ГЕНЕРАЦИЯ 3D-МОДЕЛЕЙ С ПОМОЩЬЮ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

Шафаренко Д. И.

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Красковский П. Н. – ст. преподаватель

В данной работе рассматриваются методы решения задачи генерации трехмерных моделей, в частности генерация моделей в текстовом представлении с использованием больших языковых моделей, а также предлагается метод генерации, сочетающий в себе большие языковые модели и диффузионные модели.

Прежде чем приступить к решению задачи генерации 3D-моделей, необходимо определиться с целевым представлением. Существует множество 3D-представлений, такие как воксельное преставление, сетка полигонов, многоракурсное представление, облако точек, САПР-представление, неявное представление [1]. В данной работе в качестве целевого представления была выбрана полигональная сетка как наиболее удобное для непосредственного применения на практике.

Для решения задачи генерации 3D-моделей существуют различные архитектуры нейронных сетей, обладают разными характеристиками скорости, требуемой памяти и качества генераций, а также ориентированы на разные представления 3D-моделей. Для генерации полигональных сеток чаще всего используется авторегрессионный подход, который последовательно генерирует вершины и полигоны.

В данной работе выбран подход LLaMA-Mesh [2]. Этот подход в своей основе использует большую языковую модель LLaMA-3.1 [3], которая представляет собой архитектуру трансформера и, в сущности, является авторегрессионной моделью, предсказывающей следующий токен текста на основе предыдущих. В отличие от специализированных авторегрессионных моделей, подход, основанный на большой языковой модели, открывает возможности для совмещения 3D-генерации и текстового общения.

Подход LLaMA-Mesh дообучает LLaMA-3.1 на генерацию файлов формата OBJ, который хранит координаты вершин и полигонов в текстовом человекочитаемом виде. Языковые модели не могут обрабатывать числа естественным образом, потому необходимо предварительно обработать данные для обучения, чтобы добиться лучшего качества. Для этого координаты вершин и полигонов масштабируются и округляются до определенных дискретных значений, таким образом ограничивая возможную точность генераций, но облегчая задачу для языковой модели.

В данной работе предлагается объединить генерацию 3D-объектов при помощи авторегрессионных языковых моделей с диффузионными моделями, которые отличаются высоким качеством результатов генерации. Принцип работы диффузионных моделей заключается в итеративной одновременной генерации всех вершин и полигонов из нормального шума. Подход LLaDA [4] предлагает большую языковую модель на основе дискретной диффузионной модели, которая генерирует токены не последовательно, а в произвольном порядке. Предполагается, что последовательный порядок генерации 3D-модели является ограничением, и дискретная диффузионная модель позволит генерировать 3D-модель в произвольном порядке, вместе с тем сохраняя сочетание 3D-генерации и понимания естественного языка.

Чтобы сократить объем памяти, требуемый для дообучения модели, можно использовать технику QLoRA, которая сочетает в себе использование низкоранговой матрицы для обновления весов модели и квантизацию весов до 4 или 8 бит.

В результате работы было создано программное средство, сочетающее в себе функции генерации 3D-моделей и текстового диалогового общения, что становится возможным благодаря выбранному методу генерации. Однако этот метод обладает существенными ограничениями, в частности большие требования к вычислительным ресурсам и качество генерации, уступающее специализированным методам, что может быть следствием относительно небольшого размера используемой модели, поэтому стоит провести дальнейшие исследования способностей языковых моделей к генерации 3D-объектов с использованием моделей большего размера.

Список использованных источников:

1. Advances in 3d generation: A survey / X. Li [et al.] // arXiv preprint arXiv:2401.17807. – 2024.
2. Llama-mesh: Unifying 3d mesh generation with language models / Z. Wang [et al.] // arXiv preprint arXiv:2411.09595. – 2024.
3. The llama 3 herd of models / A. Grattafiori [et al.] // arXiv preprint arXiv:2407.21783. – 2024.
4. Large Language Diffusion Models / S. Nie [et al.] // arXiv preprint arXiv:2502.09992. – 2025.
5. Qlora: Efficient finetuning of quantized llms / T. Dettmers [et al.] // Advances in neural information processing systems. – 2023. – Vol. 36. – P. 10088-10115.