

Research Article

Features, Models, and Applications of Deep Learning in Music Composition

Yanjun Chen*

College of Music and Dance, Division of Arts, Shenzhen University, Shenzhen, China

Abstract

Due to the swift advancement of artificial intelligence and deep learning technologies, computers are assuming an increasingly prominent role in the realm of music composition, thereby fueling innovations in techniques for music generation. Deep learning models such as RNNs, LSTMs, Transformers, and diffusion models have demonstrated outstanding performance in the music generation process, effectively handling temporal relationships, long-term dependencies, and complex structural issues in music. Transformers, with their self-attention mechanism, excel at capturing long-term dependencies and generating intricate melodies, while diffusion models exhibit significant advantages in audio quality, producing higher-fidelity and more natural audio. Despite these breakthroughs in generation quality and performance, challenges remain in areas such as efficiency, originality, and structural coherence. This research undertakes a comprehensive examination of the utilization of diverse and prevalent deep learning frameworks in music generation, emphasizing their respective advantages and constraints in managing temporal correlations, prolonged dependencies, and intricate structures. It aims to provide insights to address current challenges in efficiency and control capabilities. Additionally, the research explores the potential applications of these technologies in fields such as music education, therapy, and entertainment, offering theoretical and practical guidance for future music creation and applications. Furthermore, this study highlights the importance of addressing the limitations of current models, such as the computational intensity of Transformers and the slow generation speed of diffusion models, to pave the way for more efficient and creative music generation systems. Future work may focus on combining the strengths of different models to overcome these challenges and to foster greater originality and diversity in AI-generated music. By doing so, we aim to push the boundaries of what is possible in music creation, leveraging the power of AI to inspire new forms of artistic expression and enhance the creative process for musicians and composers alike.

Keywords

Artificial Intelligence, Deep Learning, Music Generation, Transformers

1. Introduction

With the rapid advancement of artificial intelligence and deep learning technologies, music generation is no longer solely reliant on the inspiration and skills of human creators—artificial intelligence has also begun to participate in the

music composition process. The application of deep learning in music generation has evolved progressively from simple to complex models [1]. Initially, methods such as Markov chains and rule-based models were primarily used to generate simple

*Corresponding author: chenyanjun2021@email.szu.edu.cn (Yanjun Chen)

Received: 20 April 2025; **Accepted:** 12 June 2025; **Published:** 15 July 2025



Copyright: © The Author(s), 2025. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

melodic sequences. While these methods achieved some success in their early stages, they exhibited clear limitations when dealing with complex musical structures. As deep learning technologies advance, the sophistication and capability of models have continually increased. For example, architectures such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks [2], possess the capability to produce musical compositions that are more intricate and expressive in nature. These models effectively capture the temporal relationships in music, thereby enhancing the coherence and expressiveness of the generated compositions. Building on this foundation, Transformer models, with their powerful self-attention mechanisms, have further advanced music generation technology. Unlike traditional RNNs and LSTMs, Transformers leverage self-attention to simultaneously consider all positions in the input sequence, avoiding the limitations of generating long-span sequences [3]. This capability makes Transformers particularly effective in capturing long-range dependencies in music, enabling the generation of more coherent and intricate melodies [4]. Furthermore, in recent years, diffusion models have also demonstrated significant applications in music generation. By employing multi-step reverse inference to re-cover clear audio signals from noise, diffusion models are capable of producing high-quality, stable audio and excel in capturing fine details in timbre and tonal variations [5].

This paper aims to explore the application of neural networks and pre-trained models in music feature extraction, with a particular focus on overcoming the limitations of traditional composition methods in generating complex musical structures and creating high-quality works. By analyzing various advanced models, such as Transformers and diffusion models, the study seeks to reveal how these models play a unique role in handling long-term dependencies in music and capturing multi-level structural features. One of the motivations of this research is to provide a comprehensive framework to better understand how different technologies can be applied to music generation and processing, especially in practical applications across education, therapy, and entertainment.

2. Theoretical Foundations of Deep Learning in Music Composition

2.1. Early Deep Learning Models

RNN is a neural network specifically designed to process sequential data, particularly well-suited for tasks with temporal dependencies [6]. It captures the dependencies between previous and subsequent time steps in the input sequence through recurrent connections, making it ideal for generating music content with temporal requirements, such as melodies and harmonies. LSTM is a specialized version of RNN, which addresses the vanishing gradient problem of traditional RNNs

by introducing a gating mechanism, enabling the network to better capture long-term dependencies [7]. LSTM effectively retains and updates long-term dependency information, enhancing the model's ability to generate complex sequences.

2.2. Transformer Model

Transformer architecture revolutionizes the framework of conventional RNNs and LSTMs by incorporating the self-attention mechanism. This mechanism facilitates interactions among all input positions within the sequence, thereby addressing the challenges faced by RNNs and LSTMs in capturing long-range dependencies. Unlike RNNs and LSTMs, Transformers do not rely on sequential processing; instead, they process inputs in parallel, significantly improving training efficiency and generation quality. Transformer is better at capturing long-range dependencies, and it demonstrates higher coherence and expressiveness, especially in generating complex melodies and multi-layered musical structures [4].

2.3. Diffusion Models

Diffusion models are a type of generative model, where the basic idea is to gradually recover the process from noise to clear signals through reverse inference during generation [8]. Through multiple steps of reverse diffusion, the model is able to generate high-quality audio signals from noise, and the generation process is more stable. Diffusion models exhibit a high level of detail when generating audio, effectively simulating changes in timbre and audio quality, making them especially suitable for generating high-quality and diverse audio content.

3. Feature Extraction

3.1. Using Neural Networks to Extract Music Features

Neural networks are widely used in music feature extraction. For example, the Piano Roll is a two-dimensional matrix representation that captures melody and chord structures. Its ability to directly map to the input and output layers of neural networks makes it highly suitable for deep learning models, facilitating the processing and generation of complex musical structures (Chen et al., 2024) [9]. VAEs generate latent space through an encoder-decoder architecture, and this latent space can represent high-level features of music. For instance, the Music VAE model uses VAE to generate multi-track music and creates new musical segments through interpolation in the latent space [10].

3.2. Using Pre-trained Models to Extract Music Features

Pre-trained models play a crucial role in music feature extraction, especially when handling large-scale music data. The Transformer model, initially used in natural language processing, also performs exceptionally well in music generation due to its attention mechanism. For example, the Music Transformer model utilizes a pre-trained Transformer to generate music with long-term structure [11]. The global relative attention mechanism calculates the relative distance between any two positions in the sequence and uses this distance information to adjust attention weights. This allows the model to capture relationships between different parts of the music, such as the repetition of melodies and the interplay of rhythms. The local relative attention mechanism divides the music sequence into smaller blocks and models the relative distances between elements within each block and between elements of the previous block. This approach reduces com-

putational complexity, enabling the model to process long-sequence music data more efficiently while still being able to recognize local musical features, such as rhythm and harmony within a measure.

Transfer learning allows fine-tuning on pre-trained models to adapt them to specific music styles or tasks. For example, the GPT-2 model, pre-trained on large-scale text data, is used to generate multi-track music [28]. The “Figure 1” illustrates the music feature extraction methods, where music is broken down into features such as notes (NOTE_ON and NOTE_OFF), time intervals (TIME_DELTA), instruments (INST), and note density (DENSITY) using MultiTrack and BarFill representations, with structural markers (BAR_START, TRACK_START) defining the organizational structure of the music. Additionally, the BarFill representation, through FILL_START and FILL_END markers, allows for precise filling and editing of music measures, enabling fine-grained control in the music generation process.

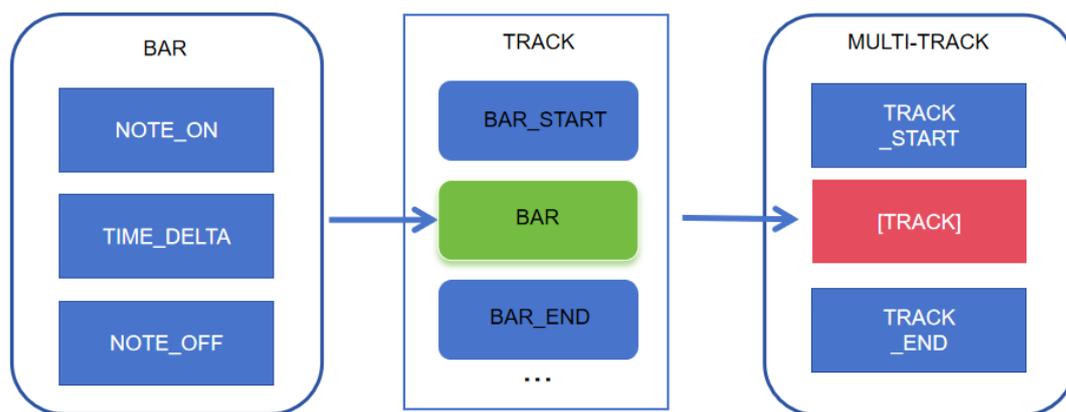


Figure 1. Overview of Multi-Track and Bar-Fill Representations in Music Generation Models [12].

4. Models

4.1. Model Based on Traditional Deep Learning

Methods such as CNN, RNN, LSTM, and GAN are widely used in the field of music generation. Oord et al. used one-dimensional convolutional layers to capture local patterns and temporal dependencies in audio signals, and by stacking multiple convolutional layers, they learned more complex feature representations, enabling the generation of high-quality audio waveforms [13]. RNNs are capable of handling sequential data, capturing temporal dependencies and structures in music, making them suitable for generating coherent music sequences. LSTM, an improvement of RNN, introduces structures like input gates, output gates, and forget gates, effectively addressing the issues of vanishing and ex-

ploding gradients, and capturing long-term dependencies in music. The use of LSTM to generate Bach-style music demonstrates its powerful capability in music generation [14]. Through adversarial training involving a generator and a discriminator, the GAN model produces realistic and high-quality musical content. The generator is tasked with synthesizing music data, whereas the discriminator serves to differentiate between synthetic and authentic musical pieces. DCGAN excels in generating high-fidelity audio, producing realistic audio waveforms [15]. WaveGAN [16] focuses on generating monophonic music and is capable of producing high-quality instrumental sounds.

4.2. Transformer-based Model

The Transformer efficiently handles sequential data using the self-attention mechanism, particularly excelling at capturing long-range dependencies and complex structures in

music composition [17]. “Figure 2” illustrates the working process of the MusicLM [18] model during training and inference. Part (1) describes the training phase, where MuLan audio labels, semantic labels, and acoustic labels are extracted from an audio dataset and predicted through a two-stage sequence-to-sequence task. Part (2) showcases the inference phase, where MuLan text labels, computed from a text prompt, are used as conditional signals. The model then generates audio labels, which are subsequently converted

into waveforms using the SoundStream decoder, resulting in music generation. The entire process utilizes multiple models, including MuLan [19], w2v-BERT [20], SoundStream [21], and the Transformer structure with a de-coder-only architecture. This approach enables MusicLM to generate music that not only considers the acoustic characteristics of the audio but also the semantic information described by the text, producing music that is highly consistent with the textual description.

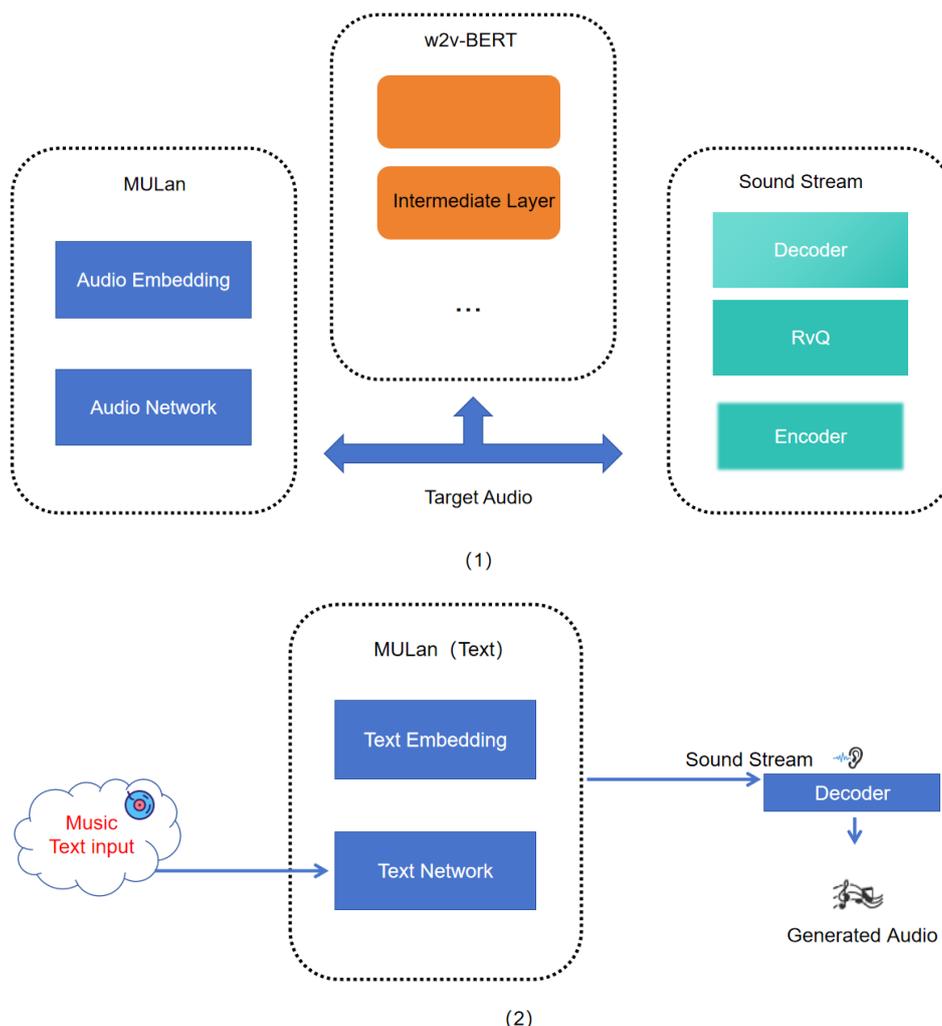


Figure 2. Workflow from Audio Feature Extraction to Text-Conditioned Music Generation [18].

4.3. Diffusion-based Model

By progressively diminishing noise, the diffusion model produces audio content of high quality, rendering it appropriate for the generation of high-fidelity music. The “Figure 3” illustrates the framework of the Moûsai model [22], which is a two-stage cascaded diffusion model for generating music

based on text descriptions. During the initial stage, a Diffusion-based Music Autoencoder (DMAE) is employed to condense the musical data, whereas in the subsequent stage, music is synthesized from the condensed representation, directed by the encoded textual description. The entire process includes text encoding, a diffusion generator, and a diffusion decoder, ultimately producing music that matches the text description.

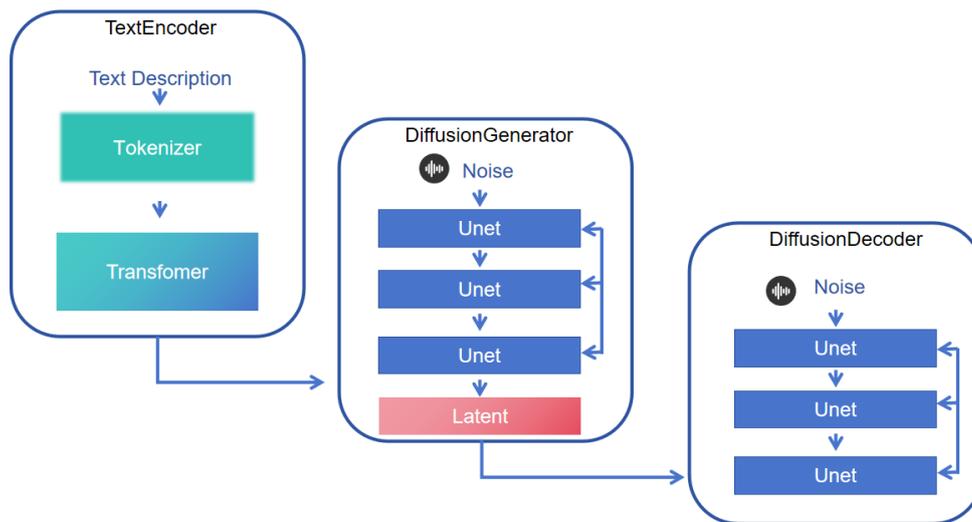


Figure 3. A text-based music generation and compression method.

Table 1. Comparison Table of Different Music Generation Models.

Base Model	Advantages	Limitations	Typical Application Examples
CNN [13]	Efficient at capturing local features, suitable for audio waveform generation	Struggles with long-range dependencies	High-fidelity audio waveform generation
RNN	Good at handling time series data, suitable for generating coherent music sequences	Vanishing/exploding gradients, long-term de-pendency issues	Coherent music sequence generation
LSTM	Effectively captures long-term dependencies, suitable for generating complex music structures	High computational complexity, but more stable than RNN	Bach Bot generating Bach-style music [14]
GAN	Generates high-quality, realistic audio content	Unstable training, may lack coherence	DCGAN [15] and Wave-GAN [16] generating high-quality audio waveforms
Transformer	Excels at capturing long-range dependencies and complex structures, can generate music based on text	Complex training process, high computational cost	MusicLM [18] generating music consistent with text descriptions
Diffusion Model	Generates high-quality audio, suitable for high-fidelity music generation	High computational cost, slower generation speed	Mo ũsai [22] generating music based on text descriptions

4.4. Comparison Between Models

As music generation technology continues to evolve, the advantages and challenges of various models have gradually emerged. Selecting an appropriate model depends not only on the specific requirements of the generation task but also on factors such as computational resources and training stability. The “Table 1” summarizes the characteristics, advantages, and disadvantages of several common music generation models. These models have their own strengths and challenges in different music generation tasks, and understanding their characteristics helps in selecting the most

suitable model for specific needs. As shown in the table, CNN is good at extracting local features from audio, making

it suitable for generating high-fidelity audio waveforms, but it has weaker capabilities in handling long-range dependencies. Both RNN and LSTM can process sequential data, with LSTM performing better at capturing long-term dependencies, making it suitable for generating coherent music sequences. GAN generates high-quality audio through adversarial training, but the training process may be unstable, and the generated music can sometimes lack coherence. Transformer excels at handling complex music structures and long-range dependencies and can generate music based on text descriptions, though it has high computational costs.

Diffusion models generate high-fidelity audio through step-by-step denoising, making them suitable for tasks requiring high-quality audio, but they have slower generation speeds and higher computational resource consumption.

5. Application

5.1. Music Education and Learning

As a teaching tool, music generation models can automatically create simple exercises to help students better understand music structure and harmony [23]. In this way, students can experience the application of theoretical knowledge through practical visual and auditory experiences, thus improving their understanding and mastery of music theory. Additionally, AI-generated music models can provide inspiration and guidance for music learners, assisting them in learning composition techniques. For example, they can generate melodies or harmonies for students to study and imitate, which not only helps them gain inspiration during the creative process but also accelerates the development of their compositional skills [24]. Therefore, AI-generated music technology has broad application potential in music education, offering students richer learning resources and creative support.

5.2. Music Therapy and Rehabilitation

AI-generated music has important applications in emotional regulation and neurorehabilitation. It can create music that helps alleviate stress and anxiety, widely used in music therapy to assist patients in regulating their emotions [25]. It can also generate customized music based on the patient's specific condition, supporting neurorehabilitation and helping restore brain function [26].

5.3. Entertainment and Gaming

Music generation in interactive entertainment applications such as videos and games greatly enhances the user experience. For example, TikTok uses music generation technology to create matching background music for video content [27], making the atmosphere of each video more fitting and enhancing the viewer's immersion. Additionally, in interactive entertainment, music generated based on the user's real-time input provides a richer interactive experience. For instance, in games, the background music dynamically adjusts according to the player's actions and emotional state [28], allowing the music to synchronize with the game context and the player's psychological state, thus enhancing the player's gaming experience and emotional resonance.

5.4. Artistic Creation and Performance

In music performance, real-time generated music can not

only provide accompaniment, enhancing the interactivity and creativity of the performance [29], but also offer performers greater freedom and space for innovation. At the same time, artists have begun using generated musical material for experimental music creation [24], exploring different forms and styles of music.

6. Limitations & Future Outlooks

Currently, the application of deep learning in music creation still faces several significant limitations. First, modeling the long-term dependencies of music is challenging, particularly in capturing aspects such as melodic coherence and harmonic stability. Although existing models have made progress in long-term sequence modeling, they still fail to fully meet the structural and logical requirements of music generation [30]. For example, while the Transformer has advantages in sequence modeling, it suffers from high computational costs, insufficient local structure modeling, and weak ability to capture dynamic changes over time, resulting in generated music lacking a sense of coherence [18]. Although Diffusion models excel in generating high-quality content, their generation process is slow, resource-intensive, and difficult to control precisely, presenting dual challenges in efficiency and control capabilities in practical applications [22]. Additionally, deep learning models also exhibit shortcomings in creativity and originality in creation, often relying on patterns found in existing data, which limits their potential to break through traditional music styles and creative paradigms.

In the future, as technology evolves, we can expect significant improvements in the use of deep learning models in music composition. One possible direction is to combine the advantages of Transformer and Diffusion models to construct a more efficient and creative hybrid generation model. The Transformer's powerful sequence modeling ability can complement the Diffusion model's advantages in generation quality, thereby addressing the current issues of efficiency, resource consumption, and creative control in music generation. At the same time, as models improve their understanding of user intentions, future music creation systems may place greater emphasis on users' personalized needs and creativity, enabling users to better engage in the creative process, breaking the limitations of traditional creation modes, and fostering diversity and innovation in music creation. Moreover, with continuous technological advancement, deep learning models may be able to generate music that showcases more originality and uniqueness in creative style, breaking traditional creative paradigms.

7. Conclusion

With the continuous advancement of artificial intelligence and deep learning, the application of deep learning in music

creation has gradually broken traditional boundaries, driving innovation and transformation in music generation technology. This paper explores in depth the application of deep learning models (such as RNN, LSTM, Transformer, and Diffusion models) in music generation, analyzing how these models address issues of temporal relationships, long-range dependencies, and complex structural problems in music.

Among them, the Transformer model stands out for its powerful self-attention mechanism, particularly in capturing long-range dependencies and generating complex melodies. It can handle information flow over long time spans, ensuring that the generated music is more coherent and layered. On the other hand, Diffusion models have a significant advantage in the quality of audio generation, producing higher-quality and more natural audio output. By comparing the characteristics of different models, we can clearly observe their respective strengths and weaknesses in terms of generation effects, efficiency, and stability.

Although existing generative models have made significant progress, they still face numerous challenges in terms of generation efficiency, originality, and the integrity of musical structure. Future research may improve generation quality and control capabilities by combining the advantages of Transformer and Diffusion models, thereby supporting more personalized and diverse creative needs. Overall, this paper not only provides a comprehensive theoretical framework for music generation technology but also offers important academic value and practical significance for advancing innovation in fields such as music creation, education, and therapy.

Abbreviations

AI	Artificial Intelligence
DL	Deep Learning
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
VAE	Variational Autoencoder

Acknowledgments

This research was supported by the Undergraduate Training Programs for Innovation and Entrepreneurship of Shenzhen University (Project Number: S202410590118). The authors sincerely thank these organizations for their generous support and encouragement, which were invaluable in the completion of this study.

Author Contributions

Yanjun Chen is the sole author. The author read and approved the final manuscript.

Funding

This work is supported by the Undergraduate Training Programs for Innovation and Entrepreneurship of Shenzhen University (Project Number: S202410590118).

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Zhu, Y., Baca, J., Rekabdar, B., & Rawassizadeh, R. (2023). A Survey of AI Music Generation Tools and Models. arXiv preprint <https://doi.org/10.48550/arXiv.2308.12982>
- [2] Zhao, J., Huang, F., Lv, J., Duan, Y., Qin, Z., Li, G., & Tian, G. (2020, November). Do RNN and LSTM have long memory. In International Conference on Machine Learning (pp. 11365-11375). PMLR.
- [3] Vaswani, A. (2017). Attention is all you need. Advances in Neural Information Processing Systems. <https://doi.org/10.48550/arXiv.1706.03762>
- [4] Vaswani, A. (2017). Attention is all you need. Advances in Neural Information Processing Systems. <https://doi.org/10.48550/arXiv.1706.03762>
- [4] Hsiao W Y, Liu J Y, Yeh Y C, et al. Compound word transformer: Learning to compose full-song music over dynamic directed hypergraphs[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(1): 178-186. <https://doi.org/10.1609/aaai.v35i1.16091>
- [5] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, 6840-6851. <https://doi.org/10.48550/arXiv.2006.11239>
- [6] Graves, A. (2013). Generating sequences with recurrent neural networks. arXiv preprint <https://doi.org/10.48550/arXiv.1308.0850>
- [7] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [8] Kong, Q., Li, B., Chen, J., & Wang, Y. (2020). Diffwave: A versatile diffusion model for audio synthesis. arXiv preprint arXiv:2009.09761. <https://doi.org/10.48550/arXiv.2009.09761>
- [9] Yang, L. (2024). The Importance of Dual Piano Performance Forms in Piano Performance in the Context of Deep Learning. Applied Mathematics and Nonlinear Sciences, 9(1). <https://doi.org/10.2478/amns.2024.9.1.123>
- [10] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In Yoshua Bengio and Yann LeCun, editors, 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings, 2014. <https://doi.org/10.48550/arXiv.1312.6114>

- [11] Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Curtis Hawthorne, Andrew M Dai, Matthew D Hoffman, and Douglas Eck. Music transformer: Generating music with long-term structure. <https://doi.org/10.48550/arXiv.1809.04281>
- [12] Jeff Ens and Philippe Pasquier. Mmm: Exploring conditional multi-track music generation with the transformer. arXiv preprint <https://doi.org/10.48550/arXiv.2008.06048>
- [13] Oord A. WaveNet: A Generative Model for Raw Audio [J]. arXiv preprint <https://doi.org/10.48550/arXiv.1609.03499>
- [14] Liang, F. (2016). Bachbot: Automatic composition in the style of bach chorales. University of Cambridge, 8(19-48), 3-1. <https://doi.org/10.48550/arXiv.1604.05450>
- [15] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training gans. Advances in neural information processing systems, 29. <https://doi.org/10.48550/arXiv.1606.03498>
- [16] Donahue, C., McAuley, J., & Puckette, M. (2018). Adversarial audio synthesis. arXiv preprint <https://doi.org/10.48550/arXiv.1802.04208>
- [17] Chen, Y., Huang, L., & Gou, T. (2024). Applications and Advances of Artificial Intelligence in Music Generation: A Review. <https://doi.org/10.48550/arXiv.2409.03715>
- [18] Agostinelli, A., Denk, T. I., Borsos, Z., Engel, J., Verzetti, M., Caillon, A.,... & Frank, C. (2023). Musiclm: Generating music from text. <https://doi.org/10.48550/arXiv.2301.11325>
- [19] Huang, Q., Jansen, A., Lee, J., Ganti, R., Li, J. Y., & Ellis, D. P. (2022). Mulan: A joint embedding of music audio and natural language. <https://doi.org/10.48550/arXiv.2208.12415>
- [20] Chung, Y. A., Zhang, Y., Han, W., Chiu, C. C., Qin, J., Pang, R., & Wu, Y. (2021, December). W2v-bert: Combining contrastive learning and masked language modeling for self-supervised speech pre-training. In 2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU) (pp. 244-250). IEEE. <https://doi.org/10.1109/ASRU51585.2021.9644252>
- [21] Zeghidour, N., Luebs, A., Omran, A., Skoglund, J., & Tagliasacchi, M. (2021). Soundstream: An end-to-end neural audio codec. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 30, 495-507. <https://doi.org/10.1109/TASLP.2021.3074799>
- [22] Schneider, F., Kamal, O., Jin, Z., & Schölkopf, B. (2024, August). Mo ſai: Efficient text-to-music diffusion models. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 8050-8068).
- [23] Pachet, F. (2003). The continuator: Musical interaction with style. Journal of New Music Research, 32(3), 333-341. <https://doi.org/10.1076/jnmr.32.3.333.25054>
- [24] Briot, J. P., Hadjeres, G., & Pachet, F. D. (2020). Deep learning techniques for music generation (Vol. 1). Heidelberg: Springer. <https://doi.org/10.1007/978-3-030-43792-3>
- [25] Aalbers, S., Fusar-Poli, L., Freeman, R. E., Spreen, M., Ket, J. C., Vink, A. C.,... & Gold, C. (2017). Music therapy for depression. Cochrane database of systematic reviews, (11). <https://doi.org/10.1002/14651858.CD004517.pub3>
- [26] Sacks, O. (2008). Musicophilia. Tales of Music and the Brain. London (Picador) 2008. <https://doi.org/10.1007/978-0-387-75773-9>
- [27] Singh, P. Media 2.0: A Journey through AI-Enhanced Communication and Content. Media and AI: Navigating, 127. <https://doi.org/10.1007/978-3-030-43792-3>
- [28] Gan, C., Huang, D., Chen, P., Tenenbaum, J. B., & Torralba, A. (2020). Foley music: Learning to generate music from videos. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16 (pp. 758-775). Springer International Publishing. https://doi.org/10.1007/978-3-030-58585-2_47
- [29] Björk, S., & Holopainen, J. (2005). Games and design patterns. The game design reader: A rules of play anthology, 410-437. <https://doi.org/10.1007/978-3-030-43792-3>
- [30] Briot, J. P., & Pachet, F. (2020). Deep learning for music generation: challenges and directions. Neural Computing and Applications, 32(4), 981-993. <https://doi.org/10.1007/s00521-020-04728-0>