Machine Learning for Music Generation

SEMINAR

JAGAN S NAIR

College of Engineering Chengannur
Master of Computer Applications

January 3, 2025

Guided by
Smt.Linnet Elsa John

Roll No: 37, S4,MCA

Table of Contents

INTRODUCTION

- Advanced music generation using hybrid models.
- Potential for nuanced and expressive compositions.
- Music composition as a traditionally human domain.
- Development of genre-spanning, high-quality compositions.
- Contribution to revolutionizing music informatics.
- Benefits: personalized experiences, collaborative creation, democratization, and cultural preservation.

OBJECTIVE

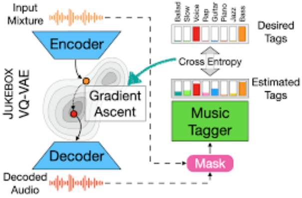
- Traditional music creation relies on manual composition by humans.
- AI-driven methods use deep learning for autonomous or collaborative music generation.
- Traditional methods depend on human creativity; AI uses computational power and data analysis.
- Challenges: complexity, originality, emotional expression, and cultural/ethical considerations.

Literature Review

Paper Name	Author(s)	Methodologies	Advantages	Disadvantages
StemGen: A Music Generation Model That Listens (2024)	Parker et al	Introduces a transformer-based model for music generation that responds to musical context.	Achieves high audio quality; exhibits strong musical coherence with context.	High computational requirements; limited control over fine-grained details.
Graph-based Polyphonic Multitrack Music Generation(2023)	Cosenza et al	Novel graph representation for music and a deep Variational Autoencoder to generate musical graphs.	Maintains tonal and rhythmic consistency; allows conditioning on specific instruments.	Complexity in modeling intricate musical structures; requires extensive training data.
JEN-1: Text-Guided Universal Music Generation with Omnidirectional Diffusion Models(2023)	Li et al	autoregressive and non-autoregressive training for text-to-music generation.	high-quality music generation; computationally efficient.	Potential challenges in capturing complex musical nuances; dependency on large datasets.
Progressive Distillation Diffusion for Raw Music Generation (2023)	Pavlova	Applies a diffusion model approach to generate raw audio files, utilizing a 1D U-Net architecture.	generates high-fidelity audio;adaptable to various music genres.	Computationally intensive; slower generation speed compared to other models.

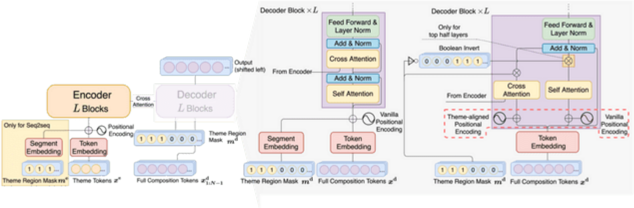
Table: Literature Review for Music Generation using ML

- **JukeBox**: Generates music and singing in raw audio using VQ-VAE and Transformers.
  - Trained on a dataset of 1 million songs.
  - Slow training process but produces remarkable results.
  - Outputs coherent music with harmony, rhythm, and singing across genres.
  - Capable of mimicking artists' styles and generating lyrics.



EXISTING SYSTEM

- **Theme Transformer: Theme-Conditioned Transformer**: Uses a Transformer encoder for theme-based music generation with cross-attention in the decoder.
  - Ensures thematic repetition and variation in the generated music.
  - Retrieves themes using contrastive representation learning and clustering.
  - Segments music into fragments, clusters them, and selects representative themes as conditions.
  - Produces musically coherent and engaging compositions compared to baseline models.



EXISTING APPLICATIONS

- **Spotify**: Popular digital music, podcast, and video streaming service with access to millions of songs and content from creators worldwide.
- **SoundHound**: Versatile music application that identifies songs, displays real-time lyrics, and offers a voice-controlled AI.

(a) Spotify

(b) SoundHound

Proposed Solutions

- **Music Notation**
  - Visual coding system to depict music using symbols.
  - Guitar tab consists of horizontal lines for strings and numbers for frets.
  - MIDI enables communication between computers and instruments, encoding details like pitch, velocity, vibrato, and volume.
  - Allows for easy transposition, time signature changes, and instrument switching.
  - Efficient and economical, with symphonies stored in lightweight files, ideal for generation tasks.

Proposed Solutions

- **GTZAN**
  - Simple, reduced size, and balanced structure.
  - Collection of 1,000 audio tracks evenly distributed across 10 popular genres, eliminating the need for preprocessing.
  - Contains 30-second audio files organized into subfolders labeled by genre, with 100 files per genre.
- **Free Music Archive (FMA)**
  - Richly annotated, high-quality, and diverse music collection for research tasks like genre tagging.
  - Provides track-level annotations, including unique ID, title, album, and release year.
  - Categorizes music into 16 top-level genres (e.g., Pop, Rock, Electronic) and 161 sub-genres (e.g., Psych-Rock, Lo-Fi, Drone).

Proposed Solutions

- **Million Song Dataset (MSD)**
  - Public collection of audio features and metadata for a million contemporary music tracks.
  - Extensive data for music research tasks, including genre classification.
  - Data computed using The Echo Nest API, providing objective song information.
  - Lacks diversity and representativeness across all world music genres.
  - Spotify API enriches research with detailed track, artist, album, playlist, and audio analysis data.

Music Generation

- AI improvises within musical parameters, making it suitable for Jazz-style generation.
- While AI lacks human emotions, it can replicate patterns to evoke specific grooves and feelings.
- Simplifies audio structure analysis by using programmatically-friendly musical notations.
- Preprocesses datasets by extracting notes from audio tracks and outputting them in musical notation.
- Model predicts new pitches for given patterns, iteratively generating sequences until the desired audio length is achieved.

Music Generation

- Deep Learning: Data Augmentation using 4 Mel spectrograms per audio track — original, frequency-masked, time-masked, and noise-added versions (mean 0, standard deviation 0.3) to increase model robustness and prevent overfitting.
- Machine Learning: Created 2 CSV files containing mean and variance of features like zero-crossing rate, spectral features, harmonics, percussive elements, and MFCC from the dataset.
- Model: Variational Autoencoder (VAE) ensures reconstructed input closely resembles the original, generating new, pleasant Jazz music.

- 2 hybrid models: VAE LSTM and Transformer.

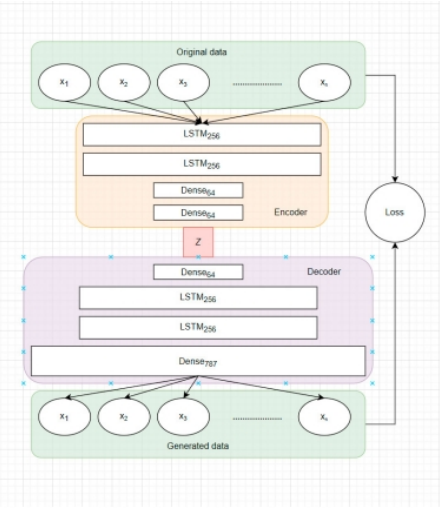
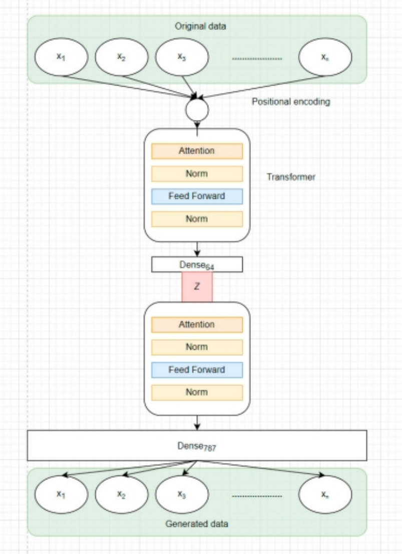


Fig. 2. Architecture of VAE that uses LSTM to encode/decode data.



Evaluation

- Classical metrics for Variational Autoencoders: reconstruction loss, KL-divergence, and total loss.
- Reconstruction loss measures how accurately the model recreates input data after encoding and decoding.
- Total loss balances reconstruction (avoids overfitting) and KL-divergence (ensures diversity but avoids low-quality outputs).
- Training time per epoch is significant, approximately 10,000 seconds.
- Algorithms used include Multilayer Perceptrons (MLP), Convolutional Neural Networks (CNN), k-Nearest Neighbors (KNN), Support Vector Machines (SVM), Random Forest, XGBoost, and ensemble methods.

Future Scope

- **Tailored Architectures for Musical Attributes:**
- **One-Shot Transfer:**
- **Diverse Evaluation Metrics:**
- **Improved Hardware for Training:**

CONCLUSION

- Models aim to mimic human music production while adding creativity.
- Future research seeks to refine music generation to better resemble human creativity and evoke artistic sensibilities.
- Focus on reducing strict dependency on input data.
- Explore "one-shot transfer" techniques.
- Aim to enhance AI-driven music generation with human emotion and artistic expression.

REFERENCES

- "STEMGEN: A Music Generation Model That Listens," by J. D. Parker et al., ICASSP 2024 - IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Korea, Republic of 2024.
- "Graph-based Polyphonic Multitrack Music Generation" by Emanuele Cosenza, Andrea Valenti, Davide Bacciu (2023), University of Pisa.
- "JEN-1: Text-Guided Universal Music Generation with Omnidirectional Diffusion Models" by Peike Patrick Li, Boyu Chen, Yao Yao, Yikai Wang, Allen Wang, Alex Wang (2024), IEEE Conference on Artificial Intelligence (CAI).
- "Progressive Distillation Diffusion for Raw Music Generation" by S. Pavlova, Svetlana (2023),

Thank you :)