

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И АРХИТЕКТУРА

Научная статья

УДК/UDC 721.012.6:004.9

DOI: 10.24412/1998-4839-2024-3-327-338

Методика использования stable diffusion для генерации вариантов фасадных решений на основе искусственного интеллекта**Павел Алексеевич Пичугов¹, Салават Галиевич Шабиев²✉**^{1,2}Южно-Уральский государственный университет, Челябинск, Россия¹pashkes421@gmail.com ²shabievsg@susu.ru

Аннотация. Возможности искусственного интеллекта увеличивают вариативность при поиске проектного решения, предоставляют архитекторам новые цифровые инструменты и способствуют формированию нового подхода в архитектурном проектировании. Статья посвящена изучению метода генерации изображений вариантов фасадов здания с использованием нейронных сетей. Демонстрируются особенности работы с моделью stable diffusion с использованием технологии LoRa, позволяющей получить множество изображений фасадов с учетом указанной архитектурной стилистики, в то время как дополнительная интеграция ControlNet улучшает управление настройками в процессе генерации результата.

Ключевые слова: архитектурная визуализация и генерация, искусственный интеллект, stable diffusion, фасадные решения

Для цитирования: Пичугов П.А. Методика использования stable diffusion для генерации вариантов фасадных решений на основе искусственного интеллекта / П.А. Пичугов, С.Г. Шабиев // Architecture and Modern Information Technologies. 2024. №3(68). С. 327-338. URL: https://marhi.ru/AMIT/2024/3kvart24/PDF/21_pichugov.pdf DOI: 10.24412/1998-4839-2024-3-327-338

INFORMATION TECHNOLOGIES AND ARCHITECTURE

Original article

Methodology of using stable diffusion for generating facade solutions based on artificial intelligence**Pavel A. Pichugov¹, Salavat G. Shabiev²✉**^{1,2}South Ural State University, Chelyabinsk, Russia¹pashkes421@gmail.com ²shabievsg@susu.ru

Abstract. The capabilities of artificial intelligence increase the variability in searching for design solutions, provide architects with new digital tools, and contribute to the formation of a new approach in architectural design. The article is dedicated to the study of a method for generating images of building facade options using neural networks. It demonstrates the features of working with the stable diffusion model using LoRa technology, which allows for the creation of numerous facade images in the specified architectural style, while the additional integration of ControlNet improves control over the settings during the generation process.

Keywords: architectural visualization and generating, artificial intelligence, stable diffusion, facade solutions

For citation: Pichugov P.A., Shabiev S.G. Methodology of using stable diffusion for generating facade solutions based on artificial intelligence. Architecture and Modern Information

^{1,2} © Пичугов П.А., Шабиев С.Г., 2024

Technologies, 2024, no. 3(68), pp. 327-338. Available at:
https://marhi.ru/AMIT/2024/3kvart24/PDF/21_pichugov.pdf DOI: 10.24412/1998-4839-2024-3-327-338

Введение

С наступлением двадцать первого века искусственный интеллект (ИИ) вступил в эпоху новаторских интеграций, пролагая путь для применения машинного обучения – фундаментальной технологии ИИ в архитектурной сфере. Архитекторы с особым интересом обращаются к генеративным состязательным сетям (GAN), которые задействуются для создания изображений фасадов зданий и архитектурных моделей, демонстрируя потенциал машинного обучения в генеративном дизайне [5,6]. Исследования выявили способность GAN, обученных на аннотированных данных, к эффективному освоению формы архитектурных элементов и их распределения по поверхности фасада здания. Контролируемые GAN, такие как Pix2Pix HD и CycleGAN, требуют условного ввода как на этапе обучения, так и в процессе генерации [7,13]. Тем не менее, адаптация к новым задачам с использованием модели GAN, предварительно обученной на специфических данных, представляет собой значительный интерес. В частности, генерация структур, отражающих различные архитектурные стили, предполагает необходимость обучения множественных моделей. Кроме того, объем выборки остается критическим ограничением. Несмотря на способность неконтролируемых моделей, таких как DCGAN, обрабатывать обширные наборы данных, выполнение последующих задач продолжает оставаться сложной задачей.

В последние годы, благодаря значительному сокращению ограничений, связанных с обучающими выборками, мультимодальная обработка задач обрела популярность в исследовательской среде, особенно в контексте генерации изображений из текстовых описаний. Модель *stable diffusion* зарекомендовала себя в качестве эффективного инструмента для трансформации текста в изображение, создавая детализированные визуализации на основе текстовых подсказок [9]. Аналогичные модели, такие как *Midjourney* и *DALLE 2*, обеспечивают сопоставимые функциональные возможности [1,2,8]. Цифровые технологии преобразования текста в изображения находят широкое применение в различных сферах, в том числе и в архитектуре. В рамках серии лекций *AI Spring*³, проводимых совместно *DigitalFUTURES* и *FIU DDEs* в 2022 году, была рассмотрена тема использования технологий генерации изображений из текста в архитектуре.

Однако модели обучения нейронных сетей показывают ограниченную адаптацию к задачам проектирования фасадов зданий, а также представляют сложности в регулировании процесса генерации результатов [10,11]. По этой причине настоящее исследование посвящено изучению особенностей работы с моделью *stable diffusion* с применением технологии *LoRa* для детализированной настройки модели на основе обработанного и структурированного массива данных (образцов фасадов зданий) с последующей интеграцией *ControlNet* для улучшения контроля процесса генерации результатов. На этапе поиска проектного решения такая комбинация цифровых инструментов предоставит архитекторам возможность использования простого в применении и перспективного в творческом плане инструмента генерации множества вариантов проектных решений фасадов, управляя процессом в виде текстовой подсказки.

³ Платформа приложений для разработки искусственного интеллекта. URL:
<https://spring.io/projects/spring-ai> (дата обращения: 11.01.2024).

Методология

Методика *stable diffusion*, базирующаяся на моделях скрытой диффузии (Latent Diffusion Models, LDMS), представляет собой передовой подход к трансформации текстовых данных в визуальные изображения. Данный метод демонстрирует превосходство над генеративно-состязательными сетями (GAN) в плане качества генерируемых изображений. Применение методики *stable diffusion* охватывает широкий спектр задач, включая синтез изображений, восстановление изображений (Inpainting), повышение разрешения изображений, а также преобразование текста в изображение. После обучения модели, случайный гауссов шум, вносимый в процесс, может быть последовательно устранен, что способствует улучшению качества визуализации.

Архитектура модели *stable diffusion* включает в себя три ключевых компонента: вариационные автоэнкодеры (VAE), U-Net и кодировщик текста. VAE состоит из кодера и декодера; кодер фиксирует существенные характеристики глубоких изображений и преобразует их в сжатое представление в скрытом пространстве, в то время как декодер воссоздает изображения из данных скрытого пространства, предназначенных для U-Net. U-Net, основанный на остаточной структуре⁴, функционирует как кодер и декодер, трансформируя изображения низкого разрешения в аналоги высокого разрешения после их сжатия кодером. Кодировщик текста конвертирует аннотированные последовательности в потенциальные последовательности для текстового встраивания, переводя входной текст в формат, понятный U-Net, и направляя последний в процессе генерации изображения.

Для упрощения процедуры загрузки модели и генерации изображений в нашем исследовании применяется веб-интерфейс *Stable Diffusion* (рис. 1). Веб-интерфейс обеспечивает интуитивно понятную пользовательскую среду, которая содержит в себе настройки преобразования текста в изображение, повышения разрешения и обучения модели, что расширяет возможности пользователя в управлении процессами генерации итоговых визуализаций.

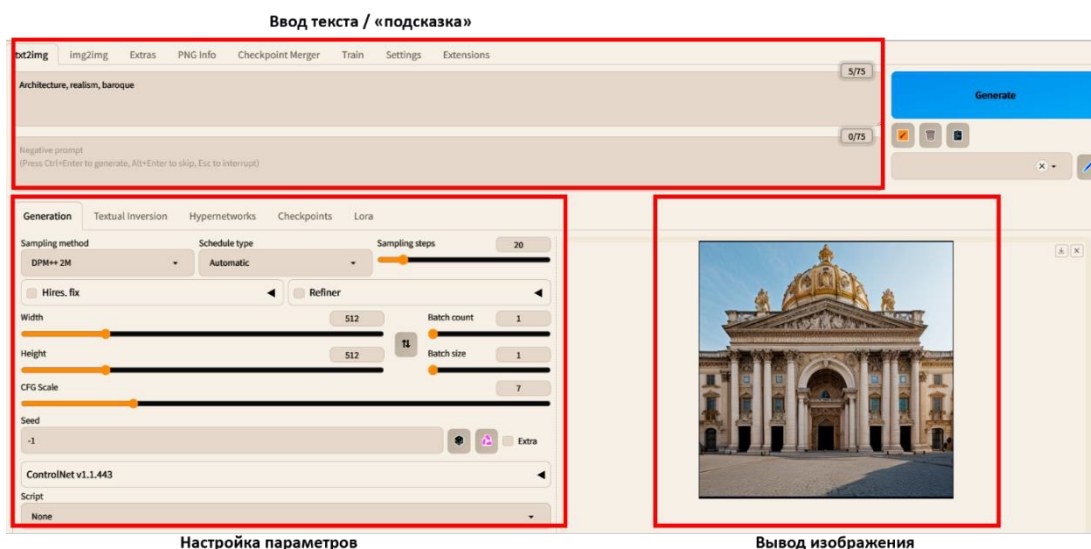


Рис. 1. Веб-интерфейс *stable diffusion*

⁴ Остаточная структура (Residual Structure) в контексте нейронных сетей и, в частности, в архитектуре U-Net, относится к концепции остаточных блоков (Residual Blocks), которые были введены для решения проблемы исчезающего градиента в глубоких нейронных сетях. Эта концепция стала основой для ResNet (Residual Network), одной из самых известных и эффективных архитектур в глубоком обучении.

Использование и настройка LoRa и ControlNet

В рамках собственных исследований Microsoft была разработана инновационная технология под названием Low-Rank Adaptation (LoRa), предназначенная для усовершенствования процесса точной настройки лингвистических моделей. В первоначальном виде модель stable diffusion характеризовалась замедленным временем просчета и значительными трудностями в настройке, поэтому появление LoRa предложило новое решение. Несмотря на то, что методы точной настройки упрощенных крупномасштабных моделей, такие как текстовая инверсия или Dreambooth, становятся всё более популярными, они по-прежнему требуют значительных вычислительных ресурсов. Метод LoRa минимизирует необходимость вычисления градиента весов модели⁵, вводя обучаемый слой в каждый блок Transformer, что существенно уменьшает объем требуемых обучающих параметров. Точная настройка с использованием LoRa осуществляется быстрее и требует меньше вычислительной мощности, сохраняя при этом качество на уровне полноценной модели.

С другой стороны, модели stable diffusion генерируют изображения с высоким уровнем случайности, что осложняет контроль над результатом. Точное регулирование сгенерированного контента с учетом информации, содержащейся в тексте, представляет определенные сложности. Более прогрессивная модель ControlNet, предлагает решение этой проблемы, позволяя управлять результатами создания изображений путем введения дополнительных условий в модель stable diffusion [12]. Это упрощает регулирование процесса генерации и снижает эффект получения случайных результатов. ControlNet включает в себя множество управляющих условий, таких как Canny Edge, карта сегментации и другие.

В данном исследовании мы применяем LoRa для оптимизации настроек модели stable diffusion, обученной на базе данных CMP Facades, после чего используем различные условия, предоставляемые моделью ControlNet, для регулирования процесса разработки фасадов зданий. Объединяющей особенностью этих технологий является возможность создания с их помощью новых идей на этапе концептуального проектирования.

Процесс обучения модели stable diffusion

В рамках проведенного исследования для начального этапа обучения нейросети было взято 200 изображений из базы данных CMP Facades, выбранных случайным отбором, демонстрирующим эффективность метода точной настройки Low-Rank Adaptation (LoRa) даже при небольшом объеме данных. Выбранные изображения были приведены к разрешению 512×512 пикселей. Впоследствии текстовые описания, соответствующие каждому изображению, использовались в качестве триггерных слов для обучающего набора.

В качестве исходной модели была выбрана версия нейросети Stable Diffusion v1-4 (Rombach), которая подверглась корректировке на графическом процессоре NVIDIA RTX 3060 с 6 ГБ памяти, при параметрах обучения: эпоха = 1, размер пакета = 20000, скорость обучения = 0.00001. Процесс обучения занял более двух часов, в результате чего размер модели составил 144 МБ и она стала доступной для скачивания в сообществе Civitai⁶.

⁵ Модель градиентного веса SD относится к методу нейронной сети для оптимизации весов модели глубокого обучения. Этот метод представляет собой разновидность градиентного спуска, популярного алгоритма оптимизации в машинном обучении. Градиентный спуск включает в себя вычисление градиента функции потерь относительно параметров модели (в данном случае весов) и обновление параметров с использованием итерационного алгоритма.

⁶ Сайт моделей stable diffusion. URL: <https://civitai.com/gallery/133518?modelId=11661&modelVersionId=13784&infinite=false&returnUrl=%2Fmodels%2F11661%2Fbuildingfacade> (дата обращения: 14.01.2024).

Дополнительно, в исследовании использовалась модель, предоставленная для ControlNet (рис. 2).

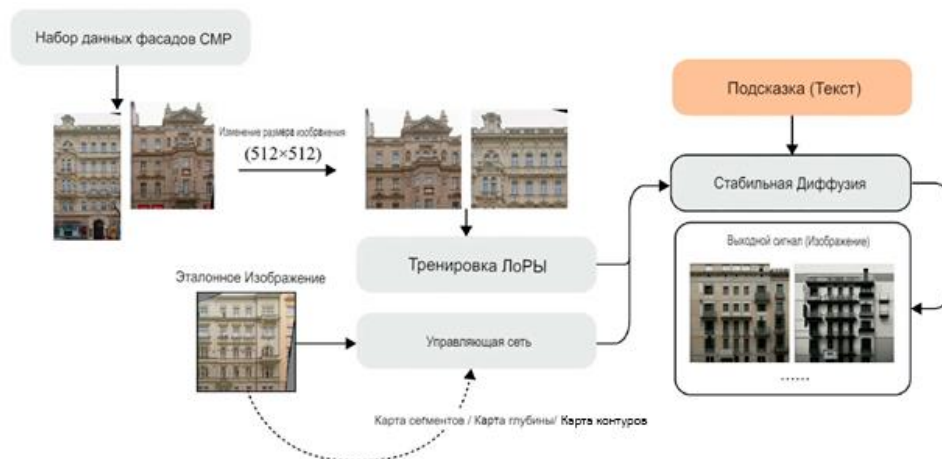


Рис. 2. Схема экспериментального рабочего процесса обучения модели stable diffusion

Генерация фасадов с другой семантической базой, основанных на LoRa

Был произведен анализ сравнения различных методик выборки изображений и этапов выборки (рис. 3), включая запросы типа:

<lora: buildingface: 0.7>, Ultra hd! редакционная статья, модель Рика Оуэнса, iso 400, сюрреалистический, 8k, полнокадровый, пейзаж, архитектура, итальянский стиль. В данном контексте, вес модели LoRa обозначается как <lora: buildingface: 0.7>, где 0.7 указывает на величину градиентного веса. CFG-шкала была установлена на уровне 7, что позволило увеличить влияние текста на генерируемые результаты при снижении значения шкалы CFG, однако это также ведет к повышению непредсказуемости. Количество этапов выборки было установлено на уровне 2, 4, 6, 8 и 32.

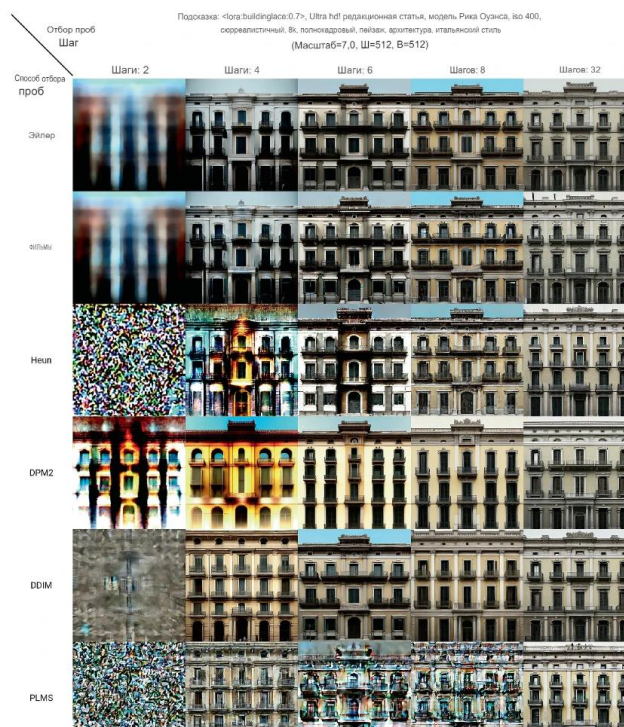


Рис. 3. Генерация результатов различных методов выборки и этапов выборки, основанных на LoRa

Исследование показало, что методы выборки Euler и LMS генерируют схожий контент на протяжении всех этапов выборки, тогда как каждый метод выборки порождает уникальный контент на различных этапах. Метод Neun схож с методом PLMS, и до момента проявления содержимого четвертого этапа, содержание шума на втором этапе выборки носит случайный характер. Следует подчеркнуть, что содержание этапов с 6 по 8 в методе PLMS меняется непредсказуемо. Также было отмечено, что результаты, полученные с использованием методов DPM2 и DDIM, удивительно схожи, несмотря на различия в подходах к выборке. Метод выборки DPM2, демонстрирующий наивысший уровень использования тегов, превышающий 80%, был выбран в качестве основы для дальнейших исследований [3,4].

Стоит отметить, что исходные изображения сами не являются идеальными образцами стилей. Генерация изображений фасадов зданий определенно относится к эклектике, но само обучение модели происходит на базе исходных стилей для добавления присущих им элементов декора.

Далее был исследован эффект генерации изображений с использованием альтернативной семантической базы архитектурного стиля в точно настроенной модели *stable diffusion*. Параметры настройки оставались неизменными, как и в предыдущем разделе. На рисунке 4 представлен процесс адаптации фасадного решения здания к различным архитектурным стилистикам в ответ на изменения текстового запроса. Например, при запросе традиционной китайской архитектурной стилистики, модель воспроизводила характерные элементы карнизов, упрощенный декор окон и дверей и т.д. Фасады, генерируемые нейронными сетями в стилистике классицизма и барокко также хорошо отличимы, и имеют свои уникальные особенности.



Рис. 4. Примеры генерации фасадов зданий различной архитектурной стилистики

В контексте применения точно настроенной модели Low-Rank Adaptation (LoRa), базирующейся на архитектуре *stable diffusion*, отмечается способность генерировать визуальный материал, который фундаментально отличается от исходных входных данных. Такая модель предоставляет большие возможности для генерации множества различных фасадных решений с учетом указанной стилистики посредством простого текстового ввода данных.

Генерация фасадов зданий с использованием ControlNet

В отсутствие управляющих ограничений модель *stable diffusion*, осуществляющая преобразование текста в изображение, характеризуется высоким уровнем

непредсказуемости результата генерации. Поэтому в модель stable diffusion были внедрены генеративные условия через ControlNet, что позволило получить набор дополнительных настроек управления процессом генерации вариантов фасадных решений здания. В данной работе мы рассматриваем использование моделей Canny Edge, Segment Map, Depth Map и MLSD в качестве условий управления и анализируем результаты генерации при различных градиентных весах ControlNet (0.2, 0.4, 0.6, 0.8 и 1.0).

Как представлено на рисунке 5, эталонным изображением для модели ControlNet служит фасад здания с разрешением 512 x 512 пикселей. Результаты исследования указывают на то, что модель Canny Edge демонстрирует наилучшую эффективность при максимальном весе сети управления ($W = 1.0$), сохраняя структурные контуры элементов фасада (Canny Edge) эталонного изображения и учитывая заданные требования к подсказкам. В то же время, расположение рельефных элементов фасада на эталонном изображении было менее схоже с результатами, полученными при использовании других моделей в аналогичных условиях. Например, результаты, полученные с использованием модели Depth Map, не соответствовали ни схеме расположения на эталонном изображении, ни художественным требованиям подсказки. Степень сходства между генерируемыми результатами была особенно высока при относительно низких весах модели ControlNet ($W = 0.2$), вопреки тому, что обе модели в итоге генерировали различные результаты. С увеличением веса до 0.4, выходные данные модели начинали демонстрировать заметные различия.

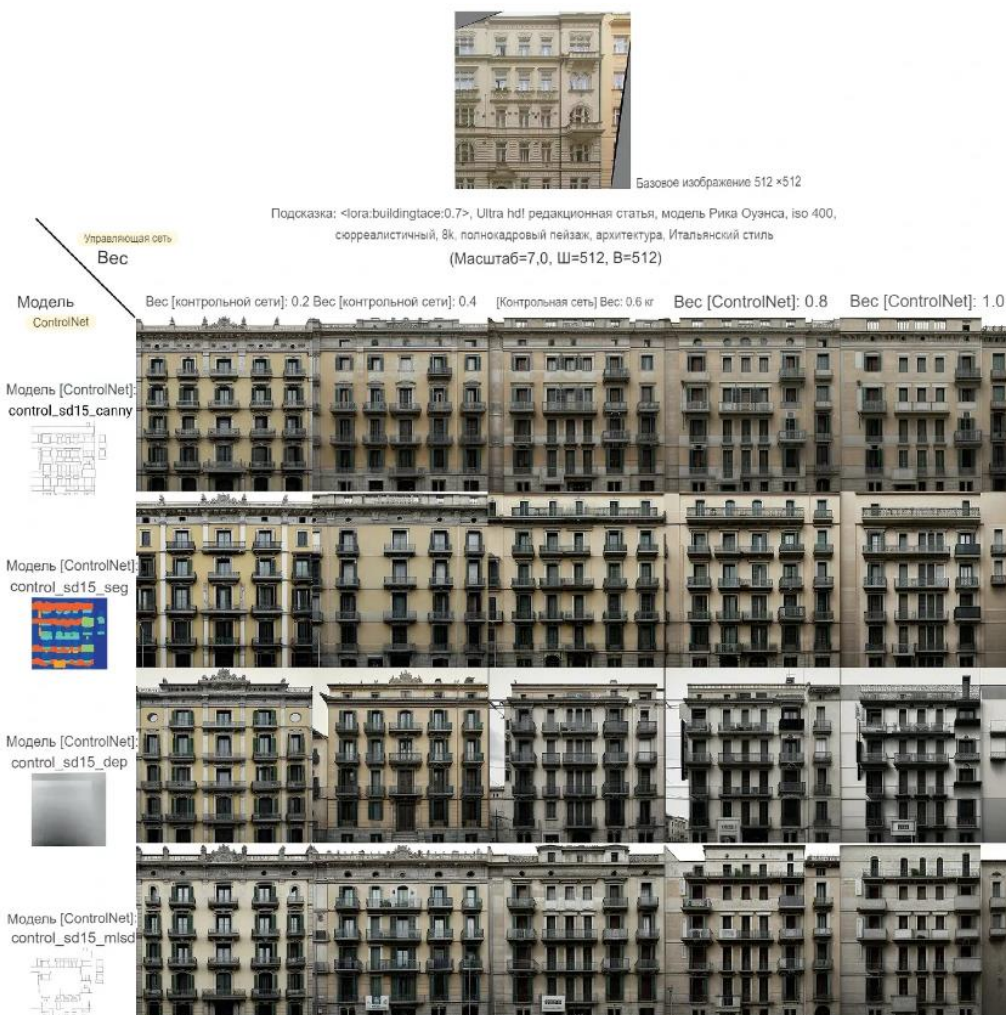


Рис. 5. Результаты генерации изображений фасадов с использованием ControlNet при разных показателях веса сети управления

В целом, различные модели управления, внедренные через ControlNet, оказывают дифференцированное воздействие на генерируемые результаты. Модель Canny Edge обеспечивает более высокую степень соответствия ожиданиям по сравнению с моделями Segment Map и MLSD. Выходные данные модели Depth Map представляют более качественное понимание пространственной ориентации. Меньшие веса сети управления способствуют генерации более разнообразных результатов при сохранении идентичных условий. Структура фасада здания становится более схожей с эталонным объектом по мере увеличения веса, однако детализация фасада зданий уменьшается. Таким образом, повышение градиента веса генерации модели, ограничивают логические вычислительные расчеты stable diffusion.

В рамках исследования была осуществлена генерация фасадов зданий с использованием модели Canny Edge, интегрированной в ControlNet, с целью оценки влияния различных значений весовых коэффициентов на конечные результаты. Как иллюстрируется на рисунке 6, наблюдается тенденция к унификации архитектурных стилей и уплотнению элементов фасада с увеличением значения веса в сети ControlNet. В контексте фасадов традиционной архитектуры Китая, при весовом значении, превышающем 0.4, отмечается сохранение отдельных элементов, тогда как характерные крупные карнизы начинают постепенно утрачиваться. При установленном весе в 1.0, элементы карниза практически исчезают, однако в верхней правой части изображения карниз фрагментарно сохраняется. ControlNet демонстрирует способность к эффективному управлению соответствием между генерируемыми результатами и исходными изображениями, однако это сопровождается низкой детализацией фасадов. Оптимальный диапазон значений градиента веса генерации в сети ControlNet определен в пределах от 0.6 до 0.8.



Рис. 6. Результаты генерации фасада здания с различным градиентным весом в сети ControlNet

Эксперименты с окончательной генерацией фасадов

В ходе экспериментов по финальной генерации были использованы оптимальные параметры настройки вспомогательных моделей (LoRa и ControlNet). Как представлено на рисунке 7, была предпринята попытка сгенерировать фасад здания в «рациональном» стиле («modern» – является семантическим определением перевода «Современный» или «Современная»), и модель stable diffusion, настроенная с помощью LoRa, демонстрировала точное понимание заданного запроса. Фасад здания сохранял согласованность с архитектурой эталонного образца под управлением ControlNet, и процесс занял всего 0.2 секунды. После добавления слов «светлый и яркий» в запрос, нейросеть предоставила белый фасад, соответствующий текстовому описанию. Добавление текстового описания позволило быстро изменить результаты выводимых изображений, что в дальнейшем может повысить эффективность при разработке фасадов зданий в процессе архитектурного проектирования.

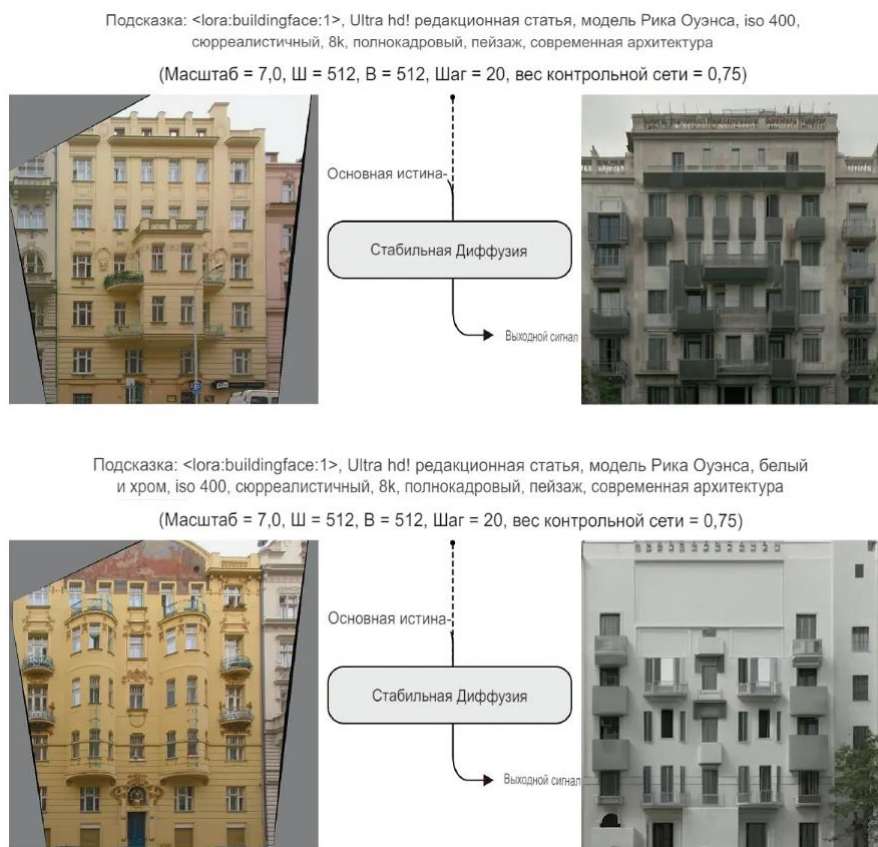


Рис. 7. Результаты генерации модели фасадного решения stable diffusion, настроенной с помощью LoRa с различным текстовым запросом

Заключение

В рамках настоящего исследования был представлен методологический подход к генерации изображений вариантов фасадов зданий, опирающийся на использование модели stable diffusion. Методика Low-Rank Adaptation (LoRa) применялась для окончательной настройки модели, предварительно обученной на изображениях фасадов разных стилей. Дополнительно, в исследовании рассматривалось применение ControlNet в качестве средства для модуляции результатов генерации. В данной работе были выполнены контрольные операции по настройке и комплексному использованию моделей нейронных сетей, улучшающие процесс семантической интерпретации текста для генерации фасадов зданий в разных стилистиках.

Основные выводы исследования

1. Выявлено, что при генеративном поиске фасадного решения, точная настройка модели stable diffusion с использованием LoRa обеспечивает снижение нагрузки на графический процессор и способствует экономии времени.
2. Модель stable diffusion, настроенная через LoRa, демонстрирует высокую гибкость в решении задач, связанных с поиском оптимального фасадного решения, или позволяет эффективно интегрировать семантические особенности различных архитектурных стилей в генерируемые результаты выводимых изображений.
3. Дополнительная надстройка модели ControlNet может быть использована для регулирования соответствия генерируемых вариантов фасадов исходного образца, однако чрезмерно высокие значения изменений модели могут ограничить разнообразие результатов выводимых изображений.

В целом, предложенный метод упрощает процесс проектирования фасадов зданий, позволяя модифицировать текст запроса и корректировать силу влияния нейросети для получения большого набора результатов изображений. Будущие исследования могут включать интеграцию морфологических генеративных алгоритмов нейронных сетей для достижения, более детализированных результатов визуализации архитектуры зданий, а также повышения количества исходных изображений при обучении модели для исключения возможных неточностей полученных результатов. Необходимы дальнейшие эксперименты для проверки стабильности и надежности предложенных методов генерации вариантов изображений архитектурных фасадных решений на основе нейронных сетей.

Источники иллюстраций

Рис. 1-7. Авторские иллюстрации.

Список источников

1. Borji A. Generated faces in the Wild: A quantitative comparison of stable diffusion, midjourney and dall-e 2 // Quintic AI. San Francisco. 2022. URL: <https://arxiv.org/pdf/2210.00586> (дата обращения: 14.01.2024).
2. Borji A. Pros and cons of gan evaluation measures // Computer Vision and Image Understanding. 2019. URL: <https://arxiv.org/pdf/1802.03446> (дата обращения: 14.01.2024).
3. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. High-resolution image synthesis with latent diffusion models // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. S. 10684-10695 (дата обращения: 14.01.2024).
4. Li X. L., Liang P. Prefix-Tuning: Optimizing Continuous Prompts for Generation // arXiv:2101.00190 [cs.CV]. 2021. URL: <http://arxiv.org/abs/2101.00190> (дата обращения: 14.01.2024).
5. Huang V., Zheng H. Recognition and generation of architectural drawings using machine learning // Proceedings of the 38th Annual Conference of the Association for Computer-Aided Design in Architecture. Mexico City, Mexico. 2018. S. 18-20. DOI: [10.52842/conf.acadia.2018.156](https://doi.org/10.52842/conf.acadia.2018.156) (дата обращения: 14.01.2024).
6. Isola P., Zhu J.-Y., Zhou T., Efros A.A. Image-to-image transformation using conditional competition networks // IEEE Conference on Computer Vision and Pattern Recognition

- (CVPR). 2017. S. 5967-5976. DOI:10.1109/CVPR.2017.632 (дата обращения: 13.01.2024).
7. Park T., Liu M.-Y., Wang T.-C., Zhu J.-Y. Semantic image synthesis with spatially adaptive normalization // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019. S. 2337-2346. DOI: 10.1109/CVPR.2019.00244 (дата обращения: 13.01.2024).
 8. Ramesh A., Dhariwal P., Nichol A., Chu C., Chen M. Hierarchical Text-Conditional Image Generation with CLIP Latents // Open AI. 2022. arXiv:2204.06125 [cs.CV]. URL: <https://doi.org/10.48550/arXiv.2204.06125> (дата обращения: 13.01.2024).
 9. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. Synthesis of high-resolution images using latent diffusion models // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. S. 10674-10685.
 10. Ruiz N., Li Y., Jampani V., Pritch Y., Rubinstein M., Aberman K. DreamBooth: fine-tuning text-to-image diffusion models for domain-oriented generation // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. S. 22500-22510.
 11. Fürst A., Rumetshofer E., Tran V. T., Ramsauer H., Tang F., Lehner J., Kreil D., Kopp M., Klambauer G., Bitto-Nemling A., Hochreiter S. CLOOB: Modern Hopfield Networks with InfoLOOB Outperform CLIP. 2022. URL: <https://openreview.net/forum?id=qw674L9PfQE> (дата обращения: 09.03.2024).
 12. Zhang L., Agrawala M. Adding conditional control to text-to-image diffusion models // arXiv:2302.05543 [cs.CV]. 2023. URL: <https://doi.org/10.48550/arXiv.2302.05543> (дата обращения: 10.03.2024).
 13. Zhu J.Y., Park T., Isola P., Efros A.A. Unpaired image-to-image translation using cyclically consistent adversarial networks // Proceedings of the IEEE International Conference on Computer Vision. 2017. S. 2223-2232 (дата обращения: 10.03.2024).

References

1. Borji A. Generated faces in the Wild: A quantitative comparison of stable diffusion, midjourney and dall-e 2. Quintic AI, San Francisco, 2022. Available at: <https://arxiv.org/pdf/2210.00586>
2. Borji A. Pros and cons of gan evaluation measures. Computer Vision and Image Understanding, 2019. Available at: <https://arxiv.org/pdf/1802.03446>
3. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. High-resolution image synthesis with latent diffusion models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 10684-10695.
4. Li X. L., Liang P. Prefix-Tuning: Optimizing Continuous Prompts for Generation. arXiv:2101.00190 [cs.CV], 2021. Available at: <http://arxiv.org/abs/2101.00190>
5. Huang V., Zheng H. Recognition and generation of architectural drawings using machine learning. Proceedings of the 38th Annual Conference of the Association for Computer-Aided Design in Architecture, Mexico City, Mexico, 2018, pp. 18-20. DOI: [10.52842/conf.acadia.2018.156](https://doi.org/10.52842/conf.acadia.2018.156)
6. Isola P., Zhu J.-Y., Zhou T., Efros A.A. Image-to-image transformation using conditional competition networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5967-5976. DOI: 10.1109/CVPR.2017.632

7. Park T., Liu M.-Y., Wang T.-C., Zhu J.-Y. Semantic image synthesis with spatially adaptive normalization. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 2337-2346. DOI: 10.1109/CVPR.2019.00244
8. Ramesh A., Dhariwal P., Nichol A., Chu C., Chen M. Hierarchical Text-Conditional Image Generation with CLIP Latents. Open AI, 2022, arXiv:2204.06125 [cs.CV]. Available at: <https://doi.org/10.48550/arXiv.2204.06125>
9. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. Synthesis of high-resolution images using latent diffusion models. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 10674-10685.
10. Ruiz N., Li Y., Jampani V., Pritch Y., Rubinstein M., Aberman K. DreamBooth: fine-tuning text-to-image diffusion models for domain-oriented generation. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, C. 22500-22510.
11. Fürst A., Rumetshofer E., Tran V. T., Ramsauer H., Tang F., Lehner J., Kreil D., Kopp M., Klambauer G., Bitto-Nemling A., Hochreiter S. CLOOB: Modern Hopfield Networks with InfoLOOB Outperform CLIP, 2022. Available at: <https://openreview.net/forum?id=qw674L9PfQE>
12. Zhang L., Agrawala M. Adding conditional control to text-to-image diffusion models. arXiv:2302.05543 [cs.CV], 2023. Available at: <https://doi.org/10.48550/arXiv.2302.05543>
13. Zhu J.-Y., Park T., Isola P., Efros A.A. Unpaired image-to-image translation using cyclically consistent adversarial networks. Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223-2232.

ОБ АВТОРАХ

Пичугов Павел Алексеевич

Аспирант кафедры «Архитектура», Южно-Уральский государственный университет, Челябинск, Россия
pashkes421@gmail.com

Шабиев Салават Галиевич

Доктор архитектуры, заведующий кафедрой «Архитектура», профессор, Южно-Уральский государственный университет, Челябинск, Россия
shabievsg@susu.ru

ABOUT THE AUTHORS

Pichugov Pavel A.

Postgraduate Student of the Department of Architecture, South Ural State University, Chelyabinsk, Russia
pashkes421@gmail.com

Shabiev Salavat G.

Doctor of Architecture, Head of the Department of Architecture, Professor, South Ural State University, Chelyabinsk, Russia
shabievsg@susu.ru

Статья поступила в редакцию 06.05.2024; одобрена после рецензирования 05.09.2024; принята к публикации 10.09.2024.