

УДК 004.8

ГЕНЕРАЦИЯ ФОТОГРАФИИ ЧЕЛОВЕКА В ПОЛНЫЙ РОСТ ПО ТЕКСТОВОМУ ОПИСАНИЮ

Крамер К.А. (Университет ИТМО), Ефимова В.А. (Университет ИТМО)

Научный руководитель – кандидат физико-математических наук, доцент Факультета информационных технологий и программирования Фильченков А.А.
(Университет ИТМО)

В рамках данной работы была исследована проблема генерации изображения человека по текстовому описанию. Изучены существующие методы решения данной задачи, проанализированы преимущества и недостатки. Реализован алгоритм, который позволяет, по фотографии человека и текстовому описанию генерировать новое изображение человека посредством трансформации краев.

Введение. Все мы можем описать изображение словами, в том числе и фотографию человека, но создание реалистичных изображений по описанию, намного сложнее и требует большого умения и опыта в графическом дизайне. Для машинного обучения это генеративная задача, которая гораздо более сложна, чем дискриминационные задачи, так как генеративная модель должна генерировать намного больше информации на основе меньших исходных данных. Данную задачу можно разбить на несколько подзадач, в числе которых будет генерация нового изображения человека по существующему. Это позволит создавать изображения несуществующих людей, не защищенные авторским правом. В последнее время появилось много видов сетей, решающих задачу генерации изображения по тексту, но они получают не реалистичные изображения низкого качества.

Основная часть. Предлагаемый алгоритм принимает на вход фотографию и текстовое описание, в виде JSON, на английском языке. После этого, первым важным этапом нашей задачи является извлечение человека из фотографии. Чтобы качественно отделить человека от остального изображения, было принято использовать alpha matting. Следующим этапом обработки изображений является получение его «ребер». В качестве модели для поиска «ребер» использовался Детектор границ Кэнни. Дальнейшим шагом является генерация новых «ребер» по описанию, в разработанном для этого алгоритме используется механизм внимания (attention mechanism). Конечным этапом является генерация нового изображения по полученным «ребрам».

Выводы. Несмотря на все сложности, генеративные модели чрезвычайно важны и полезны во многих сферах: игровая индустрия, киноиндустрия, различные арт-проекты, научные исследования и многое другое. Данный подход позволит получать изображения людей в высоком разрешении, соответствующие данному текстовому описанию.

Крамер К.А. (автор)

Фильченков А.А. (научный руководитель)