Background Research

Ben Wakefield

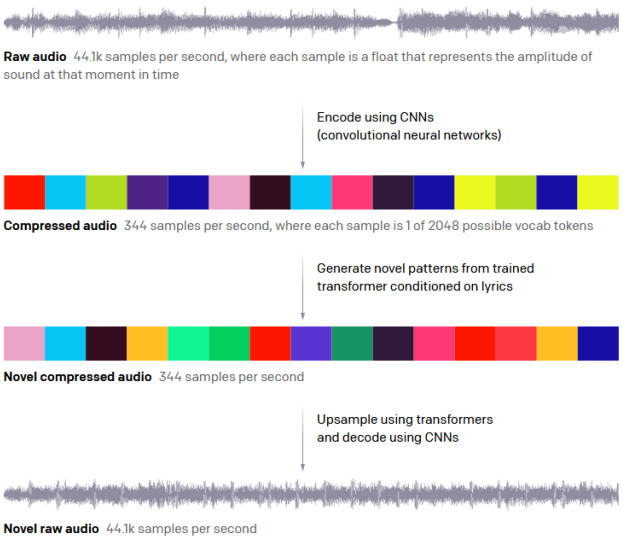
February 2023

Words: 988

1 Models for Music Generation

Jukebox, MuseNet, PerformanceRNN, and Riffusion are all AI music generation models that use deep neural networks. However, they use different types of neural network architectures and techniques to generate music, and have varying levels of usability.

Jukebox

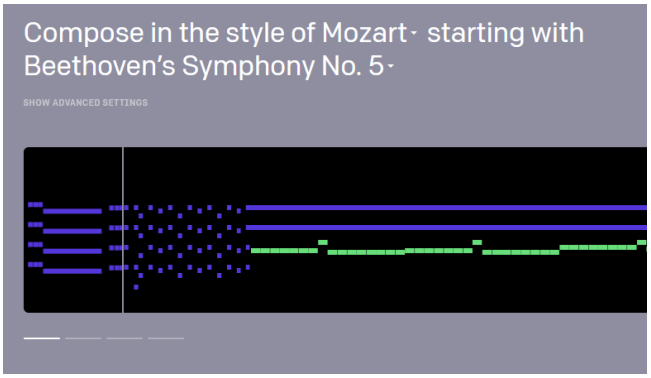


OpenAI Jukebox[1] is an AI system developed by OpenAI that uses a combination of deep neural networks and natural language processing to generate music in a variety of styles and genres. The system is based on a large corpus of music that has been preprocessed and organized into a hierarchy of musical concepts, such as scales, chords, and melody lines. When a user inputs a musical prompt, such as a short melody or a genre, the system generates a sequence of musical events that follow the input and are consistent with the hierarchical structure of the music corpus.

OpenAI Jukebox is trained on a dataset of over a million songs, and its deep neural networks have billions of parameters. The system is capable of generating original compositions in a wide range of styles and genres, from classical and

jazz to pop and hip-hop. The generated music can be customized by adjusting parameters such as the tempo, key, and instrumentation.

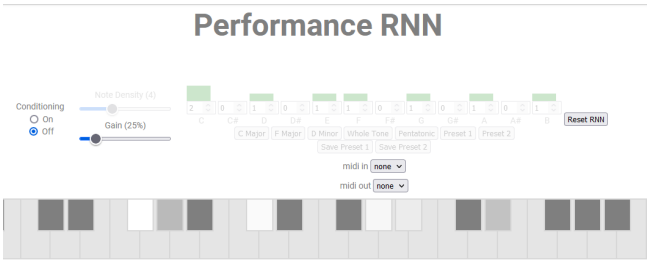
MuseNet



MuseNet[2] is an AI system developed by OpenAI that uses deep neural networks to generate music in a wide range of genres and styles. The system is based on a generative model architecture that uses a combination of hierarchical latent variables and autoregressive models to generate long sequences of musical events. MuseNet is trained on a massive dataset of MIDI files that cover a broad range of musical genres and styles, including classical, jazz, pop, and rock. The system can generate original compositions of varying lengths and complexity, and users can specify parameters such as the genre, tempo, and instrumentation to customize the output.

One of the key features of MuseNet is its ability to generate music that is both diverse and coherent. The system is able to learn and incorporate complex musical structures and patterns from the training data, enabling it to generate music that is musically consistent and coherent, while also being diverse and unpredictable.

PerformanceRNN

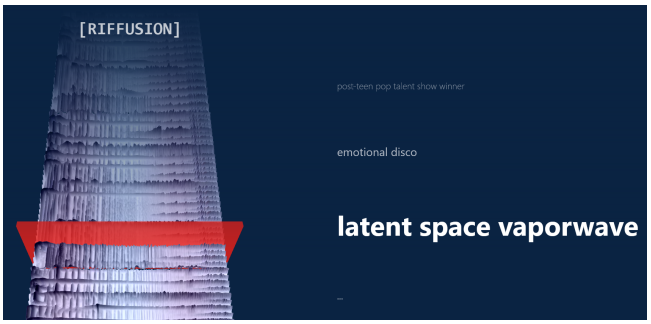


PerformanceRNN[5] is a recurrent neural network (RNN) model developed by the Magenta project at Google that is designed to generate expressive and nuanced music performances. Unlike other AI music generation models that focus on generating music notation or MIDI files, PerformanceRNN generates raw audio waveforms that capture the detailed nuances and subtleties of a musical performance, such as the timing, phrasing, and articulation.

The model is trained on a large dataset of music performances, such as piano recordings and MIDI files, and uses a combination of LSTM (Long Short-Term Memory) and WaveNet architectures to generate high-quality audio waveforms. The system is able to generate performances in a wide range of styles and genres, including classical, jazz, and pop.

PerformanceRNN is designed to be highly flexible and customizable, allowing users to specify various parameters, such as the style and genre of the music, as well as the level of expressiveness and complexity. The system can also be used for tasks such as music transcription and arrangement, as it is able to analyze and generate complex musical structures and patterns.

Riffusion



Riffusion[6] is an open-source music generation system developed by the Riffusion research group at the University of Montreal. The system uses a combination of deep neural networks and evolutionary algorithms to generate novel and creative musical sequences, such as melodies, chords, and rhythms.

The system is based on a generative model architecture that uses deep recurrent neural networks (RNNs) to learn the statistical patterns and relationships in a large corpus of MIDI data. The system also employs an evolutionary algorithm to guide the generation process, allowing the model to explore and discover new and creative musical sequences that are not present in the training data.

One of the unique features of Riffusion is its ability to generate music that is both diverse and coherent. The system can generate complex and intricate musical structures and patterns, while also maintaining a consistent musical style and coherence. The system is also highly customizable, allowing users to specify various parameters such as the style, genre, and complexity of the music.

2 Limitations

The biggest challenge with these models is that they often generate music that is repetitive or lacking in originality. This is because the models tend to learn and replicate common patterns and structures from the training data, which can lead to a lack of novelty in the generated music. Additionally, generating high-quality music requires a significant amount of computational resources, which can be a limiting factor for many users. Training and generating music with these models can be computationally intensive and require specialized hardware such as GPUs or TPUs.

Furthermore, these models are still in the early stages of development, and there is still much to be learned about how to generate truly creative and expressive music with AI. While these models have made significant progress in generating realistic and coherent music, they still have limitations in terms of their ability to generate truly original and innovative music.

Lastly, these models typically lack accessibility and generally would not be used in a typical users workflow. They require knowledge of command-line applications, installing dependencies, and creating virtual environments. MuseNet[3], Magenta[4], and Riffusion[7] provide demos for experimenting with the models, but the output is at reduced quality and the UI limits some customization options.

3 Improvements

In essence, this project will focus on a simpler user-interface. The models will also already be fine-tuned to a very specific style in order to yield the best results, without the user having to tediously experiment with parameters and wait for training to take place. Additionally, the interface will allow users to provide feedback to the model, letting them rank results on a quality scale, which will be fed back into the model for continuous improvement.

References

[1] Prafulla Dhariwal et al. *Jukebox: A Generative Model for Music*. 2020. URL: https://github.com/openai/jukebox.

[2] Curtis Hawthorne et al. "Neural Generation of Multi-track Music with Hierarchical Variational Autoencoders". In: *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS)*. 2019. URL: https://proceedings.neurips.cc/paper/2019/hash/e42a93761cbae22b58d4e5c5a924bc24-Abstract.html.

[3] *MuseNet Demo*. https://openai.com/blog/musenet/. 2019.

[4] *PerformanceRNN Demo*. https://magenta.tensorflow.org/demos/performance_rnn/. 2022.

[5] Colin Raffel et al. "Composer Identification from Polyphonic Music Using Neural Embeddings". In: *Proceedings of the 20th International Society for Music Information Retrieval Conference (ISMIR)*. 2019. URL: https://ismir2019.ewi.tudelft.nl/static/papers/144_ISMIR_2019.pdf.

[6] Zafar Rasheed et al. "Riffusion: Creative AI for Music Generation". In: *Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR)*. 2020. URL: https://ismir2020.ismir.net/wp-content/uploads/2020/10/165_Paper.pdf.

[7] *Riffusion Demo*. https://www.riffusion.com/. 2022.