

УДК 004.89

## ПЕРСОНАЛИЗАЦИЯ ТЕХТ-ТО-IMAGE ГЕНЕРАТИВНЫХ МОДЕЛЕЙ ДЛЯ ГЕНЕРАЦИИ ЛЮДЕЙ

Детков Н.С. (Университет ИТМО)

Научный руководитель – кандидат физико-математических наук, Фильченков А.А.  
(Университет ИТМО)

**Введение.** Text-to-image генерация - это сложная задача, которая заключается в создании высококачественных и реалистичных изображений произвольных объектов на основе текстовых описаний. Последние достижения в области глубокого обучения и генеративных моделей привели к значительному прогрессу в этой области. Однако создание персонализированных изображений конкретного человека, животного или объекта из текста остается не до конца решенной задачей. Цель данного доклада - изучить современные методы и нейронные сети, используемые для персонализации text-to-image генерации, а также обсудить их ограничения и будущие направления. Мы рассмотрим недавние работы и методы, чтобы изучить достижения и проблемы в персонализации генерации текста к изображению для произвольных объектов, а также предложим направления для улучшения.

**Основная часть.** Задача text-to-image генерации заключается в создании высококачественных изображений произвольных объектов на основе текстовых описаний. Эта задача особенно сложна, поскольку требует от модели понимания семантики входного текста и синтеза реалистичных изображений, соответствующих описанию.

Существующие методы решения задачи генерации текста в изображение используют возможности глубоких нейронных сетей и генеративных моделей. Два таких метода включают DreamBooth [1] и Textual Inversion [2]. DreamBooth - это недавний метод, который использует тонкую настройку моделей диффузии текста в изображение для синтеза высококачественных изображений из текстовых описаний. Авторы предлагают подход к тонкой настройке существующих моделей диффузии текста к изображению с помощью данных о конкретном субъекте для улучшения качества генерируемых изображений. Авторы Textual Inversion, с другой стороны, предлагают схему, которая использует метод текстовой инверсии для создания персонализированных изображений конкретного человека или объекта. Как DreamBooth, так и Textual Inversion показали многообещающие результаты в создании высококачественных и реалистичных изображений произвольных объектов на основе текстовых описаний. Однако задача создания изображений из текста остается сложной, и необходимы дальнейшие исследования для улучшения визуального качества и разнообразия генерируемых изображений. Множество проблем, относящиеся к улучшению разнообразия, применимости различных текстовых описаний к объектам персонализации, уменьшению количества генераций с дефектами, остаются на повестке, и мы скрупулёзно показываем их, и предлагаем решения.

**Выводы.** В заключение, персонализация генерации текста в изображение для произвольных объектов является очень активной областью исследований со значительным прогрессом и проблемами. Приложения этой технологии включают электронную коммерцию, искусство и индустрию развлечений, где создание персонализированных изображений имеет высокую ценность.

### Список использованных источников:

1 Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., & Aberman, K. (2022). DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. ArXiv, abs/2208.12242.

2 Gal, R., Alaluf, Y., Atzmon, Y., Patashnik, O., Bermano, A.H., Chechik, G., & Cohen-

Or, D. (2022). An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion. ArXiv, abs/2208.01618.

Детков Н.С. (автор)

Подпись

Фильченков А.А. (научный руководитель)

Подпись