

Генерация реалистичных 3D-моделей фасадов многоэтажных жилых зданий по заданным масс-моделям

Остапенко С.С. (Университет ИТМО)

Научный руководитель: к.т.н. Ефимова В.А. (Университет ИТМО)

Введение: Современные технологии позволяют автоматизировать всё больше задач без участия человека, и этот процесс постоянно ускоряется. Одной из активно развивающихся областей является архитектурный и урбанистический дизайн, где алгоритмы и математические методы помогают ускорять разработку проектов. Важным инструментом стала автоматическая визуализация, которая позволяет быстро оценить качество и привлекательность проекта и наглядно представить его заказчикам.

Рассматриваемая задача состоит в том, чтобы сгенерировать реалистичную 3D-модель жилого здания по заданной масс-модели и, возможно, текстовому запросу (промпту), описывающему желаемый результат. Сгенерированная модель должна иметь реалистичные трёхмерные фасадные элементы, а также соответствующие текстуры. Предполагается, что решение будет опираться на современные методы и модели машинного обучения и нейронных сетей.

Масс-модель здания – это 3D-модель, которая описывает грубую и примерную геометрию здания: форму, физические размеры, этажность. Типичные масс-модели жилых зданий: прямоугольный параллелепипед, объёмные буквы “Г”, “П”, “Ш” и т.д.

Среди существующих решений – ArcGIS CityEngine, Houdini и расширения Blender, такие как Geometry Nodes. Однако они основаны на процедурной генерации, зависят от заранее заданных правил и требуют глубоких знаний архитектуры для настройки, из-за чего плохо адаптируются к новым условиям и стилям. Также есть генератор FrankenGAN [1], который использует нейросети для создания текстур и учитывает заданный стиль по изображению, но не позволяет гибко настраивать геометрию фасадных элементов.

Таким образом, ни одна из представленных программ не решает поставленную задачу полностью, поэтому она остаётся актуальной.

Основная часть: Предложенное решение генерирует каждую стену здания отдельно. Процесс генерации каждой стены состоит из трёх частей: генерация семантической разметки элементов фасада, генерация изображения фасада по разметке его элементов и генерация 3D-модели стены здания по полученному изображению фасада. После того, как 3D-модели всех стен сгенерированы, они собираются вместе, чтобы образовать 3D-модель всего здания.

Семантическую разметку элементов фасада можно представить как изображение фасада здания, на котором прямоугольниками обозначено расположение всех фасадных элементов, таких как окна, двери, балконы, карнизы, молдинги и т.д. Генерация такой разметки происходит процедурным образом: по заданным размерам стены можно определить количество этажей в здании, а также количество рядов окон и балконов; после чего по полученной информации можно сгенерировать расположения всех окон, балконов и прочих элементов фасада случайным образом с учётом некоторых простых правил (например, что 2 балкона не могут перекрывать друг друга, и т.д.).

Для получения изображения фасада по семантической разметке используется модель из работы [2], которая представляет из себя LoRA-адаптер для модели Stable Diffusion v1.4 [3], обученный на задачу генерации изображений фасадов по входному текстовому запросу (промпту). Чтобы модель генерировала элементы фасада в соответствии с полученной ранее разметкой, используется механизм ControlNet.

Третьим шагом генерации является получение, собственно, финальной 3D-модели стены. Для этого используется TRELIS [4] – нейронная сеть для генерации 3D-моделей по заданным на вход изображениям. Для получения 3D-модели на вход подаётся полученное ранее изображение фасада, а также разметка фасадных элементов.

Поскольку TRELIS не понимает форму модели, которую нужно сгенерировать (т.к. форму фасадных элементов очень тяжело понять по одной фотографии), разметка используется в качестве подсказки при генерации. Исходно сам процесс генерации выглядит следующим образом. На первом шаге TRELIS генерирует примерную геометрию модели, используя трёхмерную сетку из вокселей размером 64 на 64 на 64: каждый воксель получает значение 0 или 1, и их комбинация описывает примерную геометрию будущей 3D-модели. На втором шаге по этим вокселям генерируются детальная геометрия и текстура, которые потом экспортируются в 3D-модель. Для более точного управления желаемой геометрией в решении был убран первый шаг генерации. Вместо этого значения вокселей устанавливаются в соответствии с полученной ранее разметкой фасадных элементов (изначально в пространстве отмечается сама стена, после чего в ней “вырезаются” окна и двери и к ней “приклеиваются” балконы и другие элементы), после чего эти значения используются ровно так же, как если бы они были получены на первом шаге генерации в TRELIS.

Выводы: Решение находится в стадии разработки, однако предварительные результаты демонстрируют, что оно предлагает более высокую гибкость и адаптивность, чем существующие аналоги, а также не требует глубокой экспертизы в области архитектурного и урбанистического дизайна. Это преимущество обеспечивается тем, что количество мест в решении, использующих процедурную генерацию, сведено к минимуму, т.к. всё, кроме семантической разметки элементов, генерируется с помощью нейронных сетей, которые могут быть обучены под новые данные для работы в новых условиях.

Среди возможных применений предложенного решения такие задачи, как визуализация архитектурных проектов и генерация трёхмерных моделей зданий, которые можно использовать в компьютерных играх и при создании анимации.

Список использованных источников:

- [1] Tom Kelly, Paul Guerrero, Anthony Steed, Peter Wonka, and Niloy J. Mitra, “FrankenGAN: guided detail synthesis for building mass models using style-synchronized GANs,” *ACM Transactions on Graphics*, vol. 37, no. 6, p. Article 216, 1-14, Dec. 2018.
- [2] Ma, H. and Zheng, H., “Text Semantics to Image Generation: A Method of Building Facades Design Base on Stable Diffusion Model,” in *CDRF 2023. Computational Design and Robotic Fabrication*, Singapore: Springer, 2024, pp. 24–34.
- [3] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer, “High-Resolution Image Synthesis With Latent Diffusion Models,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 10684–10695.
- [4] Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong and Jiaolong Yang, “Structured 3D Latents for Scalable and Versatile 3D Generation,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025, pp. 21469–21480.