

## **ДИФФУЗИЯ ИЗОБРАЖЕНИЙ: ПОДХОД К АВТОМАТИЧЕСКОЙ ГЕНЕРАЦИИ РЕАЛИСТИЧНЫХ ИНТЕРЬЕРОВ ПО ЭСКИЗУ**

**Аннотация.** Проведен сравнительный анализ моделей расширения генерации реалистичных интерьеров по эскизу ControlNet для Stable Diffusion. Выделены модели, наиболее точно генерирующие интерьер по заданному изображению и текстовому описанию.

**Ключевые слова:** генерация изображений, нейронная сеть, использование расширения ControlNet, сравнительный анализ моделей Stable Diffusion.

**Введение.** В последние годы автоматическая генерация реалистичных интерьеров активно используется в области компьютерного зрения и графики. С развитием технологий машинного обучения появилась возможность создавать детализированные изображения интерьеров, которые используются в различных сферах: виртуальная реальность, дизайн интерьеров, архитектурное моделирование и другое [1]. Однако генерация реалистичных интерьеров по эскизу все еще является сложной задачей. Существующие методы, такие как фотореалистичный рендеринг и компьютерное зрение, зачастую не способны правильно интерпретировать эскизы и создать реалистичные изображения. Эскизы, которые создают дизайнеры или пользователи, обычно имеют низкую детализацию и содержат ограниченную информацию о расположении объектов и освещении [2].

В этой статье описывается новый подход глубокого машинного обучения к генерации интерьеров, основанный на диффузии изображений с использованием Stable Diffusion. Подход позволяет создавать детализированные и реалистичные интерьеры, сохраняя стиль и требования дизайнера либо пользователя [3]. Это важный шаг в области автоматической генерации интерьеров, который имеет широкое практическое применение. Stable Diffusion — это система, состоящая из множества компонентов и моделей. В процессе обучения модели связывают слова с изображениями благодаря методике CLIP

[4]. В Stable Diffusion используется процесс скрытой диффузии, при котором модель обучается распознавать знакомые формы в поле шума и фокусировать эти формы, если они соответствуют текстовому описанию [5]. Stable Diffusion имеет расширение ControlNet, которое улучшает обучение моделей диффузии изображений. Данное расширение помогает генерировать наиболее качественное изображение, используя семь разных моделей, такие как Canny, HED, Depth Map, Normal Map, MLSD, OpenPose, Scribble.

**Проблема и цель исследования.** Исходя из вышесказанного, проблемой данной работы является сложность сохранения уникальности и стиля изображения интерьеров помещений.

Целью данной работы является сравнительный анализ моделей Stable Diffusion, подбор модели, наиболее точно генерирующей интерьер по исходному изображению и текстовому описанию.

**Материалы и методы.** Модель Canny на основании исходного изображения формирует набросок, выделяет очертание объектов, сохраняя детали. Далее набросок переносится на новое изображение. Дополнительно на генерацию влияет текстовое описание. С помощью данной модели можно заимствовать у изображения композицию и стилистику, но менять детали [6].

Модель HED позволяет генерировать неаккуратные изображения, напоминающие человеческие наброски.

По исходному изображению Модель Depth Map создает набросок с картой глубины, отображает расположение объектов [7].

Модель Normal Map определяет положение объекта в трехмерном пространстве. Может использоваться для генерации 3D изображений.

Модель MLSD точно повторяет прямые линии исходного изображения. Подходит для генерации интерьеров, зданий [8].

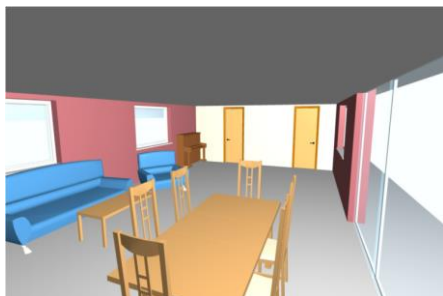
На основе исходного изображения Модель OpenPose формирует схему позы, благодаря чему наиболее точно повторяет положение частей тела. В модели существует редактор, с помощью которого можно перемещать части тела и создавать свои позы.

Модель Scribble генерирует изображение на основании исходного скетча. С помощью текстового описания можно указать, во что сгенерировать скетч [9].

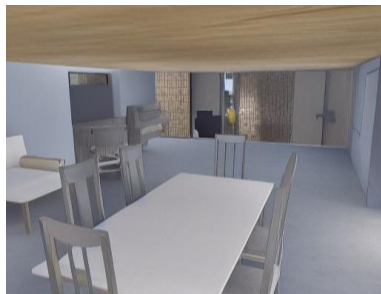
Для генерации изображений интерьеров важно сохранять детали, объем элементов и повторять их очертание [10].

**Результаты.** Проведен сравнительный анализ моделей расширения ControlNet для Stable Diffusion. Выбрано исходное изображение (рис. 1) в виде интерьера комнаты с разными предметами мебели, определено текстовое описание интерьера. Для тестирования использованы модели Canny, HED, Depth Map, MLSD, Scribble (рис. 2-6).

По результатам тестирования выделены модели, визуально наиболее точно повторяющие интерьер исходного изображения и учитывающие текстовое описание. Таковыми являются Canny и Depth Map, так как сохраняют объем элементов, стиль изображения и наиболее четко повторяют очертание объектов.



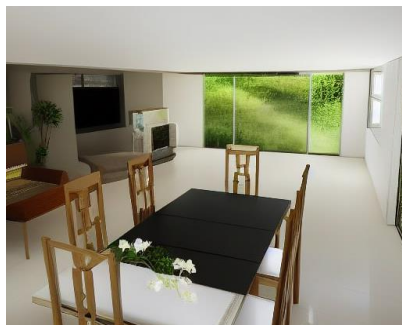
*Рис. 1.* Исходное изображение



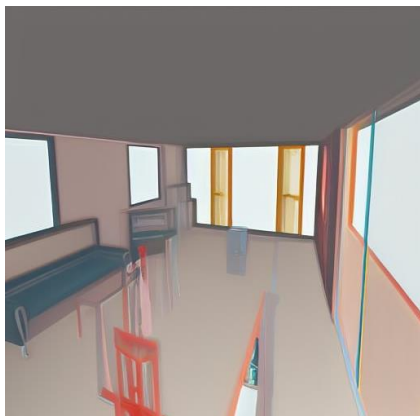
*Рис. 2.* Сгенерированное изображение с использованием модели Depth Map



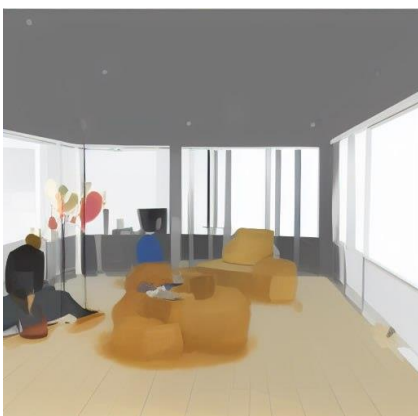
*Рис. 3.* Сгенерированное изображение с использованием модели Canny



*Рис. 4.* Сгенерированное изображение с использованием модели MLSD



*Рис. 5.* Сгенерированное изображение с использованием модели HED



*Рис. 6.* Сгенерированное изображение с использованием модели Scribble

## СПИСОК ЛИТЕРАТУРЫ

1. Багаев И. И. Анализ понятий нейронная сеть и сверточная нейронная сеть, обучение сверточной нейросети при помощи модуля TensorFlow / И. И. Багаев. — DOI 10.18503/2306-2053-2020-8-1-15-22. — Текст : электронный // Математическое и программное обеспечение систем в промышленной и социальной сферах. — 2020. — Т. 8, № 1. С. 15-22. — URL: <https://www.elibrary.ru/item.asp?id=50010096> (дата обращения 29.05.2023).
2. Богатырев С. А. Применение нейронных сетей для генерации и обработки изображений / С. А. Богатырев, М. М. Лазарева. — DOI 10.34755/IROK.2022.70.17.017. — EDN RYTQWB // Сборник материалов III Международной научно-практической конференции, Москва, 2022. — Москва: Алэф, 2022. — С. 134-143. — URL: <https://www.elibrary.ru/> (дата обращения 29.05.2023)
3. Мартыненко А. М. Анализ нейронных сетей «Stable Diffusion» для генерации фотографий, по преобразованию текста в изображение / А. М. Мартыненко, С. В. Васильев // Материалы VII Международной научной конференции. Донецк, 2022. Под редакцией С.В. Беспаловой.

- Т. 2. — Донецк: Донецкий национальный университет, 2022. — С. 265-267. — EDN SMYPLZ. — Текст : непосредственный.
4. Chen Henry Wu, Fernando De la Torre. Unifying Diffusion Models' Latent Space, with Applications to CycleDiffusion and Guidance. URL: <https://arxiv.org/abs/2210.05559> (date of the application 22.05.2023). — Text : electronic.
  5. Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H. Bermano, Gal Chechik, Daniel Cohen-Or. An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion. URL: <https://arxiv.org/abs/2208.01618> (date of the application 22.05.2023). — Text : electronic.
  6. Как работает Stable Diffusion [Сайт]. URL: <https://habr.com/ru/articles/693298/> / (дата обращения 26.05.2023).
  7. Stable Diffusion. Самая впечатляющая нейросеть. Как пользоваться новым инструментом ИИ? [Сайт]. URL: <https://vc.ru/future/506283-stable-diffusion-samaya-vpechatlayushchaya-neyroset-kak-polzovatsya-novym-instrumentom-ii> (дата обращения 26.05.2023).
  8. Andreas Blattmann, Robin Rombach, Kaan Oktay, Jonas Müller, Björn Ommer. Semi-Parametric Neural Image Synthesis]. — URL: <https://arxiv.org/abs/2204.11824> (date of the application 22.05.2023). — Text : electronic.
  9. Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, Daniel Cohen-Or. Prompt-to-Prompt Image Editing with Cross Attention Control. — URL: <https://arxiv.org/abs/2208.01626> (date of the application 22.05.2023). — Text : electronic.
  10. Weixi Feng, Xuehai He, Tsu-Jui Fu, Varun Jampani, Arjun Reddy Akula, Pradyumna Narayana, Sugato Basu, Xin Eric Wang, William Yang Wang. Prompt-to- Training-Free Structured Diffusion Guidance for Compositional Text-to-Image Synthesis. — URL: <https://openreview.net/forum?id=PUIqjT4rzq7> (date of the application 22.05.2023). — Text : electronic.