

УДК 004.89

РАЗРАБОТКА МЕТОДА ГЕНЕРАЦИИ ТРЕХМЕРНОЙ МОДЕЛИ МЕБЕЛИ ПО ТЕКСТОВОМУ ОПИСАНИЮ С УЧЕТОМ ЭРГОНОМИКИ

Попов А. С. (ИТМО)

Научный руководитель – кандидат технических наук Ефимова В. А.
(ИТМО)

Введение. Генерация трехмерных моделей на основе текстовых описаний является одним из наиболее перспективных направлений в области компьютерного зрения и искусственного интеллекта. Этот подход открывает новые возможности для автоматизации процессов проектирования и создания виртуальных объектов и сцен. Основной проблемой в области генерации трехмерных объектов является недостаток обширных баз данных, связывающих текстовые описания с соответствующими трехмерными моделями. Тем не менее, существуют модели, которые были обучены на обширных наборах данных для создания двухмерных изображений на основе текста, и знания, полученные из этих исследований, могут быть адаптированы для генерации трехмерного контента [1, 2]. В данной работе представлен подход к созданию трехмерных моделей мебели, учитывающий эргономику — ключевой аспект, который часто игнорируется в существующих методах, ориентированных преимущественно на визуальные аспекты и не придающих должного значения удобству и функциональности. При этом обучение трехмерной сцены по тексту будет производиться без предварительного дообучения модели генерации на данных, в которых сопоставляются тексты и трехмерные сцены.

Основная часть. В ходе данного исследования была автоматически собрана коллекция двухмерных изображений мебели вместе с текстовыми описаниями и различными характеристиками продуктов, такими как максимальная нагрузка, возможность регулировки высоты, форм-фактор и др.

Учитывая, что для эффективного дообучения современных моделей генерации двухмерных изображений требуются значительные вычислительные ресурсы, мы применили дообучение адаптера LoRa [3] модели Stable Diffusion XL [4]. Для сохранения представления трехмерных сцен использовалась модифицированная версия NeRF [5]. Процесс обучения базировался на дистилляции параметров случайного образца двухмерного изображения из трехмерной модели, а также параметров изображения, сгенерированного с использованием Stable Diffusion XL. Алгоритм содержит следующие основные этапы.

- Обработку входного текстового описания и извлечения необходимых характеристик.
- Генерацию случайного представления трехмерной сцены
- Минимизацию разницы между параметрами NeRF и параметрами, сгенерированными адаптером LoRa модели генератора двухмерного изображения.

Предложенный метод позволяет создавать функциональные и удобные для использования трехмерные модели мебели, которые в дальнейшем могут быть изготовлены в реальном мире. При этом значительно сокращается время, необходимое для проектирования.

Выводы. Был разработан метод генерации трехмерных моделей мебели на основе текстовых описаний, с акцентом на эргономику. Также в автоматическом режиме был собран набор данных для обучения адаптера модели генерации двухмерных снимков мебели. Предложенный подход обеспечивает существенное сокращение времени, необходимого для проектирования, увеличивает точность создаваемых моделей и обращает особое внимание на удобство использования готовых изделий.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ:

1. Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S. Sara Mahdavi, Rapha Gontijo Lopes, Tim Salimans, Jonathan Ho† , David J Fleet† , Mohammad Norouzi. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding // Google Research – 2022.
2. Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser Bjorn Ommer, High-Resolution Image Synthesis with Latent Diffusion Models – 2022.
3. Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yanzhi Li, Shean Wang, Lu Wang, Weizhu Chen. LoRA: Low-Rank Adaptation of Large Language Models – 2021.
4. Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, Robin Rombach. SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis – 2023.
5. Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis // Google Research – 2020.