

РАЗРАБОТКА ГЕНЕРАТИВНОЙ ТЕХТ-ТО-IMAGE МОДЕЛИ ДЛЯ ГЕНЕРАЦИИ МОДНОГО СТИЛЯ

Иванова Е. А. (ИТМО)

Научный руководитель – преподаватель Михайленко А. Ю. (ИТМО)

Искусственный интеллект - активно развивающаяся сфера в современном мире. На данный момент существует ИИ, генерирующий стили художников, а сетей, генерирующих (а не рекомендуемых) модный стиль пока что нет. Поэтому главная цель проекта – разработать нейронную сеть, которая будет работать, как искусственный консультант-стилист, и на основе запроса пользователя (желаемый стиль или ситуация) предлагать модный образ. Это поможет оптимизировать времени, затрачиваемое на выбор и покупку одежды. За последние годы многие отрасли в России расширили использование решений в области ИИ. В данной работе мы приближаем применимость искусственного интеллекта к повседневной жизни.

Решение задачи предполагается через разработку генеративной модели.

За основу для работы взят датасет, включающий изображения одежды, на нем планируется обучение модели. Рассмотрение статей контекстно-зависимого прогнозирования визуальной совместимости. Как определить, совместимы ли два или более предмета одежды? Частично ответ заключается в понимании визуальной эстетики и зависит от личных предпочтений, сформированных социальными установками, временем и местом. Существует решение проблемы прогнозирования совместимости, используя графовую нейронную сеть, которая учится генерировать встраивания продуктов в зависимости от их контекста [1].

Анализ имеющихся подходов (отдельно text-to-image). Выделение их недостающих частей и неточностей, изучение различных моделей генеративных сетей, их обзор и сравнение. Во всех методах основные недостатки: неточная метрика; неизвестно, как работает архитектура; не совсем точная генерация.

Изучение разных типов генеративных сетей (например, VAE и GAN) и сравнение сгенерированных стилей показывает: GAN получается добавлением в VAE дискриминатора, который обучается на определение: подан на вход сети оригинал или восстановленный объект. И с помощью GAN изображения будут более "реальными" и яркими. VAE же обычно генерируют размытые и не фотореалистичные изображения [2].

На начальном этапе главное – научиться распознавать различную одежду классическими сверточными сетями, и построить начальную версию одной из генеративных моделей на основе существующих решений.

Разработанная модель будет иметь применение на сайтах интернет-магазинов одежды, в особенности на тех, где отсутствует возможность примерки перед покупкой. Она поможет облегчить выбор одежды, а также будет хорошим помощником стилистам и упростит им работу, повысив производительность за счет разнообразия вариантов.

Список использованных источников:

1. Guillem Cucurull, Perouz Taslakian, David Vazquez. Context-Aware Visual Compatibility Prediction // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 12617-12626.
2. Chenrui Z., Yuxin P. Stacking VAE and GAN for Context-aware Text-to-Image Generation, 2018. // IEEE Fourth International Conference on Multimedia Big Data (BigMM).