

## *Retraction*

# **Retracted: Research on Chord-Constrained Two-Track Music Generation Based on Improved GAN Networks**

### **Scientific Programming**

Received 1 August 2023; Accepted 1 August 2023; Published 2 August 2023

Copyright © 2023 Scientific Programming. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] X. Li and Y. Niu, "Research on Chord-Constrained Two-Track Music Generation Based on Improved GAN Networks," *Scientific Programming*, vol. 2022, Article ID 5882468, 7 pages, 2022.

## Research Article

# Research on Chord-Constrained Two-Track Music Generation Based on Improved GAN Networks

**Xinru Li and Yizhen Niu** 

*Xinyang Vocational and Technical College, Xinyang 464000, Henan, China*

Correspondence should be addressed to Yizhen Niu; [nyz@xyvtc.edu.cn](mailto:nyz@xyvtc.edu.cn)

Received 6 December 2021; Revised 19 February 2022; Accepted 22 February 2022; Published 16 March 2022

Academic Editor: Le Sun

Copyright © 2022 Xinru Li and Yizhen Niu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chords have a role in music for emotional expression, and the generated melodies have more richness through the constraining effect of chords. In this paper, based on a GAN network music generation model based on chord features, a GRU network is used in chord feature extraction in order to autonomously learn chords at  $1:t-1$  moments and generate chords at  $t$  moments, by saving the hidden layer state of each batch and constructing a layer of GRU combined with a generator, thus achieving the effect of automatically learning the overall style of chords. The performance of the four models is gradually optimized by weighted averaging, and the melodic pleasantness generated by all four models has a significant positive correlation with musical coherence and creativity.

## 1. Introduction

Music is an important form of artistic expression in our lives, closely related to us and an integral part of our lives. It can be used to soothe the mood and express emotions, and different tunes and melodies can express different emotions, resulting in countless classics. This is why there are a wide variety of styles and genres of music, which can be divided into pop, classical and pop music (Bach, Beethoven, Mozart), jazz, R&B, rock and roll, and vocal and instrumental music [1]. In recent decades, with the rise of the Internet industry and the significant improvement in the performance of computer hardware, the cost of generating a series of digital audio files regarding audio production, audio editing, audio recording, etc. has significantly reduced the threshold for nonprofessionals to enter the field of music, causing the number of music creations to grow exponentially, and the music industry has become an important area of the cultural market, with the series of Jitterbug, Racer, etc. With the emergence of streaming media and the maturation of the Internet, the competition for music platforms has gradually changed from competition for copyright and content to competition for audio diversity and multimodal content combined with

graphics and video. Users' demand for this has also become more urgent [2, 3].

Artificial intelligence (AI) composition is a technique for generating digital music using algorithms, neural networks, and other executions [4]. AI is a technology that enables computers to learn and simulate human thought processes and behavioral capabilities through training. Stochastic algorithms, such as Gaussian noise, are often used as input in AI compositions, which undergo a finite number of state transitions and conditional constraints to generate the final sequence of notes [5]. Composition is the process by which a music creator creates music through a series of theoretical systems such as polyphony, harmonics, orchestration, etc. It is a process of expressing creative ideas [6]. Composing logarithmically is a new way of composing music. Compared with the traditional mode of music composition, it can combine human creativity, emotional expression, aesthetics, and other intelligent operations with the computing power of computers, human-computer interaction systems, automation control, and other technologies to break through the professional technical constraints of human composition and create more novel musical effects, while saving human costs and improving the efficiency of music composition.

This makes it easier for nonprofessional musicians to enter the hall of music creation and enjoy the joy of creation, while professional musicians are assisted by computer-generated musical works in different forms and styles to develop new innovative ideas and muses for composers.

Most current music generation neural network models use recurrent neural networks (RNNs) and their variants. Neural networks used for music generation usually use all the music information from previous events as a condition for generating the current music, and the generated music information will also be too repetitive, which greatly reduces the interest of the music [7]. When using only GAN to generate music, they are prone to unstable training, gradient disappearance, and pattern running, and do not take into account the time dependence, which can reduce the authenticity of the generated music. The chord feature-based GAN network music generation model (DCC\_GAN) and the overall style-based GAN network music generation model (DCG\_GAN) generate music in which the generator CNN and the moderator CNN are trained together and generate music melodies under the constraints of the chord CNN module, and the generated results are fed to the discriminator CNN, which submits the feedback to the generator [8, 9].

Music as a kind of auditory art, brings not only auditory feelings, but also can directly hit people's hearts and express their emotions, among which pop music is a style form that can fully express human emotions through popular melodies and words that can be combined with different cultural backgrounds in different countries and regions around the world to form a very different style of pop music, and with development of deep learning, China has made some achievements in natural language processing, speech recognition, image processing, etc., but the research in the field of music generation is still at an elementary level, and there is a lot of room for development [10]. In the traditional composition field, composers need to have solid musical skills and musicianship, and it takes a long time to create excellent works, which is relatively difficult for people who love music but are not strong in music. Using deep neural networks to generate music can provide a vast creative platform for people who love music and bring a vast market and economic benefits, and the future is immeasurable [11, 12].

## 2. Convolutional Adversarial Generative Network-Based Model

A convolutional GAN for symbolic-domain music generation was proposed in 2017 by Richard Yang, which is based on the principle of applying a convolutional GAN model to the music generation domain to form a convolutional GAN for symbolic-domain music generation (MidiNet). adversarial network for symbolic-domain music generation [13]. The model is fed with a pre-processed music melody dataset, trained by a generator CNN and a moderator CNN, and the generated results are fed to a discriminator CNN, which feeds the output of the discriminator CNN to the generator CNN so that the

whole model forms a game process and finally outputs a better music melody.

*2.1. Datasets.* The input to the music generation model based on convolutional adversarial generative networks is a collection of popular music melodies in nope format reprocessed to a melody bar count of 50,496 (789 MB), a chord bar count of 50,496 (5.01 MB), a memory size of 5.01 MB, a dimension of 13, and a piano roll format with 16 note units, a pitch range of C4-B5, and random noise of Gaussian white noise of length 100 [14].

*2.2. Model Structure.* The model used in this paper is based on GAN for optimization, which has opened up a new era of neural networks since Ian Goodfellow proposed GAN in 2014 [15].

An artificial neural network (ANN), referred to as a neural network (NN), is a mathematical model that mimics the behavioral characteristics of biological neural networks and processes data to achieve human artificial intelligence [16]. Machine learning techniques for human artificial intelligence [17]. A neural network is shown in Figure 1 as a typical three-layer neural network framework, including an input layer, a hidden layer, an activation layer, an output layer, and a normalization process for the output.

The neural network graph has three neurons in the input layer and four neurons in the hidden layer. An activation function is added after the hidden layer to add a nonlinear factor to the results of the matrix operations, mapping the features to a high-dimensional nonlinear interval for interpretation. The output layer has two neurons, and the output of the output layer is normalized so that the data are restricted to a certain range, thus eliminating the undesirable effects caused by odd sample data [18].

The internal structure of the neural network: this structure is shown in Figure 2 as a processing unit of the neural network,  $x_i$  is the input from the  $i$ -th neuron;  $w$  is the connection weight of the  $i$ -th neuron, equivalent to the eigenvalue. The absolute value of the weight represents the size of the influence of the input signal on the neuron,  $\theta_j$  is the bias, also known as the threshold, after the activation function to obtain the output results, the output results are shown in equation (1), [19].

$$y = f\left(\sum_{i=1}^n w_i x_i - \theta\right). \quad (1)$$

*2.3. Generative Adversarial Networks.* The GAN is primarily trained as a generator and a discriminator neural network, where the two networks are played to obtain the better result of the two networks. A high-performance discriminator is used for identification [20]. The input music, which may be generated by the generator, is identified by the discriminator, and if it is real music, the identification result is true, and if it is generated music, the identification result is false. The result of the identification by the discriminator gives

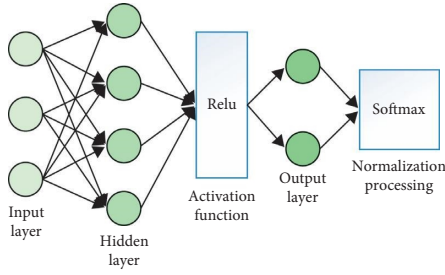


FIGURE 1: Neural network diagram.

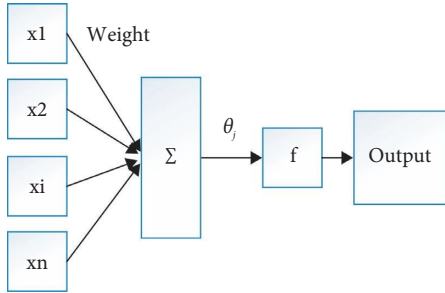


FIGURE 2: Diagram of the internal structure of the neural network.

feedback to the generator to improve its performance in generating music, and the generator also gives feedback to the discriminator to improve its performance in generating music [21]. The initial stage of the GAN network (as shown in Figure 3) is mainly used in image generation, in which the two networks play a game, each trying to beat the other to achieve its own performance improvement. The ultimate goal is to use the generator network to generate music melodies that can be faked.

#### 2.4. Music Generation Models Based on Convolutional GAN.

The GAN is a game between a generator neural network and a discriminator neural network in which the two networks are trained to give the best result of the two networks, and the identification of real and generated music is described [22]. The two networks are eventually better; the generator network is trained so that the generated music is highly similar to the real music and the discriminator network is highly discriminative. The initial stage of the GAN network was mainly used for image generation, where the generator and discriminator networks were used to generate the real music. Each network tries to beat the other to improve the performance of its own network. The ultimate goal is to use the generator network to generate music melodies that can be falsely described as real [23].

In the MidiNet model, it consists of a moderator CNN, a generator CNN, and a discriminator CNN. In the moderator CNN, the input is a two-dimensional starter bar, which is convoluted into four layers, with each layer outputting a corresponding starter bar to be combined with the generator CNN; in the generator CNN, the input is a one-dimensional chord and random noise, which is also conserved in four

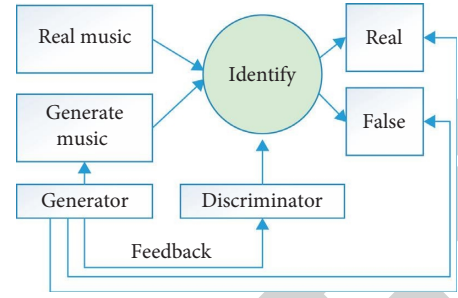


FIGURE 3: GAN network structure.

layers, with each layer combining with the starter bar generated by the moderator to generate a new melody [24]; in the discriminator CNN, the input is either a real melody or a generated In the discriminator CNN, the input is either the real melody or the generated melody, and the start bars and chords are added through two layers of convolution and one layer of full concatenation, resulting in a discriminated output.

**2.5. Model Objectives.** The total objective formula is as in equation (2). The discriminator CNN is equation (3), and the generator CNN is equation (4). Where  $x \sim \text{data}(x)$  denotes sampling from real data,  $z \sim p_z(z)$  denotes sampling from a random distribution,  $D$  denotes the discriminator network, and  $G$  denotes the generator network. In the discriminator network equation (3), the goal is to identify whether the input is a real melody or a generated melody, and the generation process is shown in Figure 4. If the data comes from real data, the discriminator probability is the maximum, and the purpose of doing log conversion is similar to log-likelihood, which does not affect the monotonicity of the function, but makes the operation more simple [25]; if the data comes from a Gaussian noise distribution, the input of the discriminator is the result generated by the generator, then the probability of the discriminator network will fall and  $1 - D(G(z))$  will rise, and then take the log conversion so that the probability of equation (3) takes the maximum value. For the generator network equation (4), the goal is to generate melodies that can fool over the discriminator network, the generation process is illustrated in Figure 5, the data  $x$  comes from the generated data i.e., the result generated by the Gaussian noise  $z$ , then the probability of  $D(G(z))$  goes up and the probability of  $\log(1 - D(G(z)))$  goes down, and finally the minimum value of the generator network is obtained.

$$\min \max_G V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log(D(x))] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))], \quad (2)$$

$$\max_D \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))], \quad (3)$$

$$\min_G \mathbb{E}_{x \sim p(x)} [\log(1 - D(G(z)))]. \quad (4)$$

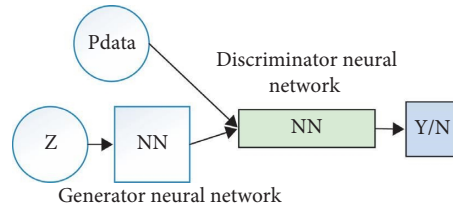


FIGURE 4: Discriminator network diagram.

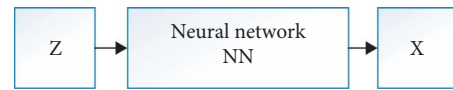


FIGURE 5: Generator network diagram.

### 3. Experimental Results and Analysis

**3.1. Experimental Results.** This paper uses the music theory rule-based music generation model, the DCC\_GAN model, and the DCG\_GAN model to generate a large number of musical melodies after training with chord constraints and time dependence. The model with chord constraints and time-dependent training to generate a large number of musical melodies. The melodies generated are more coherent, pleasant, and innovative than those generated by the baseline model. Here, the DCG\_GAN model, for example, generates melodies in the nyc format as shown in Figure 6, and performs different rounds (1 epoch, 100 epoch, and 200 epoch). (100 epoch, 200 epoch) iterations, all selecting the first two bars of the first phrase of each round for observation, and as the number of training rounds increased, the notes became more varied and the resulting melodies more diverse [26].

The generated music in midi format is displayed as a piano roll in the MidiEditor software by selecting the first four bars of each melody as shown in Figure 7.

The experimental results of the baseline model are shown in Figure 7, which shows that the generation process tends to level off in both the chord and melody sections.

**3.2. Assessment and Analysis.** There are currently no scientifically rigorous and objective evaluation criteria for music melody generation, and the main evaluation method is based on the subjective evaluation of the user. The evaluation perspective is based on the coherence, ear-friendliness, and interest of the generated music melodies, with the baseline model's generated music melodies as a control group, and the music theory rule-based music generation model, the chord feature-based GAN network music generation model, and the overall style-based GAN music generation model. The GAN network music generation models are based on chord features and the GAN network music generation models are based on overall style. The four groups of models were trained for 200 rounds, and the generated music files were processed through a python library. The generated music files were converted from the nyc format to midi format by the Python library piano roll,

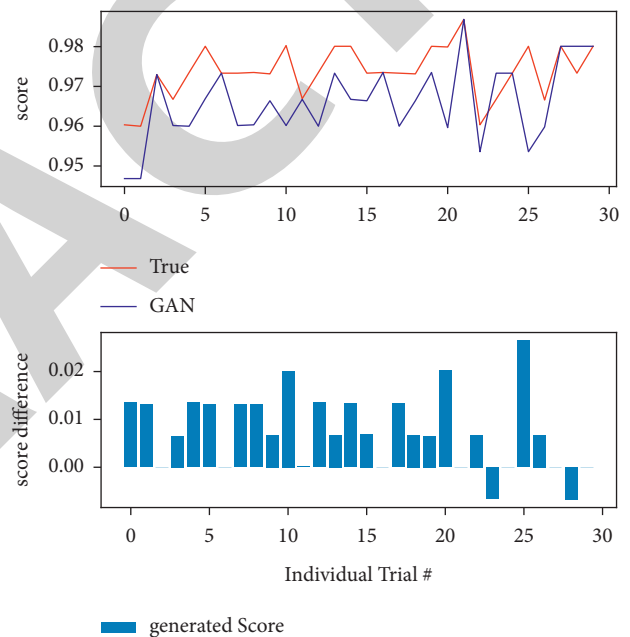


FIGURE 6: DCG\_GAN model generation results (1 epoch, 100 epoch, and 200 epoch).

and then the melodies in midi format were converted to MP3 format by MIDI 3 Pro software, and finally, the generated music melodies were evaluated and analyzed [27].

A total of 50 people was evaluated, 40 of whom were general listeners and 10 of whom were music professionals (music-related learners or instrumentalists), and the three groups of results were evaluated in terms of their melodic coherence, fearfulness, and creativity. The results were evaluated on a five-point scale, with 1 being the least effective, 5 being the most effective, and so on, and a weighted average of 50 scores produced the following results are shown in Table 1, [28].

The evaluation results were analyzed using a weighted average method in the two-track music generation based on chord-constrained GAN networks. For the four groups of models, the results were calculated as 40% for 40 general listeners and 60% for 10 music professionals, as in equation

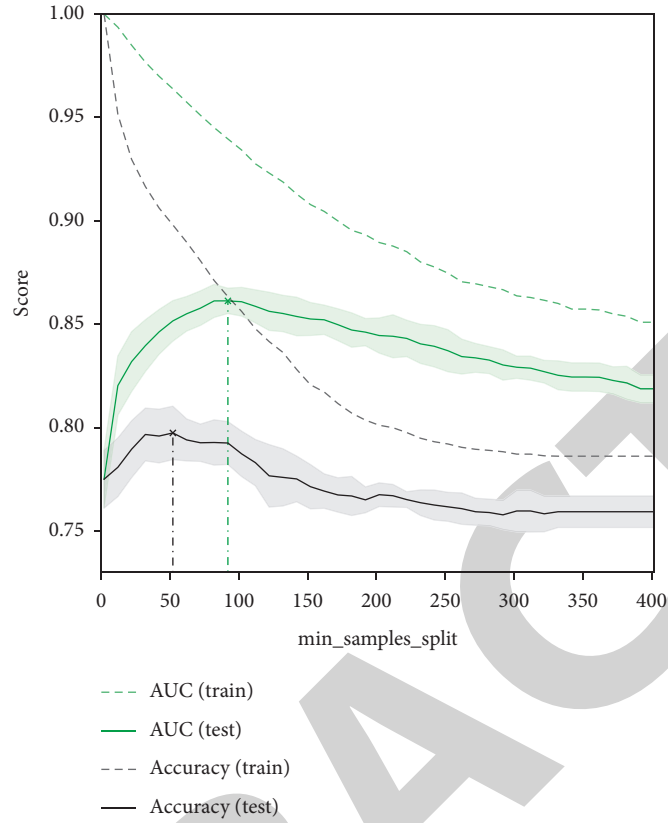


FIGURE 7: Baseline model experimental results.

TABLE 1: Subjective evaluation of experimental results of MidiNet, Rakuhari-based, DCC\_GAN, and DCG\_GAN models.

Model/Performance	Ordinary audience			Music professionals		
	Continuity	Pleasant ear	Innovation	Continuity	Pleasant ear	Innovation
MidiNet	3.0	3.1	3.2	2.2	2.5	2.1
Based on music theory	3.1	3.1	3.1	2.6	2.6	3.2
DCC_GAN	3.3	3.5	3.5	3.0	3.0	2.6
DCG_GAN	3.8	3.8	3.6	3.6	3.8	3.2

TABLE 2: Weighted average results of musical melodies generated by the four groups of models.

Model	Average audience score	Professional musician score	Total score
MidiNet	3.07	2.27	2.59
Based on music theory	3.10	2.72	2.87
DCC_GAN	3.40	2.92	3.11
DCG_GAN	3.76	3.58	3.65

(5). The weighting of the three performance evaluation metrics of coherence, earliness, and innovation was analyzed as 5:3:2, resulting in the performance analysis of the four groups of models in Table 2. The performance of the models gradually improves, and the melodies generated are more realistic and pleasing to the ear [29, 30].

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \cdots + x_k f_k}{\sum_1^k f_i} \quad (5)$$

In music, the core of the generated results is the fearfulness of the generated music. Therefore, the fearfulness of the melodies generated by the four groups of models was

TABLE 3: Analysis of the correlation between melodic pleasantness and coherence and creativity of the music generated by the four groups of models (Pearson correlation).

Model	Correlation coefficient between agreeableness and coherence	Correlation coefficient between pleasant ear and innovation
MidiNet	0.674	0.432
Based on music theory	0.698	0.385
DCC_GAN	0.571	0.475
DCG_GAN	0.788	0.514

correlated with the coherence and creativity, respectively, according to the Pearson correlation coefficient as in equation (6). The results of the analysis are shown in Table 3 [31].

$$r = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{N \sum x_i^2 - (\sum x_i)^2} \sqrt{N \sum y_i^2 - (\sum y_i)^2}} \quad (6)$$

#### 4. Conclusions

This paper introduces the baseline model in this experiment, the music generation model based on convolutional adversarial generative network, which is divided into two subsections: the first subsection introduces the dataset used in the model training process, including the data et format, data type, total amount of data, and data units; the second subsection introduces the model structure of the baseline model, including GAN. By introducing the baseline model, the baseline model can be better optimized for subsequent work.

#### Data Availability

The raw data supporting the conclusions of this article will be made available by the authors without undue reservation.

#### Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

#### References

- [1] C. H. Cao, Y. N. Tang, and D. Y. Huang, "IIBE: An improved identity-based encryption algorithm for WSN security," *Security and Communication Networks*, vol. 2021, 2021.
- [2] D. Wu, C. Zhang, L. Ji, R. Ran, H. Wu, and Y. Xu, "Forest fire recognition based on feature extraction from multi-view images," *Traitement du Signal*, vol. 38, no. 3, pp. 775–783, 2021.
- [3] L. Wang, C. Zhang, Q. Chen et al., "A communication strategy of proactive nodes based on loop theorem in wireless sensor networks," in *Proceedings of the 2018 Ninth International Conference on Intelligent Control and Information Processing (ICICIP)*, pp. 160–167, IEEE, Wanzhou, China, 2018.
- [4] H. Li, D. Zeng, and L. Chen, "Immune multipath reliable transmission with fault tolerance in wireless sensor networks," in *Proceedings of the International Conference on Bio-Inspired Computing: Theories and Applications*, pp. 513–517, Springer, Singapore, 2016.
- [5] H. Purwins, B. Li, T. Virtanen, J. Schluter, S.-Y. Chang, and T. Sainath, "Deep learning for audio signal processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 2, pp. 206–219, 2019.
- [6] S. Oore, I. Simon, S. Dieleman, D. Eck, and K. Simonyan, "This time with feeling: Learning expressive musical performance," *Neural Computing & Applications*, vol. 32, no. 4, pp. 955–967, 2020.
- [7] V. S. Parekh and M. A. Jacobs, "Deep learning and radiomics in precision medicine," *Expert review of precision medicine and drug development*, vol. 4, no. 2, pp. 59–72, 2019.
- [8] G. Brunner, Y. Wang, R. Wattenhofer, and S. Zhao, "Symbolic music genre transfer with cyclegan," in *Proceedings of the 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 786–793, IEEE, Volos, Greece, Nov. 2018.
- [9] B. L. Sturm, O. Ben-Tal, U. Monaghan et al., "Machine learning research that matters for music creation: A case study," *Journal of New Music Research*, vol. 48, no. 1, pp. 36–55, 2019.
- [10] L.-C. Yang and A. Lerch, "On the evaluation of generative models in music," *Neural Computing & Applications*, vol. 32, no. 9, pp. 4773–4784, 2020.
- [11] X. Zhao, S. Wang, and L. Sun, "Single and sequential sampling plans for multi-attribute products and multi-class lot in reliability test," *IEEE Access*, vol. 7, pp. 81145–81155, 2019.
- [12] J. Wu, X. Liu, X. Hu, and J. Zhu, "PopMNet: Generating structured pop music melodies using neural networks," *Artificial Intelligence*, vol. 286, p. 103303, 2020.
- [13] K. Tatar and P. Pasquier, "Musical agents: A typology and state of the art towards musical metacreation," *Journal of New Music Research*, vol. 48, no. 1, pp. 56–105, 2019.
- [14] C. Gan, D. Huang, P. Chen, J. B. Tenenbaum, and A. Torralba, "Foley music: Learning to generate music from videos, Computer Vision - ECCV 2020," in *Proceedings of the Computer Vision-ECCV 2020: 16th European Conference*, pp. 758–775, Springer International Publishing, Glasgow, UK, August 2020.
- [15] J. Zhang and D. Tao, "Empowering things with intelligence: A survey of the progress, challenges, and opportunities in artificial intelligence of things," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 7789–7817, 2020.
- [16] I. P. Yamshchikov and A. Tikhonov, "Music generation with variational recurrent autoencoder supported by history," *SN Applied Sciences*, vol. 2, no. 12, pp. 1–7, 2020.
- [17] W. T. Lu and L. Su, *Transferring the Style of Homophonic Music Using Recurrent Neural Networks and Autoregressive Model*, pp. 740–746, ISMIR, 2018.
- [18] D. Makris, M. Kaliakatsos-Papakostas, I. Karydis, and K. L. Keramnidis, "Conditional neural sequence learners for

- generating drums' rhythms," *Neural Computing & Applications*, vol. 31, no. 6, pp. 1793–1804, 2019.
- [19] F. Carnovalini and A. Rodà, "Computational creativity and music generation systems: An introduction to the state of the art," *Frontiers in Artificial Intelligence*, vol. 3, p. 14, 2020.
- [20] Y. Zhang, Q. Ai, H. Wang, Z. Li, and X. Zhou, "Energy theft detection in an edge data center using threshold-based abnormality detector," *International Journal of Electrical Power & Energy Systems*, vol. 121, p. 106162, 2020.
- [21] R. Kapoor, D. Sharma, and T. Gulati, "State of the art content based image retrieval techniques using deep learning: a survey," *Multimedia Tools and Applications*, vol. 80, no. 19, pp. 29561–29583, 2021.
- [22] Z. Ren, J. Li, X. Xue et al., "Reconstructing seen image from brain activity by visually-guided cognitive representation and adversarial learning," *NeuroImage*, vol. 228, p. 117602, 2021.
- [23] A. Yuniarti, A. Z. Arifin, and N. Suciati, "A 3D template-based point generation network for 3D reconstruction from single images," *Applied Soft Computing*, vol. 111, p. 107749, 2021.
- [24] M. Huzaifah and L. Wyse, "Deep generative models for musical audio synthesis," in *Proceedings of the Handbook of Artificial Intelligence for Music*, pp. 639–678, Springer, Cham, 2021.
- [25] H. Yang, X. Lu, S. H. Wang et al., "Synthesizing multi-contrast MR images via novel 3D conditional Variational auto-encoding GAN," *Mobile Networks and Applications*, vol. 26, no. 1, pp. 415–424, 2021.
- [26] J. Ahn, H. N. Loc, R. K. Balan, Y. Lee, and J. Ko, "Finding small-bowel lesions: challenges in endoscopy-image-based learning systems," *Computer*, vol. 51, no. 5, pp. 68–76, 2018.
- [27] C. Jin, Y. Tie, Y. Bai, X. Lv, and S. Liu, "A style-specific music composition neural network," *Neural Processing Letters*, vol. 52, no. 3, pp. 1893–1912, 2020.
- [28] I. C. Wei, C. W. Wu, and L. Su, *Generating Structured Drum Pattern Using Variational Autoencoder and Self-similarity Matrix*, pp. 847–854, ISMIR, 2019.
- [29] S. Latifi and N. Torres-Reyes, "Audio enhancement and synthesis using generative adversarial networks: A survey," *International Journal of Computer Application*, vol. 182, no. 35, p. 27, 2019.
- [30] J. Luo, X. Yang, S. Ji, and J. Li, "MG-VAE: Deep Chinese folk songs generation with specific regional styles," in *Proceedings of the 7th Conference on Sound and Music Technology (CSMT)*, pp. 93–106, Springer, Singapore, 2020.
- [31] Y.-C. Yeh, W.-Y. Hsiao, S. Fukayama et al., "Automatic melody harmonization with triad chords: A comparative study," *Journal of New Music Research*, vol. 50, no. 1, pp. 37–51, 2021.