related work for the project

March 26, 2023

Contents

1 Code

1.1 hugging face diffusers

- `https://github.com/huggingface/diffusers`

- Diffusers is the go-to library for state-of-the-art pretrained diffusion models for generating images, audio, and even 3D structures of molecules.

1.2 audio-diffusion-pytorch (ETH + max-planck)

- https://github.com/archinetai/audio-diffusion-pytorch

- A fully featured audio diffusion library, for PyTorch. Includes models for unconditional audio generation, text-conditional audio generation, diffusion autoencoding, upsampling, and vocoding. The provided models are waveform-based, however, the U-Net (built using a-unet), DiffusionModel, diffusion method, and diffusion samplers are both generic to any dimension and highly customizable to work on other formats. Note: no pre-trained models are provided here, this library is meant for research purposes.

1.3 dance diffusion (harmonai)

- https://colab.research.google.com/github/Harmonai-org/sample-generator/blob/main/Dance_Diffusion.ipynb?pli=1#scrollTo=lU97ZiP7nSKS

- Unconditional random audio sample generation

- Audio sample regeneration/style transfer using a single audio file or recording

- Audio interpolation between two audio files

2 Literature

2.1 Noise2Music: Text-conditioned Music Generation with Diffusion Models

- https://arxiv.org/abs/2302.03917

- https://google-research.github.io/noise2music/

- We introduce Noise2Music, where a series of diffusion models is trained to generate high-quality 30-second music clips from text prompts. Two types of diffusion models, a generator model, which generates an intermediate representation conditioned on text, and a cascader model, which generates high-fidelity audio conditioned on the intermediate

representation and possibly the text, are trained and utilized in succession to generate high-fidelity music. We explore two options for the intermediate representation, one using a spectrogram and the other using audio with lower fidelity. We find that the generated audio is not only able to faithfully reflect key elements of the text prompt such as genre, tempo, instruments, mood, and era, but goes beyond to ground fine-grained semantics of the prompt. Pretrained large language models play a key role in this story – they are used to generate paired text for the audio of the training set and to extract embeddings of the text prompts ingested by the diffusion models.

2.2 DANCE2MIDI: DANCE-DRIVEN MULTI-INSTRUMENTS MUSIC GENERATION

- `https://www.catalyzex.com/paper/arxiv:2301.09080`

- Dance-driven music generation aims to generate musical pieces conditioned on dance videos. Previous works focus on monophonic or raw audio generation, while the multi- instruments scenario is under-explored. The challenges of the dance-driven multi-instruments music (MIDI) genera- tion are two-fold: 1) no publicly available multi-instruments MIDI and video paired dataset and 2) the weak correla- tion between music and video. To tackle these challenges, we build the first multi-instruments MIDI and dance paired dataset (D2MIDI). Based on our proposed dataset, we in- troduce a multi-instruments MIDI generation framework (Dance2MIDI) conditioned on dance video. Specifically, 1) to model the correlation between music and dance, we encode the dance motion using the GCN, and 2) to generate harmo- nious and coherent music, we employ Transformer to decode the MIDI sequence. We evaluate the generated music of our framework trained on D2MIDI dataset and demonstrate that our method outperforms existing methods. The data and code are available on `https://github.com/Dance2MIDI/Dance2MIDI`

2.3 Moûsai: Text-to-Music Generation with Long-Context Latent Diffusion

- `https://arxiv.org/pdf/2301.11757.pdf`

2.4 review of music generation

- https://www.catalyzex.com/paper/arxiv:2211.09124

2.5 SYMBOLIC MUSIC GENERATION WITH DIFFU-SION MODELS

- https://arxiv.org/pdf/2103.16091.pdf

2.6 Diffusion-LM on Symbolic Music Generation with Controllability (stanford)

- http://cs230.stanford.edu/projects_fall_2022/reports/16.pdf

3 Available datasets

TODO

4 diverse

4.1 overview of different music gen methods

- https://www.catalyzex.com/s/music%20generation