

Рецензия

Jukebox: A Generative Model for Music

Добряев Иван

В чем суть?

Команда Open AI использует VQ-VAE-архитектуру, чтобы сжать аудио до дискретных кодов, а также моделирует их, используя авторегрессионные трансформеры

Авторы статьи

Команда OpenAI

- Ilya Sutskever
 - Co-Founder and Chief Scientist of OpenAI
 - h-index 75
- Alec Radford
 - Researcher
 - h-index 31
 - NLP
- Prafulla Dhariwal
 - Researcher,
 - h-index 15
 - NLP, CV, Loss function etc.
- Jong Wook Kim
 - Researcher
 - h-index 11
 - Music Information Retrieval

Основные статьи

Neural Discrete Representation Learning

Цитирование и продолжение

У данной статьи нет прямого продолжения. А также цитирование происходит как указание существующих других подходов, а не заимствование подхода.

- Msanii: High Fidelity Music Synthesis on a Shoestring Budget 2023 (WIP) diffusion- based model
- AudioLM: a Language Modeling Approach to Audio Generation Google

Сильные стороны

Новый на то время подход, что дает возможность сгенерировать достаточно высококачественную (по мнению авторов) и разнообразную музыку с корреляцией, способной продержаться в течение нескольких минут. Также можно управлять стилем музыки и вокала.

Слабые стороны

Результаты экспериментов демонстрируют достаточно высокое качество в генерации песен (со слов авторов), но в тоже время нет достаточных данных для понимания и измерения того, насколько генерируемые результаты соответствуют музыке, созданной человеком и отсутствует сравнение с предыдущими подходами на более коротких интервалах времени.

Возможные улучшения

- Для улучшения текущей работы можно было бы исследовать влияние внешних, признаков на музыкальные результаты, такие как паузы.
- Также стоит проанализировать влияние лирики на адекватность генерируемой музыки.

Привести примеры метрик. Например MOS