music remains subjective and challenging. While there are metrics like FID (Fréchet Inception Distance) or Inception Score used in image generation, similar metrics for music generation are still under development.
- Semantic Incoherence: GANs may produce music that lacks semantic coherence or meaningful structure. This means that even though the music might sound pleasant, it may not convey any deeper emotional or narrative content.

IX. CONCLUSION

Generative Adversarial Networks (GANs) have revolutionized the landscape of music composition, ushering in an era of unprecedented creative exploration and innovation. Early successes such as MIDI-Net demonstrated the remarkable ability of GANs to capture the intricate temporal structures of music, laying a solid foundation for generating coherent and compelling compositions. These early endeavors provided a glimpse into the vast potential of GANs in pushing the boundaries of musical creativity.

Subsequent research has further expanded the capabilities of GAN-based music generation by exploring new modalities and techniques. For example, the introduction of lyric-conditioned music generation models by Huang et al. has enabled the synthesis of music that is not only melodically rich but also thematically coherent, opening up new avenues for emotional expression and storytelling through music. Additionally, the development of style transfer algorithms, as exemplified by Choi et al.'s Music Style Transfer Network, has empowered musicians to experiment with blending different musical styles and genres, fostering a culture of innovation and genre fusion.

Recent advancements in GAN-based music generation, exemplified by models like WaveGAN, have significantly enhanced the realism and fidelity of generated music. By directly synthesizing raw audio waveforms, these models have overcome previous limitations and produced music that is indistinguishable from human- composed pieces in terms of audio quality and authenticity. This has paved the way for a new era of AI-generated music that is not only diverse and expressive but also remarkably lifelike

Despite these advancements, several challenges persist in the field of GAN- based music generation. One of the primary challenges is the acquisition of diverse and representative datasets that adequately capture the breadth and depth of musical styles and genres. Additionally, evaluating the quality and coherence of generated music remains a complex and subjective task, requiring the development of robust evaluation methodologies and metrics. Moreover, ethical considerations such as data privacy, copyright, and cultural representation must be carefully navigated to ensure responsible and inclusive practices in AI-generated music systems.

X. ACKNOWLEDGMENT

We would like to extend our heartfelt gratitude to Asst. Prof. Tabassum Khan, Associate Professor at the Artificial Intelligence department of G. H Raisoni Institute of Engineering and Technology, Nagpur, India. Her invaluable insights, guidance, and unwavering encouragement played a pivotal role in the successful completion of this research. Her expertise and dedication significantly influenced the direction and outcomes of our study, and we are deeply grateful for her mentorship and guidance throughout this research endeavor.

REFERENCES

[1]. Muhamed, A., Li, L., Shi, X., Yaddanapudi, S., Chi, W., Jackson, D., Suresh, R., Lipton, Z. C., & Smola, A. J. (2021, May 18). Symbolic Music Generation with Transformer-GANs. Proceedings of the . . . AAAI Conference on Artificial Intelligence.

[2]. J, S. G. F. (n.d.). Pix2Pitch: generating music from paintings by using conditionals GANs | Archivo Digital UPM.

[3]. Hernandez-Olivan, C., & Beltrán, J. R. (2022, September 23). Music Composition with Deep Learning: A Review. Signals and Communication Technology.

[4]. Mukherjee, S., & Mulimani, M. (2022, April 1). ComposeInStyle: Music composition with and without Style Transfer. Expert Systems With Applications.

[5]. Jin, C., Tie, Y., Bai, Y., Lv, X., & Liu, S. (2020, June 9). A Style-Specific Music Composition Neural Network. Neural Processing Letters

[6]. Generating Music Algorithm with Deep Convolutional Generative Adversarial Networks. (2019, May 1). IEEE Conference Publication | IEEE Xplore.

[7]. Hilmkil, A., Thomé, C., & Arpteg, A. (2020, June 11). Perceiving Music Quality with GANs. arXiv.org.

[8]. Gong, C., Liu, Y., Zhong, S. H., & Zhang, X. (2018, October 15). Musicality-Novelty Generative Adversarial Nets for Algorithmic Composition. https://doi.org/10.1145/3240508.3240604

[9]. Liu, W. (2022, November 8). Literature survey of multi-track music generation model based on generative confrontation network in intelligent composition. the Journal of Supercomputing/Journal of Supercomputing.

[10]. Generative Adversarial Networks (GANs) for Creative Applications: Exploring Art and Music Generation. (2023, October 5). https://ijmirm.com/index.php/ijmirm/article/view/43

[11]. Music Understanding LLaMA: Advancing Text-to-Music Generation with Question Answering and Captioning. (2024, April 14). IEEE Conference Publication | IEEE Xplore .https://ieeexplore.ieee.org/abstract/document/10447027

[12]. Huang, Q., Park, D. S., Wang, T., Denk, T., Ly, A., Chen, N., Zhang, Z., Zhang, Z., Yu, J., Frank, C., Engel, J., Le, Q., Chan, W., Chen, Z., & Han, W. (2023, February 8). Noise2Music: Text-conditioned Music Generation with Diffusion Models. arXiv.org

[13]. Dash, A., & Agres, K. R. (2023, January 10). AI-Based Affective Music Generation Systems: A Review of Methods, and Challenges. arXiv.org.

[14]. Prabhakar, S., & Lee, S. (2023, January 1). Holistic Approaches to Music Genre Classification using Efficient Transfer and Deep Learning Techniques. Expert Systems With Applications.

[15]. MusicLDM: Enhancing Novelty in text-to-music Generation Using Beat-Synchronous mixup Strategies. (2024, April 14). IEEE Conference Publication | IEEE Xplore.

[16]. Schneider, F., Kamal, O., Jin, Z., & Schölkopf, B. (2023, January 27). Mo\^usai: Text-to-Music Generation with Long-Context Latent Diffusion.

[17]. Garcia, H. F., Seetharaman, P., Kumar, R., & Pardo, B. (2023, July 10). VampNet: Music Generation via Masked Acoustic Token Modeling. arXiv.org. https://arxiv.org/abs/2307.04686

[18]. Wang, L., Zhao, Z., Liu, H., Pang, J., Qin, Y., & Wu, Q. (2024, February 19). A review of intelligent music generation systems. Neural Computing & Applications.

[19]. Kang, J., Poria, S., & Herremans, D. (2024, March 1). Video2Music: Suitable music generation from videos using an Affective Multimodal Transformer model. Expert Systems With Applications. https://doi.org/10.1016/j.eswa.2024.123640

[20]. Ainur: Harmonizing Speed and Quality in Deep Music Generation Through Lyrics-Audio Embeddings. (2024, April 14). IEEE Conference Publication | IEEE Xplore.

[21]. Li, S., Dong, W., Zhang, Y., Tang, F., Ma, C., Deussen, O., Lee, T. Y., & Xu, C. (2024, January 31). Dance-to-Music Generation with Encoder-based Textual Inversion of Diffusion Models.