

УДК 004.85

ГЕНЕРАТИВНАЯ СЕТЬ ДЛЯ СОЗДАНИЯ ОБЛОЖКИ МУЗЫКАЛЬНОЙ КОМПОЗИЦИИ В ВЕКТОРНОМ ФОРМАТЕ

Бизяев И.В. (федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Научный руководитель – к.ф.-м.н., доцент факультета информационных технологий и программирования (квалификационная категория «ординарный доцент»)

Фильченков А.А.

(федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

В работе изложены основные принципы предполагаемого алгоритма для генерации обложек музыкальных композиций в виде векторной графики с помощью методов машинного обучения. Описана проделанная на данный момент и предстоящая работа.

Введение. Восприятие человеком звуковых и визуальных стимулов сильно связано. При прослушивании аудиозаписи, люди нередко могут представить действия и цвета, связанные со звуком. Но, несмотря на недавние достижения в области различных генеративных моделей, основанных на текстовых описаниях, аудиовизуальных моделей было разработано немного.

Существующие модели в основном нацелены на соотнесение звуковой информации с некими реальными сценами, действиями или действующими лицами. В данной же работе предлагается генерировать для музыкальных композиций обложки альбомов на основе векторной графики.

Построение аудиовизуальной генеративной модели включает извлечение звуковых характеристик и условную генерацию изображения. Существует множество работ по условной генерации растровых изображений, но тема условной генерации векторного изображения исследована недостаточно. Предполагается, что благодаря использованию векторной графики можно будет избежать нечетких линий и артефактов, характерных для порождающих состязательных сетей, и устранить ограничение на размер в пикселях.

В качестве примеров существующих сервисов, предлагающих пользователям функционал для автоматической генерации обложек музыкальных композиций, можно привести веб-сайты artbreeder.com и ganalbum.art. Оба указанных сервиса, используя генеративно-состязательные модели (GAN), без опоры на аудиоданные выдают растровые изображения с неопределенными фигурами и с контурами, напоминающими буквы, но не являющимися осмысленным текстом, поэтому для получения результатов приемлемого качества требуется активное участие человека.

Основная часть. Цель данной работы: предложить алгоритм генерации обложки высокого качества к музыкальному треку — в векторном формате.

На данный момент собран набор данных (датасет) из 1500 треков с обложками на основе векторной графики, подготовлены скрипты для получения необходимой дополнительной информации о треках (жанров, признаков исполнителей) из публичных информационных баз, рассмотрены способы извлечения признаков из аудио. Средствами краудсорсинговой платформы Яндекс.Толока произведена разметка соответствия воспринимаемых настроений трекам из датасета. Аудиоданные музыкальных композиций представлены в виде эмбеддингов. Для работы с визуальной составляющей изучено представление векторных изображений в формате SVG, реализовано упрощенное представление подмножества этого формата, его преобразование в SVG, отрисовка и построение из эмбеддингов на выходе GAN.

Для достижения поставленной цели предстоит использовать результаты разметки эмоций, чтобы научиться автоматически выделять эмоции из музыкальных композиций. Затем требуется реализовать подходящую для генерации обложек архитектуру GAN. Предполагается использовать вложенный GAN для генерации фигур заданных типов, внешний GAN — для анализа музыкальной композиции с использованием механизма внимания и генерации обложки в упрощенном представлении формата SVG с учетом воспринимаемого настроения музыкального произведения. Затем нужно задействовать определение текста на обложках тренировочных данных и выбор положения и оформления текста на генерируемых обложках.

Необходимо обучить GAN генерировать векторные файлы, похожие на обложки собранных треков. На основе результатов этого обучения возможно оптимизировать архитектуру сети. Наконец, предстоит получить оценку генерируемых обложек с помощью ассессоров. Для сравнения можно реализовать генерацию обложки по тексту песни с помощью AttnGAN и предложить ассессорам выбрать из сгенерированных двумя способами обложек более подходящую на наборе популярных музыкальных треков.

Выводы. В данной работе рассматривается алгоритм генерации обложек музыкальных композиций в виде векторной графики на основе генеративно-состязательных сетей и с учетом воспринимаемых эмоций от прослушивания. Планируется, что предложенный алгоритм позволит улучшить качество создаваемых изображений, что, в свою очередь, сделает его использование проще, быстрее и дешевле в сравнении с существующими решениями.

Бизяев И.В. (автор)

Подпись

Фильченков А.А. (научный руководитель)

Подпись